Regression Analysis of Interval Data 4

Multi-Response Permutation Procedures (MRPP) were introduced in Chap. 2 and applied to interval-level, completely randomized data in Chap. 3. While multi-response permutation procedures are generally thought of as providing tests of differences among g treatment groups as demonstrated in Chap. 3, they also have applications in ordinary least squares (OLS) linear regression analyses with $v = 2$ and least absolute deviations (LAD) linear regression analyses with $v = 1$. In this fourth chapter of *Permutation Statistical Methods*, MRPP analyses of LAD regression residuals are illustrated with a variety of experimental designs, including one-way completely randomized with and without a covariate, one-way and two-way randomized-block, two-way factorial, Latin square, and two-factor nested analysis-of-variance designs. Also considered are multivariate multiple regression designs.

4.1 LAD Linear Regression

OLS linear regression has long been recognized as a useful tool in many fields of research. The optimal properties of OLS regression are well known when the errors are normally distributed. However, in practice the assumption of multivariate normality is rarely justified. LAD linear regression is an attractive alternative to OLS regression as it is extremely robust to deviations from normality as well as to the presence of extreme values [297, p. 172].

It is widely recognized that estimators of OLS regression parameters can be severely affected by unusual values in either the criterion variable or in one or more of the predictor variables. This is due in large part to the weight given to each data point when minimizing the sum of squared errors. In contrast, LAD regression is much less sensitive to the effects of unusual-value errors due to the fact that the errors are not squared. Moreover, LAD regression has been shown to be superior to OLS regression when errors are generated from heavy-tailed or outlier-

K.J. Berry et al., *Permutation Statistical Methods*,
DOI 10.1007/978-3-319-28770-6_4

producing distributions, such as the Cauchy and double-exponential distributions; see, for example, articles by Blattburg and Sargent [46], Dielman [94, 95], Dielman and Pfaffenberger [96], Dielman and Rose [97], Mathew and Nordström [264], Mielke, Berry, Landsea, and Gray [303], Pfaffenberger and Dinkel [337], Rice and White [346], Rosenberg and Carlson [352], Rousseeuw [355], Taylor [394], and Wilson [432].

As described by Sheynin, the initial known use of regression by Daniel Bernoulli (c. 1734) for astronomical prediction problems involved LAD regression based on ordinary Euclidean distances between the observed and predicted response values [372]. Further developments in LAD regression were due to Roger Joseph (Rogerius Josephus) Boscovich (c. 1755), Pierre-Simon Laplace (c. 1789), and Carl Friedrich Gauss (c. 1809). The American mathematician and astronomer Nathaniel Bowditch (c. 1809) was highly critical of OLS regression because, as he argued, squared regression residuals unduly emphasized questionable observations in comparison with the absolute regression residuals associated with LAD regression [372].

Consider the general multivariate regression model given by

$$\mathbf{y}_i = \boldsymbol{h}\left(\boldsymbol{\beta}, \mathbf{x}_i\right) + \mathbf{e}_i \ ,$$

where $\mathbf{y}_i' = (y_{1i}, \ldots, y_{ri})$ denotes the row vector of r observed response measurements for the ith of N objects, $\mathbf{x}_i' = (x_{1i}, \ldots, x_{si})$ is the row vector of s predictor values for the ith object, $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_t)$ is the row vector of t parameters, $\boldsymbol{h}' = (h_1, \ldots, h_r)$ is the row vector of r model functions of $\boldsymbol{\beta}$ and $\mathbf{x}_i$ for the ith object, and $\mathbf{e}_i' = (e_{1i}, \ldots, e_{ri})$ denotes the r errors between the response variables and model functions for the ith object, $i = 1, \ldots, N$ objects. The special case of a multivariate linear regression model is given by

$$\mathbf{y}_i = \mathbf{B}\boldsymbol{f}(\mathbf{x}_i) + \mathbf{e}_i \ ,$$

where $\boldsymbol{f}(\mathbf{x}_i)$ denotes a column vector of p distinct functions of s predictors ($\mathbf{x}_i$) for the ith object, $i = 1, \ldots, N$, and $\mathbf{B}$ is an $r \times p$ matrix of parameters in which $(\mathbf{B}_{j1}, \ldots, \mathbf{B}_{jp})$ is the row vector of p parameters associated with the jth response measurement, $j = 1, \ldots, r$.

Let $\mathbf{y}_i$ denote a column vector of r observed response measurement scores and let $\tilde{\mathbf{y}}_i$ denote a column vector of r predicted response values for the ith object, $i = 1, \ldots, N$. Thus, the general and linear predicted multivariate regression models are given by

$$\tilde{\mathbf{y}}_i = \boldsymbol{h}\left(\tilde{\boldsymbol{\beta}}, \mathbf{x}_i\right)$$

and

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{B}}\boldsymbol{f}\left(\mathbf{x}_i\right) \ ,$$

respectively, where $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{B}}$ are estimated parameters that are intended to provide good fits between the $\mathbf{y}_i$ and $\tilde{\mathbf{y}}_i$ values relative to a selected goodness-of-fit criterion. The null hypothesis (H_0) underlying each criterion dictates that each of the $N!$ possible, equally-likely pairings of the predicted sequential ordering $(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_N)$ with the fixed observed sequential ordering $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$ occurs with equal probability, i.e., $1/N!$.

Let $\Delta(\tilde{\mathbf{y}}_\mathbf{i}, \mathbf{y}_i)$ for $i = 1, \ldots, N$ denote the distance function between the predicted and observed response measurement values and consider the generalized Minkowski distance function given by

$$\Delta(\tilde{\mathbf{y}}_i, \mathbf{y}_i) = \left(\sum_{j=1}^{r} \left| \tilde{y}_{ij} - y_{ij} \right|^w \right)^{v/w} ,$$

where $w \geq 1$ and $v > 0$. Since $v = 1$ yields the Minkowski metric [12], the choice of $v = 1$ is preferred since $v > 1$ yields distance functions that do not satisfy the triangle inequality property of a metric. Consequently, the distance function of choice utilizes $v = 1$ and $w = 2$, i.e., an ordinary Euclidean distance function.

Let the average distance function between $(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_N)$ and $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$ be given by

$$\delta = \frac{1}{N} \sum_{i=1}^{N} \Delta\big(\tilde{\mathbf{y}}_i, \mathbf{y}_i\big) . \tag{4.1}$$

As noted previously, a distance function with $v > 1$ is not a metric function. If the distance function associated with LAD regression is squared (i.e., $v = 2$), then the estimated parameters that minimize δ yield an OLS regression model.

The criterion for fitting multivariate regression models based on δ is the chance-corrected measure of agreement between the observed and predicted response measurement values given by

$$\Re = 1 - \frac{\delta}{\mu_\delta} , \tag{4.2}$$

where μ_δ is the expected value of δ over the $N!$ possible pairings under the null hypothesis. An efficient computational expression for obtaining μ_δ that involves a sum of N^2 rather than $N!$ terms is given by

$$\mu_\delta = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \Delta\big(\tilde{\mathbf{y}}_i, \mathbf{y}_j\big) . \tag{4.3}$$

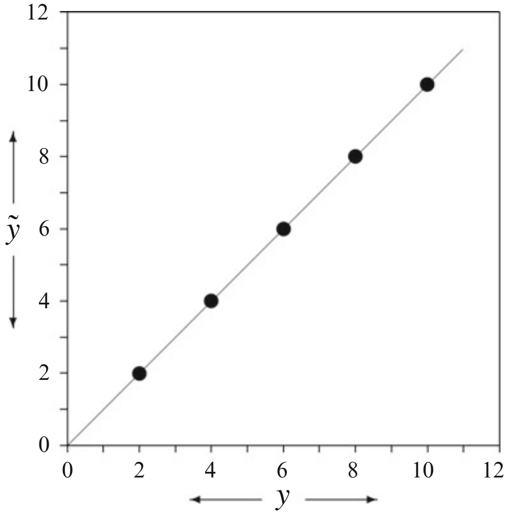

Fig. 4.1 Graphic depicting a regression line with perfect agreement between y and $\tilde{y}$ with intercept equal to 0.00 and slope equal to +1.00

4.1.1 Linear Regression and Agreement

A simple interpretation of $\Re$ can be described for $r = s = 1$ since the same interpretation holds for any r and s. In the case involving perfect agreement, $\tilde{y}_i = y_i$ for $i = 1, \ldots, N$, $\delta = 0.00$, and $\Re = 1.00$. This implies that the functional relationship between $\tilde{y}$ and y can be described by a straight line that passes through the origin with a slope of 45°, as depicted in Fig. 4.1 with $N = 5$ bivariate $(y, \tilde{y})$ values: $(2, 2)$, $(4, 4)$, $(6, 6)$, $(8, 8)$, and $(10, 10)$. For the $N = 5$ data points depicted in Fig. 4.1, the intercept is $\tilde{\beta}_0 = 0.00$, the unstandardized slope is $\tilde{\beta}_1 = +1.00$, the squared Pearson product-moment correlation coefficient is $r^2_{y\tilde{y}} = +1.00$, and the agreement percentage is also 1.00, i.e., all five of the y and $\tilde{y}$ paired values agree.

In this context, the squared Pearson product-moment correlation coefficient, $r^2_{y\tilde{y}}$, has also been used as a measure of agreement. However, $r^2_{y\tilde{y}} = +1.00$ implies a linear relationship between y and $\tilde{y}$, where both the intercept and slope are arbitrary. While perfect agreement is described by $\Re = +1.00$, $r^2_{y\tilde{y}} = +1.00$ describes a linear relationship that may or may not reflect perfect agreement as depicted in Fig. 4.2 with $N = 5$ $(y, \tilde{y})$ values: $(2, 4)$, $(4, 5)$, $(6, 6)$, $(8, 7)$, and $(10, 8)$. For the $N = 5$ bivariate data points depicted in Fig. 4.2, the intercept is $\tilde{\beta}_0 = +3.00$, the unstandardized slope is $\tilde{\beta}_1 = +0.50$, the squared Pearson product-moment correlation coefficient is $r^2_{y\tilde{y}} = +1.00$, and the agreement percentage is 0.20, i.e., only one $(6, 6)$ of the $N = 5$ y and $\tilde{y}$ paired values agree. Comparisons of $\Re$ with other measures of agreement and the advantages of $\Re$ relative to the other agreement measures were detailed in a 1996 article by Watterson [416].

While the agreement measure $\Re$ provides a description of the functional relationship between $(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_N)$ and $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$, it does not indicate how extreme an observed value of $\Re$, say $\Re_o$, is relative to the $N!$ possible values of $\Re$ under the null

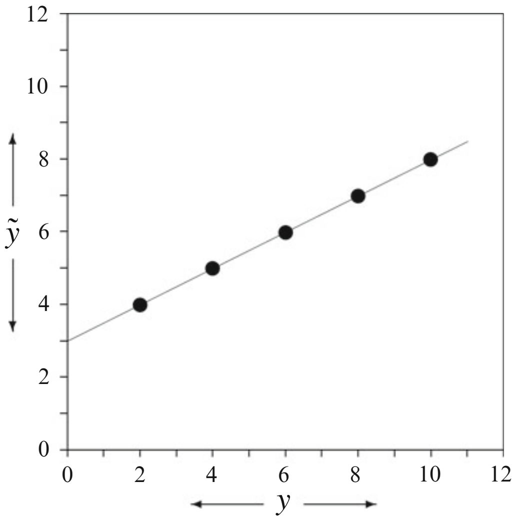

Fig. 4.2 Graphic depicting a regression line with perfect correlation between y and $\tilde{y}$ with intercept equal to +3.00 and slope equal to +0.50

hypothesis. Since μ_δ is invariant under the null hypothesis and the observed value of δ is given by

$$\delta_{\rm o} = \mu_\delta(1 - \Re_{\rm o}) \; ,$$

the exact probability value for $\Re_{\rm o}$ is given by

$$P\big(\Re \geq \Re_{\rm o}|H_0\big) = P\big(\delta \leq \delta_{\rm o}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{\rm o}}{M} \; ,$$

where $M = N!$. Because an exact probability value requires generating $N!$ arrangements of the observed data, calculation of an exact value is prohibitive even for small values of N, e.g., $M = N! = 15! = 1{,}307{,}674{,}368{,}000$.

When M is very large, an approximate probability value for δ may be obtained from a resampling permutation procedure. Let L denote a random sample of all possible arrangements of the observed data, where L is typically a large number, e.g., $L = 1{,}000{,}000$. Then, an approximate resampling probability value is given by

$$P\big(\Re \geq \Re_{\rm o}|H_0\big) = P\big(\delta \leq \delta_{\rm o}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{\rm o}}{L} \; .$$

Also, when M is very large and P is exceedingly small, a resampling-approximation permutation procedure based on fitting the first three exact moments of the discrete permutation distribution to a Pearson type III distribution provides approximate probability values, as detailed in Chap. 1, Sect. 1.2.2; see also references [284] and [300].

4.2 Example LAD Regression Analyses

In this section, example analyses illustrate the permutation approach to typical multiple regression problems. The first example analyzes a small set of multivariate response measurement scores using LAD regression and generates a resampling permutation probability value; the second example analyzes the same small set of multivariate response measurement scores using OLS regression and also generates a resampling permutation probability value; the third example analyzes the same set of multivariate response measurement scores using OLS regression, but provides a conventional approximate probability value based on Snedecor's F distribution.

4.2.1 Example Analysis 1

Consider the multiple regression data listed in Fig. 4.3 where $s = 2$ observed response measurement scores have been obtained for each of $N = 12$ objects, $y_1, \ldots, y_N$ denotes the observed response measurement scores for the N objects, and $\mathbf{x}'_i = (x_{1i}, \ldots, x_{2i})$ is the row vector of $s = 2$ predictor variables for the ith of N objects. Because there are $M = 12! = 479{,}001{,}600$ possible, equally-likely arrangements of the $N = 12$ multivariate response measurement scores in Fig. 4.3, an exact permutation approach is impractical and a resampling procedure is mandated.

A LAD regression analysis of the multivariate response measurement scores listed in Fig. 4.3 yields estimated regression coefficients of

$$\tilde{\beta}_0 = +3.8571 \,, \quad \tilde{\beta}_1 = +0.4286 \,, \quad \text{and} \quad \tilde{\beta}_2 = +0.1429 \,.$$ [1]

Fig. 4.3 Example data with $s = 2$ independent variables on $N = 12$ objects

	Variable		
Object	x_1	x_2	y
1	11	22	11
2	11	24	12
3	11	26	13
4	11	26	15
5	12	28	13
6	12	26	11
7	13	22	15
8	13	22	10
9	14	20	16
10	14	22	13
11	15	20	17
12	15	26	14

[1] For the remainder of this chapter, a tilde over a β ($\tilde{\beta}$) indicates an unstandardized LAD regression coefficient, while a caret over a β ($\hat{\beta}$) indicates an unstandardized OLS regression coefficient.

Fig. 4.4 Observed, predicted, and residual LAD regression values for the example data listed in Fig. 4.3

Object	y_i	$\tilde{y}_i$	e_i
1	11	11.7143	−0.7143
2	12	12.0000	0.0000
3	13	12.2857	+0.7143
4	15	12.2857	+2.7143
5	13	13.0000	0.0000
6	11	12.7143	−1.7143
7	15	12.5714	+2.4286
8	10	12.5714	−2.5714
9	16	12.7143	+3.2857
10	13	13.0000	0.0000
11	17	13.1429	+3.8571
12	14	14.0000	0.0000

Figure 4.4 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 12$. Following Eq. (4.1) on p. 117 with $v = 1$, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.4 is $\delta_o = 1.50$.

If all M possible arrangements of the $N = 12$ observed LAD regression residuals listed in Fig. 4.4 occur with equal chance, the approximate resampling probability value of $\delta_o = 1.50$ calculated on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals is

$$P\big(\delta \leq \delta_o | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{L} = \frac{191{,}128}{1{,}000{,}000} = 0.0191 \ .$$

Following Eq. (4.3) on p. 117, the exact expected value of the $M = 479{,}001{,}600$ δ values is $\mu_\delta = 1.8294$ and, following Eq. (4.2) on p. 117, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_o = 1 - \frac{\delta_o}{\mu_\delta} = 1 - \frac{1.50}{1.8294} = +0.1800 \ ,$$

indicating 18 % agreement between the observed and predicted y values above that expected by chance.

4.2.2 Example Analysis 2

For a second example analysis of the multivariate response measurement scores listed in Fig. 4.3 on p. 120, consider an OLS regression analysis based on a resampling permutation procedure. An OLS regression analysis of the multivariate

Fig. 4.5 Observed, predicted, and residual OLS regression values for the example data listed in Fig. 4.3

Object	y_i	$\hat{y}_i$	e_i
1	11	12.3823	−1.3823
2	12	12.2524	−0.2524
3	13	12.1226	+0.8774
4	15	12.1226	+2.8774
5	13	12.6282	+0.3718
6	11	12.7581	−1.7581
7	15	13.6534	+1.3466
8	10	13.6534	−3.6534
9	16	14.4188	+1.5812
10	13	14.2890	−1.2890
11	17	15.0544	+1.9456
12	14	14.6648	−0.6648

response measurement scores listed in Fig. 4.3 yields estimated regression coefficients of

$$\hat{\beta}_0 = +6.8198 \,, \quad \hat{\beta}_1 = +0.6356 \,, \quad \text{and} \quad \hat{\beta}_2 = -0.0649 \,.$$

Figure 4.5 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 12$.

Following Eq. (4.1) on p. 117 with $v = 2$, the observed value of the MRPP test statistic computed on the OLS regression residuals listed in Fig. 4.5 is $\delta_o = 3.1502$. If all M possible arrangements of the $N = 12$ observed OLS regression residuals listed in Fig. 4.5 occur with equal chance, the approximate resampling probability value of $\delta_o = 3.1502$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals is

$$P\big(\delta \leq \delta_o | H_0\big) = \frac{\text{number of } \delta_o \text{ values} \leq \delta_o}{L} = \frac{96{,}104}{1{,}000{,}000} = 0.0961 \,.$$

For comparison, the approximate resampling probability value based on LAD regression in Example 1 is $P = 0.0191$.

Following Eq. (4.3) on p. 117, the exact expected value of the $M = 479{,}001{,}600$ δ values is $\mu_\delta = 5.2942$ and, following Eq. (4.2) on p. 117, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_o = 1 - \frac{\delta_o}{\mu_\delta} = 1 - \frac{3.1502}{5.2942} = +0.4050 \,,$$

indicating approximately 41 % agreement between the observed and predicted y values above that expected by chance.

4.2.3 Example Analysis 3

Finally, consider a conventional OLS regression analysis of the multivariate response measurement scores listed in Fig. 4.3 on p. 120. An OLS regression analysis yields estimated regression coefficients of

$$\hat{\beta}_0 = +6.8198 \,, \quad \hat{\beta}_1 = +0.6356 \,, \quad \text{and} \quad \hat{\beta}_2 = -0.0649 \,,$$

the regression residuals are listed in Fig. 4.5, and the observed squared multiple correlation coefficient is $R^2_{y.x_1,x_2} = 0.2539$. $R^2_{y.x_1,x_2}$ may be transformed into an F-ratio by

$$F = \frac{(N-s-1)R^2_{y.x_1,x_2}}{s(1-R^2_{y.x_1,x_2})} = \frac{(12-2-1)(0.2539)}{(2)(1-0.2539)} = 1.5313 \,.$$

Assuming independence, normality, and homogeneity of variance, F is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = s = 2$ and $\nu_2 = N - s - 1 = 12 - 2 - 1 = 9$ degrees of freedom. Under the null hypothesis, the observed value of $F_\text{o} = 1.5313$ yields an approximate probability value of $P = 0.2677$.

Note that the asymptotic probability value based on OLS regression in Example 3 is $P = 0.2677$, while a resampling analysis of the same data in Example 2 yielded a probability value, again based on OLS regression, of $P = 0.0961$, a marked difference. Moreover, a LAD regression analysis of the same data in Example 1 yielded an approximate resampling probability value of $P = 0.0191$, once again demonstrating the different results possible with $v = 1$ and $v = 2$, both with and without a permutation analysis.

4.3 LAD Regression and Analysis of Variance Designs

It is well known that experimental designs that would ordinarily be analyzed by some form of analysis of variance can also be analyzed by OLS multiple regression using either dummy- or effect-coding schemes. The same is true of LAD regression. In this section a variety of analysis-of-variance designs are analyzed using MRPP, LAD regression, and either dummy or effect coding of treatment groups; included are one-way randomized, one-way randomized with a covariate, one-way randomized-block, two-way randomized-block, two-way factorial, Latin square, split-plot, and two-factor nested analysis-of-variance designs.

Fig. 4.6 Example data for a one-way randomized design with $g = 3$ treatment groups and univariate response measurement scores on $N = 26$ objects

Treatment		
1	2	3
15	17	6
18	22	9
12	15	12
12	12	11
9	20	11
10	13	8
12	15	13
20	20	30
	21	7

4.3.1 One-Way Randomized Design

Consider a one-way completely randomized experimental design with fixed effects in which $N = 26$ objects have been randomly assigned to one of $g = 3$ treatment groups with $n_1 = 8$ and $n_2 = n_3 = 9$. The design and data are adapted from Stevens [387, p. 70] and are given in Fig. 4.6.

For a one-way randomized experimental design, the appropriate regression model is given by

$$y_i = \sum_{j=1}^{m} x_{ij}\beta_j + e_i \ ,$$

where y_i denotes the ith of N responses possibly affected by a treatment; x_{ij} is the jth of m covariates associated with the ith response, where $x_{i1} = 1$ if the model includes an intercept; β_j denotes the jth of m regression parameters; and e_i designates the error associated with the ith of N responses. If the estimates of $\beta_1, \ldots, \beta_m$ that minimize

$$\sum_{i=1}^{N} |e_i|$$

are denoted by $\tilde{\beta}_1, \ldots, \tilde{\beta}_m$, then the N residuals of the LAD regression model are given by $e_i = y_i - \tilde{y}_i$ for $i = 1, \ldots, N$, where the predicted value of y_i is given by

$$\tilde{y}_i = \sum_{j=1}^{m} x_{ij}\tilde{\beta}_j \ , \qquad i = 1, \ldots, N \ .$$

In contrast, OLS regression estimators of $\beta_1, \ldots, \beta_m$ minimize

$$\sum_{i=1}^{N} e_i^2 \,,$$

the N residuals of the OLS regression model are given by $e_i = y_i - \hat{y}_i$ for $i = 1, \ldots, N$, and the predicted value of y_i is given by

$$\hat{y}_i = \sum_{j=1}^{m} x_{ij}\hat{\beta}_j \,, \qquad i = 1, \ldots, N \,.$$

If the N regression residuals are partitioned into g disjoint treatment groups of sizes $n_1, \ldots, n_g$, where $n_i \geq 2$ for $i = 1, \ldots, g$ and

$$N = \sum_{i=1}^{g} n_i \,,$$

then the permutation test depends on test statistic

$$\delta = \sum_{i=1}^{g} C_i \xi_i \,, \tag{4.4}$$

where

$$C_i = \frac{n_i}{N} \,, \qquad i = 1, \ldots, g \,,$$

is a positive weight for the ith of g treatment groups that minimizes the variability of δ,

$$\sum_{i=1}^{g} C_i = 1 \,,$$

and ξ_i is the average pairwise Euclidean difference among the n_i residuals in the ith of g treatment groups defined by

$$\xi_i = \binom{n_i}{2}^{-1} \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} \left[\left(e_j - e_k \right)^2 \right]^{v/2} \Psi_{ji}\, \Psi_{ki} \,, \tag{4.5}$$

where $v = 1$ for LAD regression and

$$\Psi_{ji} = \begin{cases} 1 & \text{if } e_i \text{ is in the } i\text{th treatment group ,} \\ 0 & \text{otherwise .} \end{cases}$$

The null hypothesis specifies that each of the

$$M = \frac{N!}{\prod_{i=1}^{g} n_i!}$$

allocations of the N residuals to the g treatment groups is equally likely with n_i, $i = 1, \ldots, g$, residuals preserved for each arrangement of the observed data. The exact probability value of an observed value of δ, δ_o, is given by

$$P\big(\delta \leq \delta_o | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{M} \; .$$

As previously, when M is large, an approximate probability value of δ may be obtained from a resampling procedure, where

$$P\big(\delta \leq \delta_o | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{L}$$

and L denotes the number of resampled test statistic values. Typically, L is set to a large number to ensure accuracy, e.g., $L = 1{,}000{,}000$. When M is very large and P is exceedingly small, a resampling-approximation permutation procedure may produce no δ values equal to or less than δ_o, even with $L = 1{,}000{,}000$, yielding an approximate resampling probability value of $P = 0.00$. In such cases, moment-approximation permutation procedures based on fitting the first three exact moments of the discrete permutation distribution to a Pearson type III distribution provide approximate probability values, as detailed in Chap. 1, Sect. 1.2.2 [284, 300].

