**Common Symbols**

n = Sample Size | N = Population Size
$\bar{x}$ = Sample Mean | μ = Population Mean
s = Sample Standard Deviation | $\sigma$ = Population Standard Deviation
$s^2$ = Sample Variance | $\sigma^2$ = Population Variance
$\hat{p}$ = Sample Proportion | p = Population Proportion
r = Sample Correlation Coefficient | $\rho$ = Population Correlation Coefficient

**Descriptive Statistics** **In Excel use Data > Data Analysis > Descriptive Statisitcs**

$\bar{x} = \frac{\Sigma x}{n}$ $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$ $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$ $z = \frac{x-\bar{x}}{s}$ Range = Max – Min

IQR=interquartile range = $Q_3 - Q_1$

Finding Outlier Limits: Lower = $Q_1$ – (1.5*IQR) Upper = $Q_3$ + (1.5*IQR)

**Probability**

Union Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ Complement Rule: $P(A^C) = 1 - P(A)$

Conditional Probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ Mutually Exclusive Events: $P(A \cap B) = 0$

Dependent Events: $P(A \cap B) = P(B) \cdot P(A|B)$ Independent Events: $P(A \cap B) = P(B) \cdot P(A)$

**Discrete Probability Distributions**

$E(x) = \mu = \Sigma x \cdot P(x)$ $\sigma^2 = \left(\Sigma x^2 \cdot P(x)\right) - \mu^2$ $\sigma = \sqrt{\left(\Sigma x^2 \cdot P(x)\right) - \mu^2}$

**Binomial Distribution**: $P(x) = {}_nC_x \cdot p^x \cdot q^{(n-x)}$ $\mu = n \cdot p$ $\sigma^2 = n \cdot p \cdot q$ $\sigma = \sqrt{n \cdot p \cdot q}$

| **Signs Are Important For Discrete Distributions!** | | |
|---|---|---|
| **P(X=x)** | **P(X≤x)** | **P(X≥x)** |
| Is | Is less than or equal to | Is greater than or equal to |
| Is equal to | Is at most | Is at least |
| Is exactly the same as | Is not greater than | Is not less than |
| Has not changed from | Within | |
| Is the same as | | |
| | | |
| =BINOM.DIST(x,n,p,false) | =BINOM.DIST(x,n,p,true) | =1-BINOM.DIST(x-1,n,p,true) |
| | | |
| | **P(X>x)** | **P(X<x)** |
| | More than | Less than |
| | Greater than | Below |
| | Above | Lower than |
| | Higher than | Shorter than |
| | Longer than | Smaller than |
| | Bigger than | Decreased |
| | Increased | Reduced |
| | Smaller | Larger |
| | | |
| | =1-BINOM.DIST(x,n,p,true) | =BINOM.DIST(x-1,n,p,true) |

**Poisson Distribution**: $P(x)=\frac{e^{-\mu}\mu^{x}}{x!}$ new $\mu$ = old $\mu \times \left(\frac{\text{new units}}{\text{old units}}\right)$

| **P(X=x)** | **P(X≤x)** | **P(X≥x)** |
|---|---|---|
| Is | Is less than or equal to | Is greater than or equal to |
| Is equal to | Is at most | Is at least |
| Is exactly the same as | Is not greater than | Is not less than |
| Has not changed from | Within | |
| Is the same as | | |
| | | |
| =POISSON.DIST(x,mean,false) | =POISSON.DIST(x,mean,true) | =1-POISSON.DIST(x-1,mean,true) |
| | | |
| | **P(X>x)** | **P(X<x)** |
| | More than | Less than |
| | Greater than | Below |
| | Above | Lower than |
| | Higher than | Shorter than |
| | Longer than | Smaller than |
| | Bigger than | Decreased |
| | Increased | Reduced |
| | Smaller | Larger |
| | | |
| | =1-POISSON.DIST(x,mean,true) | =POISSON.DIST(x-1,mean,true) |

