# Graphing calculators in teaching statistical p-values to elementary statistics students

Eric Benson
American University in Dubai

ABSTRACT

The statistical output of interest to most elementary statistics students is the p-value, outputted in computer programs like SPSS, Minitab and SAS. Statistical decisions are sometimes made using these values without understanding the meaning or how these values are calculated. Most elementary statistics textbooks calculates p-values for z-tests while omitting t-tests, c^2-tests, and F-tests because of the complexity of these distributions. All elementary statistics texts surveyed for this paper use Excel and Minitab output to report p-values for t-tests, c^2-tests, and F-tests. This paper shows an alternative for calculating p-values using the embedded statistical distributions in the graphing calculator regardless of the complexity of the statistical distributions.

Keywords: Hypothesis testing, Null hypothesis, Alternative hypothesis, p-value, Level of significance
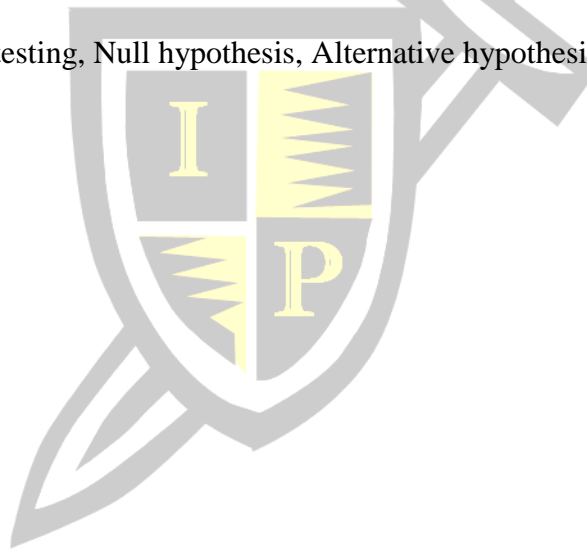
INTRODUCTION

Statistical inference methods are used for testing claims and predicting population parameters, methods that are very useful for statistical decisions. The *p*-value method is often used in null hypothesis significance testing (Gliner, Leech, and Morgan, 2002). Statistical programs on tablets, graphical calculators, Minitab and SPSS, all automatically report *p-values* for hypothesis tests. Students and researchers often report p-values without understanding the meaning or interpretation of this value (Hoekstra, Kiers, Johnson and Groenier. 2006). Statistical *p-value* (Keller, G. 2005) is the probability of exceeding the test statistic value, given that the null hypothesis is true. The decision criterion for the p-value method, is reject the null hypothesis if the *p <0.05*, or reject the null hypothesis if $p > 0.05$, 0.05 is the probability of a Type I Error (rejecting the null hypothesis given that it is true).

The purpose of this paper is to show how to use the embedded statistical distributions in a graphing calculator, for example, the TI-84 Plus to calculate p-values. This approach will serve as an alternative to the approach used by most authors of elementary statistics texts, which output computer generated *p*-value for *t*-tests, c^2-tests, and *F*-tests.

p-Value or Probability Method

Hypothesis testing is an Achilles Heel (Hogg, 1991) for most elementary students, but the concepts underling hypothesis testing is quite old and goes back to the ancient Greeks. Two hypotheses are needed to test any statistical claim, the null hypothesis (H_0) and the alternative hypothesis (H_a). Only one of these hypotheses can be true for a given situation (Keller, 2005). Two possible errors can be made in any hypothesis test, Type I error with probability a occurs when H_0 is rejected when it is true and a Type II error with probability b occurs when (H_0) is accepted when it is false.

The decision criterion for the *p-value* method, accept H_0 when $p \geq \alpha$, or reject H_0 when $p < \alpha$, where α is the significance level or size of the critical region (significant region), the size of alpha is usually decided on before the test is conducted. The size $\alpha = 0.05$ is an arbitrary value that is commonly used as a default when the size of the significance level is not given. This statistical reasoning help bridged the concept or inferential statistics (hypothesis testing and estimation) areas most undergraduate students find unattainable. Accordingly graphing calculator makes statistical reasoning accessible to all students (Burrill, 1996).

