Linear Hypothesis Testing Based on Unbiased Fuzzy Estimators and Fuzzy Significance Level

Alireza Jiryaei and Mashaallah Mashinchi

Abstract A wide variety of applied problems of statistical hypothesis testing can be treated under a general setup of the linear models which includes analysis of variance. In this study, a new method is presented to test linear hypothesis using a fuzzy test statistic produced by a set of confidence intervals with non-equal tails. Also, a fuzzy significance level is used to evaluate the linear hypothesis. The method can be used to improve linear hypothesis testing when there is a sensitively in accepting or rejecting the null hypothesis. Also, as a simple case of linear hypothesis testing, one-way analysis of variance based on fuzzy test statistic and fuzzy significance level is investigated. Numerical examples are provided for illustration.

Keywords Analysis of variance · Confidence interval · Fuzzy critical value · Fuzzy test statistic · Fuzzy significance level · Linear hypothesis · Linear model

1 Introduction and Background

Analysis of Variance (ANOVA) is a common and popular method in the analysis of experimental designs. It includes important cases such as one-way and two-way ANOVA, and one-way and two-way analysis of covariance, and it has many useful applications in industry, agriculture and social sciences [8, 12, 13]. Various aspects of this topic have been considered in a fuzzy environment. One-way and two-way ANOVA using fuzzy unbiased estimators for variance parameter are discussed based on arithmetic operations on intervals by Buckley [3]. Wu [16] presented one-way ANOVA based on several notations of the α-cuts of fuzzy random variables, optimistic and pessimistic degrees and solving an optimization problem. An approach for one-way ANOVA has been carried out by Nourbakhsh et al. [10] for fuzzy data in which Zadeh's extension principle [9, 17] plays a key role for the applied

A. Jiryaei (✉) · M. Mashinchi
Department of Statistics, Faculty of Mathematics and Computer Sciences,
Shahid Bahonar University of Kerman, Kerman, Iran
e-mail: a.jiryae@gmail.com

computing operations. A statistical technique for testing the fuzzy hypothesis of one-way ANOVA is proposed by Kalpanapriya et al. [7] using the levels of pessimistic and optimistic of the triangular fuzzy data.

Linear hypothesis testing is an extension of analysis of variance. It can test hypotheses about the unknown parameters of the linear model, such as testing the equality of the means of several random variables [12]. Sometimes the observed value of test statistic is close to the related quantiles of statistical distributions, so there is uncertainty in accepting or rejecting the null hypothesis H_0. In this paper, a method is presented for linear hypothesis testing using a fuzzy test statistic and a fuzzy significance level. Moreover, the method can be used for modelling this uncertainty using fuzzy sets theory.

A method for testing statistical hypotheses in a fuzzy environment was introduced by Buckley [2, 3]. It considers a fuzzy test statistic and fuzzy critical values produced using confidence intervals with equal tails and arithmetic operations on intervals. In Buckley's method the fuzzy estimates are developed as fuzzy numbers, and their membership functions have been derived by Falsafain et al. [5]. In [2] the non-fuzzy hypotheses are tested, and in [14] and [1] the presented approach in [2] is generalized to the case where the statistical hypotheses and the observed data are also fuzzy. When dealing with non-symmetric statistical distributions, using confidence intervals with equal tails results in producing a fuzzy estimate where the membership degree for the unbiased point estimate of the required parameter is not equal to one [4]. While we expect that the unbiased point estimate has the highest importance in the fuzzy estimate, i.e. its membership degree should be equal to one. Solutions to overcome this problem using the confidence intervals with non-equal tails are provided by Buckley [3], and Falsafain and Taheri [4]. It has been shown that the solution presented by Falsafain and Taheri [4] is reduced to Buckley's method when dealing with symmetric statistical distributions. Moreover, it is possible to obtain the membership functions of the corrected fuzzy estimates. Therefore, we use this solution in this paper.

In order to discuss linear hypothesis testing based on fuzzy test statistic and fuzzy significance level, we first recall some basic concepts of fuzzy sets theory in Sect. 2. Section 3 contains a brief review of linear model and linear hypothesis. In Sect. 4, fuzzy test statistic and fuzzy critical value are discussed and decision rules are presented. Also, one-way ANOVA as a special case of linear hypothesis testing is discussed in Sect. 5. Two numerical examples are provided in Sect. 6 to show that our approach could perform quite well in practice. A conclusion is provided in Sect. 7.

2 Preliminaries

Some concepts of fuzzy sets theory, which will be referred to throughout this paper, are discussed in this section. Let U be a universal set and $F(U) = \{\tilde{A} | \tilde{A} : U \to [0, 1]\}$. Any $\tilde{A} \in F(U)$ is called a fuzzy set on U. The α-cuts of $\tilde{A}$ is

the crisp set $\tilde{A}_\alpha = \{u \in U | \tilde{A}(u) \geq \alpha\}$, for $0 < \alpha \leq 1$. Moreover, $\tilde{A}_0$ is separately defined [2] as the closure of the union of all the $\tilde{A}_\alpha$, for $0 < \alpha \leq 1$. The value $\tilde{A}(u)$ is interpreted as the membership degree of a point u. $\tilde{A} \in F(\mathbb{R})$ is called a fuzzy number, under the following conditions:

1. There is a unique $r_0 \in \mathbb{R}$ with $\tilde{A}(r_0) = 1$,
2. The α-cuts of $\tilde{A}$ are closed and bounded intervals on $\mathbb{R}$ for any $0 \leq \alpha \leq 1$,

where $\mathbb{R}$ is the set of all real numbers. In other words for every fuzzy number $\tilde{A}$ we have $\tilde{A}_\alpha = [a_1(\alpha), a_2(\alpha)]$ for all $\alpha \in [0, 1]$ which are the closed, bounded, intervals and their bounds are as functions of α.

To continue discussions, we need to clarify the concept of an unbiased fuzzy estimator, using the following definition. Similar to conventional statistics, a fuzzy estimator is a rule for calculating a fuzzy estimate of an unknown parameter based on observed data. Thus the rule and its result (the fuzzy estimate) are distinguished.