An index of the effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is given by the chance-corrected measure

$$\Re = 1 - \frac{\delta}{\mu_\delta} \; , \tag{4.6}$$

where μ_δ is the arithmetic average of the δ values calculated on all M equally-likely arrangements of the observed response measurements, i.e.,

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i \; . \tag{4.7}$$

Fig. 4.7 Design matrix and data for a one-way randomized design with $g = 3$ treatment groups and univariate response measurement scores on $N = 26$ objects

Matrix	Score	Matrix	Score	Matrix	Score
1	15	1	17	1	16
1	18	1	22	1	9
1	12	1	15	1	12
1	12	1	12	1	11
1	9	1	20	1	11
1	10	1	14	1	8
1	12	1	15	1	13
1	20	1	20	1	30
		1	21	1	7

A design matrix of dummy codes for an MRPP regression analysis of the $N = 26$ response measurement scores in Fig. 4.6 is given in Fig. 4.7 where the first columns of 1 values provide for an intercept. The second columns contain the $N = 26$ univariate response measurement scores listed according to the original random assignment of the $N = 26$ objects to the $g = 3$ treatment groups with the first $n_1 = 8$ scores, the next $n_2 = 9$ scores, and the last $n_3 = 9$ scores associated with the first, second, and third treatment groups, respectively.

Because the purpose of the analysis is to test for possible differences among the $g = 3$ treatment groups, a reduced regression model is constructed without a variate for treatments. Therefore, for a single-factor experiment the design matrix for the reduced model is composed solely of a code for the intercept. The MRPP regression analysis examines the $N = 26$ regression residuals for possible differences among the $g = 3$ treatment levels; consequently, no dummy codes for treatments are included in Fig. 4.7 as this information is implicit in the ordering of the $g = 3$ treatment groups in the three columns labeled "Score" with $n_1 = 8$ and $n_2 = n_3 = 9$ values.

An exact permutation solution is impractical for the univariate response measurements listed in Fig. 4.7 since there are

$$M = \frac{N!}{\displaystyle\prod_{i=1}^{g} n_i!} = \frac{26!}{8!\ 9!\ 9!} = 75{,}957{,}810{,}500$$

possible, equally-likely arrangements of the $N = 26$ univariate response measurement scores; consequently, a resampling procedure is the default in this case.

LAD Regression Analysis

An MRPP resampling analysis of the LAD regression residuals calculated on the univariate response measurement scores listed in Fig. 4.7 yields an estimated LAD regression coefficient of $\tilde{\beta}_0 = +12.00$. Figure 4.8 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 26$.

Fig. 4.8 Observed, predicted, and residual LAD regression values for the example one-way randomized data listed in Fig. 4.7

Object	y_i	$\tilde{y}_i$	e_i
1	15	12.00	+3.00
2	18	12.00	+6.00
3	12	12.00	0.00
4	12	12.00	0.00
5	9	12.00	−3.00
6	10	12.00	−2.00
7	12	12.00	0.00
8	20	12.00	+8.00
9	17	12.00	+5.00
10	22	12.00	+10.00
11	15	12.00	+3.00
12	12	12.00	0.00
13	20	12.00	+8.00
14	14	12.00	+2.00
15	15	12.00	+3.00
16	20	12.00	+8.00
17	21	12.00	+9.00
18	6	12.00	−6.00
19	9	12.00	−3.00
20	12	12.00	0.00
21	11	12.00	−1.00
22	11	12.00	−1.00
23	8	12.00	−4.00
24	13	12.00	+1.00
25	30	12.00	+18.00
26	7	12.00	−5.00

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 26$ LAD regression residuals listed in Fig. 4.8 yield $g = 3$ average distance-function values of

$$\xi_1 = 4.50\ , \quad \xi_2 = 4.2222\ , \quad \text{and} \quad \xi_3 = 6.8889\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.8 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_i}{n}\ , \qquad i = 1, 2, 3\ ,$$

is

$$\delta_o = \sum_{i=1}^{g} C_i \xi_i = \frac{1}{26}\big[(8)(4.50) + (9)(4.2222) + (9)(6.8889)\big] = 5.2308\ .$$

If all M possible arrangements of the $N = 26$ observed LAD regression residuals listed in Fig. 4.8 occur with equal chance, the approximate resampling probability value of $\delta_o = 5.2308$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_1 = 8$ and $n_2 = n_3 = 9$ residuals preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{L} = \frac{12{,}062}{1{,}000{,}000} = 0.0121 \; .$$

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 6.1262$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_o = 1 - \frac{\delta_o}{\mu_\delta} = 1 - \frac{5.2308}{6.1262} = +0.1462 \; ,$$

indicating approximately 15 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP resampling analysis of OLS regression residuals calculated on the $N = 26$ univariate response measurement scores listed in Fig. 4.7 on p. 127. The MRPP regression analysis yields an estimated OLS regression coefficient of $\hat{\beta}_0 = +14.2692$. Figure 4.9 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 26$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 26$ OLS regression residuals listed in Fig. 4.9 yield $g = 3$ average distance-function values of

$$\xi_1 = 29.7143 \; , \quad \xi_2 = 25.00 \; , \quad \text{and} \quad \xi_3 = 103.2222 \; .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.9 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g} \; , \qquad i = 1, 2, 3 \; ,$$

is

$$\delta_o = \sum_{i=1}^{g} C_i \xi_i = \frac{1}{26 - 3} \big[(8-1)(29.7143) + (9-1)(25.00) + (9-1)(103.2222) \big] = 53.6425 \; .$$

Fig. 4.9 Observed, predicted, and residual OLS regression values for the example one-way randomized data listed in Fig. 4.7

Object	y_i	$\hat{y}_i$	e_i
1	15	14.2692	+0.7308
2	18	14.2692	+3.7308
3	12	14.2692	−2.2692
4	12	14.2692	−2.2692
5	9	14.2692	−5.2692
6	10	14.2692	−4.2692
7	12	14.2692	−2.2692
8	20	14.2692	+5.7308
9	17	14.2692	+2.7308
10	22	14.2692	+7.7308
11	15	14.2692	+0.7308
12	12	14.2692	−2.2692
13	20	14.2692	+5.7308
14	14	14.2692	−0.2692
15	15	14.2692	+0.7308
16	20	14.2692	+5.7308
17	21	14.2692	+6.7308
18	6	14.2692	−8.2692
19	9	14.2692	−5.2692
20	12	14.2692	−2.2692
21	11	14.2692	−3.2692
22	11	14.2692	−3.2692
23	8	14.2692	−6.2692
24	13	14.2692	−1.2692
25	30	14.2692	+15.7308
26	7	14.2692	−7.2692

If all M possible arrangements of the $N = 26$ observed OLS regression residuals listed in Fig. 4.9 occur with equal chance, the approximate resampling probability value of $\delta_o = 53.6425$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_1 = 8$ and $n_2 = n_3 = 9$ residuals preserved for each arrangement is

$$P(\delta \leq \delta_o | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{L} = \frac{91{,}842}{1{,}000{,}000} = 0.0918 \; .$$

For comparison, the approximate resampling probability value based LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_i/N$ for $i = 1, 2, 3$ is $P = 0.0121$.

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 75{,}957{,}810{,}500$ δ values is $\mu_\delta = 60.5692$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_o = 1 - \frac{\delta_o}{\mu_\delta} = 1 - \frac{53.6425}{60.5692} = +0.1144 \; ,$$

indicating approximately 11 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional fixed-effects one-way analysis of variance calculated on the $N = 26$ univariate response measurement scores listed in Fig. 4.6 on p. 124 yields an observed F-ratio of $F_o = 2.6141$. Assuming independence, normality, and homogeneity of variance, F is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = g - 1 = 3 - 1 = 2$ and $\nu_2 = N - g = 26 - 3 = 23$ degrees of freedom. Under the null hypothesis, the observed value of $F_o = 2.6141$ yields an approximate probability value of $P = 0.0948$, which is similar to that produced by the MRPP resampling analysis of the OLS regression residuals.

4.3.2 One-Way Randomized Design with a Covariate

A covariate experimental design permits the testing of differences among the treatment groups after the effect of the covariate has been removed from the analysis. Consider a one-way completely randomized design with a covariate in which $N = 47$ objects are randomly assigned to one of $g = 5$ treatment groups. The experimental data are listed in Table 4.1 and are adapted from a 1984 study by Conti and Musty [78].

A design matrix of dummy codes for analyzing treatments is given in Fig. 4.10, where the first column of 1 values provides for an intercept, the second column contains the covariate (Pre-test) values, and the third column contains the (Post-test) scores listed according to the original random assignment of the $N = 47$ objects to

Table 4.1 Example data for a one-way randomized design with a covariate, consisting of pre-test (Pre) and post-test (Post) response measurement scores on $N = 47$ randomly assigned objects to $g = 5$ treatment groups

Treatment									
1		2		3		4		5	
Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
4.34	1.30	1.55	0.93	7.18	5.10	6.94	2.29	4.00	2.93
3.50	0.94	10.56	4.44	8.33	4.16	6.10	4.75	4.10	1.11
4.33	2.25	8.39	4.03	4.05	1.54	4.90	3.48	3.62	2.17
2.76	1.05	3.70	1.92	10.78	6.36	3.69	2.76	3.92	2.00
4.62	0.92	2.40	0.67	6.09	3.96	4.76	1.67	2.90	0.84
5.40	1.90	1.83	1.70	7.78	4.51	4.30	1.51	2.90	0.99
3.95	0.32	2.40	0.77	5.08	3.76	2.32	1.07	1.82	0.44
1.55	0.64	7.67	3.53	2.86	1.92	7.35	2.35	4.94	0.84
1.42	0.69	5.79	3.65	6.30	3.84			5.69	2.84
1.90	0.93	9.58	4.22					5.54	2.93

Fig. 4.10 Design matrix and data, consisting of an intercept and pre- and post-test measurement scores for a one-way randomized design with a covariate

Matrix	Pre	Post	Matrix	Pre	Post
1	4.34	1.30	1	6.94	2.29
1	3.50	0.94	1	6.10	4.75
1	4.33	2.25	1	4.90	3.48
1	2.76	1.05	1	3.69	2.76
1	4.62	0.92	1	4.76	1.67
1	5.40	1.90	1	4.30	1.51
1	3.95	0.32	1	2.32	1.07
1	1.55	0.64	1	7.35	2.35
1	1.42	0.69			
1	1.90	0.93	1	4.00	1.44
			1	4.10	1.11
1	1.55	0.93	1	3.62	2.17
1	10.56	4.44	1	3.92	2.00
1	8.39	4.03	1	2.90	0.84
1	3.70	1.92	1	2.90	0.99
1	2.40	0.67	1	1.82	0.44
1	1.83	1.70	1	4.94	0.84
1	2.40	0.77	1	5.69	2.84
1	7.67	3.53	1	5.54	2.93
1	5.79	3.65			
1	9.58	4.22			
1	7.18	5.10			
1	8.33	4.16			
1	4.05	1.54			
1	10.78	6.36			
1	6.09	3.96			
1	7.78	4.51			
1	5.08	3.76			
1	2.86	1.92			
1	6.30	3.84			

the $g = 5$ treatment groups with the first $n_1 = 10$ scores, the next $n_2 = 10$ scores, the next $n_3 = 9$ scores, the next $n_4 = 8$ scores, and the last $n_5 = 10$ scores associated with the $g = 5$ treatment groups, respectively.

The MRPP regression analysis examines the $N = 47$ regression residuals for possible differences among the $g = 5$ treatment levels; consequently, no dummy codes for treatments are included in Fig. 4.10 as this information is implicit in the ordering of the $g = 5$ treatment groups in the two paired columns labeled "Pre" and "Post."

Because there are

$$M = \frac{N!}{\prod_{i=1}^{g} n_i!} = \frac{47!}{10!\ 10!\ 9!\ 8!\ 10!} = 369{,}908{,}998{,}147{,}203{,}213{,}613{,}129{,}815{,}600$$

possible, equally-likely arrangements of the $N = 47$ univariate response measurement scores listed in Table 4.1, an exact permutation approach is not possible and a resampling analysis is mandated.

LAD Regression Analysis

An MRPP resampling analysis of the LAD regression residuals calculated on the $N = 47$ response measurement scores listed in Fig. 4.10 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = -0.1282 \quad \text{and} \quad \tilde{\beta}_1 = +0.4956 \,.$$

Table 4.2 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 47$.

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals $v = 1$, the LAD regression residuals listed in Table 4.2 yield

Table 4.2 Observed, predicted, and residual LAD regression values for the example covariate data listed in Fig. 4.10

Object	y_i	$\tilde{y}_i$	e_i	Object	y_i	$\tilde{y}_i$	e_i
1	1.30	2.0228	−0.7228	25	3.96	2.8901	+1.0699
2	0.94	1.6064	−0.6664	26	4.51	3.7277	+0.7823
3	2.25	2.0178	+0.2322	27	3.76	2.3895	+1.3705
4	1.05	1.2397	−0.1897	28	1.92	1.2893	+0.6307
5	0.92	2.1615	−1.2415	29	3.84	2.9942	+0.8458
6	1.90	2.5481	−0.6481	30	2.29	3.3114	−1.0214
7	0.32	1.8295	−1.5095	31	4.75	2.8950	+1.8550
8	0.64	0.6400	0.0000	32	3.48	2.3003	+1.1797
9	0.69	0.5756	+0.1144	33	2.76	1.7006	+1.0594
10	0.93	0.8135	+0.1165	34	1.67	2.2309	−0.5609
11	0.93	0.6400	+0.2900	35	1.51	2.0029	−0.4929
12	4.44	5.1055	−0.6655	36	1.07	1.0216	+0.0484
13	4.03	4.0300	0.0000	37	2.35	3.5146	−1.1646
14	1.92	1.7056	+0.2144	38	1.44	1.8543	−0.4143
15	0.67	1.0613	−0.3913	39	1.11	1.9038	−0.7938
16	1.70	0.7788	+0.9212	40	2.17	1.6659	+0.5041
17	0.77	1.0613	−0.2913	41	2.00	1.8146	+0.1854
18	3.53	3.6732	−0.1432	42	0.84	1.3091	−0.4691
19	3.65	2.7414	+0.9086	43	0.99	1.3091	−0.3191
20	4.22	4.6198	−0.3998	44	0.44	0.7738	−0.3338
21	5.10	3.4303	+1.6697	45	0.84	2.3201	−1.4801
22	4.16	4.0003	+0.1597	46	2.84	2.6918	+0.1482
23	1.54	1.8790	−0.3390	47	2.93	2.6175	+0.3125
24	6.36	5.2145	+1.1455				

$g = 5$ average distance-function values of

$$\xi_1 = 0.7072\ ,\quad \xi_2 = 0.6335\ ,\quad \xi_3 = 0.7213\ ,\quad \xi_4 = 1.3409\ ,\ \text{and}\ \ \xi_5 = 0.6795\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Table 4.2 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_i}{N}\ , \qquad i = 1, \ldots, 5\ ,$$

is

$$\delta_{\text{o}} = \sum_{i=1}^{g} C_i \xi_i = \frac{1}{47}\big[(10)(0.7072) + (10)(0.6335) + (9)(0.7213) + (8)(1.3409) + (10)(0.6795)\big] = 0.7962\ .$$

If all M possible arrangements of the observed LAD regression residuals listed in Table 4.2 occur with equal chance, the approximate resampling probability value of $\delta_{\text{o}} = 0.7962$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_1 = n_2 = n_5 = 10$, $n_3 = 9$, and $n_4 = 8$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_{\text{o}}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{\text{o}}}{L} = \frac{4{,}095}{1{,}000{,}000} = 0.0041\ .$$

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 0.9178$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{\text{o}} = 1 - \frac{\delta_{\text{o}}}{\mu_\delta} = 1 - \frac{0.7962}{0.9178} = +0.1326\ ,$$

indicating approximately 13 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP resampling analysis of the OLS regression residuals calculated on the $N = 47$ univariate response measurement scores listed in Fig. 4.10 on p. 132. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = -0.2667 \quad \text{and} \quad \hat{\beta}_1 = +0.5311\ .$$

Table 4.3 Observed, predicted, and residual OLS regression values for the example covariate data listed in Fig. 4.10

Object	y_i	$\hat{y}_i$	e_i	Object	y_i	$\hat{y}_i$	e_i
1	1.30	2.0383	−0.7383	25	3.96	2.9677	+0.9923
2	0.94	1.5922	−0.6522	26	4.51	3.8652	+0.6448
3	2.25	2.0330	+0.2170	27	3.76	2.4313	+1.3287
4	1.05	1.1991	−0.1491	28	1.92	1.2523	+0.6677
5	0.92	2.1870	−1.2670	29	3.84	3.0792	+0.7608
6	1.90	2.6012	−0.7012	30	2.29	3.4191	−1.1291
7	0.32	1.8311	−1.5111	31	4.75	2.9730	+1.7770
8	0.64	0.5565	+0.0835	32	3.48	2.3357	+1.1443
9	0.69	0.4875	+0.2025	33	2.76	1.6931	+1.0669
10	0.93	0.7424	+0.1876	34	1.67	2.2613	−0.5913
11	0.93	0.5565	+0.3735	35	1.51	2.0170	−0.5070
12	4.44	5.3417	−0.9017	36	1.07	0.9655	+0.1045
13	4.03	4.1892	−0.1592	37	2.35	3.6369	−1.2869
14	1.92	1.6984	+0.2216	38	1.44	1.8577	−0.4177
15	0.67	1.0080	−0.3380	39	1.11	1.9108	−0.8008
16	1.70	0.7052	+0.9948	40	2.17	1.6559	+0.5141
17	0.77	1.0080	−0.2380	41	2.00	1.8152	+0.1848
18	3.53	3.8068	−0.2768	42	0.84	1.2735	−0.4335
19	3.65	2.8084	+0.8416	43	0.99	1.2735	−0.2835
20	4.22	4.8212	−0.6012	44	0.44	0.6999	−0.2599
21	5.10	3.5466	+1.5534	45	0.84	2.3569	−1.5169
22	4.16	4.1573	+0.0027	46	2.84	2.7553	+0.0847
23	1.54	1.8843	−0.3443	47	2.93	2.6756	+0.2544
24	6.36	5.4585	+0.9015				

Table 4.3 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 47$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the OLS regression residuals listed in Table 4.3 yield $g = 5$ average distance-function values of

$$\xi_1 = 0.8067\,, \quad \xi_2 = 0.7407\,, \quad \xi_3 = 0.7073\,, \quad \xi_4 = 2.6035\,, \quad \text{and} \quad \xi_5 = 0.6906\,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Table 4.3 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_i - 1}{N - g}\,, \qquad i = 1, \ldots, 5\,,$$

is

$$\delta_{\rm o} = \sum_{i=1}^{g} C_i \xi_i = \frac{1}{47-5}\big[(10-1)(0.8067) + (10-1)(0.7407) + (9-1)(0.7073) + (8-1)(2.6035) + (10-1)(0.6906)\big] = 1.0482\ .$$

If all M possible arrangements of the $N = 47$ observed OLS regression residuals listed in Table 4.3 occur with equal chance, the approximate resampling probability value of $\delta_{\rm o} = 1.0482$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_1 = n_2 = n_5 = 10$, $n_3 = 9$, and $n_4 = 8$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_{\rm o}|H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_{\rm o}}{L} = \frac{15{,}301}{1{,}000{,}000} = 0.0153\ .$$

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_i/N$ for $i = 1, \ldots, 5$ is $P = 0.0041$.

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 1.2761$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{\rm o} = 1 - \frac{\delta_{\rm o}}{\mu_\delta} = 1 - \frac{1.0482}{1.2761} = +0.1785\ ,$$

indicating approximately 18 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional fixed-effects one-way analysis of covariance calculated on the $N = 47$ univariate response measurement scores listed in Table 4.1 on p. 131 yields an observed F-ratio of $F_{\rm o} = 4.6978$. Assuming independence, normality, and homogeneity of variance, F is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = g - 1 = 5 - 1 = 4$ and $\nu_2 = N - g - 1 = 47 - 5 - 1 = 41$ degrees of freedom. Under the null hypothesis, the observed value of $F_{\rm o} = 4.6978$ yields an approximate probability value of $P = 0.0033$.

4.3.3 One-Way Randomized-Block Design

One-way randomized-block designs are common in experimental research and have long been valuable statistical tools in such fields as agriculture and genetics. E.J.G. Pitman, for example, developed a permutation approach for one-way randomized-block designs in 1938 [342]. With modern developments in embryo transplants and cloning where subjects can be genetically matched on a large number of important characteristics, randomized-block designs have become very practical and efficient.[2]

Consider a one-way randomized-block design where $b = 6$ objects (blocks) are evaluated over $a = 3$ treatments with $r = 1$ response measurement. The design and data are adapted from a study by Anderson, Sweeney, and Williams [9, p. 471] and are given in Fig. 4.11.

A design matrix of dummy codes for an MRPP regression analysis is given in Fig. 4.12, where the first column of 1 values provides for an intercept, the next five columns contain dummy codes for the $b = 6$ blocks, and the last column contains the univariate response measurement scores listed according to the original random assignment of the $N = 18$ objects to the $a = 3$ treatment levels of Factor A with the first $n_{A_1} = 6$ objects, the next $n_{A_2} = 6$ objects, and the last $n_{A_3} = 6$ objects

Fig. 4.11 Example data for a one-way randomized-block design with $b = 6$ blocks, $a = 3$ treatments, and $r = 1$ response measurement

	Factor A		
Object	A_1	A_2	A_3
1	15	15	18
2	14	14	14
3	10	11	15
4	13	12	17
5	16	13	16
6	13	13	13

Fig. 4.12 Design matrix and data for a one-way randomized-block design with $b = 6$ blocks, $a = 3$ treatments, and $r = 1$ response measurement

Matrix						Score
1	0	0	0	0	0	15
1	1	0	0	0	0	14
1	0	1	0	0	0	10
1	0	0	1	0	0	13
1	0	0	0	1	0	16
1	0	0	0	0	1	13
1	0	0	0	0	0	15
1	1	0	0	0	0	14
1	0	1	0	0	0	11
1	0	0	1	0	0	12
1	0	0	0	1	0	13
1	0	0	0	0	1	13
1	0	0	0	0	0	18
1	1	0	0	0	0	14
1	0	1	0	0	0	15
1	0	0	1	0	0	17
1	0	0	0	1	0	16
1	0	0	0	0	1	13

[2] All the biologically inherited information is not carried in the genes of a cell's nucleus. A small number of genes are carried by intra-cellular bodies, the mitochondria. Thus, the result of cloning is not, strictly speaking, a perfect genetic clone of the donor organism.

associated with treatment levels A_1, A_2, and A_3, respectively. The MRPP regression analysis examines the $N = 18$ regression residuals for possible differences in the $a = 3$ treatment levels; consequently, there are no dummy codes for treatments in Fig. 4.12 as this information is implicit in the ordering of the $a = 3$ treatment levels of Factor A in the last column.

Because there are

$$M = \frac{N!}{\prod_{i=1}^{a} n_{A_i}!} = \frac{18!}{(6!)^3} = 17{,}153{,}136$$

possible, equally-likely arrangements of the $N = 18$ univariate response measurement scores listed in Fig. 4.11, an exact permutation approach is not practical.

LAD Regression Analysis

An MRPP resampling analysis of the LAD regression residuals calculated on the univariate response measurement scores listed in Fig. 4.12 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +15.00\,, \quad \tilde{\beta}_1 = -1.00\,, \quad \tilde{\beta}_2 = -4.00\,, \quad \tilde{\beta}_3 = -2.00\,,$$
$$\tilde{\beta}_4 = +1.00\,, \quad \text{and} \quad \tilde{\beta}_5 = -2.00$$

for Factor A. Figure 4.13 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 18$.

