Sample Size Estimation for Statistical Comparative Test of Training by Using Augmented Reality via Theoretical Formula and OCC Graphs: Aeronautical Case of a Component Assemblage

Fernando Suárez-Warden, Yocelin Cervantes-Gloria,
and Eduardo González-Mendívil

Instituto Tecnológico y de Estudios Superiores de Monterrey -
Monterrey campus (Monterrey Tech)
Centro de innovación en diseño y tecnología (CIDYT)
Aerospace Competence Development Center (ACDC)
Cátedra de investigación en Máquinas Inteligentes
Av. Eugenio Garza Sada 2501, Monterrey, NL. 64849. Mexico
fersuarezw@gmail.com, yocelincervantes@gmail.com, egm@itesm.mx

Abstract. Advances in Augmented Reality applied to learning the assembly operation in terms of productivity, must certainly be evaluated. We propose a congruent sequence of statistical procedures that will lead to determine the estimated work sample size ($\hat{n}$) according to the level of significance required by the aeronautical sector or a justified similar one and the estimated (sometimes preconceived) value of the plus-minus error (E or E±). We used the Kolmogorov-Smirnov test to verify that a normal distribution fits, it is a nonparametric one (free-distribution). And by taking into account normal population, confidence interval is determined using the Student's t distribution for (n-1) degrees of freedom. We have gotten E error and obtained various sample sizes via statistical formula. Additionally, we proceeded to utilize both an alpha α significance level and a beta β power of the test selected for the aeronautical segment to estimate the size of the sample via application of Operating Characteristic Curves (CCO), being this one of the ways with statistical high rigor. Several scenarios with different $\hat{n}$ values make up the outcome, herein. We disclosed diverse options for the different manners of estimation.

Keywords: Augmented Reality (AR), plus-minus error or margin of error, confidence interval (CI), Operating Characteristic Curves (OCC).

1 Introduction

This document exposes how to determine the work sample size; it consists of a sequence of steps required to find a satisfactory estimated sample size value.

Aeronautical maintenance training is a strategic area within which technicians must go carefully and where people have to take advantage of intelligent technologies such as AR.

R. Shumaker (Ed.): Virtual and Mixed Reality, Part II, HCII 2011, LNCS 6774, pp. 80–89, 2011.

Advances in AR, directed to training for the assembly operations in terms of productivity, need to be evaluated. Since this viewpoint of performance, we think about time invested in training and its return when we use AR to compare with the required standard rate.

Having this in mind, we propose a consistent sequence of statistical steps to get our target of estimating the size of the sample $\hat{n}$ for the level of significance (α) used by the aeronautical sector or a justified similar one and the estimated (properly preconceived) value of the plus-minus error or margin of error (E).

At first we start with the aim of searching such an estimate for this sample size, so one via that we use is a nonparametric method (free-distribution) named Kolmogorov-Smirnov Test, which is performed to verify that a normal distribution fits.

We execute this critical test because it is possible to utilize some known distributions for testing only under normal behavior case, as Statistical Theory states.

After that, taking into account normal population, the confidence interval (CI) is determined by using the Student's **t** distribution for (n-1) degrees of freedom. Work sample sizes for initialization are considered to attain the combined standard deviation $\hat{\mathbf{s}}$ which is used in Student´s **t** distribution to find out several values of the E error.

Consequently we have obtained the E error to get the estimated sample size $\hat{n}$ via statistical formula utilized in the theory developed for these cases. We can hold the above paragraph when we have a small sample size case. And this is the situation here.

We are going to generate several scenarios with the same level of significance or probabilities of accomplish type I error (α) and confident values of E for the previous $\hat{n}$ calculations in order to make up the searched outcome.

Then we should be careful in estimating the sample size to be representative and avoid fall not only in lack but also in excess.

1.1 We Propose to Use via Theoretical Formula and OCC Graphs

Being rigorist, we build a methodology for providing two different ways:

a) Formula derived from the confidence interval assuming normal population
b) Graphs of Operating Characteristic Curves (OCC)

According to Kazmier [4], for normal population we use theoretical formula for sample size (n) based on the population variance and margin of error, but first we must do an adaptation to the estimation of n and the estimated value of sampling standard deviation **S**.

Besides, we must meet the requirements for alpha α significance level (or probability of type I error) and beta β power of the test (or probability of type II error) selected for the aeronautical segment to state the estimated sample size via graphics of Operating Characteristic Curves (CCO), being this, one of the techniques of high statistical rigor.