Definition 2.1 A fuzzy number $\tilde{\theta}$ is an unbiased fuzzy estimator for parameter θ from a statistical distribution if:

1. The α-cuts of $\tilde{\theta}$ are $(1 - \alpha)100\%$ confidence intervals for θ, with $\alpha \in [0.01, 1]$ and $\tilde{\theta}_\alpha = \tilde{\theta}_{0.01}$ for $\alpha \in [0, 0.01)$.
2. If $\hat{\theta}$ is an unbiased point estimator for θ then $\tilde{\theta}\left(\hat{\theta}\right) = 1$.

An explicit and unique membership function is given for a fuzzy estimate by the following theorem.

Theorem 2.1 [5] *Suppose that $X_1, X_2, \ldots, X_n$ is a random sample of size n from a distribution with unknown parameter θ. If, based on observations $x_1, x_2, \ldots, x_n$, we consider $\tilde{A}_\alpha = [\theta_1(\alpha), \theta_2(\alpha)]$ as a $(1 - \alpha)$ 100 % confidence interval for θ, then the fuzzy estimate of θ is a fuzzy set with the following unique membership function:*

$$\tilde{\theta}(u) = min\left\{\theta_1^{-1}(u), [-\theta_2]^{-1}(-u), 1\right\}.$$

To end this section, we give an introduction to interval arithmetic. Let $I = [a, b]$ and $J = [c, d]$ be two closed intervals. Then based on the interval arithmetic, we have

$$I + J = [a + c, b + d],$$

$$I - J = [a - d, b - c],$$

$$I \times J = [\alpha, \beta], \ \alpha = min\{ac, ad, bc, bd\} \beta = max\{ac, ad, bc, bd\}$$

and

$$I/J = [a, b] \times [1/d, 1/c],$$

where zero does not belong to $J = [c, d]$ in the last case.

3 Linear Hypothesis Testing

In this section we give a brief review of linear hypothesis testing, for more details see [12, 13]. The concepts of linear model and linear hypothesis are given in Definition 3.1. The process of linear hypothesis testing is presented in Theorem 3.1.

Definition 3.1 Let $Y = (Y_1 Y_2 \ldots Y_n)'$ be a random column vector and $\mathbf{X}$ be a $n \times k$ matrix of full rank $k < n$ and known constants $x_{ij}, i = 1, 2, \ldots, n; j = 1, 2, \ldots, k$. It is said that the distribution of Y satisfies a *linear model* if $E(Y) = \mathbf{X}\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_1 \beta_2 \ldots \beta_k)'$ is vector of unknown (scalar) parameters $\beta_1, \beta_2, \ldots, \beta_k$, where $\beta_j \in \mathbb{R}$ for $j = 1, 2, \ldots, k$. It is convenient to write $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = (\epsilon_1 \epsilon_2 \ldots \epsilon_n)'$ is a vector of non-observable independent normal random variables with common variance σ^2 and $E(\epsilon_j) = 0;\ j = 1, 2, \ldots, n$. Relation $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ is known as a linear model. The linear hypothesis concerns $\boldsymbol{\beta}$, such that $\boldsymbol{\beta}$ satisfies $H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{0}$, where $\mathbf{H}$ is a known $r \times k$ matrix of full rank $r \leq k$.

Theorem 3.1 *Consider the linear model $Y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. The generalized likelihood ratio (GLR) test for testing the linear hypothesis $H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{0}$ is to reject H_0 at significance level γ if $F \geq F_{1-\gamma, r, n-k}$, where $P_{H_0}(F < F_{1-\gamma, r, n-k}) = 1 - \gamma$ and F is the random variable given by*

$$F = \frac{SS^*/r}{SS/(n-k)}$$

where,

$$SS^* = \left(Y - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}\right)' \left(Y - \mathbf{X}\hat{\hat{\boldsymbol{\beta}}}\right) - \left(Y - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \left(Y - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

and

$$SS = \left(Y - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \left(Y - \mathbf{X}\hat{\boldsymbol{\beta}}\right),$$

$\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ and $\hat{\hat{\boldsymbol{\beta}}}$ is the MLE of $\boldsymbol{\beta}$ under H_0. Moreover, under H_0 the random variable F has the F-distribution with r and $(n-k)$ degrees of freedom.

Note 3.1 As the result of Theorem 3.1, it can be shown that the pivotal quantity SS/σ^2 has the distribution χ^2 with $(n-k)$ degrees of freedom and SS^*/σ^2 has the distribution χ^2 with r degrees of freedom, under the null hypothesis H_0. So both of these pivotal quantities can be used to produce confidence intervals for σ^2. It is clear that the statistics $SS/(n-k)$ and SS^*/r (under H_0) are the unbiased point estimators for the unknown parameter σ^2.

4 Linear Hypothesis Testing Based on Fuzzy Test Statistic

4.1 Testing at Precise Significance Level

In this section, taking into account Buckley's method in [2] and its modifications in [4], we consider testing the linear hypothesis based on a fuzzy test statistic and a fuzzy significance level. Because we could obtain a fuzzy test statistics to evaluate the linear hypothesis, we give several theorems sequentially. Also, we obtain a fuzzy critical value using α-cuts of a considered fuzzy significance level. Next we make two decision rules to the cases where the critical value is either crisp or fuzzy. In the rest of this paper, the symbols $\chi^2_{\xi,\upsilon}$ and $F_{\xi,\upsilon_1,\upsilon_2}$ will be used to represent the ξ'th quantile of the distribution χ^2 with υ degrees of freedom and the ξth quantile of the distribution F with υ_1 and υ_2 degrees of freedom, respectively.