Four of the most commonly used statistical tests in elementary statistics (*z*-test, *t*-tests, c^2-tests, and *F*-tests) were chosen from (Keller, 2005) and (Black, 2008) to demonstrate the calculation of the p-values with the use of the TI-84 Plus embedded statistical functions. Comparatively Minitab outputted results are also given as they appear in most of the elementary statistics texts sampled for this paper.

Example 1: z-test for one population proportion

This example is from Keller (2005, p. 390). Calculate the *p*-value of the test of the following hypotheses given that $\hat{p} = 0.63$ and $n = 100$: $H_0 : p = 0.60$ vs $H_1 : p > 0.60$. The test

statistic for this test is $z_0 = 0.6123$. Using the definition $p\text{-value} = \Pr(z > z_0)$, this probability statement is easily calculated with the Ti-84 Plus using the following keys strokes:
For this example the null hypothesis is accept because p-value ^{3} 0.05.

Table 2 gives the p-value probability definitions (Hogg and Tanis, 2001), and it's TI-84 plus representation for the symmetric normal and t distributions.

Example 2: **t-test for one population mean**

This example is from Black (2008, p. 314). A random sample of size 20 is taken, resulting in a sample mean of 16.45 and a sample standard deviation of 3.59. Assume x is normally distributed and use this information and $a = .05$ to test the following hypotheses.

$$H_0 : m = 16 \qquad H_a : m \; ^1 \; 16$$

$$t_0 = \frac{16.45 - 16}{3.59 / \sqrt{20}} = 0.5606$$

$$p\text{-value} = 2\Pr(t \; ^3 \; 0.5605) = 0.5816$$

For this example the null hypothesis is accept because p-value ^{3} 0.05.

Table 4 gives the p-value probability definitions (Hogg and Tanis, 2001), for the test of one population variance using Chi-Squared distribution and it's TI-84 Plus representation.

Example 3: c^2-**test for one population variance**

This example is from Black (2008, p. 325). Previous experience shows the variance of a given process to be 14. Researchers are testing to determine whether this value has changed. They gather the following dozen measurement of the process. Use these data and $a = .05$ to test the null hypothesis about the variance. Assume the measurements are normally distributed.
52   44   51   58   48   49   38   49   50   42   55   41

$$H_0 : s^2 = 14 \qquad H_a : s^2 \; ^1 \; 14$$

$$c_0^2 = \frac{(n-1)s^2}{s^2} = \frac{(11)5.8846^2}{14} = 27.2081$$

$$p\text{-value} = 2\Pr(c^2 \; ^3 \; 27.2081) = 0.00854$$

The null hypothesis for this example is rejected because, $p - value < 0.05$. Based of sample evidence the population variance of the process has changed from 14, is the conclusion that has to be made. This test is significant while the pervious two examples were non-significant, at the alpha equals 0.05 significance level.

Table 6 gives the p-value probability definitions for testing the ratio of two population variances using the F distribution and it's TI-84 Plus representation.

Example 4: **F-test for the ratio of two variances**

This example is from Keller (2005, p. 451). A statistics professor hypothesized that not only would the means vary but also so would the variance if the business statistics course was

taught in two different ways but had the same final exam. He organized and experiment wherein on section of the course was taught using detailed PowerPoint slides whereas the other required students to read the book and answer questions in class discussions. A sample of the marks was recorded and lasted next. Can we infer that the variances of the marks differ between the two sections?

| **Class 1** | 64 | 85 | 80 | 64 | 48 | 62 | 75 | 77 | 50 | 81 | 90 |
| **Class 2** | 73 | 78 | 66 | 69 | 79 | 81 | 74 | 59 | 83 | 79 | 84 |

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \qquad H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

$$F = \frac{s_1^2}{s_2^2} = \frac{193.673}{60} = 3.2280$$

$$p - value = 2\Pr\left(F \geq 3.2280\right) = 0.07833$$

Based on the sample evidence the two population variances do not differ, because the p-value of this test is larger than the 5% significance level.