2 Paper Preparation

Now we acquire the data achieved from Mercado [6]: the assemblage times by using RA for a wing of an aircraft RV-10 kit. Below is a table with the results obtained in his thesis.

Table 1. Results from the AR method

EXPERIMENT	X = TIME (min)	ERRORS (qty)	QUESTIONS (qty)
1	248	3	4
2	210	2	3
3	236	2	3
4	**284**	2	**5**
5	197	0	1
6	166	0	2

Source: Mercado[6]. *Improving Training and Maintenance. Operations in Aeronautic Related Processes Using Augmented Reality.*

2.1 Data Analysis and Filtering

With the aim of avoid slanted sampling or skewed data generation that would lead to a bad estimation of S, we will accomplish the next data analysis:

Mercado [6] shows in pp. 32 his *Table 3* where only the variables *Time* and *Questions* are centered on *Boxplot diagram* (from Minitab) while the variable *Errors* is biased.

Also, in his *Table 1*, he only take into account the columns *Time* and *Questions* that appear as Graphics 1st. and 3rd. in *Table 2* (pp. 30 and 31 respectively) and here is where we note with interest the last point that corresponds to his: EXPERIMENT #4 with *Time 284, Questions 5.*

Filtering. Because that is the farthest point from the straight line of the statistical trial of Kolmogorov-Smirnov Normality test and therefore we will remove it and determine well, in a filtered form, a more representative average than that obtained by Mercado [6]. In a second step we will eliminate and substitute such a point.

2.2 Formulas

Permit us present several formulas:

1. We obtain the average of the data Xav with the following formula

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (1)$$

2. The sample variance

$$s^2 = \frac{\sum_{i=1}^{n}\left\{(x_i - \bar{x})^2\right\}}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2}{n-1} \qquad (2)$$

3. And based on discrete uniform distribution, from Eq. (3.7) in Appendix:

$$u^2(x_i) = \frac{1}{12}(a_+ - a_-)^2 \qquad (3)$$

where: α_+ = MAX , $\alpha\_$ = MIN, and $\hat{s}^2 = u^2$

4. The sampling standard deviation

$$s = \sqrt{s^2} \qquad (4)$$

5. The Combined Std. Deviation

$$\mathbf{\hat{s}_D} = \mathbf{\hat{s}} / n^{½} \qquad (5)$$

6. Weighed Combined Std. Deviation

$$\mathbf{\hat{s}_{Xav1 - Xav2}} = \mathbf{\hat{s}_1} (\mathbf{1/n_1 + 1/n_2})^{½} \qquad (6)$$

7. In order to obtain the estimation of **d**:

$$đ = (1/2\ \mathbf{\hat{s}_D})\ |\ Xav1 - Xav2| \qquad (7)$$

8. From OCC graph we get $\hat{n} = (n^* + 1) / 2$ (8)

for a given value of β (= 0.01 in aeronautical sector).

Other formulas shall appear according to paper development. For the aviation sector, we take the level of significance α = 1% = 0.01, thus: α/2= .005

From statistical tables for normal distribution, we find $z_{\alpha/2}$ $Z_{.005} = 2.58$

To begin with, we initially make calculations, by mean of previous formulas, using filtered Mercado´s data (see **Filtering** at section **2.1**), which are shown below:

Reminder: we removed the farthest point from the straight line of the statistical tests of Kolmogorov-Smirnov Normality test and therefore the average changes to:

Xav =Average ($\bar{x}$)	211.4
Variance (s^2)	1053.8
Std. Deviation (s)	32.462
Combined Std. Deviation ($\mathbf{\hat{s}_D}$)	= 27/ (5)½ = 12
d estimated = đ	= [(½)/ (12)] I 243 – 211.4I = 1.32

Where we considered X = TIME.

2.3 Kolmogorov- Smirnov Test

We will validate if the experimental data model fits the normal distribution statistical model, by executing the Fit goodness Kolmogorov- Smirnov Test:

Ho: The spent time follows a normal distribution

Hi: The spent time does not follow a normal distribution

To determine should the obtained and filtered data follow a normal distribution and are significant accomplish the Kolmogorov- Smirnov (K-S) test. This K-S test indicates whether and how well the model fits the data.

The K-S statistical test calculates the distance between the theoretical normal distribution and the distribution estimated from the residuals, adjusted for suspensions, and reports the p- value (specific probability value). When the p-value is less than α (1%) the null hypothesis is rejected, and accepted otherwise.