Theorem 4.1.1 *In a linear model consider $SS/(n-k)$ as an unbiased point estimator for parameter σ^2. Then an unbiased fuzzy estimator for σ^2 is $\widetilde{\sigma^2}$ with the following α-cuts*

$$\left(\widetilde{\sigma^2}\right)_\alpha = \begin{cases} \left[SS/\chi^2_{1-\alpha+\alpha p',(n-k)}, SS/\chi^2_{\alpha p',(n-k)}\right] & 0.01 \le \alpha \le 1 \\ \left(\widetilde{\sigma^2}\right)_{0.01} & 0 \le \alpha < 0.01 \end{cases}$$

in which p' is obtained from the relation $\chi^2_{p',(n-k)} = n-k$.

Proof Based on the pivotal quantity SS/σ^2, a $(1-\alpha)\,100\,\%$ confidence interval for σ^2 is $\left[SS/\chi^2_{1-\alpha+\alpha p,(n-k)}, SS/\chi^2_{\alpha p,(n-k)}\right]$ for any $0<\alpha<1$ and $0<p<1$. When $\alpha=1$ and $p=p'$, satisfying $\chi^2_{p',(n-k)} = n-k$, this interval becomes the point $SS/(n-k)$ which is unbiased point estimator for σ^2. Now fixing $p=p'$ and varying α from 0.01 to 1 we obtain nested intervals which are the α - cuts of a fuzzy number, say $\widetilde{\sigma^2}$. Finally, $\left(\widetilde{\sigma^2}\right)_\alpha = \left(\widetilde{\sigma^2}\right)_{0.01}$ for $0 \le \alpha < 0.01$. So, we have the unbiased fuzzy estimator $\widetilde{\sigma^2}$ for σ^2. $\qquad\qquad\square$

Lemma 4.1.1 *The membership function of fuzzy estimator $\widetilde{\sigma^2}$ in Theorem 4.1.1 is as follows:*

$$
\widetilde{\sigma^2}(u) = \begin{cases} \frac{1-G(SS/u)}{1-p'} & \frac{SS}{\chi^2_{0.99+0.01p',(n-k)}} \leq u \leq \frac{SS}{n-k} \\ \frac{G(SS/u)}{p'} & \frac{SS}{n-k} \leq u \leq \frac{SS}{\chi^2_{0.01p',(n-k)}} \\ 0 & otherwise, \end{cases}
$$

where G is the cumulative distribution function of the χ^2 variable with $(n-k)$ degrees of freedom.

Proof By Theorem 4.1.1, we have $\theta_1(\alpha) = SS/\chi^2_{1-\alpha+\alpha p',(n-k)}$ for $0.01 \leq \alpha \leq 1$. Hence, $\theta_1^{-1}(u) = [1 - G(ss/u)]/(1 - p')$. Also $\theta_2(\alpha) = SS/\chi^2_{\alpha p',(n-k)}$, therefore $[-\theta_2]^{-1}(-u) = [G(ss/u)]/p'$ for $0.01 \leq \alpha \leq 1$. Based on Theorem 2.1 $\widetilde{\sigma^2}(u) = \min\{\theta_1^{-1}(u), [-\theta_2]^{-1}(-u), 1\}$. So, the proof follows. $\square$

Theorem 4.1.2 *Consider SS^*/r as an unbiased point estimator for parameter σ^2 under the null hypothesis $H_0 : \mathbf{H\beta} = \mathbf{0}$. Then, an unbiased fuzzy estimator for σ^2 is $\widetilde{\sigma^2_{H_0}}$ with α-cuts $(\widetilde{\sigma^2_{H_0}})_\alpha$, where*

$$
(\widetilde{\sigma^2_{H_0}})_\alpha = \begin{cases} \left[SS^*/\chi^2_{1-\alpha+\alpha p'',r}, SS^*/\chi^2_{\alpha p'',r} \right] & 0.01 \leq \alpha \leq 1 \\ (\widetilde{\sigma^2_{H_0}})_{0.01} & 0 \leq \alpha < 0.01 \end{cases}
$$

in which p'' is obtained from the relation $\chi^2_{p'',r} = r$.

Proof Consider the pivotal quantity SS^*/σ^2. Now the proof is similar to that of Theorem 4.1.1. $\square$

Notice that, similar to Lemma 4.1.1, one can derive the membership function of fuzzy estimator $\widetilde{\sigma^2_{H_0}}$, but under $H_0 : \mathbf{H\beta} = 0$.

Lemma 4.1.2 *The membership function of fuzzy estimator $\widetilde{\sigma^2_{H_0}}$ in Theorem 4.1.2 is as follows:*

$$
\widetilde{\sigma^2_{H_0}}(u) = \begin{cases} \frac{1-G(SS^*/u)}{1-p''} & \frac{SS^*}{\chi^2_{0.99+0.01p'',r}} \leq u \leq \frac{SS^*}{r} \\ \frac{G(SS^*/u)}{p''} & \frac{SS^*}{r} \leq u \leq \frac{SS^*}{\chi^2_{0.01p'',r}} \\ 0 & otherwise, \end{cases}
$$

where G is the cumulative distribution function of the χ^2 variable with r degrees of freedom.

Proof By Theorem 4.1.2, the proof is similar to that of Lemma 4.1.1. $\square$

Remark 4.1.1 Theorems 4.1.1 and 4.1.2 define unbiased fuzzy estimators for σ^2 under null hypothesis H_0. Moreover, Lemmas 4.1.1 and 4.1.2 provide the membership functions of these two estimators.