Fig. 4.13 Observed, predicted, and residual LAD regression values for the example randomized-block data listed in Fig. 4.12

Object	y_i	$\tilde{y}_i$	e_i
1	15	15.00	0.00
2	14	14.00	0.00
3	10	11.00	−1.00
4	13	13.00	0.00
5	16	16.00	0.00
6	13	13.00	0.00
7	15	15.00	0.00
8	14	14.00	0.00
9	11	11.00	0.00
10	12	13.00	−1.00
11	13	16.00	−3.00
12	13	13.00	0.00
13	18	15.00	+3.00
14	14	14.00	0.00
15	15	11.00	+4.00
16	17	13.00	+4.00
17	16	16.00	0.00
18	13	13.00	0.00

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 18$ LAD regression residuals listed in Fig. 4.13 yield $a = 3$ average distance-function values of

$$\xi_{A_1} = 0.3333 \,, \quad \xi_{A_2} = 1.20 \,, \quad \text{and} \quad \xi_{A_3} = 2.3333 \,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.13 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{A_i}}{N} \,, \qquad i = 1, 2, 3 \,,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{6}{18}\big(0.3333 + 1.20 + 2.3333\big) = 1.2889 \,.$$

If all M possible arrangements of the $N = 18$ observed LAD regression residuals listed in Fig. 4.13 occur with equal chance, the approximate resampling probability value of $\delta_A = 1.2889$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 6$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_A}{L} = \frac{56{,}035}{1{,}000{,}000} = 0.0560 \,.$$

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 17{,}153{,}136$ δ values is $\mu_\delta = 1.6078$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{1.2889}{1.6078} = +0.1984 \,,$$

indicating approximately 20 % agreement between the observed and predicted y values above that expected by chance.

An Exact Test

Although an exact permutation analysis of the $N = 18$ LAD regression residuals listed in Fig. 4.13 is impractical, it is not impossible. In fact, exact permutation methods are oftentimes more efficient than resampling permutation methods because the $L = 1{,}000{,}000$ calls to a pseudorandom number generator, necessary for a resampling test, are not required by an exact test.

Following Eq. (4.5) on p. 125, an exact permutation analysis of the $N = 18$ LAD regression residuals listed in Fig. 4.13 yields $a = 3$ average distance-function values of

$$\xi_{A_1} = 0.3333 \;, \quad \xi_{A_2} = 1.20 \;, \quad \text{and} \quad \xi_{A_3} = 2.3333 \;.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic based on $v = 1$ and treatment-group weights

$$C_i = \frac{n_{A_i}}{N} \;, \qquad i = 1, 2, 3 \;,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{6}{18}\big(0.3333 + 1.20 + 2.3333\big) = 1.2889 \;.$$

If all arrangements of the $N = 18$ observed LAD regression residuals listed in Fig. 4.13 occur with equal chance, the exact probability value of $\delta_A = 1.2889$ computed on the $M = 17{,}153{,}136$ possible arrangements of the observed LAD regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 6$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_A}{M} = \frac{961{,}884}{17{,}153{,}136} = 0.0561 \;.$$

For comparison, the resampling probability value computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals listed in Fig. 4.13 is $P = 0.0560$.

OLS Regression Analysis

For comparison, consider an MRPP resampling analysis of OLS regression residuals calculated on the $N = 18$ univariate response measurement scores listed in Fig. 4.12 on p. 137. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +16.00 \;, \quad \hat{\beta}_1 = -2.00 \;, \quad \hat{\beta}_2 = -4.00 \;, \quad \hat{\beta}_3 = -2.00 \;,$$
$$\hat{\beta}_4 = -1.00 \;, \quad \text{and} \quad \hat{\beta}_5 = -3.00$$

for Factor A. Figure 4.14 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 18$.

Fig. 4.14 Observed, predicted, and residual OLS regression values for the example randomized-block data listed in Fig. 4.12

Object	y_i	$\hat{y}_i$	e_i
1	15	16.00	−1.00
2	14	14.00	0.00
3	10	12.00	−2.00
4	13	14.00	−1.00
5	16	15.00	+1.00
6	13	13.00	0.00
7	15	16.00	−1.00
8	14	14.00	0.00
9	11	12.00	−1.00
10	12	14.00	−2.00
11	13	15.00	−2.00
12	13	13.00	0.00
13	18	16.00	+2.00
14	14	14.00	0.00
15	15	12.00	+3.00
16	17	14.00	+3.00
17	16	15.00	+1.00
18	13	13.00	0.00

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 2$, the $N = 18$ OLS regression residuals listed in Fig. 4.14 yield $a = 3$ average distance-function values of

$$\xi_{A_1} = 2.20 \,, \quad \xi_{A_2} = 1.60 \,, \quad \text{and} \quad \xi_{A_3} = 3.80 \,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.14 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{A_i} - 1}{N - a} \,, \qquad i = 1, 2, 3 \,,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{6-1}{18-3}\big(2.20 + 1.60 + 3.80\big) = 2.5333 \,.$$

If all M possible arrangements of the $N = 18$ observed OLS regression residuals listed in Fig. 4.14 occur with equal chance, the approximate resampling probability value of $\delta_A = 2.5333$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 6$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_A}{L} = \frac{4{,}974}{1{,}000{,}000} = 0.0050 \,.$$

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{A_i}/N$ for $i = 1, 2, 3$ is $P = 0.0560$.

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 17{,}153{,}136$ δ values is $\mu_\delta = 5.5556$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_o}{\mu_\delta} = 1 - \frac{2.5333}{5.5556} = +0.5440 \,,$$

indicating approximately 54 % agreement between the observed and predicted y values above that expected by chance.

An Exact Test

Although an exact permutation analysis of the $N = 18$ OLS regression residuals listed in Fig. 4.14 is impractical, it is not impossible. Following Eq. (4.5) on p. 125, an exact permutation analysis of the $N = 18$ OLS regression residuals listed in Fig. 4.14 yields $a = 3$ average distance-function values of

$$\xi_{A_1} = 2.20 \,, \quad \xi_{A_2} = 1.60 \,, \quad \text{and} \quad \xi_{A_3} = 3.80 \,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic based on $v = 2$ and treatment-group weights

$$C_i = \frac{n_{A_i} - 1}{N - a} \,, \qquad i = 1, 2, 3 \,,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{6-1}{18-3}\big(2.20 + 1.60 + 3.80\big) = 2.5333 \,.$$

If all arrangements of the $N = 18$ observed OLS regression residuals listed in Fig. 4.14 occur with equal chance, the exact probability value of $\delta_A = 2.5333$ computed on the $M = 17{,}153{,}136$ possible arrangements of the observed OLS regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 6$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_A}{M} = \frac{85{,}188}{17{,}153{,}136} = 0.0050 \,.$$

For comparison, the approximate resampling probability value computed on $L =$ 1,000,000 random arrangements of the observed OLS regression residuals listed in Fig. 4.14 is also $P = 0.0050$.

Conventional ANOVA Analysis

A conventional randomized-block analysis of variance calculated on the $N = 18$ univariate response measurement scores listed in Fig. 4.11 on p. 137 yields an observed F-ratio of $F_A = 5.5263$. Assuming independence and normality, F_A is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = a - 1 = 3 - 1 = 2$ and $\nu_2 = (b - 1)(a - 1) = (6 - 1)(3 - 1) = 10$ degrees of freedom. Under the null hypothesis, the observed value of $F_A = 5.5263$ yields an approximate probability value of $P = 0.0242$.

4.3.4 Two-Way Randomized-Block Design

Consider a balanced two-way randomized-block design in which $n = 3$ subjects (S) are tested over $a = 3$ levels of Factor A and the experiment is repeated $b = 3$ times for Factor B. The design and data are adapted from Myers and Well [315, p. 260] and are given in Table 4.4. A complete permutation analysis of a two-way randomized-block design requires three separate analyses comprised of (1) the main effect of Factor A, (2) the main effect of Factor B, and (3) the $A{\times}B$ interaction effect.

Analysis of Factor A

A design matrix of dummy codes for analyzing Factor A is given on the left side of Table 4.5, where the first column of 1 values provides for an intercept and the second and third columns contain dummy codes for Factor B. The last column on the left side of Table 4.5 lists the $N = 9$ response measurement summations over the $b = 3$ levels of Factor B (e.g., $3.10 + 1.90 + 1.60 = 6.60$) and ordered by the $a = 3$ treatment levels of Factor A with the first $n_{A_1} = 3$ summations, the next $n_{A_2} = 3$ summations, and the last $n_{A_3} = 3$ summations associated with treatment levels A_1, A_2, and A_3, respectively. The MRPP regression analysis examines the $N = 9$ regression residuals for possible differences in the $a = 3$ treatment levels of Factor A; consequently, no dummy codes are provided for Factor A as this information is

Table 4.4 Example univariate data for a balanced two-way randomized-block design with $n = 3$ subjects, $a = 3$ levels of Factor A, and $b = 3$ levels of Factor B

	B_1			B_2			B_3		
Subject	A_1	A_2	A_3	A_1	A_2	A_3	A_1	A_2	A_3
S_1	3.10	2.90	2.40	1.90	2.00	1.70	1.60	1.90	1.50
S_2	5.70	6.80	5.30	4.50	5.70	4.40	4.40	5.30	3.90
S_3	9.70	10.90	8.00	7.40	10.50	6.60	6.90	8.90	6.00

Table 4.5 Design matrices and summation data for Factors A and B in a two-way analysis of variance randomized-block design

Factor A				Factor B			
Matrix			Sum over B	Matrix			Sum over A
1	0	0	6.60	1	0	0	8.40
1	1	0	14.60	1	1	0	17.80
1	0	1	24.00	1	0	1	28.60
1	0	0	6.80	1	0	0	5.60
1	1	0	17.80	1	1	0	14.60
1	0	1	30.30	1	0	1	24.50
1	0	0	5.60	1	0	0	5.00
1	1	0	13.60	1	1	0	13.60
1	0	1	20.60	1	0	1	21.80

implicit in the ordering of the $a = 3$ treatment levels of Factor A in the last column on the left side of Table 4.5.

An exact permutation solution is reasonable for the response measurement summations listed on the left side of Table 4.5 since there are only

$$M = \frac{N!}{\prod_{i=1}^{a} n_{A_i}!} = \frac{9!}{(3!)^3} = 1{,}680$$

possible, equally-likely arrangements of the $N = 9$ response measurement summations for Factor A with $n_{A_1} = n_{A_2} = n_{A_3} = 3$ response measurement summations preserved for each arrangement of the observed data.

LAD Regression Analysis

An MRPP analysis of the LAD regression residuals calculated on the $N = 9$ response measurement summations on the left side of Table 4.5 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +6.60\ , \quad \tilde{\beta}_1 = +8.00\ , \quad \text{and} \quad \tilde{\beta}_2 = +17.40$$

for Factor A. Figure 4.15 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 9$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 1$, the $N = 9$ LAD regression residuals listed in Fig. 4.15 yield $a = 3$ average distance-function values of

$$\xi_{A_1} = 0.00\ , \quad \xi_{A_2} = 4.0667\ , \quad \text{and} \quad \xi_{A_3} = 1.60\ .$$

Fig. 4.15 Observed, predicted, and residual LAD regression values for the summations over Factor B on the left side of Table 4.5

Object	y_i	$\tilde{y}_i$	e_i
1	6.60	6.60	0.00
2	14.60	14.00	0.00
3	24.00	24.00	0.00
4	6.80	6.80	+0.20
5	17.70	17.80	+3.20
6	30.30	30.30	+6.30
7	5.60	5.60	−1.00
8	13.60	13.60	−1.00
9	20.60	20.60	−3.40

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.15 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{A_i}}{N} , \qquad i = 1, 2, 3 ,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{3}{9}\big(0.00 + 4.0667 + 1.60\big) = 1.8889 .$$

If all arrangements of the $N = 9$ observed LAD regression residuals listed in Fig. 4.15 occur with equal chance, the exact probability value of $\delta_A = 1.8889$ computed on the $M = 1{,}680$ possible arrangements of the observed LAD regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 3$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_A}{M} = \frac{6}{1{,}680} = 0.0036 .$$

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 1{,}680$ δ values is $\mu_\delta = 2.9889$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{1.8889}{2.9889} = +0.3680 ,$$

indicating approximately 37 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP analysis of OLS regression residuals calculated on the $N = 9$ response measurement summations for Factor A listed on the left

Fig. 4.16 Observed, predicted, and residual OLS regression values for the summations over Factor *B* on the left side of Table 4.5

Object	y_i	$\hat{y}_i$	e_i
1	6.60	6.3333	+0.2667
2	14.60	15.3333	−0.7333
3	24.00	24.9667	−0.9667
4	6.80	6.3333	+0.4667
5	17.80	15.3333	+2.4667
6	30.30	24.9667	+5.3333
7	5.60	6.3333	−0.7333
8	13.60	15.3333	−1.7333
9	20.60	24.9667	−4.3667

side of Table 4.5. Again, since there are only $M = 1{,}680$ possible arrangements of the response measurement summations, an exact permutation test is selected. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +6.3333 \;, \quad \hat{\beta}_1 = +9.00 \;, \quad \text{and} \quad \hat{\beta}_2 = +18.6333$$

for Factor A. Figure 4.16 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 9$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 9$ OLS regression residuals listed in Fig. 4.16 yield $a = 3$ average distance-function values of

$$\xi_{A_1} = 0.8585 \;, \quad \xi_{A_2} = 11.9674 \;, \quad \text{and} \quad \xi_{A_3} = 7.0452 \;.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.16 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{A_i} - 1}{N - a} \;, \qquad i = 1, 2, 3 \;,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{3-1}{9-3}\big(0.8585 + 11.9674 + 7.0452\big) = 6.6237 \;.$$

If all arrangements of the $N = 9$ observed OLS regression residuals listed in Fig. 4.16 occur with equal chance, the exact probability value of $\delta_A = 6.6237$ computed on the $M = 1{,}680$ possible arrangements of the observed OLS regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 3$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_A}{M} = \frac{18}{1{,}680} = 0.0107 \;.$$

For comparison, the exact probability value based on LAD regression, $v = 1$, $M = 1{,}680$, and $C_i = n_{A_i}/N$ for $i = 1, 2, 3$ is $P = 0.0036$.

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 1{,}680$ δ values is $\mu_\delta = 14.7250$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{6.6237}{14.7250} = +0.5512 \,,$$

indicating approximately 55 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional randomized-block analysis of variance calculated on the $N = 27$ univariate response measurement scores for Factor A in Table 4.4 on p. 143 yields an observed F-ratio of $F_A = 3.9282$. Assuming independence and normality, F_A is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = a - 1 = 3 - 1 = 2$ and $\nu_2 = (n-1)(a-1) = (3-1)(3-1) = 4$ degrees of freedom. Under the null hypothesis, the observed value of $F_A = 3.9282$ yields an approximate probability value of $P = 0.1138$.

Analysis of Factor *B*

The right side of Table 4.5 on p. 144 contains a design matrix of dummy codes for analyzing Factor B, where the first column of 1 values provides for an intercept and the next two columns contain dummy codes for Factor A. The last column on the right side of Table 4.5 lists the $N = 9$ response measurement summations over the $a = 3$ levels of Factor A (e.g., $3.10 + 2.90 + 2.40 = 8.40$) and ordered by the $b = 3$ treatment levels with the first $n_{B_1} = 3$ summations, the next $n_{B_2} = 3$ summations, and the last $n_{B_3} = 3$ summations associated with treatment levels, B_1, B_2, and B_3, respectively. The MRPP regression analysis examines the $N = 9$ regression residuals for possible differences among the $b = 3$ treatment levels of Factor B; consequently, no dummy codes are provided for Factor B as this information is implicit in the ordering of the $b = 3$ treatment levels of Factor B in the last column on the right side of Table 4.5.

An exact permutation solution is ideal for the response measurement summations on the right side of Table 4.5 since there are only

$$M = \frac{N!}{\displaystyle\prod_{i=1}^{b} n_{B_i}!} = \frac{9!}{(3!)^3} = 1{,}680$$

possible, equally-likely arrangements of the $N = 9$ response measurement summations for Factor B with $n_{B_1} = n_{B_2} = n_{B_3}$ response measurement summations preserved for each arrangement of the observed data.

Fig. 4.17 Observed, predicted, and residual LAD regression values for the summations over Factor *A* on the right side of Table 4.5

Object	y_i	$\tilde{y}_i$	e_i
1	8.40	5.60	+2.80
2	17.80	14.60	+3.20
3	28.60	24.50	+4.10
4	5.60	5.60	0.00
5	14.60	14.60	0.00
6	24.50	24.50	0.00
7	5.00	5.60	−0.60
8	13.60	14.60	−1.00
9	21.80	24.50	−2.70

LAD Regression Analysis

An MRPP analysis of the LAD regression residuals calculated on the $N = 9$ response measurement summations on the right side of Table 4.5 on p. 144 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +5.60\ , \quad \hat{\beta}_1 = +9.00\ , \quad \text{and} \quad \tilde{\beta}_2 = +18.90$$

for Factor *B*. Figure 4.17 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 9$.

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 9$ LAD regression residuals listed in Fig. 4.17 yield $b = 3$ average distance-function values of

$$\xi_{B_1} = 0.8667\ , \quad \xi_{B_2} = 0.00\ , \quad \text{and} \quad \xi_{B_3} = 1.40\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.17 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{B_i}}{N}\ , \qquad i = 1, 2, 3\ ,$$

is

$$\delta_B = \sum_{i=1}^{b} C_i \xi_i = \frac{3}{9}\big(0.8667 + 0.00 + 1.40\big) = 0.7556\ .$$

If all arrangements of the $N = 9$ observed LAD regression residuals listed in Fig. 4.17 occur with equal chance, the exact probability value of $\delta_B = 0.7556$ computed on the $M = 1{,}680$ possible arrangements of the observed LAD regression residuals with $n_{B_1} = n_{B_2} = n_{B_3} = 3$ residuals preserved for each arrangement is

$$P(\delta \leq \delta_B | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_B}{M} = \frac{6}{1{,}680} = 0.0036 \, .$$

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 1{,}680$ δ values is $\mu_\delta = 2.5889$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_B = 1 - \frac{\delta_B}{\mu_\delta} = 1 - \frac{0.7556}{2.5889} = +0.7082 \, ,$$

indicating approximately 71 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP analysis of OLS regression residuals calculated on the $N = 9$ response measurement summations for Factor B listed on the right side of Table 4.5 on p. 144. Again, since there are only $M = 1{,}680$ possible arrangements of the response measurement summations, an exact permutation test is preferred. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +6.3333 \, , \quad \hat{\beta}_1 = +9.00 \, , \quad \text{and} \quad \hat{\beta}_2 = +18.6333$$

for Factor B. Figure 4.18 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 9$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 9$ OLS regression residuals listed in Fig. 4.18 yield $b = 3$ average distance-function values of

$$\xi_{B_1} = 1.3252 \, , \quad \xi_{B_2} = 0.0474 \, , \quad \text{and} \quad \xi_{B_3} = 1.8585 \, .$$

Fig. 4.18 Observed, predicted, and residual OLS regression values for the summations over Factor A on the right side of Table 4.5

Object	y_i	$\hat{y}_i$	e_i
1	8.40	6.3333	+2.0667
2	17.80	15.3333	+2.4667
3	28.60	24.9667	+3.6333
4	5.60	6.3333	− 0.7333
5	14.60	15.3333	− 0.7333
6	24.50	24.9667	− 0.4667
7	5.00	6.3333	− 1.3333
8	13.60	15.3333	− 1.7333
9	21.80	24.9667	− 3.1667

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.18 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{B_i} - 1}{N - b} , \qquad i = 1, 2, 3 ,$$

is

$$\delta_B = \sum_{i=1}^{b} C_i \xi_i = \frac{3-1}{9-3}\big(1.3252 + 0.0474 + 1.8585\big) = 1.0770 .$$

If all arrangements of the $N = 9$ observed OLS regression residuals listed in Fig. 4.18 occur with equal chance, the exact probability value of $\delta_B = 1.0770$ computed on the $M = 1{,}680$ possible arrangements of the observed OLS regression residuals with $n_{B_1} = n_{B_2} = n_{B_3} = 3$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_B | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_B}{M} = \frac{6}{1{,}680} = 0.0036 .$$

For comparison, the exact probability value based on LAD regression, $v = 1$, $M = 1{,}680$, and $C_i = n_{B_i}/N$ for $i = 1, 2, 3$ is also $P = 0.0036$.

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 1{,}680$ δ values is $\mu_\delta = 9.9150$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_B = 1 - \frac{\delta_B}{\mu_\delta} = 1 - \frac{1.0770}{9.9150} = +0.8914 ,$$

indicating approximately 89 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional randomized-block analysis of variance calculated on the $N = 27$ univariate response measurement scores for Factor B listed in Table 4.4 on p. 143 yields an observed F-ratio of $F_B = 22.5488$. Assuming independence and normality, F_B is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = b - 1 = 3 - 1 = 2$ and $\nu_2 = (n-1)(b-1) = (3-1)(3-1) = 4$ degrees of freedom. Under the null hypothesis, the observed value of $F_B = 22.5488$ yields an approximate probability value of $P = 0.0066$, which is similar to the LAD and OLS regression probability value of $P = 0.0036$.

Analysis of the $A \times B$ Interaction

A design matrix of dummy codes for analyzing the interaction of Factors A and B is given in Table 4.6, where the first column of 1 values provides for an intercept and

Table 4.6 Design matrix and univariate response measurement scores for the interaction of Factors A and B in a two-way randomized-block design with $N = 27$ objects

Matrix															Score
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3.10
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5.70
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	9.70
1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2.90
1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	6.80
1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	10.90
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2.40
1	1	0	0	1	0	0	0	1	0	0	0	0	0	0	5.30
1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	8.00
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1.90
1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	4.50
1	0	1	0	0	1	0	0	0	0	0	0	0	1	0	7.40
1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	2.00
1	1	0	1	0	1	0	1	0	0	0	1	0	0	0	5.70
1	0	1	1	0	1	0	0	0	1	0	0	0	1	0	10.50
1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1.70
1	1	0	0	1	1	0	0	1	0	0	1	0	0	0	4.40
1	0	1	0	1	1	0	0	0	0	1	0	0	1	0	6.60
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1.60
1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	4.40
1	0	1	0	0	0	1	0	0	0	0	0	0	0	1	6.90
1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1.90
1	1	0	1	0	0	1	1	0	0	0	0	1	0	0	5.30
1	0	1	1	0	0	1	0	0	1	0	0	0	0	1	8.90
1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1.50
1	1	0	0	1	0	1	0	1	0	0	0	1	0	0	3.90
1	0	1	0	1	0	1	0	0	0	1	0	0	0	1	6.00

the second and third columns contain dummy codes for subjects (S). The fourth and fifth columns contain dummy codes for Factor A, the sixth and seventh columns contain dummy codes for Factor B, and the next eight columns contain dummy codes for the $S{\times}A$ and $S{\times}B$ interactions. The last column in Table 4.6 lists the response measurement scores ordered by the $ab = (3)(3) = 9$ levels of the $A{\times}B$ interaction.

The MRPP regression analysis examines the $N = 27$ regression residuals for possible differences among the nine treatment levels of the $A{\times}B$ interaction; consequently, no dummy codes are provided for the $A{\times}B$ interaction as this information is implicit in the ordering of the treatment levels of the $A{\times}B$ interaction in the last column of Table 4.6.

Because there are

$$M = \frac{N!}{\displaystyle\prod_{i=1}^{ab} n_{(A \times B)_i}!} = \frac{27!}{(3!)^9} = 1{,}080{,}491{,}954{,}750{,}208{,}000{,}000$$

possible, equally-likely arrangements of the $N = 27$ univariate response measurement scores for the $A \times B$ interaction listed in Table 4.6, an exact permutation solution is not possible.