Once the K-S test was applied to several selected cases to be discarded someone or various, we achieved an acceptable fit goodness for: Normal distribution and Gamma distribution.

In our case, the results were a p-value > α (= 1%) for the K-S fit goodness test:

Ho: normal distribution fits
H1: normal distribution does not fit

And therefore we could not reject Ho, so we got to assume normality. Please, see Hogg and Tanis [3] who developed examples of K-S test. Note: Gamma distribution fits too.

2.4 Confidence Interval

According to Kazmier [4] we can apply t-student test in order to determine the confidence interval (CI), by using the next expression for experimental values (texp):

$$\text{texp} = (\bar{x}_1 - \bar{x}_2) \,/\, \hat{s}\,(1/n_1 + 1/n_2)^{½} \tag{9}$$

And we can build the confidence interval CI for small random sample using the t distribution with degrees of freedom ν = n1 + n2 - 2 (assuming and / or having tested normal population by mean of fit goodness via K-S test). From:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\nu,\alpha}\ \hat{s}\,(1/n_1 + 1/n_2)^{1/2} \tag{10}$$

we get:

$$E = CI / 2 = t_{\nu,\alpha}\ \hat{s}\,(1/n_1 + 1/n_2)^{1/2} \tag{11}$$

3 Via Theoretical Formula and OCC Graphs

After it has been proved our population or universe were normal it should be used with all forcefulness the corresponding theoretical formula of Kazmier [4]:

$$n = (z\,\sigma / E)^{2} \quad (12)$$

which we adapt for dealing with an estimated value:

$$\hat{n} = (z\,\hat{s} / E)^{2} \quad (13)$$

where $\hat{s}$ is the estimator for σ, E is the margin of error and $\hat{n}$ is the estimator for n.

3.1 Scenarios Developed by Formula

It was stated that: $\hat{n} = (z\,\hat{s} / E)^2 = \hat{s}^{2} (z / E)^2$ from Eq. (13) We knew the result $\hat{s}$ = 32.462 based on Std. Deviation (s) formula.

The Confidence Interval CI is depicted for small random sample (assuming and / or having tested normal population fit goodness, via K-S Test).

1. Reminder 1: we removed the farthest point from the straight line of the K-S normality test (see **Filtering** at section **2.1**) to get a 1st new margin of error E. Such as Eq. (10):

$$\bar{x} \pm t_{v,\alpha}\ \hat{s}/n^{1/2} \quad (14)$$

$\bar{x} \pm t_{4,\,0.01}\ \hat{s}/n^{1/2} = 211 \pm (4.604)\ 32.462\ /5^{1/2} = 211 \pm 66.83$; E = 66.83

By introducing attained data into adapted formula:

$\hat{n} = (z\,\hat{s} / E)^{2} = \hat{s}^2\ (z / E)^{2} = 1053.8(2.58 /66.83)^2 = 1.57 \approx 2$; $\hat{n}$ is an integer.

2. Reminder 2: we removed the farthest point from the straight line of the K-S normality test (see **Filtering**) and substituted it to get a modified average and a 2nd representative margin of error E:

$\bar{x} \pm t_{5,\,0.01}\ \hat{s}/n^{1/2} = 211 \pm (4.032)\ 23.67\ /6^{1/2} = 211 \pm 38.95$; E = 38.95

By introducing attained data into adapted formula:

$\hat{n} = (z\,\hat{s} / E)^2 = 560.3\ (2.58 /38.95)^2 = 2.45 \approx 3$

3. For two depending samples (associated pairs), where $\nu = n_1 + n_2 - 2 = 5 + 5 - 2 = 8$

$\bar{x} \pm t_{8,\,0.01}\ \hat{\mathbf{s}}_1\ (1/ n_1 + 1/ n_2)^{½} = 211 \pm (3.5)\ 14.97 = 211 \pm 52.4$; E = 52.4

By bringing together obtained into adapted formula:

$\hat{n} = (z\,\hat{s} / E)^2 = 560.3\ (2.58 /52.4)^2 = 1.35 \approx 2$

3.2 Scenarios Developed by OCC Charts

4. With the resultant quantities gotten by formulas (2.2 section), we reached the first amount for đ (= 1.32). And placing this estimated value of **d** as input into CCO chart:

For β = .01 from OCC graph we get $\tilde{n}^* = 15$

Then, $\hat{n} = (\tilde{n}^* + 1) / 2 = (15 + 1) / 2 = 8$ by Eq. (8)