Theorem 4.1.3 *The fuzzy test statistic for testing* $H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{0}$ *is* $\tilde{F}$ *with* α-*cuts*

$$\tilde{F}_\alpha = \begin{cases} [f_1(\alpha)F, f_2(\alpha)F] & 0.01 \le \alpha \le 1 \\ \tilde{F}_{0.01} & 0 \le \alpha < 0.01, \end{cases}$$

where

$$F = \frac{SS^*/r}{SS/(n-k)},$$

$$f_1(\alpha) = \left[(r)\chi^2_{\alpha p',(n-k)}\right] / \left[(n-k)\chi^2_{1-\alpha+\alpha p'',r}\right]$$

and

$$f_2(\alpha) = \left[(r)\chi^2_{1-\alpha+\alpha p',(n-k)}\right] / \left[(n-k)\chi^2_{\alpha p'',r}\right].$$

Proof Using the equality $\tilde{F}_\alpha = (\tilde{\sigma}^2_{H_0})_\alpha / (\tilde{\sigma}^2)_\alpha$ and interval arithmetic, the fuzzy test statistic follows from Buckley's method. □

Decision rule 4.1.1 After observing the data and crisp significance level γ, a typical method for rejecting or accepting the null hypothesis $H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{0}$ can be made as follows. First we calculate the ratio $A_R/(A_R + A_L)$, where A_R (A_L) is area under the graph of the fuzzy test statistic $\tilde{F}$, but to the right (left) of the vertical line through $F_{1-\gamma,r,n-k}$ (see Fig. 1). Note that Fig. 1 just illustrates the sketch of A_R and A_L since the sides of $\tilde{F}$ are curves, not straight line segments. Next we choose a value for the credit level φ from $(0, 1]$, [1]. Finally, our decision rule at significance level γ is:

1. if $A_R/(A_R + A_L) \ge \varphi$, then reject the hypothesis $H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{0}$,
2. if $A_R/(A_R + A_L) < \varphi$, then accept H_0.

Fig. 1 The areas A_R and A_L
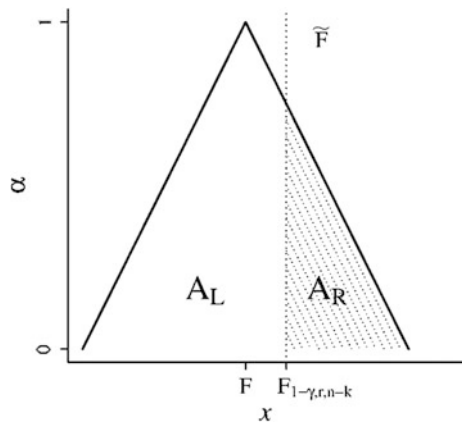
Remark 4.1.2 The presented decision rule 4.1.1 is reasonable since one can see that, by choosing any $\alpha \in [0, 1]$ and any $F \in \tilde{F}_\alpha$, this F is some value of the test statistic corresponding to this α which relates back to confidence intervals for σ^2. Therefore, if point (F, α) is in the region A_R then H_0 is rejected because $F \geq F_{1-\gamma,r,n-k}$, and if point (F, α) is in the region A_L then H_0 is accepted since $F < F_{1-\gamma,r,n-k}$.

Remark 4.1.3 In Decision rule 4.1.1, φ and γ are criterions which control possibilistic and probabilistic errors, respectively. Indeed they unify the concepts of randomness and fuzziness. The selected value of φ is more or less subjective and depends on the decision maker desire.

4.2 Testing at Fuzzy Significance Level

The approach for accepting or rejecting the null hypothesis in Subsection 4.1 is on the basis of comparing the observed fuzzy test statistic $\tilde{F}$ with the crisp critical value $F_{1-\gamma,r,n-k}$ at a crisp significance level γ. In practice it is more natural to consider the significance level as a fuzzy set since the test statistic is fuzzy. In fact, a fuzzy significance level is considered as a fuzzy number on $(0, 1)$, [6, 15]. Subsequently, we define a fuzzy significance level as a fuzzy number. We obtain a fuzzy critical value to evaluate the linear hypothesis using α-cuts of the defined fuzzy significance level. Finally, we provide a decision rule to decide whether to reject or accept the null hypothesis $H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{0}$.

Definition 4.2.1 A fuzzy significance level is a fuzzy number with the following α-cuts

$$\tilde{\gamma}_\alpha = \begin{cases} [\gamma_1 + (\gamma - \gamma_1)\alpha, \gamma_2 - (\gamma_2 - \gamma)\alpha] & 0.01 \leq \alpha \leq 1 \\ \tilde{\gamma}_{0.01} & 0 \leq \alpha < 0.01, \end{cases}$$

where $0 < \gamma_1 \leq \gamma \leq \gamma_2 < 1$.

Theorem 4.2.1 *In linear hypothesis testing based on fuzzy test statistics at the introduced fuzzy significance level in Definition 4.2.1, the critical value is a fuzzy number with the following α-cuts*

$$(\widetilde{cv})_\alpha = \begin{cases} \left[F_{(1-\gamma_2 + (\gamma_2-\gamma)\alpha),r,n-k}, F_{(1-\gamma_1-(\gamma-\gamma_1)\alpha),r,n-k} \right] & 0.01 \leq \alpha \leq 1 \\ (\widetilde{cv})_{0.01} & 0 \leq \alpha < 0.01. \end{cases}$$

Proof The proof follows by substituting α-cuts of the fuzzy significance level $\tilde{\gamma}$ for crisp one γ in the crisp critical level and by using the interval arithmetic. □

Example 4.2.1 Consider a linear hypothesis testing at the significance level $\gamma = 0.05$ where $n = 25$, $r = 3$ and $k = 4$. Assume that the significance level is a fuzzy number with the following α-cuts

$$\tilde{\gamma}_\alpha = \begin{cases} [0.03 + (0.05 - 0.03)\alpha, 0.07 - (0.07 - 0.05)\alpha] & 0.01 \leq \alpha \leq 1 \\ \tilde{\gamma}_{0.01} & 0 \leq \alpha < 0.01. \end{cases}$$

Then based on Theorem 4.2.1 the fuzzy critical value is a fuzzy number with α-cuts $(\widetilde{cv})_\alpha$ as follows

$$(\widetilde{cv})_\alpha = \begin{cases} \left[F_{(0.93 + 0.02\alpha),3,21}, \quad F_{(0.97-0.02\alpha),3,21} \right] & 0.01 \leq \alpha \leq 1 \\ (\widetilde{cv})_{0.01} & 0 \leq \alpha < 0.01. \end{cases}$$

Figure 2 shows the fuzzy numbers $\tilde{\gamma}$ and $\widetilde{cv}$.