CONCLUSION

This article discusses how embedded statistical distributions in the graphing calculator can be used as an alternative to computer output use in elementary statistics texts books to report statistical p-values for t-tests, c^2-tests, and F-tests. Using the graphing calculator help students visualize and better understand concepts in probability and statistics (Chance, Ben-Zvi, Garfield & Median 2007), the structure of the alternative hypothesis that leads directly to the probability definition of p-values (Tables 2, 4 and 6). Consider a general left-tail t-test $H_0 : m = m_0$ vs $H_a : m < m_0$, students have to know the direction of the alternative hypothesis to arrived at the appropriate probability statement, $\Pr\left(t < t_0\right)$ to calculate the p-value for this test. Minitab on the other hand have dialog boxes for students to select right-tail, left-tail or two-tail tests and output the p-value automatically. Using the method presented in this paper, students will be able to understand the statistical and probability concepts needed to calculate p-values instead of clicking dialog boxes as with most statistical programs.

In industry because of the large amount of data and responsibilities managers have, the use of dialog boxes in computer program is necessary. Students on the other hand need to learn the material from the theorems and definitions governing the statistical methods.

REFERENCES

Berger, J. O (2003). "Could Fisher, Jefferys and Neyman Have Agreed on Testing?," *Statistical Science* , Vol. 18, No. 1, 1-32

Black, K. (2008), Business Statistics for Contemporary Decision Making, 5th Ed. John Wiley & Sons, Inc.

Burrill, G. (1996). "*Graphing Calculators and their Potential for Teaching and Learning Statistics*", Proceedings of the 1996 IASE (International Association of Statistics Education) Round Table Conference, 15 – 28.

Chance, B., Ben-Zvi, D., Garfield, J. Medina, E. (2007). "The Role of Technology in Improving Student Learning of Statistics", *Technology Innovations in Statistics Education*, 1(1).

Gliner, J. A., Leech, N.L., Morgan, G. A. (2002). " Problems with Null Hypothesis Significance Testing (NHST): What Do the Text Book Say?", The Journal of Experimental Education, 71(1), 83-92.

Hoekatra, R., Kiers, H. A. L., Johnson, A. and Groenier, M. (2006). "Problems when interpreting research results using only *p-value* and sample size." International Conference on Teaching Statistics-7

Hogg, R. V. "Statistical Education: Improvements are Badly Needed,*"* The American Statistician, 45(4), 342 – 343 (1991)

Hogg, R.V. and Tanis, E. A. (2001), Probability and Statistical Inference, 6th Ed., Prentice Hall

Keller, G (2005). Statistics for Management and Economics. 7th Ed. Thomson Brooks/Cole.

Peck, R., Devore, J. L. (2012). Statistics: The Exploration & Analysis of Data. 7th Ed. Brooks/Cole, Cengage Learning
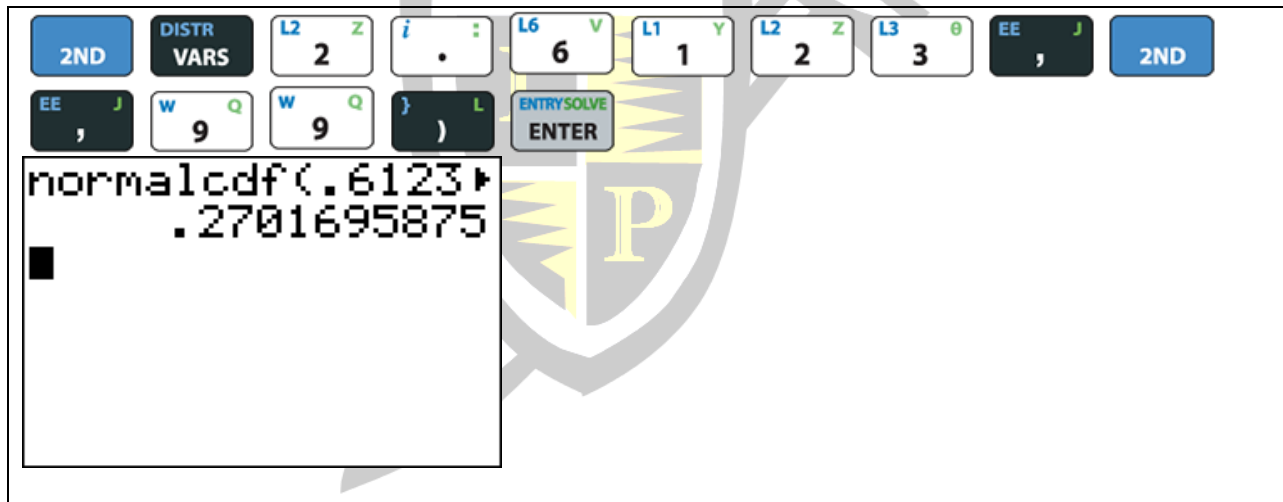
APPENDIX:

Figure 1: Screen display and keys strokes for the z-test TI-84 Plus

Test and CI for One Proportion
Test of $p = 0.6$ vs $p > 0.6$

Sample	X	N	Sample p	95% Lower Bound	Z-Value	P-Value
1	63	100	0.630000	0.550586	0.61	0.270

Using the normal approximation.

Table 1: Minitab Output for the z-test

Case 1 (Right tail test)

$H_0 : m \pounds m_0$

$H_1 : m > m_0$

$p - value = \Pr\left(z \, {}^3 \, z_0\right)$

$p - value = normalcdf\left(z_0, E99\right)$ TI-84 Plus

$p - value = tcdf\left(t_0, E99, df\right)$ TI-84 Plus

Case 2 (Left tail test)

$H_0 : m \, {}^3 \, m_0$

$H_1 : m < m_0$

$p - value = \Pr\left(z \, \pounds \, z_0\right)$

$p - value = normalcdf\left(-E99, z_0\right)$ TI-84 Plus

$p - value = tcdf\left(-E99, t_0, df\right)$ TI-84 Plus

Case 3 (Two tails test)

$H_0 : m = m_0$

$H_1 : m \, {}^1 \, m_0$

$p - value = 2\Pr\left(z \, {}^3 \, |z_0|\right)$

$p - value = 2normalcdf\left(|z_0|, E99\right)$ TI-84 Plus

$p - value = 2tcdf\left(|t_0|, E99, df\right)$ TI-84 Plus

Table 2: z-test and t-test p-value definitions
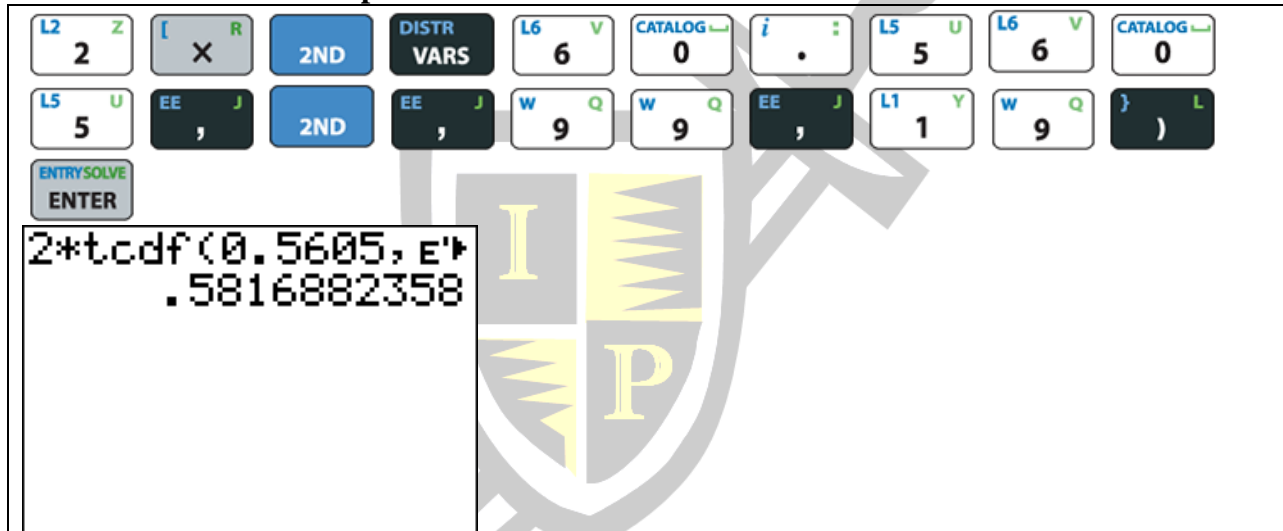