**Hypergeometric Distribution**

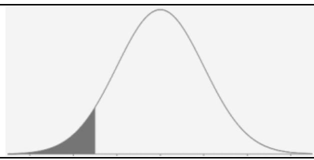
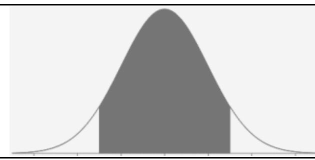
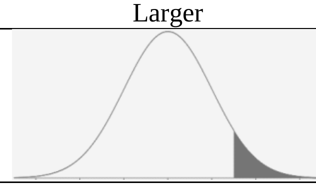
| **P(X=x)** | **P(X≤x)** | **P(X≥x)** |
|---|---|---|
| Is | Is less than or equal to | Is greater than or equal to |
| Is equal to | Is at most | Is at least |
| Is exactly the same as | Is not greater than | Is not less than |
| Has not changed from | Within | |
| Is the same as | | |
| | | |
| =HYPGEOM.DIST(x,n,R,N,false) | = HYPGEOM.DIST(x, n,R,N,true) | =1- HYPGEOM.DIST(x-1, n,R,N,true) |
| | | |
| | **P(X>x)** | **P(X<x)** |
| | More than | Less than |
| | Greater than | Below |
| | Above | Lower than |
| | Higher than | Shorter than |
| | Longer than | Smaller than |
| | Bigger than | Decreased |
| | Increased | Reduced |
| | Smaller | Larger |
| | | |
| | =1- HYPGEOM.DIST(x, n,R,N,true) | = HYPGEOM.DIST(x-1, n,R,N,true) |

**Continuous Distributions**

$z=\frac{x-\mu}{\sigma}$ $x=z\cdot\sigma+\mu$ Central Limit Theorem: $z=\frac{\bar{x}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$

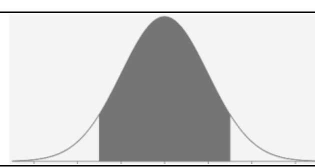
Note that for a continuous distribution there is no area at a line under the curve, so ≥ and > will have the same probability and use the same Excel commands.

## Normal Distribution

| Normal Distribution Finding a Probability | | |
|---|---|---|
| **$P(X \le x)$ or $P(X<x)$** | **$P(x_1<X<x_2)$ or $P(x_1 \le X \le x_2)$** | **$P(X \ge x)$ or $P(X>x)$** |
| Is less than or equal to | Between | Is greater than or equal to |
| Is at most | | Is at least |
| Is not greater than | | Is not less than |
| Within | | More than |
| Less than | | Greater than |
| Below | | Above |
| Lower than | | Higher than |
| Shorter than | | Longer than |
| Smaller than | | Bigger than |
| Decreased | | Increased |
| Reduced | | Larger |
|  |  |  |
| =NORM.DIST (x,μ ,σ,true) | =NORM.DIST ($x_2$,μ ,σ,true)-NORM.DIST ($x_1$,μ ,σ,true) | =1- NORM.DIST (x,μ ,σ,true) |

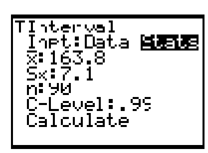
Note that the NORM.S.DIST function is for a standard normal when μ=0 and σ=1.

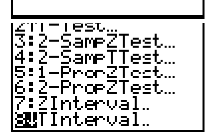
## Inverse Normal Distribution

| Normal Distribution Finding an X-value Given an Area or Probability | | |
|---|---|---|
| **$P(X \le x)$ or $P(X<x)$** | **$P(x_1<X<x_2)$ or $P(x_1 \le X \le x_2)$** | **$P(X \ge x)$ or $P(X>x)$** |
| Lower | Between | Upper |
| Bottom | | Top |
| Below | | Above |
| Reduced | | More than |
| Less than | | Greater than |
| Lower than | | Larger |
| Shorter than | | Higher than |
| Smaller than | | Longer than |
| Decreased | | Bigger than |
| | | Increased |
| | | |
|  |  |  |
| =NORM.INV(area,μ,σ) | $x_1$ =NORM.INV(1-area/2,μ ,σ) $x_2$ = \| $x_1$\| | =NORM.INV(1-area, μ,σ) |

**Confidence Intervals** **When n < 30 the variable must be approximately normally distributed.**

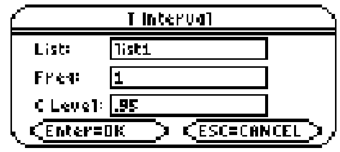
**t-Interval** **The 100(1 - α)% confidence interval for μ, σ is unknown, is** $\bar{x} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$.



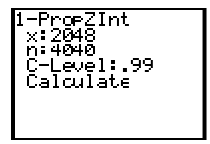

- On the TI-83 you can find a confidence interval using the statistics menu. Press the [STAT] key, arrow over to the [TESTS] menu, arrow down to the [8:TInterval] option and press the [ENTER] key. Arrow over to the [Stats] menu and press the [ENTER] key. Then type in the mean, sample standard deviation, sample size and confidence level, arrow down to [Calculate] and press the [ENTER] key. The calculator returns the answer in interval notation. Be careful, if you accidentally use the [7:ZInterval] option you would get the wrong answer.