Decision rule 4.2.1 After observing the data, the final decision rule is derived by comparing two fuzzy numbers $\widetilde{cv}$ and $\tilde{F}$. Here a way is provided to decide whether to reject or accept the null hypothesis $H_0 : \mathbf{H\beta} = \mathbf{0}$. First we calculate the ratio $A_R/(A_R + A_L)$, where A_R and A_L are depicted in Fig. 3. Note that Fig. 3 just illustrates the sketch of A_R and A_L since the sides of $\tilde{F}$ and $\widetilde{cv}$ are curves, not straight line segments. Next we choose a value for the credit level φ from $(0, 1]$. Finally, our decision rule at significance level γ is as follows:

1. if $A_R/(A_R + A_L) \geq \varphi$, then reject the hypothesis $H_0 : \mathbf{H\beta} = \mathbf{0}$,
2. if $A_R/(A_R + A_L) < \varphi$, then accept H_0.

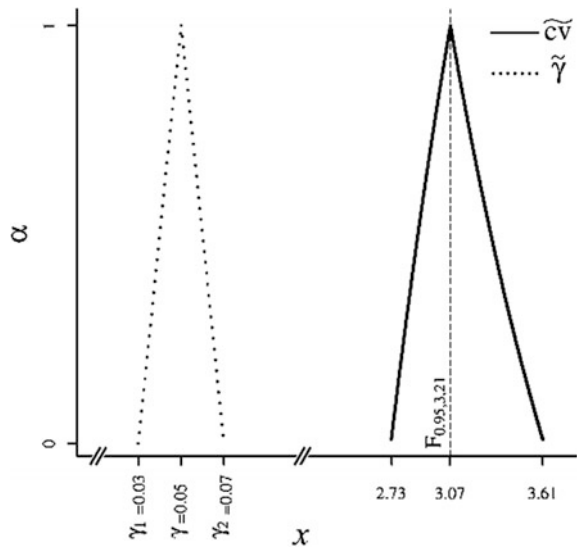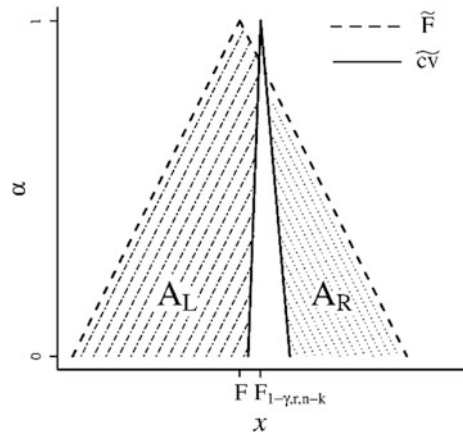**Fig. 2** The fuzzy numbers $\tilde{\gamma}$ and $\widetilde{cv}$

Fig. 3 The areas A_R and A_L

Remark 4.2.1 Decision rule 4.2.1 is reasonable since one can see that, by choosing any $\alpha \in [0,1]$, any $F \in \tilde{F}_\alpha$ and any $F_{1-\gamma,r,n-k} \in (\widetilde{cv})_\alpha$, $F_{1-\gamma,r,n-k}$ and F are some values of the test statistic and the critical level corresponding to this α which relates back to confidence intervals for σ^2 and α-cuts for the fuzzy significance level, respectively. Therefore, if point (F,α) is in the region A_R then we reject H_0 because $F \geq \max\{F_{1-\gamma,r,n-k} : F_{1-\gamma,r,n-k} \in (\widetilde{cv})_\alpha\}$, if point (F,α) is in the region A_L then we decide to accept H_0 since $F < \min\{F_{1-\gamma,r,n-k} : F_{1-\gamma,r,n-k} \in (\widetilde{cv})_\alpha\}$, and finally if point (F,α) is not in the region A_L or A_R then we do not make any decision on H_0. To this end, we have not shared point (F,α) in the final decision in Decision rule 4.2.1.

Remark 4.2.2 While, Buckley [2, 3] and Taheri et al. [1, 14] consider the problem of testing hypothesis based on a fuzzy test statistic and a crisp significance level, we assume that the significance level is fuzzy. Our method is, therefore, more convenient in real world studies.

5 One-Way ANOVA: A Simple Case of Linear Hypothesis Testing

In this section, one-way ANOVA is considered taking into account the method of linear hypothesis testing based on the fuzzy test statistic and fuzzy significance level. However, we must mention that the procedure proposed in this article is still applicable for any case of linear hypothesis testing. We now give a brief review of one-way ANOVA. For more detail refer to [8, 12]. Consider the linear model

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad j = 1, 2, \ldots, n_i; \quad i = 1, 2, \ldots, k,$$

where ϵ_{ij}'s have a normal distribution with an unknown variance σ^2, zero mean and μ_i's are unknown parameters. We are interested in testing the linear hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$. To simplify the discussion we use the following notations

$$SSTr = \sum_{i=1}^{k} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2, \; SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2, \; n = \sum_{i=1}^{k} n_i \text{ and}$$

$$F1 = \frac{SSTr/(k-1)}{SSE/(n-k)},$$

where $\bar{Y}_{i.} = \sum_{j=1}^{n_i} Y_{ij}/n_i$ and $\bar{Y}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}/n$. By replacing $\boldsymbol{\beta} = (\mu_1 \mu_2 \ldots \mu_k)'$ in Theorem 3.1, it can be shown that $SS = SSE, SS^* = SSTr, r = (k-1), F = F1$ and the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ is rejected if the observed value of $F1$ statistic is greater than or equal to $F_{1-\gamma,k-1,n-k}$. The case described above is referred to as a one-way analysis of variance which is a very simple case of linear hypothesis testing. One-way ANOVA has many applications in agricultural and engineering sciences.

6 Illustrative Examples

Example 6.1 An experiment is conducted to determine if there is a difference in the breaking strength of a monofilament fibre produced by four different machines for a textile company. Also it is known that all fibres are of equal thickness. A random sample is selected from each machine. The fibre strength y for each specimen is shown in Table 1. The one-way ANOVA model is $Y_{ij} = \mu_i + \epsilon_{ij}, j = 1, 2, \ldots, n_i;$ $i = 1, 2, 3, 4$. We are going to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. All computations are done by R software [11].