Figure 2: Screen display and keys strokes for the t-test TI-84 Plus

One-Sample T

Test of mu = 16 vs not = 16

N	Mean	StDev	SE Mean	95% CI	T	P
20	16.450	3.590	0.803	(14.770, 18.130)	0.56	0.582

Table 3: Minitab Output for the t-test

Case 1 (Right tail test)

$H_0 : S^2 \pounds S_0^2$

$H_1 : S^2 > S_0^2$

$p - value = \Pr\left(C^2 \;^3\; C_0^2\right)$

$p - value = C^2 cdf\left(C_0^2, E99, df\right)$ TI-84 Plus

Case 2 (Left tail test)

$H_0 : S^2 \;^3\; S_0^2$

$H_1 : S^2 < S_0^2$

$p - value = \Pr\left(C^2 \pounds C_0^2\right)$

$p - value = C^2 cdf\left(0, C_0^2, df\right)$ TI- 84 Plus

Case 3 (Two tails test)

$H_0 : S^2 = S_0^2$

$H_1 : S^2 \neq S_0^2$

$$p - value = \begin{cases} 2\Pr\left(C^2 \geq C_0^2\right) \text{ if } C_0^2 > df = n - 1 \\ 2\Pr\left(C^2 \leq C^2\right) \text{ if } C_0^2 < df = n - 1 \\ 2\,C^2 cdf\left(C_0^2, E99, df\right) \quad \text{TI-84 Plus} \\ 2\,C^2 cdf\left(0, C_0^2, df\right) \quad \text{TI-84 Plus} \end{cases}$$

Table 4: C^2 - *test* **p-value definitions**
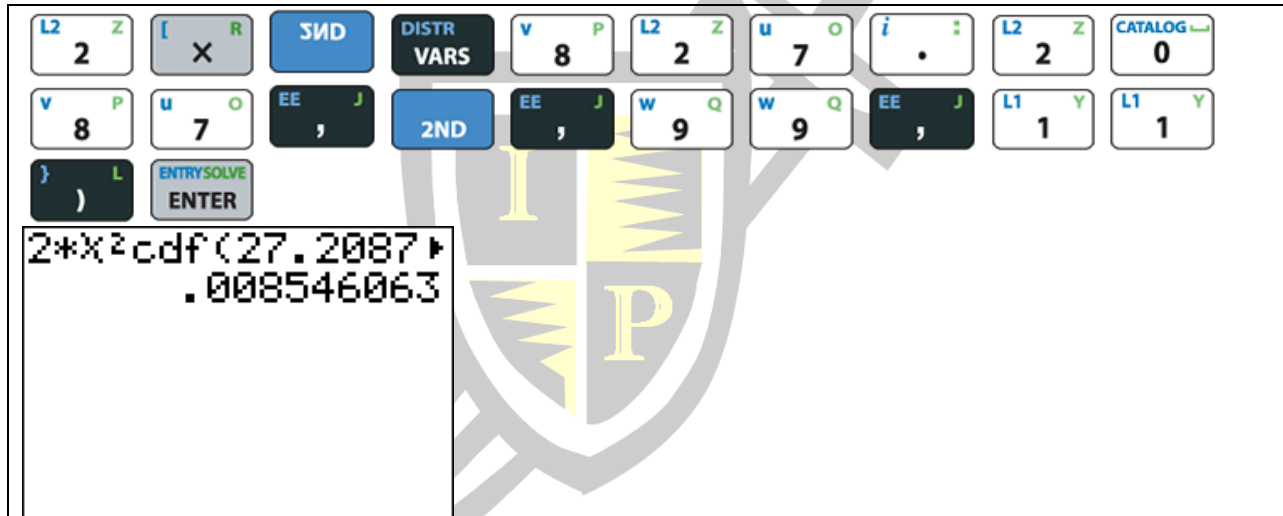

Figure 3: Screen display and keys strokes for the C^2**-test**

Test and CI for One Variance: x
Method

Null hypothesis   Sigma-squared = 14
Alternative hypothesis Sigma-squared not = 14

The chi-square method is only for the normal distribution.