LAD Regression Analysis

An MRPP resampling analysis of the LAD regression residuals calculated on the $N = 27$ univariate response measurement scores listed in Table 4.6 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +2.70\,, \quad \tilde{\beta}_1 = +3.00\,, \quad \tilde{\beta}_2 = +6.20\,, \quad \tilde{\beta}_3 = +0.20\,,$$
$$\tilde{\beta}_4 = -0.20\,, \quad \tilde{\beta}_5 = -0.80\,, \quad \tilde{\beta}_6 = -1.00\,, \quad \tilde{\beta}_7 = +0.90\,,$$
$$\tilde{\beta}_8 = -0.20\,, \quad \tilde{\beta}_9 = +1.80\,, \quad \tilde{\beta}_{10} = -0.70\,, \quad \tilde{\beta}_{11} = -0.30\,,$$
$$\tilde{\beta}_{12} = -0.40\,, \quad \tilde{\beta}_{13} = -0.60\,, \quad \text{and} \quad \tilde{\beta}_{14} = -1.00$$

for the interaction of Factors A and B. Figure 4.19 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 27$.

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 27$ LAD regression residuals listed in Fig. 4.19 yield $ab = (3)(3) = 9$ average distance-function values of

$$\xi_{(A\times B)_1} = 0.5333\,, \quad \xi_{(A\times B)_2} = 0.00\,, \quad \xi_{(A\times B)_3} = \xi_{(A\times B)_4} = 0.0667\,,$$
$$\xi_{(A\times B)_5} = 0.7333\,, \quad \xi_{(A\times B)_6} = \xi_{(A\times B)_7} = 0.1333\,, \quad \xi_{(A\times B)_8} = 0.0667\,,$$
$$\text{and} \quad \xi_{(A\times B)_9} = 0.00\,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.19 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{(A\times B)_i}}{N}\,, \qquad i = 1, \ldots, 9\,,$$

is

$$\delta_{A\times B} = \sum_{i=1}^{ab} C_i \xi_i = \frac{3}{9}\big(0.5333 + 0.00 + \cdots + 0.0667 + 0.00\big) = 0.1926\,.$$

Fig. 4.19 Observed, predicted, and residual LAD regression values for the univariate response measurement scores listed in Table 4.6

Object	y_i	$\tilde{y}_i$	e_i
1	3.10	2.70	+0.40
2	5.70	5.70	0.00
3	9.70	8.90	+0.80
4	2.90	2.90	0.00
5	6.80	6.80	0.00
6	10.90	10.90	0.00
7	2.40	2.50	−0.10
8	5.30	5.30	0.00
9	8.00	8.00	0.00
10	1.90	1.90	0.00
11	4.50	4.60	−0.10
12	7.40	7.50	−0.10
13	2.00	2.10	−0.10
14	5.70	5.70	0.00
15	10.50	9.50	+1.00
16	1.70	1.70	0.00
17	4.40	4.20	+0.20
18	6.60	6.60	0.00
19	1.60	1.70	−0.10
20	4.40	4.30	+0.10
21	6.90	6.90	0.00
22	1.90	1.90	0.00
23	5.30	5.40	−0.10
24	8.90	8.90	0.00
25	1.50	1.50	0.00
26	3.90	3.90	0.00
27	6.00	6.00	0.00

If all M possible arrangements of the $N = 27$ observed LAD regression residuals listed in Fig. 4.19 occur with equal chance, the approximate resampling probability value of $\delta_{A\times B} = 0.1926$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{(A\times B)_1} = \cdots = n_{(A\times B)_9} = 3$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_{A\times B}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{A\times B}}{L} = \frac{235{,}542}{1{,}000{,}000} = 0.2355 \ .$$

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 0.2063$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{A\times B} = 1 - \frac{\delta_{A\times B}}{\mu_\delta} = 1 - \frac{0.1926}{0.2063} = +0.0663 \ ,$$

indicating approximately 7 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP analysis of OLS regression residuals calculated on the $N = 27$ univariate response measurement scores for the $A \times B$ interaction listed in Table 4.6. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +2.8889\,, \quad \hat{\beta}_1 = +2.80\,, \quad \hat{\beta}_2 = +6.3222\,, \quad \hat{\beta}_3 = +0.0667\,,$$

$$\hat{\beta}_4 = -0.3333\,, \quad \hat{\beta}_5 = -0.9333\,, \quad \hat{\beta}_6 = -1.1333\,, \quad \hat{\beta}_7 = +1.00\,,$$

$$\hat{\beta}_8 = 0.00\,, \quad \hat{\beta}_9 = +2.0333\,, \quad \hat{\beta}_{10} = -0.80\,, \quad \hat{\beta}_{11} = -0.1333\,,$$

$$\hat{\beta}_{12} = -0.2667\,, \quad \hat{\beta}_{13} = -0.4333\,, \quad \text{and} \quad \hat{\beta}_{14} = -1.1333$$

for the interaction of Factors A and B. Figure 4.20 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 27$.

Fig. 4.20 Observed, predicted, and residual OLS regression values for the univariate response measurement scores listed in Table 4.6

Object	y_i	$\hat{y}_i$	e_i
1	3.10	2.8889	+0.2111
2	5.70	5.6889	+0.0111
3	9.70	9.2111	+0.4889
4	2.90	2.9556	−0.0556
5	6.80	6.7556	+0.0444
6	10.90	11.3111	−0.4111
7	2.40	2.5556	−0.1556
8	5.30	5.3556	−0.0556
9	8.00	8.0778	−0.0778
10	1.90	1.9556	−0.0556
11	4.50	4.6222	−0.1222
12	7.40	7.8444	−0.4444
13	2.00	2.0222	−0.2222
14	5.70	5.6889	+0.0111
15	10.50	9.9444	+0.5556
16	1.70	1.6222	+0.0778
17	4.40	4.2889	+0.1111
18	6.60	6.7111	−0.1111
19	1.60	1.7556	−0.1556
20	4.40	4.2889	+0.1111
21	6.90	6.9444	−0.0444
22	1.90	1.8222	+0.0778
23	5.30	5.3556	−0.0556
24	8.90	9.0444	−0.1444
25	1.50	1.4222	+0.0778
26	3.90	3.9556	−0.0556
27	6.00	5.8111	+0.1889

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 27$ OLS regression residuals listed in Fig. 4.20 yield $ab = (3)(3) = 9$ average distance-function values of

$$\xi_{(A\times B)_1} = 0.1151 \,, \quad \xi_{(A\times B)_2} = 0.1147 \,, \quad \xi_{(A\times B)_3} = 0.0055 \,,$$
$$\xi_{(A\times B)_4} = 0.0865 \,, \quad \xi_{(A\times B)_5} = 0.2105 \,, \quad \xi_{(A\times B)_6} = 0.0287 \,,$$
$$\xi_{(A\times B)_7} = 0.0359 \,, \quad \xi_{(A\times B)_8} = 0.0250 \,, \quad \text{and} \quad \xi_{(A\times B)_9} = 0.0300 \,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.20 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{(A\times B)_i} - 1}{N - ab} \,, \qquad i = 1, \ldots, 9 \,,$$

is

$$\delta_{A\times B} = \sum_{i=1}^{ab} C_i \xi_i = \frac{3-1}{27-9}\big(0.1151 + 0.1147 + 0.0055 + \cdots + 0.0250 + 0.0300\big) = 0.0724 \,.$$

If all M possible arrangements of the $N = 27$ observed OLS regression residuals listed in Fig. 4.20 occur with equal chance, the approximate resampling probability value of $\delta_{A\times B} = 0.0724$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_{(A\times B)_1} = \cdots = n_{(A\times B)_9} = 3$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_{A\times B}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{A\times B}}{L} = \frac{141{,}960}{1{,}000{,}000} = 0.1420 \,.$$

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{(A\times B)_i}/N$ for $i = 1, \ldots, 9$ is $P = 0.2355$.

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 8.9231$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{A\times B} = 1 - \frac{\delta_{A\times B}}{\mu_\delta} = 1 - \frac{0.0724}{8.9231} = +0.1883 \,,$$

indicating approximately 19 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional randomized-block analysis of variance calculated on the $N = 27$ response measurement scores for the $A{\times}B$ interaction listed in Table 4.4 on p. 143 yields an observed F-ratio of $F_{A\times B} = 1.5591$. Assuming independence and normality, $F_{A\times B}$ is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = (a-1)(b-1) = (3-1)(3-1) = 4$ and $\nu_2 = (n-1)(a-1)(b-1) = (3-1)(3-1)(3-1) = 8$ degrees of freedom. Under the null hypothesis, the observed value of $F_{A\times B} = 1.5591$ yields an approximate probability value of $P = 0.2744$.

4.3.5 Two-Way Factorial Design

Consider a 2×3 fixed-effects factorial design with $n = 4$ subjects in each treatment combination for a total of $N = 24$ subjects. The univariate response measurement scores for Factors A and B are listed in Fig. 4.21, and the design matrices and data for Factors A and B are given in Table 4.7; the design and data are adapted from Keppel [214, p. 197]. While design matrices of either dummy or effect codes are appropriate for one-way completely randomized and randomized-block designs, the main effects of factorial designs are best analyzed with effect codes when estimation of the effects of each factor is adjusted for all other factors in the model to obtain the unique contribution of each factor [31,37,294].[3] A permutation analysis of factorial designs requires three separate analyses comprising (1) the main effect of Factor A, (2) the main effect of Factor B, and (3) the $A{\times}B$ interaction effect.

Fig. 4.21 Example univariate response measurement scores for Factors A and B in a two-way factorial design

Factor A		Factor B		
A_1	A_2	B_1	B_2	B_3
1	15	1	13	9
4	6	1	5	16
0	10	0	7	18
7	13	7	15	13
13	6	15	6	14
5	18	6	18	7
7	9	10	9	6
15	15	13	15	13
9	14			
16	7			
18	6			
13	13			

[3] This method of estimation is known as Method I as presented in a seminal article by Overall and Spiegel in 1969 [330].

Table 4.7 Design matrices and univariate response measurement scores for the main effects of Factors *A* and *B* in a two-way factorial design with $N = 24$ subjects

Factor *A*						Factor *B*				
Matrix					Score	Matrix				Score
1	1	0	1	0	1	1	1	1	0	1
1	1	0	1	0	4	1	1	1	0	4
1	1	0	1	0	0	1	1	1	0	0
1	1	0	1	0	7	1	1	1	0	7
1	0	1	0	1	13	1	−1	−1	0	15
1	0	1	0	1	5	1	−1	−1	0	6
1	0	1	0	1	7	1	−1	−1	0	10
1	0	1	0	1	15	1	−1	−1	0	13
1	−1	−1	−1	−1	9					
1	−1	−1	−1	−1	16	1	1	0	1	13
1	−1	−1	−1	−1	18	1	1	0	1	5
1	−1	−1	−1	−1	13	1	1	0	1	7
						1	1	0	1	15
1	1	0	−1	0	15	1	−1	0	−1	6
1	1	0	−1	0	6	1	−1	0	−1	18
1	1	0	−1	0	10	1	−1	0	−1	9
1	1	0	−1	0	13	1	−1	0	−1	15
1	0	1	0	−1	6					
1	0	1	0	−1	18	1	1	−1	−1	9
1	0	1	0	−1	9	1	1	−1	−1	16
1	0	1	0	−1	15	1	1	−1	−1	18
1	−1	−1	1	1	14	1	1	−1	−1	13
1	−1	−1	1	1	7	1	−1	1	1	14
1	−1	−1	1	1	6	1	−1	1	1	7
1	−1	−1	1	1	13	1	−1	1	1	6
				1	−1	1	1	13		

Analysis of Factor *A*

A design matrix of effect codes for analyzing Factor *A* is given on the left side of Table 4.7, where the first column of 1 values provides for an intercept. The second and third columns contain effect codes for Factor *B*, the fourth and fifth columns contain effect codes for the $A{\times}B$ interaction, and the last column on the left side of Table 4.7 contains the $N = 24$ univariate response measurement scores listed according to the original random assignment of the subjects to the $a = 2$ levels of Factor *A* with the first $n_{A_1} = 12$ scores and the last $n_{A_2} = 12$ scores associated with treatment levels A_1 and A_2, respectively. The MRPP regression analysis examines the $N = 24$ regression residuals for possible differences between the $a = 2$ treatment levels of Factor *A*; consequently, no effect codes are provided for Factor *A* as this information is implicit in the ordering of the $a = 2$ treatment levels of Factor *A* in the last column on the left side of Table 4.7.

An exact permutation solution is feasible for the univariate response measurement scores listed on the left side of Table 4.7 since there are only

$$M = \frac{N!}{\prod_{i=1}^{a} n_{A_i}!} = \frac{24!}{(12!)^2} = 2{,}704{,}156$$

possible, equally-likely arrangements of the $N = 24$ response measurement scores for Factor A.

LAD Regression Analysis

An MRPP analysis of the LAD regression residuals calculated on the $N = 24$ univariate response measurement scores on the left side of Table 4.7 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +9.6667\,, \quad \tilde{\beta}_1 = -1.1667\,, \quad \tilde{\beta}_2 = +0.8333\,, \quad \tilde{\beta}_3 = -4.50\,, \quad \text{and}$$

$$\tilde{\beta}_4 = +1.50$$

for Factor A. Figure 4.22 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 24$.

Fig. 4.22 Observed, predicted, and residual LAD regression values for the univariate response measurement scores listed on the left side of Table 4.7

Object	y_i	$\tilde{y}_i$	e_i
1	1	4.00	−3.00
2	4	4.00	0.00
3	0	4.00	−4.00
4	7	4.00	+3.00
5	13	12.00	+1.00
6	5	12.00	−7.00
7	7	12.00	−5.00
8	15	12.00	+3.00
9	9	13.00	−4.00
10	16	13.00	+3.00
11	18	13.00	+5.00
12	13	13.00	0.00
13	15	13.00	+2.00
14	6	13.00	−7.00
15	10	13.00	−3.00
16	13	13.00	0.00
17	6	9.00	−3.00
18	18	9.00	+9.00
19	9	9.00	0.00
20	15	9.00	+6.00
21	14	7.00	+7.00
22	7	7.00	0.00
23	6	7.00	−1.00
24	13	7.00	+6.00

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 24$ LAD regression residuals listed in Fig. 4.22 yield $a = 2$ average distance-function values of

$$\xi_{A_1} = 4.5455 \quad \text{and} \quad \xi_{A_2} = 5.6061 \ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.22 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{A_i}}{N} \ , \qquad i = 1, 2 \ ,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{12}{24}\big(4.5455 + 5.6061\big) = 5.0758 \ .$$

If all arrangements of the $N = 24$ observed LAD regression residuals listed in Fig. 4.22 occur with equal chance, the exact probability value of $\delta_A = 5.0758$ computed on the $M = 2{,}704{,}156$ possible arrangements of the observed LAD regression residuals with $n_{A_1} = n_{A_2} = 12$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_A}{M} = \frac{1{,}039{,}084}{2{,}704{,}156} = 0.3843 \ .$$

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 2{,}704{,}156$ δ values is $\mu_\delta = 5.0725$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{5.0758}{5.0725} = -0.6494 \times 10^{-3} \ ,$$

indicating slightly less than chance agreement between the observed and predicted y values.

OLS Regression Analysis

For comparison, consider an MRPP analysis of OLS regression residuals calculated on the $N = 24$ univariate response measurement scores for Factor A on the left side of Table 4.7. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +10.00 \ , \ \ \hat{\beta}_1 = -3.00 \ , \ \ \hat{\beta}_2 = +1.00 \ , \ \ \hat{\beta}_3 = -3.00 \ , \ \text{and} \ \ \hat{\beta}_4 = 0.00$$

Fig. 4.23 Observed, predicted, and residual OLS regression values for the univariate response measurement scores listed on the left side of Table 4.7

Object	y_i	$\hat{y}_i$	e_i
1	1	4.00	−3.00
2	4	4.00	0.00
3	0	4.00	−4.00
4	7	4.00	+3.00
5	13	11.00	+2.00
6	5	11.00	−6.00
7	7	11.00	−4.00
8	15	11.00	+4.00
9	9	15.00	−6.00
10	16	15.00	+1.00
11	18	15.00	+3.00
12	13	15.00	−2.00
13	15	10.00	+5.00
14	6	10.00	−4.00
15	10	10.00	0.00
16	13	10.00	+3.00
17	6	11.00	−5.00
18	18	11.00	+7.00
19	9	11.00	−2.00
20	15	11.00	+4.00
21	14	9.00	+5.00
22	7	9.00	−2.00
23	6	9.00	−3.00
24	13	9.00	+4.00

for Factor A. Figure 4.23 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 24$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 24$ OLS regression residuals listed in Fig. 4.23 yield $a = 2$ average distance-function values of

$$\xi_{A_1} = 26.1818 \quad \text{and} \quad \xi_{A_2} = 33.8182 \,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.23 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{A_i} - 1}{N - a} \,, \qquad i = 1, 2 \,,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{12 - 1}{24 - 2} \big(26.1818 + 33.8182\big) = 30.00 \,.$$

If all arrangements of the $N = 24$ observed OLS regression residuals listed in Fig. 4.23 occur with equal chance, the exact probability value of $\delta_A = 30.00$ computed on the $M = 2{,}704{,}156$ possible arrangements of the observed OLS regression residuals with $n_{A_1} = n_{A_2} = 12$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_A}{M} = \frac{637{,}454}{2{,}704{,}156} = 0.2357\ .$$

For comparison, the exact probability value based on LAD regression, $v = 1$, $M = 2{,}704{,}156$, and $C_i = n_{A_i}/N$ for $i = 1, 2$ is $P = 0.3843$.

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 2{,}704{,}156$ δ values is $\mu_\delta = 30.7826$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{30.00}{30.7826} = +0.0254\ ,$$

indicating approximately 3 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional fixed-effects factorial analysis of variance calculated on the $N = 24$ Factor A response measurement scores listed in Fig. 4.21 on p. 156 yields an observed F-ratio of $F_A = 1.3091$. Assuming independence, normality, and homogeneity of variance, F_A is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = a - 1 = 2 - 1 = 1$ and $\nu_2 = N - ab = 24 - (2)(3) = 18$ degrees of freedom. Under the null hypothesis, the observed value of $F_A = 1.3091$ yields an approximate probability value of $P = 0.2675$, which is similar to the OLS regression probability value of $P = 0.2357$.

Analysis of Factor *B*

The right side of Table 4.7 on p. 157 contains a design matrix of effect codes for analyzing Factor B, where the first column of 1 values provides for an intercept. The second column contains effect codes for Factor A, the third and fourth columns contain effect codes for the $A \times B$ interaction, and the last column on the right side of Table 4.7 contains the $N = 24$ univariate response measurement scores listed according to the original random assignment of the subjects to the $b = 3$ levels of Factor B with the first $n_{B_1} = 8$ scores, the next $n_{B_2} = 8$ scores, and the last $n_{B_3} = 8$ scores associated with treatment levels, B_1, B_2, and B_3, respectively. The MRPP regression analysis examines the $N = 24$ regression residuals for possible differences among the $b = 3$ treatment levels of Factor B; consequently, no effect codes are provided for Factor B as this information is implicit in the ordering of the $b = 3$ treatment levels of Factor B in the last column on the right side of Table 4.7.

Because there are

$$M = \frac{N!}{\prod_{i=1}^{b} n_{B_i}!} = \frac{24!}{(8!)^3} = 9{,}465{,}511{,}770$$

possible, equally-likely arrangements of the $N = 24$ response measurement scores for Factor B listed on the right side of Table 4.7, an exact permutation approach is not practical.

LAD Regression Analysis

An MRPP resampling analysis of the $N = 24$ LAD regression residuals calculated on the univariate response measurement scores on the right side of Table 4.7 on p. 157 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +9.50\ , \quad \tilde{\beta}_1 = +0.1667\ , \quad \tilde{\beta}_2 = -3.6667\ , \quad \text{and} \quad \tilde{\beta}_3 = +0.3333$$

for Factor B. Figure 4.24 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 24$.

Fig. 4.24 Observed, predicted, and residual LAD regression values for the univariate response measurement scores listed on the right side of Table 4.7

Object	y_i	$\tilde{y}_i$	e_i
1	1	6.00	− 5.00
2	4	6.00	− 2.00
3	0	6.00	− 6.00
4	7	6.00	+1.00
5	15	13.00	+2.00
6	6	13.00	−7.00
7	10	13.00	−3.00
8	13	13.00	0.00
9	13	10.00	+3.00
10	5	10.00	−5.00
11	7	10.00	−3.00
12	15	10.00	+5.00
13	6	9.00	−3.00
14	18	9.00	+9.00
15	9	9.00	0.00
16	15	9.00	+6.00
17	9	13.00	−4.00
18	16	13.00	+3.00
19	18	13.00	+5.00
20	13	13.00	0.00
21	14	6.00	+8.00
22	7	6.00	+1.00
23	6	6.00	0.00
24	13	6.00	+7.00

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 24$ LAD regression residuals listed in Fig. 4.24 yield $b = 3$ average distance-function values of

$$\xi_{B_1} = 4.0714\ , \quad \xi_{B_2} = 6.0714\ , \quad \text{and} \quad \xi_{B_3} = 4.8571\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.24 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{B_i}}{N}\ , \qquad i = 1, 2, 3\ ,$$

is

$$\delta_B = \sum_{i=1}^{b} C_i \xi_i = \frac{12}{24}\big(4.0714 + 6.0714 + 4.8571\big) = 5.00\ .$$

If all M possible arrangements of the $N = 24$ observed LAD regression residuals listed in Fig. 4.24 occur with equal chance, the approximate resampling probability value of $\delta_B = 5.00$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{B_1} = n_{B_2} = n_{B_3} = 8$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_B | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_B}{L} = \frac{125{,}031}{1{,}000{,}000} = 0.1250\ .$$

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 5.3333$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_B = 1 - \frac{\delta_B}{\mu_\delta} = 1 - \frac{5.00}{5.3333} = +0.0625\ ,$$

indicating approximately 6 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP analysis of OLS regression residuals calculated on the $N = 24$ univariate response measurement scores for Factor B listed on the right side of Table 4.7 on p. 157. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +10.00\ , \quad \hat{\beta}_1 = -1.00\ , \quad \hat{\beta}_2 = -3.00\ , \quad \text{and} \quad \hat{\beta}_3 = 0.00$$

Fig. 4.25 Observed, predicted, and residual OLS regression values for the univariate response measurement scores listed on the right side of Table 4.7

Object	y_i	$\hat{y}_i$	e_i
1	1	6.00	−5.00
2	4	6.00	−2.00
3	0	6.00	−6.00
4	7	6.00	+1.00
5	15	14.00	+1.00
6	6	14.00	−8.00
7	10	14.00	−4.00
8	13	14.00	−1.00
9	13	9.00	+4.00
10	5	9.00	−4.00
11	7	9.00	−2.00
12	15	9.00	+6.00
13	6	11.00	−5.00
14	18	11.00	+7.00
15	9	11.00	−2.00
16	15	11.00	+4.00
17	9	12.00	−3.00
18	16	12.00	+4.00
19	18	12.00	+6.00
20	13	12.00	+1.00
21	14	8.00	+6.00
22	7	8.00	−1.00
23	6	8.00	−2.00
24	13	8.00	+5.00

for Factor B. Figure 4.25 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 24$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 24$ OLS regression residuals listed in Fig. 4.25 yield $b = 3$ average distance-function values of

$$\xi_{B_1} = 21.7143\,, \quad \xi_{B_2} = 45.1429\,, \quad \text{and} \quad \xi_{B_3} = 27.4286\,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.25 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{B_i} - 1}{N - b}\,, \qquad i = 1, 2, 3\,,$$

is

$$\delta_B = \sum_{i=1}^{b} C_i \xi_i = \frac{8-1}{24-3}\big(21.7143 + 45.1429 + 27.4286\big) = 31.4286\,.$$

If all M possible arrangements of the $N = 24$ observed OLS regression residuals listed in Fig. 4.25 occur with equal chance, the approximate resampling probability value of $\delta_B = 31.4286$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_{B_1} = n_{B_2} = n_{B_3} = 8$ residuals preserved for each arrangement is

$$P(\delta \leq \delta_B | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_B}{L} = \frac{49{,}168}{1{,}000{,}000} = 0.0492 \; .$$

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{B_i}/N$ for $i = 1, 2, 3$ is $P = 0.1250$.

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 9{,}465{,}511{,}770$ δ values is $\mu_\delta = 38.4348$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_B = 1 - \frac{\delta_B}{\mu_\delta} = 1 - \frac{31.4286}{38.4348} = +0.1823 \; ,$$

indicating approximately 18 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional fixed-effects factorial analysis of variance calculated on the $N = 24$ Factor B response measurement scores listed in Fig. 4.21 on p. 156 yields an observed F-ratio of $F_B = 3.0545$. Assuming independence, normality, and homogeneity of variance, F_B is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = b - 1 = 3 - 1 = 2$ and $\nu_2 = N - ab = 24 - (2)(3) = 18$ degrees of freedom. Under the null hypothesis, the observed value of $F_B = 3.0545$ yields an approximate probability value of $P = 0.0721$.