In the traditional statistics point of view and based on Theorem 3.1, we have $F1 = 2.789$ and $F_{0.95,3,32} = 2.901$. Therefore we accept H_0 at the crisp significance level $\gamma = 0.05$ because $F1 < F_{0.95,3.32}$. In other words, there is not any difference at significance level 0.05 in the breaking strength of a monofilament fibre produced by

Table 1 Breaking strength data where y is strength in pounds

Machine 1	$y_{11} = 37$	$y_{12} = 41$	$y_{13} = 40$	$y_{14} = 40$	$y_{15} = 39$
	$y_{16} = 35$	$y_{17} = 39$	$y_{18} = 39$	$y_{19} = 40$	$y_{110} = 43$
Machine 2	$y_{21} = 39$	$y_{22} = 41$	$y_{23} = 41$	$y_{24} = 40$	$y_{25} = 43$
	$y_{26} = 41$	$y_{27} = 42$	$y_{28} = 38$	$y_{29} = 40$	
Machine 3	$y_{31} = 45$	$y_{32} = 42$	$y_{33} = 40$	$y_{34} = 41$	$y_{35} = 40$
	$y_{36} = 40$	$y_{37} = 41$	$y_{38} = 40$		
Machine 4	$y_{41} = 41$	$y_{42} = 44$	$y_{43} = 42$	$y_{44} = 41$	$y_{45} = 41$
	$y_{46} = 41$	$y_{47} = 43$	$y_{48} = 39$	$y_{49} = 41$	

four machines. In follows, consider two situations to understand the need of presenting fuzzy-decision-based approach:

1. Let us only change $y_{11} = 37$, in Table 1, to $y_{11} = 36$. Now we have $F1 = 2.907$ and $F_{0.95,3,32} = 2.901$. So we reject $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, for $\gamma = 0.05$ because $F1 \geq F_{0.95,3,32}$ (i.e. there is a difference in the breaking strength of a monofilament fibre produced by four machines).
2. Reconsider the observations of the experiment in Table 1, if we change the crisp significance level $\gamma = 0.05$ to $\gamma = 0.06$ then the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is rejected since $F1 = 2.789$ is greater than $F_{0.94,3,32} = 2.732$.

Therefore, we are not sure whether to accept or reject H_0, since the values of $F_{1-\gamma,k-1,n-k}$ and $F1$ are close to each other, based on data in Table 1 and $\gamma = 0.05$. We overcome the sensitivity of this test by using linear hypothesis testing based on a fuzzy test statistic and a fuzzy significance level. In this example we have $SS = SSE = 97.419$, $SS^* = SSTr = 25.469$, $r = (k - 1) = 3$, $F = F1 = 2.789$ and $n = 36$. Based on Theorem 4.1.1 a fuzzy estimate for σ^2 is a fuzzy number with α-cuts

$$(\widetilde{\sigma^2})_\alpha = \begin{cases} \left[97.419/\chi^2_{1-\alpha+\alpha(0.533),32}, 97.419/\chi^2_{\alpha(0.533),32} \right] & 0.01 \leq \alpha \leq 1 \\ (\widetilde{\sigma^2})_{0.01} & 0 \leq \alpha < 0.01 \end{cases}$$

where $p' = 0.533$ is obtained from the relation $\chi^2_{p',32} = 32$. By Lemma 4.1.1, the membership function of this fuzzy estimate can be given by

$$\widetilde{\sigma^2}(u) = \begin{cases} \dfrac{1-G(97.419/u)}{1-0.533} & \dfrac{97.419}{\chi^2_{0.99+0.01(0.533),32}} \leq u \leq \dfrac{97.419}{32} \\ \dfrac{G(97.419/u)}{0.533} & \dfrac{97.419}{32} \leq u \leq \dfrac{97.419}{\chi^2_{0.01(0.533),32}} \\ 0 & otherwise, \end{cases}$$

where G is the cumulative distribution function of a χ^2 variable with 32 degrees of freedom, as depicted in Fig. 4. Also, under the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ an unbiased fuzzy estimate for σ^2 based on Theorem 4.1.2 is a fuzzy number with the following α - cuts:

$$(\widetilde{\sigma^2_{H_0}})_\alpha = \begin{cases} \left[25.469/\chi^2_{1-\alpha+\alpha(0.608),3}, 25.469/\chi^2_{\alpha(0.608),3} \right] & 0.01 \leq \alpha \leq 1 \\ (\widetilde{\sigma^2_{H_0}})_{0.01} & 0 \leq \alpha < 0.01 \end{cases}$$

and $p'' = 0.608$ is obtained from the relation $\chi^2_{p'',3} = 3$. Now by Theorem 4.1.3, the observed value of the fuzzy test statistic $\tilde{F}$ is a fuzzy number with the following α-cuts (see Fig. 5):