5. After that, we detached the farthest point from the straight line of the K-S normality test (see **Filtering**) and therefore other representative average resulted:

Xav modified = 211.4 and $\hat{\mathbf{s}}_2$ modified = 32.462

$\hat{\mathbf{s}}_1$ = 23.67 based on discrete uniform distribution (rectangular distribution)

And we assumed $\hat{\mathbf{s}}_1 \approx \hat{\mathbf{s}}_2 = \hat{\mathbf{s}} = 23.67$

On the other hand, Combined Std. Deviation $\hat{s}_D = \hat{s} / n^{½} = 23.67/ (5)^{½} = 10.6$
So: $đ = [\ 1/ 2(10.6)]\ |243 - 211.4| = 1.4925 \approx 1.5$
With this value as input in CCO chart, for a beta β = .01, from OCC graph:
We get $ñ^* = 14$ and $\hat{n} = (ñ^* + 1) / 2 = (14 + 1) / 2 = 7.5 \approx 8$ by Eq. (8)

6. We removed the farthest point from the straight line of the K-S normality test (see **Filtering**) and substitute it for the previous average to get another new Xav (= 211.3)

And another new Combined Std. Deviation $\hat{s}_D = \hat{s} / n^{½} = 23.67/ (6)^{½} = 9.66$
So, we have a new: $đ = [1/ 2\ (9.66)]\ |243 - 211.3| = 1.6384$
With this value as input in CCO chart, for a beta: β = .01 from OCC graph:
We get $ñ^* = 8$ and $\hat{n} = (ñ^* + 1) / 2 = (8 + 1) / 2 = 4.5 \approx 5$ by Eq. (8)

7. Also, for depending samples (associated pairs), a weighed Combined Std. Deviation:

$$\hat{s}_{Xav1 - Xav2} = \hat{s}_1 \, (1/n_1 + 1/n_2)^{½} = 14.97 \qquad \text{from Eq. (6)}$$

Thus we exploit it in determining $đ = 1/ 2\ (14.97)]\ |243 - 211.3| = 1.058$
For β =.01, from OCC graph $ñ^* = 21$, so $\hat{n} = (ñ^* + 1) / 2 = (21 + 1) / 2 = 11$ by Eq.(8)

4 Results

The scenarios we obtained for $\hat{n}$ were:

Table 2. Various scenarios of estimated sample size (for depending and independent sample types)

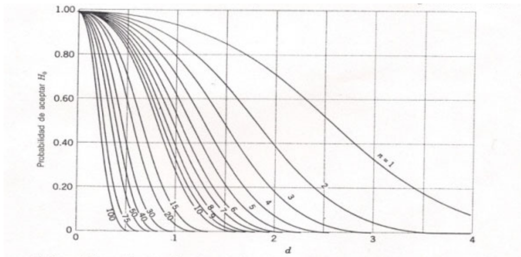
Scenario	**Sample type**	**Via**	$\hat{s}$	**z**	**E error**	**Estimation $\hat{n}$**	**α**	**β**
1.-	independent	Formula	32.462	2.58	66.83	2	.01	-
2.-	independent	Formula	23.67	2.58	38.95	3	.01	-
3.-	depending	Formula	14.97	2.58	52.4	2	.01	-
Scenario	**Sample type**	**Via**	$\hat{s}_D$		**d**	**Estimation $\hat{n}$**	**α**	**β**
4.-	independent	CCO	12	-	1.32	8	.01	.01
5.-	independent	CCO	10.6	-	1.4925	8	.01	.01
6.-	independent	CCO	9.66	-	1.6384	5	.01	.01
7.-	depending	CCO	14.97	-	1.058	11	.01	.01

By the next manage approvals, we briefly want to clarify the OCC charts use. Input:

a) Set α = 0.01 (aeronautical sector), hence select the correspondent OCC chart
b) Enter to horizontal axis by addressing the estimated value for **d:**

$$đ = (1/ 2\hat{s}_D)\ |Xav1 - Xav2| \qquad \text{from Eq. (7)}$$

c) Select the β = 0.01 (aeronautical sector) on the vertical axis of the OCC chart.



b) Curvas CO para diferentes valores de n para la prueba normal bilateral con nivel de significancia α = 0.01.

Fig. 1. Operating Characteristic Curves (OCC). Source: Hines and Montgomery [2]

Output:

- ñ* value in OCC chart
- with ñ* thus we estimate the work sample size $\hat{n} = (ñ^* + 1) / 2$ by Eq. (8)

5 Conclusion

We learn to solve assessment sample size problems by doing research in advance so necessary, especially when the situation demands to know a good approximation of the corresponding magnitude before running the experiment. This is our study case, herein.