Analysis of the $A{\times}B$ Interaction

A design matrix of effect codes for analyzing the $A{\times}B$ interaction of the data listed in Fig. 4.21 on p. 156 is given in Fig. 4.26, where the first column of 1 values provides for an intercept, the second column contains effect codes for Factor A, the third and fourth columns contain effect codes for Factor B, and the last column lists the $N = 24$ univariate response measurement scores listed according to the original random assignment of the subjects to the $ab = (2)(3) = 6$ levels of the $A{\times}B$ interaction. The MRPP regression analysis examines the $N = 24$ regression residuals for possible differences among the six treatment levels of the $A{\times}B$ interaction; consequently, no effect codes are provided for the $A{\times}B$ interaction as this information is implicit in the ordering of the treatment levels of the $A{\times}B$ interaction in the last column of Fig. 4.26.

Fig. 4.26 Design matrix and univariate response measurement scores for the $A{\times}B$ interaction in a 2×3 factorial design with $N = 24$ subjects

Matrix				Score
1	1	1	0	1
1	1	1	0	4
1	1	1	0	0
1	1	1	0	7
1	−1	1	0	15
1	−1	1	0	6
1	−1	1	0	10
1	−1	1	0	13
1	1	0	1	13
1	1	0	1	5
1	1	0	1	7
1	1	0	1	15
1	−1	0	1	6
1	−1	0	1	18
1	−1	0	1	9
1	−1	0	1	15
1	1	−1	−1	9
1	1	−1	−1	16
1	1	−1	−1	18
1	1	−1	−1	13
1	−1	−1	−1	14
1	−1	−1	−1	7
1	−1	−1	−1	6
1	−1	−1	−1	13

Because there are

$$M = \frac{N!}{\displaystyle\prod_{i=1}^{ab} n_{(A\times B)_i}!} = \frac{24!}{(4!)^6} = 118{,}569{,}536{,}025{,}665{,}614{,}982{,}267{,}535{,}360{,}000$$

possible, equally-likely arrangements of the $N = 24$ univariate response measurement scores for the $A{\times}B$ interaction listed in Fig. 4.26, an exact permutation approach is clearly not possible.

LAD Regression Analysis

An MRPP resampling analysis of the LAD regression residuals calculated on the univariate response measurement scores in Fig. 4.26 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +8.3333\ , \quad \tilde{\beta}_1 = -1.00\ , \quad \tilde{\beta}_2 = -3.3333\ , \quad \text{and} \quad \tilde{\beta}_3 = -0.3333$$

for the interaction of Factors A and B. Figure 4.27 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 24$.

Fig. 4.27 Observed, predicted, and residual LAD regression values for the univariate response measurement scores listed in Fig. 4.26

Object	y_i	$\tilde{y}_i$	e_i
1	1	4.00	−3.00
2	4	4.00	0.00
3	0	4.00	−4.00
4	7	4.00	+3.00
5	15	6.00	+9.00
6	6	6.00	0.00
7	10	6.00	+4.00
8	13	6.00	+7.00
9	13	7.00	+6.00
10	5	7.00	−2.00
11	7	7.00	0.00
12	15	7.00	+8.00
13	6	9.00	−3.00
14	18	9.00	+9.00
15	9	9.00	0.00
16	15	9.00	+6.00
17	9	11.00	−2.00
18	16	11.00	+5.00
19	18	11.00	+7.00
20	13	11.00	+2.00
21	14	13.00	+1.00
22	7	13.00	−6.00
23	6	13.00	−7.00
24	13	13.00	0.00

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 24$ LAD regression residuals listed in Fig. 4.27 yield $ab = (2)(3) = 6$ average distance-function values of

$$\xi_{(A\times B)_1} = 4.00\ , \quad \xi_{(A\times B)_2} = 5.00\ , \quad \xi_{(A\times B)_3} = 6.00\ , \quad \xi_{(A\times B)_4} = 7.00\ ,$$
$$\text{and} \quad \xi_{(A\times B)_5} = \xi_{(A\times B)_6} = 5.00\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.27 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{(A\times B)_i}}{N}\ , \qquad i = 1, \ldots, 6\ ,$$

is

$$\delta_{A\times B} = \sum_{i=1}^{ab} C_i \xi_i = \frac{4}{24}\big(4.00 + 5.00 + 6.00 + 7.00 + 5.00 + 5.00\big) = 5.3333\ .$$

If all M possible arrangements of the $N = 24$ observed LAD regression residuals listed in Fig. 4.27 occur with equal chance, the approximate resampling probability value of $\delta_{A \times B} = 5.3333$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{(A \times B)_1} = \cdots = n_{(A \times B)_6} = 4$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_{A \times B} | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{A \times B}}{L} = \frac{347{,}675}{1{,}000{,}000} = 0.3477 \ .$$

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 5.50$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{A \times B} = 1 - \frac{\delta_{A \times B}}{\mu_\delta} = 1 - \frac{5.3333}{5.50} = +0.0303 \ ,$$

indicating approximately 3 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP analysis of OLS regression residuals calculated on the $N = 24$ univariate response measurement scores of the $A \times B$ interaction listed in Fig. 4.26. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +10.00 \ , \quad \hat{\beta}_1 = -1.00 \ , \quad \hat{\beta}_2 = -3.00 \ , \quad \text{and} \quad \hat{\beta}_3 = +1.00$$

for the interaction of Factors A and B. Figure 4.28 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 24$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 24$ OLS regression residuals listed in Fig. 4.28 yield $ab = (2)(3) = 6$ average distance-function values of

$$\xi_{(A \times B)_1} = 20.00 \ , \quad \xi_{(A \times B)_2} = 30.6667 \ , \quad \xi_{(A \times B)_3} = 45.3333 \ ,$$
$$\xi_{(A \times B)_4} = 60.00 \ , \quad \xi_{(A \times B)_5} = 30.6667 \ , \quad \text{and} \quad \xi_{(A \times B)_6} = 33.3333 \ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.28 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{(A \times B)_i} - 1}{N - ab} \ , \qquad i = 1, \ldots, 6 \ ,$$

Fig. 4.28 Observed, predicted, and residual OLS regression values for the univariate response measurement scores listed in Fig. 4.26

Object	y_i	$\hat{y}_i$	e_i
1	1	6.00	−5.00
2	4	6.00	−2.00
3	0	6.00	−6.00
4	7	6.00	+1.00
5	15	8.00	+7.00
6	6	8.00	−2.00
7	10	8.00	+2.00
8	13	8.00	+5.00
9	13	10.00	+3.00
10	5	10.00	−5.00
11	7	10.00	−3.00
12	15	10.00	+5.00
13	6	12.00	−6.00
14	18	12.00	+6.00
15	9	12.00	−3.00
16	15	12.00	+3.00
17	9	11.00	−2.00
18	16	11.00	+5.00
19	18	11.00	+7.00
20	13	11.00	+2.00
21	14	13.00	+1.00
22	7	13.00	−6.00
23	6	13.00	−7.00
24	13	13.00	0.00

is

$$\delta_{A\times B} = \sum_{i=1}^{ab} C_i \xi_i = \frac{4-1}{24-6}\big(20.00 + 30.6667 + 45.3333 + 60.00 + 30.6667 + 33.3333\big) = 36.6667 \,.$$

If all M possible arrangements of the observed OLS regression residuals listed in Fig. 4.28 occur with equal chance, the approximate resampling probability value of $\delta_{A\times B} = 36.6666$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_{(A\times B)_1} = \cdots = n_{(A\times B)_6} = 4$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_{A\times B}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{A\times B}}{L} = \frac{224{,}204}{1{,}000{,}000} = 0.2242 \,.$$

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{(A\times B)_i}/N$ for $i = 1, \ldots, 6$ is $P = 0.3477$.

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 41.2174$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{A\times B} = 1 - \frac{\delta_{A\times B}}{\mu_\delta} = 1 - \frac{36.6667}{41.2174} = +0.1104 \,,$$

indicating approximately 11 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional fixed-effects factorial analysis of variance calculated on the $N = 24$ univariate response measurement scores listed in Fig. 4.21 on p. 156 yields an observed F-ratio of $F_{A\times B} = 3.9273$. Assuming independence, normality, and homogeneity of variance, $F_{A\times B}$ is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = (a-1)(b-1) = (2-1)(3-1) = 2$ and $\nu_2 = ab(n-1) = (2)(3)(4-1) = 18$ degrees of freedom. Under the null hypothesis, the observed value of $F_{A\times B} = 3.9273$ yields an approximate probability value of $P = 0.0384$, which differs greatly from the LAD and OLS regression probability values of $P = 0.3477$ and $P = 0.2242$, respectively.

4.3.6 Latin Square Design

A Latin square experimental design assigns treatments to subjects so the treatments occur in a balanced fashion within a square block or field; thus, n treatments appear once in each of n rows and n columns. The Latin square is the design of choice when controlling for two blocking factors. Consider an ordinary balanced Latin square experiment involving repeated measurements in which $n = 4$ subjects (S) are each tested $b = 4$ times on Factor A. The design and data are adapted from Ferguson [115, p. 349] and are given in Table 4.8, where B refers to the ordinal position in which the levels of Factor A are administered. Thus, the first subject, S_1, receives the $b = 4$ treatments in the order A_2, A_4, A_1, A_3, and so on. Due to the balanced nature of Latin square designs, the assumption is that there is no interaction between blocking Factors A and B, or between either blocking factor and the treatments.

Table 4.8 Design and data for a Latin square design with four subjects (S), four treatments (A), and four orders (B)

	Design					Scores			
Subject	B_1	B_2	B_3	B_4	Subject	B_1	B_2	B_3	B_4
S_1	A_2	A_4	A_1	A_3	S_1	10	21	5	14
S_2	A_3	A_1	A_2	A_4	S_2	12	7	11	19
S_3	A_1	A_3	A_4	A_2	S_3	6	16	24	12
S_4	A_4	A_2	A_3	A_1	S_4	22	8	17	9

Fig. 4.29 Design matrix and univariate response measurement scores for treatment (*A*) in a Latin square design

Matrix							Score
1	0	0	0	0	1	0	5
1	1	0	0	1	0	0	7
1	0	1	0	0	0	0	6
1	0	0	1	0	0	1	9
1	0	0	0	0	0	0	10
1	1	0	0	0	1	0	11
1	0	1	0	0	0	1	12
1	0	0	1	1	0	0	8
1	0	0	0	0	0	1	14
1	1	0	0	0	0	0	12
1	0	1	0	1	0	0	16
1	0	0	1	0	1	0	17
1	0	0	0	1	0	0	21
1	1	0	0	0	0	1	19
1	0	1	0	0	1	0	24
1	0	0	1	0	0	0	22

Analysis of Factor *A*

A design matrix of dummy codes for analyzing Factor A is given in Fig. 4.29, where the first column of 1 values provides for an intercept, the second through fourth columns contain dummy codes for Subjects, the fifth through seventh columns contain dummy codes for Factor B, and the last column lists the univariate response measurement scores ordered by the $a = 4$ levels of Factor A, with the first $n_{A_1} = 4$ scores, the next $n_{A_2} = 4$ scores, the next $n_{A_3} = 4$ scores, and the last $n_{A_4} = 4$ scores associated with treatment levels A_1, A_2, A_3, and A_4, respectively. The MRPP regression analysis examines the $N = 16$ regression residuals for possible differences among the $a = 4$ treatment levels of Factor A; consequently, no dummy codes are provided for Factor A as this information is implicit in the ordering of the $a = 4$ treatment levels of Factor A in the last column of Fig. 4.29.

Because there are

$$M = \frac{N!}{\prod_{i=1}^{a} n_{A_i}!} = \frac{16!}{(4!)^4} = 63{,}063{,}000$$

possible, equally-likely arrangements of the $N = 16$ univariate response measurement scores listed in Fig. 4.29, an exact permutation approach is not practical.

LAD Regression Analysis

An MRPP resampling analysis of the $N = 16$ LAD regression residuals calculated on the univariate response measurement scores listed in Fig. 4.29 yields estimated

Fig. 4.30 Observed, predicted, and residual LAD regression values for the univariate response measurement scores listed in Fig. 4.29

Object	y_i	$\tilde{y}_i$	e_i
1	5	22.00	−17.00
2	7	20.00	−13.00
3	6	8.00	−2.00
4	9	9.00	0.00
5	10	10.00	0.00
6	11	24.00	−13.00
7	12	12.00	0.00
8	8	13.00	−5.00
9	14	14.00	0.00
10	12	12.00	0.00
11	16	16.00	0.00
12	17	17.00	0.00
13	21	18.00	+3.00
14	19	16.00	+3.00
15	24	20.00	+4.00
16	22	5.00	+17.00

LAD regression coefficients of

$$\tilde{\beta}_0 = +10.00\,, \quad \tilde{\beta}_1 = +2.00\,, \quad \tilde{\beta}_2 = -2.00\,, \quad \tilde{\beta}_3 = -5.00\,,$$
$$\tilde{\beta}_4 = +8.00\,, \quad \tilde{\beta}_5 = +12.00\,, \quad \text{and} \quad \tilde{\beta}_6 = +4.00$$

for Factor A. Figure 4.30 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 16$.

Following (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 16$ LAD regression residuals listed in Fig. 4.30 yield $a = 4$ average distance-function values of

$$\xi_{A_1} = 10.3333\,, \quad \xi_{A_2} = 7.3333\,, \quad \xi_{A_3} = 0.00\,, \quad \text{and} \quad \xi_{A_4} = 7.1667\,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.30 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{A_i}}{N}\,, \qquad i = 1, \ldots, 4\,,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{4}{16}\big(10.3333 + 7.3333 + 0.00 + 7.1667\big) = 6.2083\,.$$

If all M possible arrangements of the $N = 16$ observed LAD regression residuals listed in Fig. 4.30 occur with equal chance, the approximate resampling probability value of $\delta_A = 6.2083$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{A_1} = \cdots = n_{A_4} = 4$ residuals preserved for each arrangement is

$$P(\delta \leq \delta_A | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_A}{L} = \frac{27{,}289}{1{,}000{,}000} = 0.0273 \ .$$

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 63{,}063{,}000$ δ values is $\mu_\delta = 8.2750$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{6.2083}{8.2750} = +0.2497 \ ,$$

indicating approximately 25 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP resampling analysis of the OLS regression residuals calculated on the $N = 16$ univariate response measurement scores listed in Fig. 4.29 on p. 171. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +11.6875 \ , \quad \hat{\beta}_1 = -0.2500 \ , \quad \hat{\beta}_2 = +2.00 \ , \quad \hat{\beta}_3 = +1.50 \ ,$$
$$\hat{\beta}_4 = +0.50 \ , \quad \hat{\beta}_5 = +1.7500 \ , \quad \text{and} \quad \hat{\beta}_6 = +1.00$$

for Factor A. Figure 4.31 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 16$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 16$ OLS regression residuals listed in Fig. 4.31 yield $a = 4$ average distance-function values of

$$\xi_{A_1} = 6.2083 \ , \quad \xi_{A_2} = 6.4583 \ , \quad \xi_{A_3} = 0.8750 \ , \quad \text{and} \quad \xi_{A_4} = 2.3750 \ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.31 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{A_i} - 1}{N - a} \ , \qquad i = 1, \ldots, 4 \ ,$$

Fig. 4.31 Observed, predicted, and residual LAD regression values for the univariate response measurement scores listed in Fig. 4.29

Object	y_i	$\hat{y}_i$	e_i
1	5	13.4375	−8.4375
2	7	11.9375	−4.9375
3	6	13.6875	−7.6875
4	9	14.1875	−5.1875
5	10	11.6875	−1.6875
6	11	13.1875	−2.1875
7	12	14.6875	−2.6875
8	8	13.6875	−5.6875
9	14	12.6875	+1.3125
10	12	11.4375	+0.5625
11	16	14.1875	+1.8125
12	17	14.9375	+2.0625
13	21	12.1875	+8.8125
14	19	12.4375	+6.5625
15	24	15.4375	+8.5625
16	22	13.1875	+8.8125

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{4-1}{16-4}\big(6.2083 + 6.4583 + 0.8750 + 2.3750\big) = 3.9792\ .$$

If all M possible arrangements of the $N = 16$ observed OLS regression residuals listed in Fig. 4.31 occur with equal chance, the approximate resampling probability value of $\delta_A = 3.9792$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_{A_1} = \cdots = n_{A_4} = 4$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_A}{L} = \frac{1}{1{,}000{,}000} = 0.10 \times 10^{-5}\ .$$

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{A_i}/N$ for $i = 1, \ldots, 4$ is $P = 0.0273$.

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 63{,}063{,}000$ δ values is $\mu_\delta = 68.0083$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{3.9792}{68.0083} = +0.9415\ ,$$

indicating approximately 95 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional Latin square analysis of variance calculated on the $N = 16$ univariate response measurement scores listed in Table 4.8 on p. 170 yields an observed F-ratio of $F_A = 40.7277$. Assuming independence and normality, F_A is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = a - 1 = 4 - 1 = 3$ and $\nu_2 = (a-2)(a-1) = (4-2)(4-1) = 6$ degrees of freedom. Under the null hypothesis, the observed value of $F_A = 40.7277$ yields an approximate probability value of $P = 0.2204 \times 10^{-3}$.

Analysis of Factor *B*

A design matrix of dummy codes for analyzing Factor B is given in Fig. 4.32, where the first column of 1 values provides for an intercept, the second through fourth columns contain dummy codes for Subjects, the fifth through seventh columns contain dummy codes for Factor A, and the last column lists the univariate response measurement scores ordered by the $b = 4$ treatment levels of Factor B, with the first $n_{B_1} = 4$ scores, the next $n_{B_2} = 4$ scores, the next $n_{B_3} = 4$ scores, and the last $n_{B_4} = 4$ associated with treatment levels B_1, B_2, B_3, and B_4, respectively. The MRPP regression analysis examines LAD regression residuals for possible differences among the $b = 4$ treatment levels of Factor B; consequently, no dummy codes are provided for Factor B as this information is implicit in the ordering of the $b = 4$ treatment levels of Factor B in the last column of Fig. 4.32.

Because there are

$$M = \frac{N!}{\prod_{i=1}^{b} n_{B_i}!} = \frac{16!}{(4!)^4} = 63{,}063{,}000$$

Fig. 4.32 Design matrix and univariate response measurement scores for order (B) in a Latin square design

Matrix							Score
1	0	0	0	1	0	0	10
1	1	0	0	0	1	0	12
1	0	1	0	0	0	0	6
1	0	0	1	0	0	1	22
1	0	0	0	0	0	1	21
1	1	0	0	0	0	0	7
1	0	1	0	0	1	0	16
1	0	0	1	1	0	0	8
1	0	0	0	0	0	0	5
1	1	0	0	1	0	0	11
1	0	1	0	0	0	1	24
1	0	0	1	0	1	0	17
1	0	0	0	0	1	0	14
1	1	0	0	0	0	1	19
1	0	1	0	1	0	0	12
1	0	0	1	0	0	0	9

possible, equally-likely arrangements of the $N = 16$ univariate response measurement scores listed in Fig. 4.32, an exact permutation approach is not practical.

LAD Regression Analysis

An MRPP resampling analysis of the $N = 16$ LAD regression residuals calculated on the univariate response measurement scores in Fig. 4.32 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +21.00\ , \quad \tilde{\beta}_1 = -2.00\ , \quad \tilde{\beta}_2 = +2.00\ , \quad \tilde{\beta}_3 = +1.00\ ,$$
$$\tilde{\beta}_4 = -13.00\ , \quad \tilde{\beta}_5 = -11.00\ , \quad \text{and} \quad \tilde{\beta}_6 = -7.00$$

for Factor B. Figure 4.33 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 16$.

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 16$ LAD regression residuals listed in Fig. 4.33 yield $b = 4$ average distance-function values of

$$\xi_{B_1} = 2.00\ , \quad \xi_{B_2} = 2.00\ , \quad \xi_{B_3} = 3.1667\ , \quad \text{and} \quad \xi_{B_4} = 0.00\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.33 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{B_i}}{N}\ , \qquad i = 1, \ldots, 4\ ,$$

Fig. 4.33 Observed, predicted, and residual LAD regression values for the univariate response measurement scores listed in Fig. 4.32

Object	y_i	$\tilde{y}_i$	e_i
1	10	10.00	0.00
2	12	12.00	0.00
3	6	10.00	−4.00
4	22	22.00	0.00
5	21	21.00	0.00
6	7	6.00	+1.00
7	16	16.00	0.00
8	8	11.00	−3.00
9	5	8.00	−3.00
10	11	8.00	+3.00
11	24	23.00	+1.00
12	17	15.00	+2.00
13	14	14.00	0.00
14	19	19.00	0.00
15	12	12.00	0.00
16	9	9.00	0.00

is

$$\delta_B = \sum_{i=1}^{b} C_i \xi_i = \frac{4}{16}\big(2.00 + 2.00 + 3.1667 + 0.00\big) = 1.7917 \ .$$

If all M possible arrangements of the $N = 16$ observed LAD regression residuals listed in Fig. 4.33 occur with equal chance, the approximate resampling probability value of $\delta_B = 1.7917$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{B_1} = \cdots = n_{B_4} = 4$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_B | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_B}{L} = \frac{495{,}269}{1{,}000{,}000} = 0.4953 \ .$$

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 63{,}063{,}000$ δ values is $\mu_\delta = 1.8583$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$ is

$$\Re_B = 1 - \frac{\delta_B}{\mu_\delta} = 1 - \frac{1.7917}{1.8583} = +0.0359 \ ,$$

indicating approximately 4 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP resampling analysis of the OLS regression residuals calculated on the $N = 16$ univariate response measurement scores listed in Fig. 4.29 on p. 171. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +20.6875 \ , \quad \hat{\beta}_1 = -0.2500 \ , \quad \hat{\beta}_2 = +2.00 \ , \quad \hat{\beta}_3 = +1.50 \ ,$$
$$\hat{\beta}_4 = -14.7500 \ , \quad \hat{\beta}_5 = -11.2500 \ , \quad \text{and} \quad \hat{\beta}_6 = -6.7500$$

for Factor B. Figure 4.34 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 16$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 16$ OLS regression residuals listed in Fig. 4.34 yield $b = 4$ average distance-function values of

$$\xi_{B_1} = 2.8750 \ , \quad \xi_{B_2} = 6.7083 \ , \quad \xi_{B_3} = 3.2083 \ , \quad \text{and} \quad \xi_{B_4} = 3.1250 \ .$$

Fig. 4.34 Observed, predicted, and residual OLS regression values for the univariate response measurement scores listed in Fig. 4.32

Object	y_i	$\hat{y}_i$	e_i
1	10	9.4375	+0.5625
2	12	13.6875	−1.6875
3	6	7.9375	−1.9375
4	22	22.1875	−0.1875
5	21	20.6875	+0.3125
6	7	5.6875	+1.3125
7	16	15.9375	+0.0625
8	8	10.9375	−2.9375
9	5	5.9375	−0.9375
10	11	9.1875	+1.8125
11	24	22.6875	+1.3125
12	17	15.4375	+1.5625
13	14	13.9375	+0.0625
14	19	20.4375	−1.4375
15	12	11.4375	+0.5625
16	9	7.4375	+1.5625

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.34 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{B_i} - 1}{N - b} \;, \qquad i = 1, \ldots, 4 \;,$$

is

$$\delta_B = \sum_{i=1}^{b} C_i \xi_i = \frac{4-1}{16-4}\big(2.8750 + 6.7083 + 3.2083 + 3.1250\big) = 3.9792 \;.$$

If all M possible arrangements of the $N = 16$ observed OLS regression residuals listed in Fig. 4.34 occur with equal chance, the approximate resampling probability value of $\delta_B = 3.9792$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_{B_1} = n_{B_2} = n_{B_3} = n_{B_4} = 4$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_B | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_B}{L} = \frac{378{,}875}{1{,}000{,}000} = 0.3789 \;.$$

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{B_i}/N$ for $i = 1, \ldots, 4$ is $P = 0.4953$.