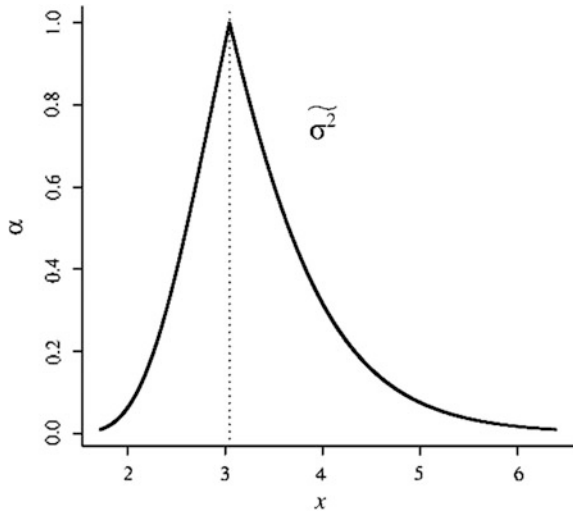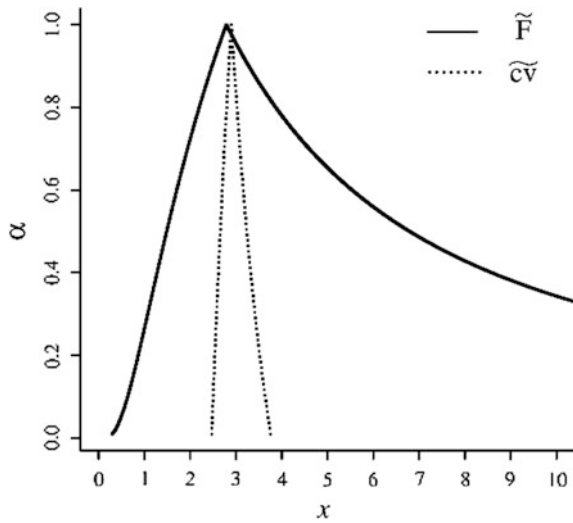**Fig. 4** The fuzzy estimator for σ^2

Fig. 5 The fuzzy numbers $\tilde{F}$ and $\widetilde{cv}$

$$\tilde{F}_\alpha = \begin{cases} \left[\dfrac{\chi^2_{\alpha(0.533),32}}{\chi^2_{1-\alpha+\alpha(0.608),3}} 0.261, \dfrac{\chi^2_{1-\alpha+\alpha(0.533),32}}{\chi^2_{\alpha(0.608),3}} 0.261 \right] & 0.01 \leq \alpha \leq 1 \\ \tilde{F}_{0.01} & 0 \leq \alpha < 0.01. \end{cases}$$

By using Definition 4.2.1 we consider the fuzzy significance level as a fuzzy number with the following α-cuts:

Table 2 Quantity of dissolved oxygen (%)

Location 1	$y_{11} = 7.8$	$y_{12} = 6.4$	$y_{13} = 8.2$	$y_{14} = 6.9$		
Location 2	$y_{21} = 6.7$	$y_{22} = 6.8$	$y_{23} = 7.1$	$y_{24} = 6.9$	$y_{25} = 7.3$	
Location 3	$y_{31} = 7.2$	$y_{32} = 7.4$	$y_{33} = 6.9$	$y_{34} = 6.4$	$y_{35} = 6.5$	
Location 4	$y_{41} = 6$	$y_{42} = 7.4$	$y_{43} = 6.5$	$y_{44} = 6.9$	$y_{45} = 7.2$	$y_{46} = 6.8$

$$\tilde{\gamma}_\alpha = \begin{cases} [0.02 + 0.03\alpha, 0.08 - 0.03\alpha] & 0.01 \le \alpha \le 1 \\ \tilde{\gamma}_{0.01} & 0 \le \alpha < 0.01, \end{cases}$$

and by Theorem 4.2.1 one can obtain α-cuts of the fuzzy critical value as follows:

$$(\widetilde{cv})_\alpha = \begin{cases} \left[F_{(0.92 + 0.03\alpha),3,32}, F_{(0.98 - 0.03\alpha),3,32} \right] & 0.01 \le \alpha \le 1 \\ (\widetilde{cv})_{0.01} & 0 \le \alpha < 0.01. \end{cases}$$

The graphs of the fuzzy numbers $\widetilde{cv}$ and $\tilde{F}$ are shown in Fig. 5. Finally, by Decision rule 4.2.1, since $A_R/(A_R + A_L) = 0.8971$ where $A_R = 9.7892$ and $A_L = 1.1230$, the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is rejected for every credit level $\varphi \in (0, 0.8971]$. In fact it is possible for us to reject H_0 for a high level of credit, since high ratio of observed values of the test statistic lead to reject H_0.

Example 6.2 The quantity of oxygen dissolved in water is used as a measure of water pollution. Samples are taken at four locations in a lake and the quantity of dissolved oxygen is recorded in [12] as in Table 2 (lower reading corresponds to greater pollution). We would like to see that whether the data indicate a significant difference in the average amount of dissolved oxygen for the four location based on a fuzzy test statistic and the crisp significance level $\gamma = 0.05$. The one-way ANOVA model is $Y_{ij} = \mu_i + \epsilon_{ij}, j = 1, 2, \ldots, n_i; i = 1, 2, 3, 4$.
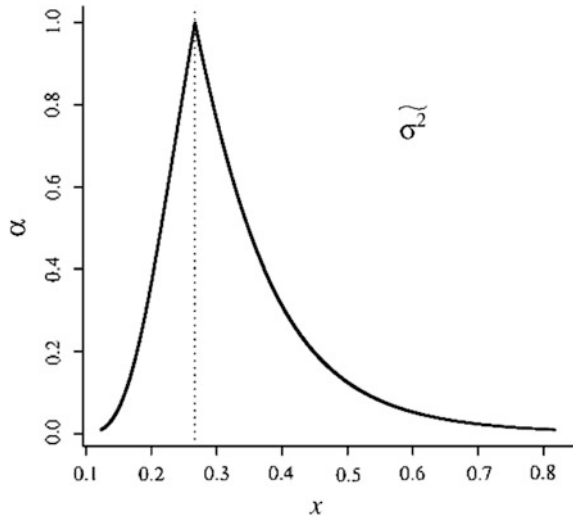
In this example we have $SS = SSE = 4.267$, $SS^* = SSTr = 0.718$, $r = (k - 1) = 3$, $F = F1 = 0.897$, $n = 20$ and $F_{0.95,3,16} = 3.239$. By Lemma 4.1.1, the membership function of the fuzzy estimate for σ^2 is given as follows:

$$\tilde{\sigma}^2(u) = \begin{cases} \frac{1 - G(4.267/u)}{1 - 0.547} & \frac{4.267}{\chi^2_{0.99 + 0.01(0.547),16}} \le u \le \frac{4.267}{16} \\ \frac{G(4.267/u)}{0.547} & \frac{4.267}{16} \le u \le \frac{4.267}{\chi^2_{0.01(0.547),16}} \\ 0 & otherwise, \end{cases}$$

where G is the cumulative distribution function of a χ^2 variable with 16 degrees of freedom, as depicted in Fig. 6.