Several scenarios of estimated sample size could be construed from two points of view.

We think about different manners for getting the two outcome types. However it is crucial to consider the beta probability (β) to achieve our target given that we also must try to avoid falling in the type II error. For this reason, the second way gave more credible results for the proposed sizes at 4, 5, 6 and 7 rows of Table 2 by utilizing the CCO via.

We invite you to take advantage of this study which is focused on estimating the size for small work samples in order to cut costs within an adequate criterion.

Acknowledgments. We want to thanks to professors Salvador García-Lumbreras, Ph.D., José Guadalupe Ríos, Ph.D. and María del Carmen Temblador, Ph.D. from Monterrey Tech for their support and recommendations.

References

1. Campbell, J.D., Jardine, A.K.S.: Maintenance Excellence: Optimizing Life-Cycle Decisions. Marcel Dekker, USA (2001)
2. Hines, W.W., Montgomery, D.C.: Probabilidad y Estadística para ingeniería y administración. In: CECSA (2005)
3. Hogg, R.V., Tanis, E.A.: Probability and Statistical Inference. Macmillan, NYC (1977)
4. Kazmier, L.: Estadística Aplicada a la Administración y la Economía. McGrawHill, NewYork (1993)

5. Mendenhall, W., Schaffer.: Estadística Matemática con Aplicaciones. Grupo Editorial Iberoamérica (1990)
6. Mercado Field, E.R.: Improving Training and Maintenancce Operations in Aeronautic Related Processes Using Augmented Reality. ITESM- Monterrey. México (2010)
7. Organización Internacional para la Estandarización (ISO). Estándar Int. ISO 3534-1 Estadísticas –Vocabulario y símbolos-Parte I: Probabilidad y Estadística en gral., 1a ed. Génova, Suiza (1993)

Appendix: Discrete Uniform Model (Rectangular Distribution)

Organización Internacional para la Estandarización (ISO) [7].

Disclaimer. This section is composed by the international standard ISO 3534-1 Vocabulary and Symbols-Statistics, Part I. It is an official translation from ISO organization, so sorry we are not authorized for converting to English. Let us use Spanish.

Distribución uniforme discreta (distribución rectangular). (*)El uso correcto de la información disponible para la evaluación Tipo B de la incertidumbre estándar en la medición tiene un adentramiento basado en la experiencia y conocimiento general. Es una habilidad que puede ser aprendida con la práctica. Una evaluación de Tipo B de la incertidumbre bien basada puede ser tan confiable como la de Tipo A, especialmente en una situación de medición donde la valoración Tipo A es fundada sólo en un pequeño número comparativo de observaciones estadísticamente independientes. Casos a ver:

a) Sólo un *valor singular* es conocido por la cantidad X_i, p. ej. un valor sencillo medido, un valor resultante de una medición previa, un valor de referencia de literatura o un valor de corrección; usado para x_i. La incertidumbre estándar $u(x_i)$ asociada con x_i debe ser adoptada cuando es dada. De otro modo, es calculada de datos de incertidumbre inequívocos, pero si no están disponibles, ésta debe evaluarse basándose en la experiencia.

b) Cuando la *distribución de probabilidad* puede ser asumida de la cantidad X_i, basada en teoría o experiencia, el valor esperado y la raíz cuadrada de la varianza deben ser tomados como un estimado de x_i y la incertidumbre estándar asociada $u(x_i)$ respectivamente.

c) Si sólo los *límites inferior y superior* a_+ y a_- pueden ser estimados para el valor de la cantidad X_i (por ejemplo, las especificaciones de la manufacturera de un instrumento de medición, un rango de temperatura, un error de redondeo o de truncado resultado de una reducción automática), una distribución de probabilidad con una densidad de probabilidad constante entre esos límites (distribución de probabilidad rectangular) tiene que ser asumida para la posible variabilidad de la cantidad de entrada X_i. Según el caso (**b**):

$$x_i = \frac{1}{2}\left(a_+ + a_-\right)$$ para el valor estimado Xav o Xprom; y: (3.6)

$$u^2\left(x_i\right) = \frac{1}{12}\left(a_+ - a_-\right)^2 \quad (3.7)$$

para el cuadrado de la incertidumbre estándar $\hat{s}^2$. Donde: $a+$ = MAX , $a-$ = MIN.