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 63{,}063{,}000$ δ values is $\mu_\delta = 4.0750$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_B = 1 - \frac{\delta_B}{\mu_\delta} = 1 - \frac{1.7917}{4.0750} = +0.0235 ,$$

indicating only approximately 2 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional Latin square analysis of variance calculated on the $N = 16$ univariate response measurement scores listed in Table 4.8 on p. 170 yields an observed F-ratio of $F_B = 0.5602$. Assuming independence and normality, F_B is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = b - 1 = 4 - 1 = 3$ and $\nu_2 = (b-2)(b-1) = (4-2)(4-1) = 6$ degrees of freedom. Under the null hypothesis, the observed value of $F_B = 0.5602$ yields an approximate probability value of $P = 0.6606$. The LAD regression, OLS regression, and F-ratio probability values of $P = 0.4953$, $P = 0.3789$, and $P = 0.6606$, respectively, all indicate that the order in which the treatments were distributed did not matter.

4.3.7 Split-Plot Design

Imagine a testing experiment with two treatment factors, A and B, with a and b treatment levels, respectively, so that there are ab treatment combinations. If each testing session requires h hours of a subject's time and every subject is to be treated under all treatment conditions, each subject will require ab testing sessions and abh hours of testing time. When this is unreasonable, then with S subjects available, assign $n = S/A$ subjects to each level of Factor A and test each subject under all levels of Factor B. The design is a repeated-measures split-plot design in which subjects are randomly assigned to the a treatment levels of Factor A (i.e., plots), and each subject is then tested under all b levels of Factor B (i.e., subplots). The design is also called a mixed factorial design with one between-subjects factor (A) and one within-subjects factor (B), or an $A\times(B\times S)$ design [214].

Consider a split-plot experiment in which Factor A has $a = 3$ treatment levels, Factor B has $b = 3$ treatment levels, $n = 12$ subjects are randomly assigned to each of the $a = 3$ levels of Factor A, and each subject is tested at all $b = 3$ levels of Factor B. The design and data are adapted from Keppel and Zedeck and are given in Fig. 4.35 [215, p. 303].

Analysis of Factor *A*

A design matrix of effect codes for an MRPP regression analysis of Factor A is given in Fig. 4.36, where the first column of 1 values provides for an intercept and the second column lists the total of response measurement summations over the b

Fig. 4.35 Example univariate response measurements for a split-plot design

Factor A	Subject	Factor B B_1	B_2	B_3
A_1	S_1	53	51	35
	S_2	49	34	18
	S_3	47	44	32
	S_4	42	48	27
A_2	S_5	47	42	16
	S_6	42	33	10
	S_7	39	13	11
	S_8	37	16	6
A_3	S_9	45	35	29
	S_{10}	41	33	21
	S_{11}	38	46	30
	S_{12}	36	40	20

Fig. 4.36 Design matrix and response measurement summations for the main effects of Factor A in a split-plot design

Matrix	Sum over B
1	139
1	101
1	123
1	117
1	105
1	85
1	63
1	59
1	109
1	95
1	114
1	96

levels of Factor B (e.g., $53 + 51 + 35 = 139$). The summations are ordered by the $a = 3$ treatment levels of Factor A with the first $n_{A_1} = 4$ summations, the second $n_{A_2} = 4$ summations, and the last $n_{A_3} = 4$ summations associated with treatment levels A_1, A_2, and A_3, respectively. The MRPP regression analysis examines the $N = 12$ regression residuals for possible differences among the $a = 3$ treatment levels of Factor A; consequently, no effect codes are provided for Factor A as this information is implicit in the ordering of the $a = 3$ treatment levels of Factor A in the second column of Fig. 4.36.

An exact permutation solution is feasible for the response measurement summations listed in Fig. 4.36 since there are only

$$M = \frac{N!}{\prod_{i=1}^{a} n_{A_i}!} = \frac{12!}{(4!)^3} = 34{,}650$$

Fig. 4.37 Observed, predicted, and residual LAD regression values for the response measurement summations listed in Fig. 4.36

Object	y_i	$\tilde{y}_i$	e_i
1	139	101.00	+38.00
2	101	101.00	0.00
3	123	101.00	+22.00
4	117	101.00	+16.00
5	105	101.00	+4.00
6	85	101.00	−16.00
7	63	101.00	−38.00
8	59	101.00	−42.00
9	109	101.00	+8.00
10	95	101.00	−6.00
11	114	101.00	+13.00
12	96	101.00	−5.00

possible, equally-likely arrangements of the $N = 12$ response measurement summations for Factor A.

LAD Regression Analysis

An MRPP analysis of the $N = 12$ LAD regression residuals calculated on the response measurement summations in Fig. 4.36 yields an estimated LAD regression coefficient of $\tilde{\beta}_0 = +101.00$ for Factor A. Figure 4.37 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 12$.

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 12$ LAD regression residuals listed in Fig. 4.37 yield $a = 3$ average distance-function values of

$$\xi_{A_1} = 20.00\ , \quad \xi_{A_2} = 26.6667\ , \quad \text{and} \quad \xi_{A_3} = 11.6667\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.37 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{A_i}}{N}\ , \quad i = 1, 2, 3\ ,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{4}{12}\big(20.00 + 26.6667 + 11.6667\big) = 19.4444\ .$$

If all arrangements of the $N = 16$ observed LAD regression residuals listed in Fig. 4.37 occur with equal chance, the exact probability value of $\delta_A = 19.4444$ calculated on the $M = 34{,}650$ possible arrangements of the observed LAD regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 4$ residuals preserved for each arrangement is

$$P\left(\delta \leq \delta_A | H_0\right) = \frac{\text{number of } \delta \text{ values} \leq \delta_A}{M} = \frac{672}{34{,}650} = 0.0194 \,.$$

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 34{,}650$ δ values is $\mu_\delta = 27.00$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{19.4444}{27.00} = +0.2798 \,,$$

indicating approximately 28 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression

For comparison, consider an MRPP analysis of OLS regression residuals calculated on the response measurement summations for Factor A in Fig. 4.36. The MRPP regression analysis yields an estimated OLS regression coefficient of $\hat{\beta}_0 = +100.50$ for Factor A. Figure 4.38 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 12$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 12$ OLS regression residuals listed in Fig. 4.38 yield $a = 3$ average distance-function values of

$$\xi_{A_1} = 493.3333 \,, \quad \xi_{A_2} = 909.3333 \,, \quad \text{and} \quad \xi_{A_3} = 179.3333 \,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.38 with $v = 2$ and

Fig. 4.38 Observed, predicted, and residual OLS regression values for the response measurement summations listed in Fig. 4.36

Object	y_i	$\hat{y}_i$	e_i
1	139	100.50	+38.50
2	101	100.50	+0.50
3	123	100.50	+22.50
4	117	100.50	+16.50
5	105	100.50	+4.50
6	85	100.50	−15.50
7	63	100.50	−37.50
8	59	100.50	−41.50
9	109	100.50	+8.50
10	95	100.50	−5.50
11	114	100.50	+13.50
12	96	100.50	−4.50

treatment-group weights

$$C_i = \frac{n_{A_i} - 1}{N - a} , \qquad i = 1, 2, 3 ,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{4-1}{12-3}\big(493.3333 + 909.3333 + 179.3333\big) = 527.3333 .$$

If all arrangements of the $N = 12$ observed OLS regression residuals listed in Fig. 4.38 occur with equal chance, the exact probability value of $\delta_A = 527.3333$ computed on the $M = 34{,}650$ possible arrangements of the observed OLS regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 4$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_A}{M} = \frac{564}{34{,}650} = 0.0163 .$$

For comparison, the exact probability value based on LAD regression, $v = 1$, $M = 34{,}650$, and $C_i = n_{A_i}/N$ for $i = 1, 2, 3$ is $P = 0.0194$.

Following Eq. (4.7) on p. 126, the exact expected value of the $M = 34{,}650$ δ values is $\mu_\delta = 1{,}082.7273$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{527.3333}{1{,}082.7273} = +0.5130 ,$$

indicating approximately 51 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional split-plot analysis of variance calculated on the $N = 12$ univariate response measurement scores listed in Fig. 4.35 on p. 180 yields an observed F-ratio of $F_A = 6.7927$. Assuming independence, normality, and homogeneity of variance, F_A is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = a - 1 = 3 - 1 = 2$ and $\nu_2 = a(n-1) = 3(4-1) = 9$ degrees of freedom. Under the null hypothesis, the observed value of $F_A = 6.7927$ yields an approximate probability value of $P = 0.0159$.

Analysis of Factor *B*

A design matrix of effect codes for an MRPP regression analysis of Factor B is given in Table 4.9, where the first column of 1 values provides for an intercept, the next 11 columns contain effect codes for Subjects nested within Factor A, and the next four columns contain effect codes for the $A{\times}B$ interaction. The last column lists the $N =$

Table 4.9 Design matrix and univariate response measurement scores for the main effects of Factor B

Matrix																Score
1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	53
1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	49
1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	47
1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	42
1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	47
1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	42
1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	39
1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	37
1	0	0	0	0	0	0	0	0	1	0	0	−1	−1	0	0	45
1	0	0	0	0	0	0	0	0	0	1	0	−1	−1	0	0	41
1	0	0	0	0	0	0	0	0	0	0	1	−1	−1	0	0	38
1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	0	0	36
1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	51
1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	34
1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	44
1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	48
1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	42
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	33
1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	13
1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	16
1	0	0	0	0	0	0	0	0	1	0	0	0	0	−1	−1	35
1	0	0	0	0	0	0	0	0	0	1	0	0	0	−1	−1	33
1	0	0	0	0	0	0	0	0	0	0	1	0	0	−1	−1	46
1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	0	0	−1	−1	40
1	1	0	0	0	0	0	0	0	0	0	0	−1	0	−1	0	35
1	0	1	0	0	0	0	0	0	0	0	0	−1	0	−1	0	18
1	0	0	1	0	0	0	0	0	0	0	0	−1	0	−1	0	32
1	0	0	0	1	0	0	0	0	0	0	0	−1	0	−1	0	27
1	0	0	0	0	1	0	0	0	0	0	0	0	−1	0	−1	16
1	0	0	0	0	0	1	0	0	0	0	0	0	−1	0	−1	10
1	0	0	0	0	0	0	1	0	0	0	0	0	−1	0	−1	11
1	0	0	0	0	0	0	0	1	0	0	0	0	−1	0	−1	6
1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	29
1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	21
1	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	30
1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	1	1	1	1	20

36 univariate response measurement scores ordered by the $b = 3$ treatment levels of Factor B, with the first $n_{B_1} = 12$ scores, the next $n_{B_2} = 12$ scores, and the last $n_{B_3} = 12$ scores associated with treatment levels B_1, B_2, and B_3, respectively. The

MRPP regression analysis examines the $N = 36$ regression residuals for possible differences among the $b = 3$ treatment levels of Factor B; consequently, no effect codes are provided for Factor B as this information is implicit in the ordering of the $b = 3$ treatment levels of Factor B in the last column of Table 4.9.

Because there are

$$M = \frac{N!}{\prod_{i=1}^{b} n_{B_i}!} = \frac{36!}{(12!)^3} = 3{,}384{,}731{,}762{,}521{,}200$$

possible, equally-likely arrangements of the $N = 36$ univariate response measurement scores listed in Table 4.9, an exact permutation approach is not possible.

LAD Regression Analysis

An MRPP resampling analysis of the LAD regression residuals calculated on the $N = 36$ univariate response measurement scores in Table 4.9 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +35.50\ , \quad \tilde{\beta}_1 = +9.8333\ , \quad \tilde{\beta}_2 = -7.1667\ , \quad \tilde{\beta}_3 = +2.8333\ ,$$
$$\tilde{\beta}_4 = +5.8333\ , \quad \tilde{\beta}_5 = +4.8333\ , \quad \tilde{\beta}_6 = -0.1667\ , \quad \tilde{\beta}_7 = -20.1667\ ,$$
$$\tilde{\beta}_8 = -17.1667\ , \quad \tilde{\beta}_9 = +2.8333\ , \quad \tilde{\beta}_{10} = +0.8333\ , \quad \tilde{\beta}_{11} = +9.8333\ ,$$
$$\tilde{\beta}_{12} = +0.6667\ , \quad \tilde{\beta}_{13} = +6.6667\ , \quad \tilde{\beta}_{14} = +5.6667\ , \quad \text{and}$$
$$\tilde{\beta}_{15} = -2.3333$$

for Factor B. Figure 4.39 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 36$.

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 36$ LAD regression residuals listed in Fig. 4.39 yield $b = 3$ average distance-function values of

$$\xi_{B_1} = 8.6061\ , \quad \xi_{B_2} = 1.3182\ , \quad \text{and} \quad \xi_{B_3} = 13.5606\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.39 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{B_i}}{N}\ , \qquad i = 1, 2, 3\ ,$$

Fig. 4.39 Observed, predicted, and residual LAD regression values for the univariate response measurement scores listed in Table 4.9

Object	y_i	$\tilde{y}_i$	e_i
1	53	46.00	+7.00
2	49	29.00	+20.00
3	47	39.00	+8.00
4	42	42.00	0.00
5	47	47.00	0.00
6	42	42.00	0.00
7	39	22.00	+17.00
8	37	25.00	+12.00
9	45	31.00	+14.00
10	41	29.00	+12.00
11	38	38.00	0.00
12	36	36.00	0.00
13	51	51.00	0.00
14	34	34.00	0.00
15	44	44.00	0.00
16	48	47.00	+1.00
17	42	38.00	+4.00
18	33	33.00	0.00
19	13	13.00	0.00
20	16	16.00	0.00
21	35	35.00	0.00
22	33	33.00	0.00
23	46	42.00	+4.00
24	40	40.00	0.00
25	35	39.00	−4.00
26	18	22.00	−4.00
27	32	32.00	0.00
28	27	35.00	−8.00
29	16	36.00	−20.00
30	10	31.00	−21.00
31	11	11.00	0.00
32	6	14.00	−8.00
33	29	49.00	−20.00
34	21	47.00	−26.00
35	30	56.00	−26.00
36	20	54.00	−34.00

is

$$\delta_B = \sum_{i=1}^{b} C_i \xi_i = \frac{12}{36}\big(8.6061 + 1.3182 + 13.5606\big) = 7.8283 \, .$$

If all M possible arrangements of the $N = 36$ observed LAD regression residuals listed in Fig. 4.39 occur with equal chance, the approximate resampling probability value of $\delta_B = 7.8283$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{B_1} = n_{B_2} = n_{B_3} = 12$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_B | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_B}{L} = \frac{0}{1{,}000{,}000} = 0.00 \; ,$$

which may be interpreted as a probability of less than one in a million.

When M is very large and the probability of an observed δ is extremely small, as in this case, resampling permutation procedures sometimes result in zero probability, even with $L = 1{,}000{,}000$ random arrangements of the observed regression residuals. A reanalysis of Factor B using $L = 10{,}000{,}000$ random arrangements of the observed data yielded an identical resampling probability value of $P = 0.00$. Moment-approximation permutation procedures, described briefly in Chap. 1, Sect. 1.2.2, can often provide results in these extreme situations. The moment-approximation of a test statistic requires computation of the exact moments of the test statistic, assuming equally-likely arrangements of the observed regression residuals [284, 300]. Usually, the first three exact moments are used: the exact mean, μ_δ, the exact variance, σ^2_δ, and the exact skewness, γ_δ, of δ. The three moments are then used to fit a specified distribution, such as a Pearson type III distribution, that approximates the underlying discrete permutation distribution and provides an approximate probability value. For Factor B, a moment-approximation procedure yields $\delta_B = 7.8283$, $\mu_\delta = 12.5460$, $\sigma^2_\delta = 0.1675$, $\gamma_\delta = -1.3580$, a standardized test statistic of

$$T_B = \frac{\delta_B - \mu_\delta}{\sigma_\delta} = \frac{7.8283 - 12.5460}{\sqrt{0.1675}} = -11.5272 \; ,$$

and a Pearson type III approximate probability value of $P = 0.1495{\times}10^{-6}$.

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 12.5460$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_B = 1 - \frac{\delta_B}{\mu_\delta} = 1 - \frac{7.8283}{12.5460} = +0.3760 \; ,$$

indicating approximately 38 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP analysis of OLS regression residuals calculated on the $N = 36$ response measurement summations for Factor B in Table 4.9

on p. 184. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +33.50\,, \quad \hat{\beta}_1 = +12.8333\,, \quad \hat{\beta}_2 = +0.1667\,, \quad \hat{\beta}_3 = +7.50\,,$$

$$\hat{\beta}_4 = +5.50\,, \quad \hat{\beta}_5 = +1.50\,, \quad \hat{\beta}_6 = -5.1667\,, \quad \hat{\beta}_7 = -12.50\,,$$

$$\hat{\beta}_8 = -13.8333\,, \quad \hat{\beta}_9 = +2.8333\,, \quad \hat{\beta}_{10} = -1.8333\,, \quad \hat{\beta}_{11} = +4.50\,,$$

$$\hat{\beta}_{12} = -1.7500\,, \quad \hat{\beta}_{13} = +5.7500\,, \quad \hat{\beta}_{14} = +1.50\,, \quad \text{and}$$

$$\hat{\beta}_{15} = -2.7500$$

for Factor B. Figure 4.40 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 36$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 36$ OLS regression residuals listed in Fig. 4.40 yield $b = 3$ average distance-function values of

$$\xi_{B_1} = 30.2727\,, \quad \xi_{B_2} = 46.4394\,, \quad \text{and} \quad \xi_{B_3} = 16.5606\,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.40 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{B_i} - 1}{N - b}\,, \qquad i = 1, 2, 3\,,$$

is

$$\delta_B = \sum_{i=1}^{b} C_i \xi_i = \frac{12 - 1}{36 - 3}\big(30.2727 + 46.4394 + 16.5606\big) = 31.0909\,.$$

If all M possible arrangements of the $N = 36$ observed OLS regression residuals listed in Fig. 4.40 occur with equal chance, the approximate resampling probability value of $\delta_B = 31.0909$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_{B_1} = n_{B_2} = n_{B_3} = 12$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_B | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_B}{L} = \frac{0}{1{,}000{,}000} = 0.00\,,$$

i.e., a probability of less than one in a million. For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{B_i}/N$ for $i = 1, 2, 3$ is also $P = 0.00$.

As with the analysis of the LAD regression residuals listed in Fig. 4.39 on p. 186, when M is large and the probability of an observed δ is very small, an alternative

Fig. 4.40 Observed, predicted, and residual OLS regression values for the response measurement scores listed in Table 4.9

Object	y_i	$\hat{y}_i$	e_i
1	53	44.5833	+8.4167
2	49	31.9167	+17.0833
3	47	39.2500	+7.7500
4	42	37.2500	+4.7500
5	47	40.7500	+6.2500
6	42	34.0833	+7.9167
7	39	26.7500	+12.2500
8	37	25.4167	+11.5833
9	45	32.3333	+12.6667
10	41	27.6667	+13.3333
11	38	34.0000	+4.0000
12	36	28.0000	+8.0000
13	51	47.8333	+3.1667
14	34	35.1667	−1.1667
15	44	42.5000	+1.5000
16	48	40.5000	+7.5000
17	42	32.2500	+9.7500
18	33	25.5833	+7.4167
19	13	18.2500	−5.2500
20	16	16.9167	−0.9167
21	35	37.5833	−2.5833
22	33	32.9167	+0.0833
23	46	39.2500	+6.7500
24	40	33.2500	+6.7500
25	35	46.5833	−11.5833
26	18	33.9167	−15.9167
27	32	41.2500	−9.2500
28	27	39.2500	−12.2500
29	16	32.0000	−16.0000
30	10	25.3333	−15.3333
31	11	18.0000	−7.0000
32	6	16.6667	−10.6667
33	29	39.0833	−10.0833
34	21	34.4167	−13.4167
35	30	40.7500	−10.7500
36	20	34.7500	−14.7500

moment procedure based on the exact mean, μ_δ, exact variance, σ_δ^2, and exact skewness, γ_δ, of δ can be employed to obtain approximate probability values; see Chap. 1, Sect. 1.2.2. For Factor B, a moment-approximation procedure yields $\delta_B = 31.0909$, $\mu_\delta = 199.2857$, $\sigma_\delta^2 = 134.8578$, $\gamma_\delta = -1.7697$, a standardized test statistic of

$$T_B = \frac{\delta_B - \mu_\delta}{\sigma_\delta} = \frac{31.0909 - 199.2857}{\sqrt{134.8578}} = -14.4835 \,,$$

and a Pearson type III approximate probability value of $P = 0.5420 \times 10^{-7}$.

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 199.2857$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_B = 1 - \frac{\delta_B}{\mu_\delta} = 1 - \frac{31.0909}{199.2857} = +0.8440 ,$$

indicating approximately 84 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional split-plot analysis of variance calculated on the $N = 36$ univariate response measurement scores listed in Fig. 4.35 on p. 180 yields an observed F-ratio of $F_B = 52.1842$. Assuming independence and normality, F_B is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = b - 1 = 3 - 1 = 2$ and $\nu_2 = a(n-1)(b-1) = 3(4-1)(3-1) = 18$ degrees of freedom. Under the null hypothesis, the observed value of $F_B = 52.1842$ yields an approximate probability value of $P = 0.3224 \times 10^{-7}$.

Analysis of the $A \times B$ Interaction

A design matrix of effect codes for an MRPP regression analysis of the $A \times B$ interaction is given in Table 4.10, where the first column of 1 values provides for an intercept, the next 11 columns contain effect codes for Subjects nested within Factor A, and the next two columns contain effect codes for Factor B. The last column lists the $N = 36$ univariate response measurement scores ordered by the $ab = (3)(3) = 9$ levels of the $A \times B$ interaction. The MRPP regression analysis examines the $N = 36$ regression residuals for possible differences among the nine treatment levels of the $A \times B$ interaction; consequently, no effect codes are provided for the $A \times B$ interaction as this information is implicit in the ordering of the treatment levels of the $A \times B$ interaction in the last column of Table 4.10.

Because there are

$$M = \frac{N!}{\prod_{i=1}^{ab} n_{(A \times B)_i}!} = \frac{36!}{(4!)^9} = 140{,}810{,}154{,}080{,}474{,}667{,}338{,}550{,}000{,}000$$

possible, equally-likely arrangements of the $N = 36$ univariate response measurement scores listed in Table 4.10, an exact permutation approach is not possible.

Table 4.10 Design matrix and univariate response measurement scores for the interaction effects of Factors *A* and *B*

Matrix														Score
1	1	0	0	0	0	0	0	0	0	0	0	1	0	53
1	0	1	0	0	0	0	0	0	0	0	0	1	0	49
1	0	0	1	0	0	0	0	0	0	0	0	1	0	47
1	0	0	0	1	0	0	0	0	0	0	0	1	0	42
1	0	0	0	0	1	0	0	0	0	0	0	1	0	47
1	0	0	0	0	0	1	0	0	0	0	0	1	0	42
1	0	0	0	0	0	0	1	0	0	0	0	1	0	39
1	0	0	0	0	0	0	0	1	0	0	0	1	0	37
1	0	0	0	0	0	0	0	0	1	0	0	1	0	45
1	0	0	0	0	0	0	0	0	0	1	0	1	0	41
1	0	0	0	0	0	0	0	0	0	0	1	1	0	38
1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	1	0	36
1	1	0	0	0	0	0	0	0	0	0	0	0	1	51
1	0	1	0	0	0	0	0	0	0	0	0	0	1	34
1	0	0	1	0	0	0	0	0	0	0	0	0	1	44
1	0	0	0	1	0	0	0	0	0	0	0	0	1	48
1	0	0	0	0	1	0	0	0	0	0	0	0	1	42
1	0	0	0	0	0	1	0	0	0	0	0	0	1	33
1	0	0	0	0	0	0	1	0	0	0	0	0	1	13
1	0	0	0	0	0	0	0	1	0	0	0	0	1	16
1	0	0	0	0	0	0	0	0	1	0	0	0	1	35
1	0	0	0	0	0	0	0	0	0	1	0	0	1	33
1	0	0	0	0	0	0	0	0	0	0	1	0	1	46
1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	0	1	40
1	1	0	0	0	0	0	0	0	0	0	0	−1	−1	35
1	0	1	0	0	0	0	0	0	0	0	0	−1	−1	18
1	0	0	1	0	0	0	0	0	0	0	0	−1	−1	32
1	0	0	0	1	0	0	0	0	0	0	0	−1	−1	27
1	0	0	0	0	1	0	0	0	0	0	0	−1	−1	16
1	0	0	0	0	0	1	0	0	0	0	0	−1	−1	10
1	0	0	0	0	0	0	1	0	0	0	0	−1	−1	11
1	0	0	0	0	0	0	0	1	0	0	0	−1	−1	6
1	0	0	0	0	0	0	0	0	1	0	0	−1	−1	29
1	0	0	0	0	0	0	0	0	0	1	0	−1	−1	21
1	0	0	0	0	0	0	0	0	0	0	1	−1	−1	30
1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	20

LAD Regression Analysis

An MRPP resampling analysis of the $N = 36$ LAD regression residuals calculated on the univariate response measurement scores in Table 4.10 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +34.00\,,\quad \tilde{\beta}_1 = +12.6667\,,\quad \tilde{\beta}_2 = -3.3333\,,\quad \tilde{\beta}_3 = +6.6667\,,$$
$$\tilde{\beta}_4 = +4.6667\,,\quad \tilde{\beta}_5 = +4.6667\,,\quad \tilde{\beta}_6 = -4.3333\,,\quad \tilde{\beta}_7 = -11.3333\,,$$
$$\tilde{\beta}_8 = -16.3333\,,\quad \tilde{\beta}_9 = +2.6667\,,\quad \tilde{\beta}_{10} = -1.3333\,,\quad \tilde{\beta}_{11} = +7.6667\,,$$
$$\tilde{\beta}_{12} = +8.3333\,,\quad \text{and}\quad \tilde{\beta}_{13} = +3.3333$$

for the interaction of Factors A and B. Figure 4.41 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 36$.