By Theorem 4.1.3, the observed value of the fuzzy test statistic $\tilde{F}$ is a fuzzy number with the following α-cuts:

Fig. 6 The fuzzy estimator for σ^2

$$\tilde{F}_\alpha = \begin{cases} \left[\dfrac{\chi^2_{\alpha(0.547),16}}{\chi^2_{1-\alpha+\alpha(0.608),3}} 0.168, \ \dfrac{\chi^2_{1-\alpha+\alpha(0.547),16}}{\chi^2_{\alpha(0.608),3}} 0.168 \right] & 0.01 \leq \alpha \leq 1 \\ \tilde{F}_{0.01} & 0 \leq \alpha < 0.01. \end{cases}$$

The graph of the fuzzy test statistic $\tilde{F}$ is shown in Fig. 7. Finally, by Decision rule 4.1.1, since $A_R/(A_R + A_L) = 0.578$ where $A_R = 2.467$ and $A_L = 1.797$, the

Fig. 7 The fuzzy test statistic $\tilde{F}$
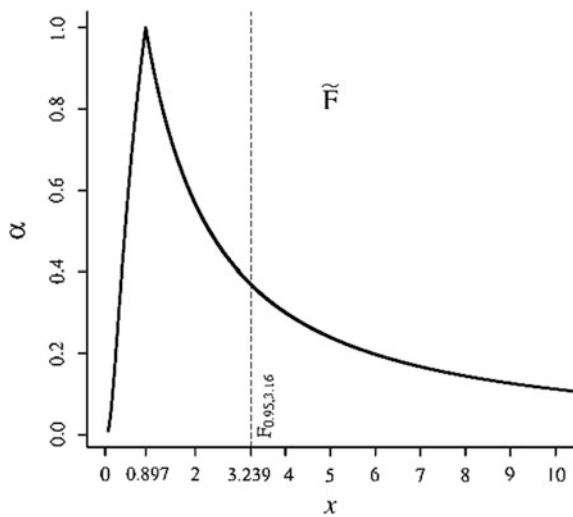
null hypothe-sis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ is accepted for every credit level $\varphi \in (0.578, 1]$. In other words, there is not any difference at significance level 0.05 in the average amount of dissolved oxygen for the four location for every credit level $\varphi \in (0.578, 1]$.

7 Conclusions

We have applied fuzzy techniques to linear hypothesis testing in this paper. Basically, in this method a set of $(1 - \alpha)$ 100 % confidence intervals, for all $0.01 \leq \alpha \leq 1$, are employed to produce the notion of the fuzzy test statistic. Also the concept of fuzzy critical value is derived based on α-cuts of a defined fuzzy significance level. Then, decision rules are provided based on these notions. Employing all the confidence intervals from the 99 % to the 0 % rather than only a single confidence interval results in using far more information in data for the statistical inference. Moreover, this method improves the statistical hypotheses testing when there is an uncertainty in accepting or rejecting the hypotheses. This issue is clarified by practical examples. As a simple case of the linear hypothesis testing, one-way analysis of variance based on fuzzy test statistic and fuzzy significance level is discussed. Nevertheless, as a matter of fact, the proposed method in this article is still applicable to other cases of linear hypothesis testing. An interesting topic for future research is the study of the proposed method on the linear hypothesis testing when the hypotheses are fuzzy rather than crisp. Also, one can consider this problem based on fuzzy data with crisp/fuzzy parameters.

References

1. Arefi, M., Taheri, S.M.: Testing fuzzy hypotheses using fuzzy data based on fuzzy test statistic. J. Uncertain Syst. **5**, 45–61 (2011)
2. Buckley, J.J.: Fuzzy statistics: hypothesis testing. Soft. Comput. **9**, 512–518 (2005)
3. Buckley, J.J.: Fuzzy Probability and Statistics. Springer, Berlin Heidelberg (2006)
4. Falsafain, A., Taheri, S.M.: On Buckley's approach to fuzzy estimation. Soft. Comput. **15**, 345–349 (2011)
5. Falsafain, A., Taheri, S.M., Mashinchi, M.: Fuzzy estimation of parameter in statistical model. Int. J. Comput. Math. Sci. **2**, 79–85 (2008)
6. Holena, M.: Fuzzy Hypothesis testing in a framework of fuzzy logic. Int. J. Intell. Syst. **24**, 529–539 (2009)
7. Kalpanapriya, D., Pandian, P.: Int. J. Mod. Eng. Res. **2**, 2951–2956 (2012)
8. Montgomery, D.C.: Design and Analysis of Experiments, 3rd edn. Wiley, New York (1991)
9. Nguyen, H.T., Walker, E.A.: A First Course in Fuzzy Logic, 3rd ed. Paris: Chapman Hall/CRC (2005)
10. Nourbakhsh, M.R., Parchami, A., Mashinchi, M.: Analysis of variance based on fuzzy observations. Int. J. Syst. Sci. **44**(4), 714–726 (2013)
11. Rizzo, M.L.: Statistical Computing with R. Chapman Hall/CRC, Paris (2008)

12. Rohatgi, V.K., Ehsanes Saleh, A.K.M.: An Introduction to Probability and Statistics, 2nd edn. Wiley, New York (2001)
13. Searl, S.R.: Linear Models. Wiley, New York (1971)
14. Taheri, S.M., Arefi, M.: Testing fuzzy hypotheses based on fuzzy test statistic. Soft. Comput. **13**, 617–625 (2009)
15. Taheri, S.M., Hesamian, G.: Goodman-Kruskal measure of assocition for fuzzy-categorized variables. Kybernetika **47**(110), 122 (2011)
16. Wu, H.C.: Analysis of variance for fuzzy data. Int. J. Syst. Sci. **38**, 235–246 (2007)
17. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–359 (1965)