Statistics

Variable N StDev Variance
x   12 5.88   34.6

	Tests			
		Test		
Variable	Method	Statistic	DF	P-Value
x	Chi-Square	27.21	11	<u>0.009</u>

Table 5: Minitab Output for the C^2-test

Case 1 (Left tail test)

$$H_0 : \frac{S_1^2}{S_2^2} \geq 1$$

$$H_1 : \frac{S_1^2}{S_2^2} < 1$$

$$p\text{-}value = \Pr\left(F \leq F_0\right)$$

$$p\text{-}value = Fcdf\left(0, F_0, df_1, df_2\right) \text{ TI-84 Plus}$$

Case 2 (Right tail test)

$$H_0 : \frac{S_1^2}{S_2^2} \leq 1$$

$$H_1 : \frac{S_1^2}{S_2^2} > 1$$

$$p\text{-}value = \Pr\left(F \geq F_0\right)$$

$$p\text{-}value = Fcdf\left(F_0, E99, df_1, df_2\right) \text{ TI-84 Plus}$$

Case 3 (Two tails test)

$$H_0 : \frac{S_1^2}{S_2^2} = 1$$

$$H_1 : \frac{S_1^2}{S_2^2} \neq 1$$

$$p\text{-}value = \begin{cases} 2\Pr\left(F \geq F_0\right) & \text{if } F_0 > 1 \\ 2\Pr\left(F \leq F_0\right) & \text{if } F_0 < 1 \\ 2Fcdf\left(F_0, E99, df_1, df_2\right) & \text{TI-84 Plus} \\ 2Fcdf\left(0, F_0, df_1, df_2\right) & \text{TI-84 Plus} \end{cases}$$

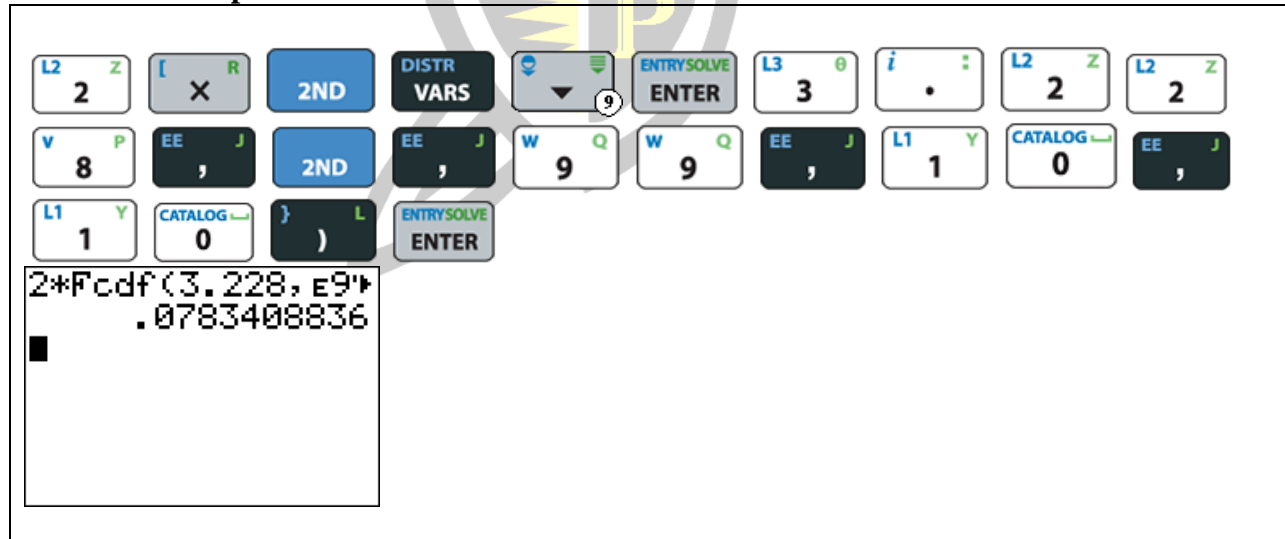**Table 6: F-test p-value definitions**

Figure 4: Screen display and keys strokes for the F-test

Test and CI for Two Variances: x1, x2
Method

Null hypothesis   Variance(x1) / Variance(x2) = 1
Alternative hypothesis Variance(x1) / Variance(x2) not = 1
Significance level   Alpha = 0.05
Statistics

Method	DF1	DF2	Test Statistic	P-Value
F Test (normal)	10	10	3.23	0.078

Table 7: Minitab Output for the F-test