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 36$ LAD regression residuals listed in Fig. 4.41 yield $ab = (3)(3) = 9$ average distance-function values of

$$\xi_{(A\times B)_1} = 7.50\,,\quad \xi_{(A\times B)_2} = 6.1667\,,\quad \xi_{(A\times B)_3} = 6.6667\,,$$
$$\xi_{(A\times B)_4} = 3.1667\,,\quad \xi_{(A\times B)_5} = 7.3333\,,\quad \xi_{(A\times B)_6} = 5.6667\,,$$
$$\xi_{(A\times B)_7} = 2.00\,,\quad \xi_{(A\times B)_8} = 6.8333\,,\quad \text{and}\quad \xi_{(A\times B)_9} = 2.00\,.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.41 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{(A\times B)_i}}{N}\,,\qquad i = 1, \ldots, 9\,,$$

is

$$\delta_{A\times B} = \sum_{i=1}^{ab} C_i \xi_i = \frac{4}{36}\big(7.50 + 6.1667 + 6.6667 + \cdots + 6.8333 + 2.00\big) = 5.2593\,.$$

If all M possible arrangements of the $N = 36$ observed LAD regression residuals listed in Fig. 4.41 occur with equal chance, the approximate resampling probability value of $\delta_{A\times B} = 5.2593$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{(A\times B)_1} = \cdots = n_{(A\times B)_9} = 4$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_{A\times B}|H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_{A\times B}}{L} = \frac{140{,}219}{1{,}000{,}000} = 0.1402\,.$$

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 5.6825$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure

Fig. 4.41 Observed, predicted, and residual LAD regression values for the univariate response measurement scores listed in Table 4.10

Object	y_i	$\tilde{y}_i$	e_i
1	53	55.00	−2.00
2	49	39.00	+10.00
3	47	49.00	−2.00
4	42	47.00	−5.00
5	47	47.00	0.00
6	42	38.00	+4.00
7	39	31.00	+8.00
8	37	26.00	+11.00
9	45	45.00	0.00
10	41	41.00	0.00
11	38	50.00	−12.00
12	36	40.00	−4.00
13	51	50.00	+1.00
14	34	34.00	0.00
15	44	44.00	0.00
16	48	42.00	+6.00
17	42	42.00	0.00
18	33	33.00	0.00
19	13	26.00	−13.00
20	16	21.00	−5.00
21	35	40.00	−5.00
22	33	36.00	−3.00
23	46	45.00	+1.00
24	40	35.00	+5.00
25	35	35.00	0.00
26	18	19.00	−1.00
27	32	29.00	+3.00
28	27	27.00	0.00
29	16	27.00	−11.00
30	10	18.00	−8.00
31	11	11.00	0.00
32	6	6.00	0.00
33	29	25.00	+4.00
34	21	21.00	0.00
35	30	30.00	0.00
36	20	20.00	0.00

of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{A\times B} = 1 - \frac{\delta_{A\times B}}{\mu_\delta} = 1 - \frac{5.2593}{5.6825} = +0.0745 \,,$$

indicating approximately 7 % agreement between the observed and predicted y values above that expected by chance.

OLS Regression Analysis

For comparison, consider an MRPP analysis of OLS regression residuals calculated on the $N = 36$ response measurement scores for the $A\times B$ interaction in Table 4.10. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +33.50\ , \quad \hat{\beta}_1 = +12.8333\ , \quad \hat{\beta}_2 = +0.1667\ , \quad \hat{\beta}_3 = +7.50\ ,$$
$$\hat{\beta}_4 = +5.50\ , \quad \hat{\beta}_5 = +1.50\ , \quad \hat{\beta}_6 = -5.1667\ , \quad \hat{\beta}_7 = -12.50\ ,$$
$$\hat{\beta}_8 = -13.8333\ , \quad \hat{\beta}_9 = +2.8333\ , \quad \hat{\beta}_{10} = -1.8333\ , \quad \hat{\beta}_{11} = +4.50\ ,$$
$$\hat{\beta}_{12} = +9.50\ , \quad \text{and} \quad \hat{\beta}_{13} = +2.7500$$

for the interaction of Factors A and B. Figure 4.42 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 36$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 36$ OLS regression residuals listed in Fig. 4.42 yield $ab = (3)(3) = 9$ average distance-function values of

$$\xi_{(A\times B)_1} = 56.2037\ , \quad \xi_{(A\times B)_2} = 16.6481\ , \quad \xi_{(A\times B)_3} = 38.1481\ ,$$
$$\xi_{(A\times B)_4} = 26.4259\ , \quad \xi_{(A\times B)_5} = 98.8148\ , \quad \xi_{(A\times B)_6} = 45.0370\ ,$$
$$\xi_{(A\times B)_7} = 15.2593\ , \quad \xi_{(A\times B)_8} = 35.7593\ , \quad \text{and} \quad \xi_{(A\times B)_9} = 9.7037\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.42 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{(A\times B)_i} - 1}{N - ab}\ , \qquad i = 1, \ldots, 9\ ,$$

is

$$\delta_{A\times B} = \sum_{i=1}^{ab} C_i \xi_i = \frac{4-1}{36-9}\big(56.2037 + 16.6481 + 38.1481 + \cdots + 35.7593 + 9.7037\big) = 38.00\ .$$

If all M possible arrangements of the $N = 36$ observed OLS regression residuals listed in Fig. 4.42 occur with equal chance, the approximate resampling probability value of $\delta_{A\times B} = 38.00$ calculated on $L = 1{,}000{,}000$ random arrangements of

Fig. 4.42 Observed, predicted, and residual OLS regression values for the univariate response measurement scores listed in Table 4.10

Object	y_i	$\hat{y}_i$	e_i
1	53	55.8333	−2.8333
2	49	43.1667	+5.8333
3	47	50.5000	−3.5000
4	42	48.5000	−6.5000
5	47	44.5000	+2.5000
6	42	37.8333	+4.1667
7	39	30.5000	+8.5000
8	37	29.1667	+7.8333
9	45	45.8333	−0.8333
10	41	41.1667	−0.1667
11	38	47.5000	−9.5000
12	36	41.5000	−5.5000
13	51	49.0833	+1.9167
14	34	36.4167	−2.4167
15	44	43.7500	+0.2500
16	48	41.7500	+6.2500
17	42	37.7500	+4.2500
18	33	31.0833	+1.9167
19	13	23.7500	−10.7500
20	16	22.4167	−6.4167
21	35	39.0833	−4.0833
22	33	34.4167	−1.4167
23	46	40.7500	+5.2500
24	40	34.7500	+5.2500
25	35	34.0833	+0.9167
26	18	21.4167	−3.4167
27	32	28.7500	+3.2500
28	27	26.7500	+0.2500
29	16	22.7500	−6.7500
30	10	16.0833	−6.0833
31	11	8.7500	+2.2500
32	6	7.4167	−1.4167
33	29	24.0833	+4.9167
34	21	19.4167	+1.5833
35	30	25.7500	+4.2500
36	20	19.7500	+0.2500

the observed OLS regression residuals with $n_{(A\times B)_1} = \cdots = n_{(A\times B)_9} = 4$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_{A\times B}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{A\times B}}{L} = \frac{72{,}276}{1{,}000{,}000} = 0.0723\ .$$

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{(A\times B)_i}/N$ for $i = 1, \ldots, 9$ is $P = 0.1402$.

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 47.6286$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{A\times B} = 1 - \frac{\delta_{A\times B}}{\mu_\delta} = 1 - \frac{38.00}{47.6286} = +0.2022 ,$$

indicating approximately 20 % agreement between the observed and predicted y values above that expected by chance.

Conventional ANOVA Analysis

A conventional split-plot analysis of variance calculated on the $N = 36$ univariate response measurement scores listed in Fig. 4.35 on p. 180 yields an observed F-ratio of $F_{A\times B} = 2.8114$. Assuming independence, normality, and homogeneity of variance, $F_{A\times B}$ is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = (a-1)(b-1) = (3-1)(3-1) = 4$ and $\nu_2 = a(n-1)(b-1) = 3(4-1)(3-1) = 18$ degrees of freedom. Under the null hypothesis, the observed value of $F_{A\times B} = 2.8114$ yields an approximate probability value of $P = 0.0565$, which is similar to the probability value of $P = 0.0723$ obtained with the OLS regression analysis.

4.3.8 Nested Design

It is sometimes necessary to compare treatment groups when one independent variable is nested under a second independent variable. Two-factor nested analysis-of-variance designs occur whenever one factor is not completely crossed with the second factor. Consider a nested design to compare $a = 3$ levels of Factor A on scores obtained from $b = 3$ levels of Factor B, with B_1, B_2, and B_3 of Factor B in level A_1; B_4, B_5, and B_6 of Factor B in level A_2, and B_7, B_8, and B_9 of Factor B in level A_3. Thus, Factor B is said to be nested under Factor A. The univariate data for this example are listed in Table 4.11 for a sample of $n = 4$ objects randomly chosen from each of the $ab = (3)(3) = 9$ levels of Factors A and B.

Table 4.11 Example univariate response measurement scores for a nested design with $b = 3$ levels of Factor B nested under $a = 3$ levels of Factor A

A_1			A_2			A_3		
B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9
29	30	28	27	33	30	31	27	35
31	32	30	29	35	32	33	29	37
31	32	30	29	35	36	33	29	37
33	34	32	31	37	30	35	31	39

Table 4.12 Design matrix and univariate response measurement scores for an analysis of Factor A with Factor B nested under Factor A

Level A_1				Level A_2				Level A_3			
Matrix			Score	Matrix			Score	Matrix			Score
1	1	0	29	1	0	1	27	1	−1	−1	31
1	1	0	31	1	0	1	29	1	−1	−1	33
1	1	0	31	1	0	1	29	1	−1	−1	33
1	1	0	33	1	0	1	31	1	−1	−1	35
1	1	0	30	1	0	1	33	1	−1	−1	27
1	1	0	32	1	0	1	35	1	−1	−1	29
1	1	0	32	1	0	1	35	1	−1	−1	29
1	1	0	34	1	0	1	37	1	−1	−1	31
1	1	0	28	1	0	1	30	1	−1	−1	35
1	1	0	30	1	0	1	32	1	−1	−1	37
1	1	0	30	1	0	1	32	1	−1	−1	37
1	1	0	32	1	0	1	34	1	−1	−1	39

Analysis of Factor A

A design matrix of effect codes for an MRPP regression analysis of Factor A is given in Table 4.12, where the first column of 1 values provides for an intercept, the next two columns contain the effect codes for Factor B, and the third column contains the univariate response measurement scores listed according to the original random assignment of the $n = 36$ objects to the $a = 3$ levels of Factor A with the first $n_{A_1} = 12$ scores, the next $n_{A_2} = 12$ scores, and the last $n_{A_3} = 12$ scores associated with the $a = 3$ levels of Factor A, respectively. The MRPP regression analysis examines the $N = 36$ regression residuals for possible differences among the $a = 3$ treatment levels of Factor A; consequently, no effect codes are provided for Factor A as this information is implicit in the ordering of the $a = 3$ levels of Factor A in the rightmost columns of Table 4.12.

Because there are

$$M = \frac{N!}{\prod_{i=1}^{a} n_{A_i}!} = \frac{36!}{(12!)^3} = 3{,}384{,}731{,}762{,}521{,}200$$

possible, equally-likely arrangements of the $N = 36$ univariate response measurement scores listed in Table 4.11, an exact permutation approach is not possible.

LAD Regression Analysis

An MRPP resampling analysis of the $N = 36$ LAD regression residuals calculated on the univariate response measurement scores listed in Table 4.12 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +32.00\,, \quad \tilde{\beta}_1 = -1.00\,, \quad \text{and} \quad \tilde{\beta}_2 = 0.00$$

Fig. 4.43 Observed, predicted, and residual LAD regression values for the nested response measurement scores listed in Table 4.12

Object	y_i	$\tilde{y}_i$	e_i
1	29	31.00	−2.00
2	31	31.00	0.00
3	31	31.00	0.00
4	33	31.00	+2.00
5	30	31.00	−1.00
6	32	31.00	+1.00
7	32	31.00	+1.00
8	34	31.00	+3.00
9	28	31.00	−3.00
10	30	31.00	−1.00
11	30	31.00	−1.00
12	32	31.00	+1.00
13	27	32.00	−5.00
14	29	32.00	−3.00
15	29	32.00	−3.00
16	31	32.00	−1.00
17	33	32.00	+1.00
18	35	32.00	+3.00
19	35	32.00	+3.00
20	37	32.00	+5.00
21	30	32.00	−2.00
22	32	32.00	0.00
23	32	32.00	0.00
24	34	32.00	+2.00
25	31	33.00	−2.00
26	33	33.00	0.00
27	33	33.00	0.00
28	35	33.00	+2.00
29	27	33.00	−6.00
30	29	33.00	−4.00
31	29	33.00	−4.00
32	31	33.00	−2.00
33	35	33.00	+2.00
34	37	33.00	+4.00
35	37	33.00	+4.00
36	39	33.00	+6.00

for Factor A. Figure 4.43 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 36$.

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 36$ LAD regression residuals listed in Fig. 4.43 yield $a = 3$ average distance-function values of

$$\xi_{A_1} = 2.00 \;, \quad \xi_{A_2} = 3.5152 \;, \quad \text{and} \quad \xi_{A_3} = 4.4242 \;.$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.43 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{A_i}}{N} \;, \qquad i = 1, 2, 3 \;,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{12}{36}\big(2.00 + 3.5152 + 4.4242\big) = 3.3131 \;.$$

If all M possible arrangements of the $N = 36$ observed LAD regression residuals listed in Fig. 4.43 occur with equal chance, the approximate resampling probability value of $\delta_A = 3.3131$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 12$ residuals preserved for each arrangement is

$$P\big(\delta \le \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \le \delta_A}{L} = \frac{704{,}848}{1{,}000{,}000} = 0.7048 \;.$$

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 3.2508$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size between the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{3.3131}{3.2508} = -0.0192 \;,$$

indicating slightly less than chance agreement between the observed and predicted y values.

OLS Regression Analysis

For comparison, consider an MRPP resampling analysis of OLS regression residuals calculated on the $N = 36$ univariate response measurement scores listed in Table 4.12. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +32.00 \;, \quad \hat{\beta}_1 = -1.00 \;, \quad \text{and} \quad \hat{\beta}_2 = 0.00$$

for Factor A. Figure 4.44 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 36$.[4]

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 36$ OLS regression residuals listed in

[4]Note that in the case of Factor A, LAD regression and OLS regression yield the same regression coefficients. Therefore, the observed regression residuals are the same for both analyses.

Fig. 4.44 Observed, predicted, and residual OLS regression values for the nested response measurement scores listed in Table 4.12

Object	y_i	$\hat{y}_i$	e_i
1	29	31.00	−2.00
2	31	31.00	0.00
3	31	31.00	0.00
4	33	31.00	+2.00
5	30	31.00	−1.00
6	32	31.00	+1.00
7	32	31.00	+1.00
8	34	31.00	+3.00
9	28	31.00	−3.00
10	30	31.00	−1.00
11	30	31.00	−1.00
12	32	31.00	+1.00
13	27	32.00	−5.00
14	29	32.00	−3.00
15	29	32.00	−3.00
16	31	32.00	−1.00
17	33	32.00	+1.00
18	35	32.00	+3.00
19	35	32.00	+3.00
20	37	32.00	+5.00
21	30	32.00	−2.00
22	32	32.00	0.00
23	32	32.00	0.00
24	34	32.00	+2.00
25	31	33.00	−2.00
26	33	33.00	0.00
27	33	33.00	0.00
28	35	33.00	+2.00
29	27	33.00	−6.00
30	29	33.00	−4.00
31	29	33.00	−4.00
32	31	33.00	−2.00
33	35	33.00	+2.00
34	37	33.00	+4.00
35	37	33.00	+4.00
36	39	33.00	+6.00

Fig. 4.44 yield $a = 3$ average distance-function values of

$$\xi_{A_1} = 5.8182 \, , \quad \xi_{A_2} = 17.4545 \, , \quad \text{and} \quad \xi_{A_3} = 27.6364 \, .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.44 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{A_i} - 1}{N - a} \, , \qquad i = 1, 2, 3 \, ,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{12-1}{36-3}(5.8182 + 17.4545 + 27.6364) = 16.9697.$$

If all M possible arrangements of the $N = 36$ observed OLS regression residuals listed in Fig. 4.44 occur with equal chance, the approximate resampling probability value of $\delta_A = 16.9697$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_{A_1} = n_{A_2} = n_{A_3} = 12$ residuals preserved for each arrangement is

$$P\big(\delta \leq \delta_A | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_A}{L} = \frac{1{,}000{,}000}{1{,}000{,}000} = 1.00 \ .$$

A reanalysis of the data based on $L = 10{,}000{,}000$ random arrangements of the $N = 36$ observed regression residuals listed in Fig. 4.44 with $n_{A_1} = n_{A_2} = n_{A_3} = 12$ residuals preserved for each arrangement also yields an approximate resampling probability value of $P = 1.00$.

A probability value of $P = 1.00$ is not very informative. In such cases, an alternative moment procedure based on the exact mean, μ_δ, exact variance, σ_δ^2, and exact skewness, γ_δ, of δ can be employed to obtain approximate probability values; see Chap. 1, Sect. 1.2.2. For Factor A, a moment-approximation procedure yields $\delta_A = 16.9697$, $\mu_\delta = 16.00$, $\sigma_\delta^2 = 0.8472$, $\gamma_\delta = -1.7012$, an observed standardized test statistic of

$$T_B = \frac{\delta_B - \mu_\delta}{\sigma_\delta} = \frac{16.9697 - 16.00}{\sqrt{0.8472}} = +0.0535 \ ,$$

and a Pearson type III approximate probability value of $P = 0.9487$.

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{A_i}/N$ for $i = 1, \ldots, a$ is $P = 0.7048$.

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 16.00$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{16.9697}{16.00} = -0.0606 \ ,$$

indicating slightly less than chance agreement between the observed and predicted y values.

Conventional ANOVA Analysis

A conventional fixed-effects nested analysis of variance calculated on the $N = 36$ response measurement scores for Factor A listed in Table 4.11 on p. 196 yields an observed F-ratio of $F_A = 3.6818$. Assuming independence, normality, and homogeneity of variance, F_A is approximately distributed as Snedecor's F under the null

Fig. 4.45 Design matrix and univariate response measurement scores for an analysis of Factor *B* with Factor *B* nested under Factor *A*

Matrix							Score
1	1	0	1	0	0	0	29
1	1	0	1	0	0	0	31
1	1	0	1	0	0	0	31
1	1	0	1	0	0	0	33
1	1	0	0	0	1	0	30
1	1	0	0	0	1	0	32
1	1	0	0	0	1	0	32
1	1	0	0	0	1	0	34
1	1	0	−1	0	−1	0	28
1	1	0	−1	0	−1	0	30
1	1	0	−1	0	−1	0	30
1	1	0	−1	0	−1	0	32
1	0	1	0	1	0	0	27
1	0	1	0	1	0	0	29
1	0	1	0	1	0	0	29
1	0	1	0	1	0	0	31
1	0	1	0	0	0	1	33
1	0	1	0	0	0	1	35
1	0	1	0	0	0	1	35
1	0	1	0	0	0	1	37
1	0	1	0	−1	0	−1	30
1	0	1	0	−1	0	−1	32
1	0	1	0	−1	0	−1	36
1	0	1	0	−1	0	−1	30
1	−1	−1	−1	−1	0	0	31
1	−1	−1	−1	−1	0	0	33
1	−1	−1	−1	−1	0	0	33
1	−1	−1	−1	−1	0	0	35
1	−1	−1	0	0	−1	−1	27
1	−1	−1	0	0	−1	−1	29
1	−1	−1	0	0	−1	−1	29
1	−1	−1	0	0	−1	−1	31
1	−1	−1	1	1	1	1	35
1	−1	−1	1	1	1	1	37
1	−1	−1	1	1	1	1	37
1	−1	−1	1	1	1	1	39

hypothesis with $\nu_1 = a - 1 = 3 - 1 = 2$ and $\nu_2 = ab(n-1) = (3)(3)(4-1) = 27$ degrees of freedom. Under the null hypothesis, the observed value of $F_A = 3.6818$ yields an approximate probability value of $P = 0.0386$.

Analysis of Factor $B|A$

A design matrix of effect codes for an MRPP regression analysis of Factor *B*, nested under Factor *A*, is given in Fig. 4.45, where the first column of 1 values provides for an intercept, the next two columns contain effect codes for Factor *A*, the next four columns contain effect codes for the *A*×*B* interaction, and the last column contains the univariate response measurement scores listed according to the $b = 3$ levels of

Factor B with the first $n_{B|A_1} = 12$ scores, the next $n_{B|A_2} = 12$ scores, and the last $n_{B|A_3} = 12$ scores associated with the $b = 3$ levels of Factor B, respectively. The MRPP regression analysis examines the $N = 36$ regression residuals for possible differences among the $b = 3$ treatment levels of Factor B; consequently, no effect codes are provided for Factor B as this information is implicit in the ordering of the $b = 3$ levels of Factor B in the last column of Fig. 4.45.

LAD Regression Analysis

Again, because there are

$$M = \frac{N!}{\prod_{i=1}^{b} n_{B|A_i}!} = \frac{36!}{(12!)^3} = 3{,}384{,}731{,}762{,}521{,}200$$

possible, equally-likely arrangements of the $N = 36$ response measurement scores listed in Fig. 4.45, an exact permutation approach is not possible. An MRPP resampling analysis of the $N = 36$ LAD regression residuals calculated on the univariate response measurement scores in Fig. 4.45 yields estimated LAD regression coefficients of

$$\tilde{\beta}_0 = +32.00\ , \quad \tilde{\beta}_1 = -1.00\ , \quad \tilde{\beta}_2 = 0.6667\ , \quad \tilde{\beta}_3 = +1.00\ ,$$
$$\tilde{\beta}_4 = -1.6667\ , \quad \tilde{\beta}_5 = +1.00\ , \quad \text{and} \quad \tilde{\beta}_6 = +2.3333$$

for Factor $B|A$. Figure 4.46 lists the observed y_i values, LAD predicted $\tilde{y}_i$ values, and residual e_i values for $i = 1, \ldots, 36$.

Following Eq. (4.5) on p. 125 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 36$ LAD regression residuals listed in Fig. 4.46 yield $a = 3$ average distance-function values of

$$\xi_{B|A_1} = 2.00\ , \quad \xi_{B|A_2} = 2.00\ , \quad \text{and} \quad \xi_{B|A_3} = 2.00\ .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.46 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{B|A_i}}{N}\ , \qquad i = 1, 2, 3\ ,$$

is

$$\delta_{B|A} = \sum_{i=1}^{b} C_i \xi_i = \frac{12}{36}\big(2.00 + 2.00 + 2.00\big) = 2.00\ .$$

Fig. 4.46 Observed, predicted, and residual LAD regression values for the nested response measurement scores listed in Table 4.12

Object	y_i	$\tilde{y}_i$	e_i
1	29	32.00	−3.00
2	31	32.00	−1.00
3	31	32.00	−1.00
4	33	32.00	+1.00
5	30	32.00	−2.00
6	32	32.00	0.00
7	32	32.00	0.00
8	34	32.00	+2.00
9	28	29.00	−1.00
10	30	29.00	+1.00
11	30	29.00	+1.00
12	32	29.00	+3.00
13	27	31.00	−4.00
14	29	31.00	−2.00
15	29	31.00	−2.00
16	31	31.00	0.00
17	33	35.00	−2.00
18	35	35.00	0.00
19	35	35.00	0.00
20	37	35.00	+2.00
21	30	32.00	−2.00
22	32	32.00	0.00
23	32	32.00	0.00
24	34	32.00	+2.00
25	31	33.00	−2.00
26	33	33.00	0.00
27	33	33.00	0.00
28	35	33.00	+2.00
29	27	29.00	−2.00
30	29	29.00	0.00
31	29	29.00	0.00
32	31	29.00	+2.00
33	35	35.00	0.00
34	37	35.00	+2.00
35	37	35.00	+2.00
36	39	35.00	+4.00

If all M possible arrangements of the $N = 36$ observed LAD regression residuals listed in Fig. 4.46 occur with equal chance, the approximate resampling probability value of $\delta_{B|A} = 2.00$ computed on $L = 1{,}000{,}000$ random arrangements of the observed LAD regression residuals with $n_{B|A_1} = n_{B|A_2} = n_{B|A_3} = 12$ residuals preserved for each arrangement is

$$P(\delta \leq \delta_{B|A}|H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_{B|A}}{L} = \frac{361{,}575}{1{,}000{,}000} = 0.3616 \; .$$

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 2.0127$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{B|A} = 1 - \frac{\delta_{B|A}}{\mu_\delta} = 1 - \frac{2.00}{2.0127} = +0.0063 ,$$

indicating approximately chance agreement between the observed and predicted y values.

OLS Regression Analysis

For comparison, consider an MRPP resampling analysis of OLS regression residuals calculated on the $N = 36$ response measurement scores listed in Table 4.12 on p. 197. The MRPP regression analysis yields estimated OLS regression coefficients of

$$\hat{\beta}_0 = +32.00 , \quad \hat{\beta}_1 = -1.00 , \quad \hat{\beta}_2 = 0.00 , \quad \hat{\beta}_3 = +1.00 ,$$
$$\hat{\beta}_4 = -2.00 , \quad \hat{\beta}_5 = +1.00 , \quad \text{and} \quad \hat{\beta}_6 = +3.00$$

for Factor $B|A$. Figure 4.47 lists the observed y_i values, OLS predicted $\hat{y}_i$ values, and residual e_i values for $i = 1, \ldots, 36$.

Following Eq. (4.5) on p. 125 and employing squared Euclidean distance between residuals with $v = 2$, the $N = 36$ OLS regression residuals listed in Fig. 4.47 yield $a = 3$ average distance-function values of

$$\xi_{B|A_1} = 5.8182 , \quad \xi_{B|A_2} = 5.8182 , \quad \text{and} \quad \xi_{B|A_3} = 5.8182 .$$

Following Eq. (4.4) on p. 125, the observed value of the MRPP test statistic calculated on the OLS regression residuals listed in Fig. 4.47 with $v = 2$ and treatment-group weights

$$C_i = \frac{n_{B|A_i} - 1}{N - b} , \qquad i = 1, 2, 3 ,$$

is

$$\delta_{B|A} = \sum_{i=1}^{b} C_i \xi_i = \frac{12 - 1}{36 - 3}\big(5.8182 + 5.8182 + 5.8182\big) = 5.8182 .$$

If all M possible arrangements of the $N = 36$ observed OLS regression residuals listed in Fig. 4.47 occur with equal chance, the approximate resampling probability value of $\delta_{B|A} = 5.8182$ computed on $L = 1{,}000{,}000$ random arrangements of the observed OLS regression residuals with $n_{B|A_1} = n_{B|A_2} = n_{B|A_3} = 12$ residuals pre-

Fig. 4.47 Observed, predicted, and residual OLS regression values for the nested response measurement scores listed in Table 4.12

Object	y_i	$\hat{y}_i$	e_i
1	29	32.00	−3.00
2	31	32.00	−1.00
3	31	32.00	−1.00
4	33	32.00	+1.00
5	30	32.00	−2.00
6	32	32.00	0.00
7	32	32.00	0.00
8	34	32.00	+2.00
9	28	29.00	−1.00
10	30	29.00	+1.00
11	30	29.00	+1.00
12	32	29.00	+3.00
13	27	30.00	−3.00
14	29	30.00	−1.00
15	29	30.00	−1.00
16	31	30.00	+1.00
17	33	35.00	−2.00
18	35	35.00	0.00
19	35	35.00	0.00
20	37	35.00	+2.00
21	30	31.00	−1.00
22	32	31.00	+1.00
23	32	31.00	+1.00
24	34	31.00	+3.00
25	31	34.00	−3.00
26	33	34.00	−1.00
27	33	34.00	−1.00
28	35	34.00	+1.00
29	27	29.00	−2.00
30	29	29.00	0.00
31	29	29.00	0.00
32	31	29.00	+2.00
33	35	36.00	−1.00
34	37	36.00	+1.00
35	37	36.00	+1.00
36	39	36.00	+3.00

served for each arrangement is

$$P(\delta \leq \delta_{B|A}|H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_{B|A}}{L} = \frac{7{,}600}{1{,}000{,}000} = 0.0076 \ .$$

For comparison, the approximate resampling probability value based on LAD regression, $v = 1$, $L = 1{,}000{,}000$, and $C_i = n_{B|A_i}/N$ for $i = 1, 2, 3$ is $P = 0.3616$.

Following Eq. (4.7) on p. 126, the exact expected value of the M δ values is $\mu_\delta = 5.4857$ and, following Eq. (4.6) on p. 126, the observed chance-corrected measure of effect size for the y_i and $\hat{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_{B|A} = 1 - \frac{\delta_{B|A}}{\mu_\delta} = 1 - \frac{5.8182}{5.4857} = -0.0606 ,$$

indicating slightly less than chance agreement between the observed and predicted y values.

Conventional ANOVA Analysis

A conventional fixed-effects nested analysis of variance calculated on the $N = 36$ univariate response measurement scores for Factor $B|A$ listed in Table 4.11 on p. 196 yields an observed F-ratio of $F_{B|A} = 10.6362$. Assuming independence, normality, and homogeneity of variance, $F_{B|A}$ is approximately distributed as Snedecor's F under the null hypothesis with $\nu_1 = a(b-1) = 3(3-1) = 6$ and $\nu_2 = ab(n-1) = (3)(3)(4-1) = 27$ degrees of freedom. Under the null hypothesis, the observed value of $F_{B|A} = 10.6362$ yields an approximate probability value of $P = 4.5461{\times}10^{-6}$.

4.4 Multivariate Multiple Regression Designs

An extension of LAD multiple regression to include multiple dependent variables, as well as multiple independent variables, i.e., multivariate multiple LAD regression, is developed in this section. The extension was prompted by a multivariate Least Sum (of) Euclidean Distances (LSED) algorithm developed by Kaufman, Taylor, Mielke, and Berry in 2002 [198].

Consider the multivariate multiple regression model given by

$$y_{ik} = \sum_{j=1}^{m} x_{ij}\beta_{jk} + e_{ik}$$

for $i = 1, \ldots, N$ and $k = 1, \ldots, r$, where y_{ik} represents the ith of N measurements for the kth of r response variables, possibly affected by a treatment; x_{ij} is the jth of m covariates associated with the ith response, where $x_{i1} = 1$ if the model includes an intercept; β_{jk} denotes the jth of m regression parameters for the kth of r response variables; and e_{ik} designates the error associated with the ith of N measurements for the k of r response variables.

If estimates of β_{jk} that minimize

$$\sum_{i=1}^{N}\left(\sum_{k=1}^{r} e_{ik}^2\right)^{1/2}$$

are denoted by $\tilde{\beta}_{jk}$ for $j = 1, \ldots, m$ and $k = 1, \ldots, r$, then the N r-dimensional residuals of the LSED multivariate multiple regression model are given by

$$e_{ik} = y_{ik} - \sum_{j=1}^{m} x_{ij}\tilde{\beta}_{jk}$$

for $i = 1, \ldots, N$ and $k = 1, \ldots, r$.

Let the N r-dimensional residuals, $(e_{i1}, \ldots, e_{ir})$ for $i = 1, \ldots, N$ obtained from an LSED multivariate multiple regression model, be partitioned into g treatment groups of sizes $n_1, \ldots, n_g$, where $n_i \geq 2$ for $i = 1, \ldots, g$ and

$$N = \sum_{i=1}^{g} n_i \ .$$

The MRPP analysis of the multivariate multiple regression residuals depends on statistic

$$\delta = \sum_{i=1}^{g} C_i \xi_i \ , \tag{4.8}$$

where $C_i = n_i/N$ is a positive weight for the ith of g treatment groups and ξ_i is the average pairwise Euclidean distance among the n_i r-dimensional residuals in the ith of g treatment groups defined by

$$\xi_i = \binom{n_i}{2}^{-1} \sum_{k=1}^{N-1} \sum_{l=k+1}^{N} \left[\sum_{j=1}^{r} \left(e_{kj} - e_{lj} \right)^2 \right]^{1/2} \Psi_{ki}\, \Psi_{li} \ , \tag{4.9}$$

where

$$\Psi_{ki} = \begin{cases} 1 & \text{if } (e_{k1}, \ldots, e_{kr}) \text{ is in the } i\text{th treatment group} \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

The null hypothesis specifies that each of the

$$M = \frac{N!}{\prod_{i=1}^{g} n_i!}$$

possible allocations of the N r-dimensional residuals to the g treatment groups is equally likely. An exact MRPP probability value associated with the observed value of δ, δ_o, is given by

$$P\big(\delta \leq \delta_o | H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_o}{M} \ .$$

As previously, when M is large an approximate probability value may be obtained from a resampling permutation procedure. Let L denote a large random sample drawn from all M possible arrangements of the observed data, then an approximate resampling probability value is given by

$$P\big(\delta \leq \delta_{\text{o}}|H_0\big) = \frac{\text{number of } \delta \text{ values} \leq \delta_{\text{o}}}{L} \ .$$

As with univariate multiple regression models, the criterion for fitting multivariate multiple regression models based on δ is the chance-corrected measure of effect size between the observed and predicted response measurement values given by

$$\Re = 1 - \frac{\delta}{\mu_\delta} \ , \tag{4.10}$$

where μ_δ is the expected value of δ over the $N!$ possible pairings under the null hypothesis, given by

$$\mu_\delta = \frac{1}{M} \sum_{i=1}^{M} \delta_i \ . \tag{4.11}$$

4.4.1 Example Analysis

To illustrate a multivariate LSED multiple regression analysis, consider an unbalanced two-way randomized-block experimental design in which $N = 16$ subjects (S) are tested over $a = 3$ levels of Factor A, the experiment is repeated $b = 2$ times for Factor B, and there are $r = 2$ response measurement scores for each subject. The design and data are adapted from Mielke and Berry [297, p. 184] and are given in Fig. 4.48. The design is intentionally kept small to illustrate the multivariate multiple regression procedure.

Analysis of Factor A

A design matrix of dummy codes for an MRPP regression analysis of Factor A is given in Fig. 4.49, where the first column of 1 values provides for an intercept, the next column contains the dummy codes for Factor B, and the third and fourth columns contain the bivariate response measurement scores listed according to the original random assignment of the $N = 16$ subjects to the $a = 3$ levels of Factor A with the first $n_{A_1} = 5$ scores, the next $n_{A_2} = 7$ scores, and the last $n_{A_3} = 4$ scores associated with the $a = 3$ levels of Factor A, respectively. The MRPP regression analysis examines the $N = 16$ regression residuals for possible differences among the $a = 3$ treatment levels of Factor A; consequently, no dummy codes are provided

Fig. 4.48 Example data for a two-way randomized-block design with $a = 3$ blocks and $b = 2$ treatments

Factor B	Factor A: A_1	A_2	A_3
B_1	(49, 102)	(63, 84)	(45, 107)
		(60, 89)	(50, 100)
			(42, 111)
			(46, 104)
B_2	(48, 103)	(27, 114)	
	(58, 94)	(66, 83)	
	(51, 100)	(74, 79)	
	(55, 97)	(69, 88)	
		(71, 82)	

Fig. 4.49 Example design matrix and bivariate response measurement scores for a multivariate LSED multiple regression analysis of Factor A with $N = 16$

Matrix		Scores	
1	1	49	102
1	0	48	103
1	0	58	94
1	0	51	100
1	0	55	97
1	1	63	84
1	1	60	89
1	0	27	114
1	0	66	83
1	0	74	79
1	0	69	88
1	0	71	82
1	1	45	107
1	1	50	100
1	1	42	111
1	1	46	104

for Factor A as this information is implicit in the ordering of the $a = 3$ levels of Factor A in the last two columns of Fig. 4.49.

Because there are only

$$M = \frac{N!}{\prod_{i=1}^{a} n_{A_i}!} = \frac{16!}{5!\ 7!\ 4!} = 1{,}441{,}440$$

possible, equally-likely arrangements of the $N = 16$ bivariate response measurement scores listed in Fig. 4.49, an exact permutation approach is feasible. An MRPP analysis of the $N = 16$ LAD regression residuals calculated on the bivariate response measurements for Factor A in Fig. 4.49 yields estimated LAD regression coefficients of

$$\tilde{\beta}_{1,1} = +58.00\ ,\ \tilde{\beta}_{2,1} = -9.00\ ,\ \tilde{\beta}_{1,2} = +94.00\ ,\ \text{and}\ \tilde{\beta}_{2,2} = +8.00$$

Fig. 4.50 Observed, predicted, and residual values for a multivariate LSED multiple regression analysis of Factor A with $N = 16$

y_{i1}	y_{i2}	$\tilde{y}_{i1}$	$\tilde{y}_{i2}$	e_{i1}	e_{i2}
49	102	49.00	102.00	0.00	0.00
48	103	58.00	94.00	−10.00	+9.00
58	94	58.00	94.00	0.00	0.00
51	100	58.00	94.00	−7.00	+6.00
55	97	58.00	94.00	−3.00	+3.00
63	84	49.00	102.00	+14.00	−18.00
60	89	49.00	102.00	+11.00	−13.00
27	114	58.00	94.00	−31.00	+20.00
66	83	58.00	94.00	+8.00	−11.00
74	79	58.00	94.00	+16.00	−15.00
69	88	58.00	94.00	+11.00	−6.00
71	82	58.00	94.00	+13.00	−12.00
45	107	49.00	102.00	−4.00	+5.00
50	100	49.00	102.00	+1.00	−2.00
42	111	49.00	102.00	−7.00	+9.00
46	104	49.00	102.00	−3.00	+2.00

for Factor A. Figure 4.50 lists the observed y_{ik} values, LAD predicted $\tilde{y}_{ik}$ values, and residual e_{ik} values for $i = 1, \ldots, 16$ and $k = 1, 2$.

Following Eq. (4.9) on p. 208 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 16$ LAD regression residuals listed in Fig. 4.50 yield $a = 3$ average distance-function values of

$$\xi_{A_1} = 7.2294 \,, \quad \xi_{A_2} = 20.0289 \,, \quad \text{and} \quad \xi_{A_3} = 7.3475 \,.$$

Following Eq. (4.8) on p. 208, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.50 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{A_i}}{N} \,, \qquad i = 1, 2, 3 \,,$$

is

$$\delta_A = \sum_{i=1}^{a} C_i \xi_i = \frac{1}{16}\big[(5)(7.2294) + (7)(20.0289) + (4)(7.3475)\big] = 12.8587 \,.$$

If all arrangements of the $N = 16$ observed LAD regression residuals listed in Fig. 4.50 occur with equal chance, the exact probability value of $\delta_A = 12.8587$ computed on the $M = 1{,}441{,}440$ possible arrangements of the observed LAD regression residuals with $n_{A_1} = 5$, $n_{A_2} = 7$, and $n_{A_3} = 4$ residuals preserved for each arrangement is

$$P(\delta \leq \delta_A | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_A}{M} = \frac{6{,}676}{1{,}441{,}440} = 0.0046 \,.$$

Following Eq. (4.11) on p. 209, the exact expected value of the $M = 1{,}441{,}440$ δ values is $\mu_\delta = 18.1020$ and, following Eq. (4.10) on p. 209, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_A = 1 - \frac{\delta_A}{\mu_\delta} = 1 - \frac{12.8587}{18.1020} = +0.2897 \,,$$

indicating approximately 29 % agreement between the observed and predicted values above that expected by chance.

Analysis of Factor *B*

A design matrix of dummy codes for an MRPP regression analysis of Factor *B* is given in Fig. 4.51, where the first column of 1 values provides for an intercept, the next two columns contain the dummy codes for Factor *A*, and the fourth and fifth columns contain the bivariate response measurement scores listed according to the original random assignment of the $N = 16$ subjects to the $b = 2$ levels of Factor *B* with the first $n_{B_1} = 7$ scores and the last $n_{B_2} = 9$ scores associated with the $b = 2$ levels of Factor *B*, respectively. The MRPP regression analysis examines the $N = 16$ regression residuals for possible differences between the $b = 2$ treatment levels of Factor *B*; consequently, no dummy codes are provided for Factor *B* as this information is implicit in the ordering of the $b = 2$ levels of Factor *B* in the last two columns of Fig. 4.51.

Because there are only

$$M = \frac{N!}{\displaystyle\prod_{i=1}^{b} n_{B_i}!} = \frac{16!}{7!\;9!} = 11{,}440$$

Fig. 4.51 Example design matrix and bivariate response measurement scores for a multivariate LSED multiple regression analysis of Factor *B* with $N = 16$

Matrix			Scores	
1	1	0	49	102
1	0	1	63	84
1	0	1	60	89
1	0	0	45	107
1	0	0	50	100
1	0	0	42	111
1	0	0	46	104
1	1	0	48	103
1	1	0	58	94
1	1	0	51	100
1	1	0	55	97
1	0	1	27	114
1	0	1	66	83
1	0	1	74	79
1	0	1	69	88
1	0	1	71	82

Fig. 4.52 Observed, predicted, and residual values for a multivariate LSED multiple regression analysis of Factor *B* with $N = 16$

y_{i1}	y_{i2}	$\tilde{y}_{i1}$	$\tilde{y}_{i2}$	e_{i1}	e_{i2}
49	102	51.00	100.00	−2.00	+2.00
63	84	66.00	84.00	−3.00	0.00
60	89	66.00	84.00	−6.00	+5.00
45	107	46.00	104.00	−1.00	+3.00
50	100	46.00	104.00	+4.00	−4.00
42	111	46.00	104.00	−4.00	+7.00
46	104	46.00	104.00	0.00	0.00
48	103	51.00	100.00	−3.00	+3.00
58	94	51.00	100.00	+7.00	−6.00
51	100	51.00	100.00	0.00	0.00
55	97	51.00	100.00	+4.00	−3.00
27	114	66.00	84.00	−39.00	+30.00
66	83	66.00	84.00	0.00	−1.00
74	79	66.00	84.00	−8.00	−5.00
69	88	66.00	84.00	+3.00	+4.00
71	82	66.00	84.00	+5.00	−2.00

possible, equally-likely arrangements of the $N = 16$ response measurement scores listed in Fig. 4.51, an exact permutation approach is feasible. An MRPP analysis of the $N = 16$ LAD regression residuals calculated on the bivariate response measurements for Factor *B* in Fig. 4.51 yields estimated LAD regression coefficients of

$$\tilde{\beta}_{1,1} = +46.00\ , \quad \tilde{\beta}_{2,1} = +5.00\ , \quad \tilde{\beta}_{3,1} = +20.00\ , \quad \tilde{\beta}_{1,2} = +104.00\ ,$$
$$\tilde{\beta}_{2,2} = -4.00\ , \quad \text{and} \quad \tilde{\beta}_{3,2} = -20.00$$

for Factor *B*. Figure 4.52 lists the observed y_{ik} values, LAD predicted $\tilde{y}_{ik}$ values, and residual e_{ik} values for $i = 1, \ldots, 16$ and $k = 1, 2$.

Following Eq. (4.9) on p. 208 and employing ordinary Euclidean distance between residuals with $v = 1$, the $N = 16$ LAD regression residuals listed in Fig. 4.52 yield $b = 2$ average distance-function values of

$$\xi_{B_1} = 6.0229 \quad \text{and} \quad \xi_{B_2} = 16.7440\ .$$

Following Eq. (4.4) on p. 208, the observed value of the MRPP test statistic calculated on the LAD regression residuals listed in Fig. 4.52 with $v = 1$ and treatment-group weights

$$C_i = \frac{n_{B_i}}{N}\ , \qquad i = 1, 2\ ,$$

is

$$\delta_B = \sum_{i=1}^{b} C_i \xi_i = \frac{1}{16}\big[(7)(6.0229) + (9)(16.7440)\big] = 12.0535\ .$$

If all arrangements of the $N = 16$ observed LAD regression residuals listed in Fig. 4.52 occur with equal chance, the exact probability value of $\delta_B = 12.0535$ computed on the $M = 11{,}440$ possible arrangements of the observed LAD regression residuals with $n_{B_1} = 7$ and $n_{B_2} = 9$ residuals preserved for each arrangement is

$$P(\delta \leq \delta_B | H_0) = \frac{\text{number of } \delta \text{ values} \leq \delta_B}{M} = \frac{2{,}090}{11{,}440} = 0.1827 \, .$$

Following Eq. (4.11) on p. 209, the exact expected value of the $M = 11{,}440$ δ values is $\mu_\delta = 12.2923$ and, following Eq. (4.10) on p. 209, the observed chance-corrected measure of effect size for the y_i and $\tilde{y}_i$ values, $i = 1, \ldots, N$, is

$$\Re_B = 1 - \frac{\delta_B}{\mu_\delta} = 1 - \frac{12.0535}{12.2923} = +0.0194 \, ,$$

indicating approximately 2 % agreement between the observed and predicted values above that expected by chance.

4.5 Coda

Chapter 4 applied the Multi-Response Permutation Procedures (MRPP) developed in Chap. 2 to interval-level response measurements, utilizing dummy and effect coding of treatment groups to generate regression residuals from LAD regression models, subsequently analyzed with MRPP. Considered in this chapter were one-way randomized, one-way randomized with a covariate, one-way randomized-block, two-way randomized-block, two-way factorial, Latin square, split-plot, and two-factor nested designs. Chapter 4 concluded with example multivariate multiple regression designs.

Comparisons of permutation-based LAD regression with ordinary Euclidean distance between response measurements, permutation-based OLS regression with squared Euclidean distance between response measurements, and conventional OLS regression with squared Euclidean distance between response measurements in Chap. 4, revealed that considerable differences can exist among the three approaches that are not systematic. Oftentimes, one of the three approaches yielded the lowest of the three probability values, while other times the same approach yielded the highest probability value. Sometimes the three approaches yielded the same, or nearly the same, probability value, as was the case with the analysis of Factor B in the two-way randomized-block design example, and other times the three probability values were markedly different, as was the case with the analysis of the $A \times B$ interaction in the two-way factorial design example. In general, permutation-based LAD regression, coupled with MRPP and ordinary Euclidean distance between response measurements, is recommended due to the lack of restrictive assumptions and robustness that is possible with extreme values.

Chapter 5

Chapter 5 establishes the relationships between the MRPP test statistics, δ and $\Re$, and selected conventional tests and measures designed for the analysis of completely randomized data at the ordinal level of measurement. Considered in Chap. 5 are the Wilcoxon two-sample rank-sum test, the Kruskal–Wallis multiple-sample rank-sum test, the Mood rank-sum test for dispersion, the Brown–Mood median test, the Mielke power-of-rank functions, the Whitfield two-sample rank-sum test, and the Cureton rank-biserial test.