- *Or* (If you have raw data in list one) Arrow over to the [Data] menu and press the [ENTER] key. Then type in the list name, $L_1$, leave Freq:1 alone, enter the confidence level, arrow down to [Calculate] and press the [ENTER] key.
- On the TI-89 go to the [Apps] **Stat/List Editor**, then select 2nd then F7 [Ints], then select **1:TInterval**. Choose the input method, data is when you have entered **data** into a list previously or **stats** when you are given the mean and standard deviation already. Type in the mean, standard deviation, sample size (or list name (list1), and Freq: 1) and confidence level, and press the [ENTER] key. The calculator returns the answer in interval notation. Be careful, if you accidentally use the [1:ZInterval] option you would get the wrong answer.

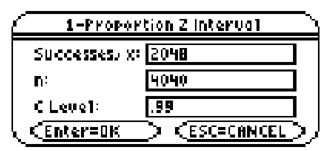



**1 Proportion z-Interval** $\hat{p} \pm z_{\alpha/2}\sqrt{\left(\frac{\hat{p}(1-\hat{p})}{n}\right)}$

- On the TI-84 press the [STAT] key, arrow over to the [TESTS] menu, arrow down to the [A:1-PropZInterval] option and press the [ENTER] key. Then type in the values for X, sample size and confidence level, arrow down to [Calculate] and press the [ENTER] key. The calculator returns the answer in interval notation. Note: sometimes you are not given the x value but a percentage instead. To find the x to use in the calculator, multiply $\hat{p}$ by the sample size and round off to the nearest integer. The calculator will give you an error message if you put in a decimal for x or n. For example if $\hat{p}$ = .22 and n = 124 then .22*124 = 27.28, so use x = 27.



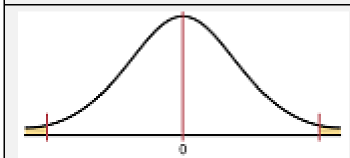
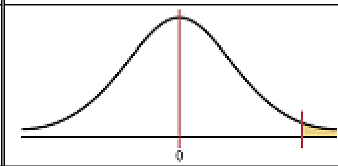
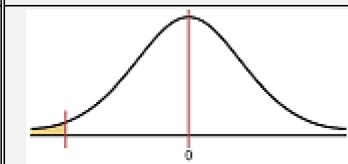

- On the TI-89 go to the [Apps] **Stat/List Editor**, then select 2nd then F7 [Ints], then select **5: 1-PropZInt**. Type in the values for X, sample size and confidence level, and press the [ENTER] key. The calculator returns the answer in interval notation. Note: sometimes you are not given the x value but a percentage instead. To find the x value to use in the calculator, multiply $\overline{p}$ by the sample size and round off to the nearest integer. The calculator will give you an error message if you put in a decimal for x or n. For example if $\overline{p}$ = .22 and n = 124 then .22*124 = 27.28, so use x = 27.




**Sample Size** $n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$ $n = \hat{p}(1-\hat{p})\left(\frac{z_{\alpha/2}}{E}\right)^2$ Always round n <u>up</u> to the next integer.

## Hypothesis Testing

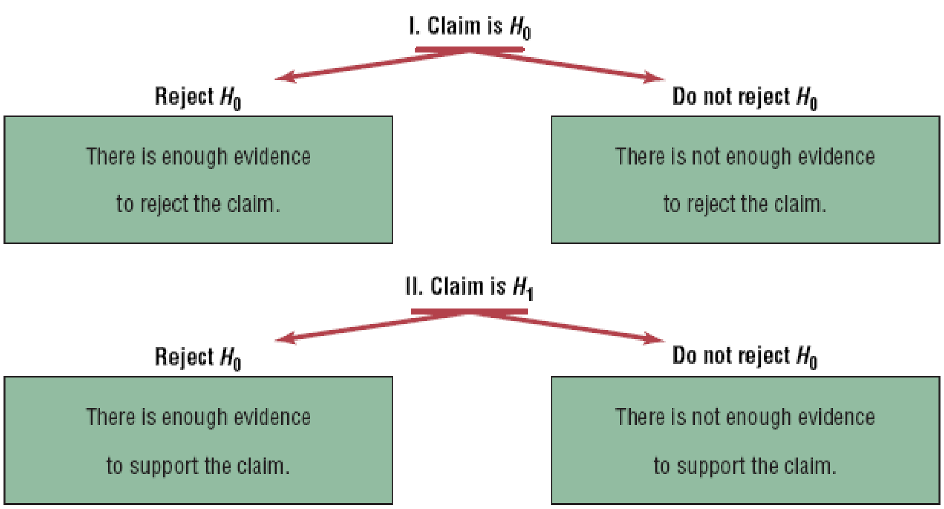
### Look for these key words to help set up your hypotheses:

| Two-tailed Test | Right-tailed Test | Left-tailed Test |
|---|---|---|
| **Null Hypothesis** | | |
| = | ≤ | ≥ |
| Is | Is less than or equal to | Is greater than or equal to |
| Is equal to | Is at most | Is at least |
| Is exactly the same as | Is not more than | Is not less than |
| Has not changed from | Within | |
| Is the same as | | |
|  |  |  |
| **Alternative Hypothesis** | | |
| ≠ | > | < |
| Is not | More than | Less than |
| Is not equal to | Greater than | Below |
| Is different from | Above | Lower than |
| Has changed from | Higher than | Shorter than |
| Is not the same as | Longer than | Smaller than |
| | Bigger than | Decreased |
| | Increased | Reduced |
| | Exceed | |

### The rejection rule:

- p-value method, reject $H_0$ when the p-value $\leq \alpha$ .
- Critical value method, reject $H_0$ when the test statistic is in the critical tail(s).

### How to start your conclusion for a hypothesis test:




Finish conclusion with context and units from question.

### One Sample Tests:

**1-Sample Mean t-test:** $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$

Test statistic when σ is unknown: $t = \dfrac{\bar{x} - \mu_0}{\left(s/\sqrt{n}\right)}$ with df= n – 1

**1-Sample Proportion z-test:** $\begin{matrix} H_0: p = p_0 \\ H_1: p \neq p_0 \end{matrix}$ Test statistic is $z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p \cdot q}{n}}}$ where $\hat{p} = \dfrac{x}{n}$

**Two Sample Tests:**

**2 Means – Independent Populations Assuming Equal Variances**

$\begin{matrix} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{matrix}$ or $\begin{matrix} H_0: \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0 \\ H_1: \mu_1 - \mu_2 \neq (\mu_1 - \mu_2)_0 \end{matrix}$ where usually you have $\begin{matrix} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{matrix}$

Test statistic when $\sigma_1^2$ and $\sigma_2^2$ are unknown: $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\left(s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)\right)}}$ Note that $(\mu_1 - \mu_2)_0 = 0$ in most cases

and the pooled variance is $s_p^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$.

**2 Means – Dependent Populations**

Find the difference (d) between each matched pairs. $\begin{matrix} H_0: \mu_D = 0 \\ H_1: \mu_D \neq 0 \end{matrix}$ Test statistic: $t = \dfrac{\bar{d} - \mu_0}{\left(s_d / \sqrt{n}\right)}$

**2 Proportions**

$\begin{matrix} H_0: p_1 = p_2 \\ H_1: p_1 \neq p_2 \end{matrix}$ Test statistic $z = \dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, (usually $(p_1 - p_2)_0 = 0$) where $\hat{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$, $\hat{q} = 1 - \hat{p}$

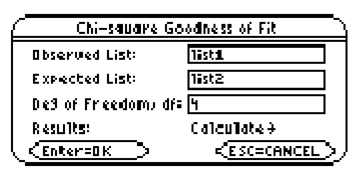
**Goodness of Fit Test**

"There is no preference" w/ 4 groups $H_0$: $p_1$=.25, $p_2$=.25, $p_3$=.25, $p_4$=.25 "There is not a preference"

$H_1$: At least one proportion is different. "There is a preference"

Proportions are 1/k or different percentages for each group given in the problem. If %'s are given use those decimals for each $p_i$. Expected values are found by taking each group's proportion times the sample size ($p_i \times n$), df=k–1. Test statistic: $\chi^2 = \sum \dfrac{(\text{O} - \text{E})^2}{\text{E}}$

- For the TI-84: Enter the observed frequencies in List1, the expected frequency in List2, press [STAT] key, arrow over to [TESTS] menu, arrow down to [$\chi^2$GOF–Test] (not available on the TI-83 and some TI-84 calculators)
- For the TI-89: Hypothesis test for three or more proportions (goodness of fit test). Go to the [Apps] **Stat/List Editor**, then type in the observed values into list 1, and the expected values into list 2. Select 2nd then F6 [Tests], then select **7: Chi-2GOF**. Type in the list names and the degrees of freedom (df = k-1). Then press the [ENTER] key to calculate. The calculator returns the $\chi^2$-test statistic and the p-value.

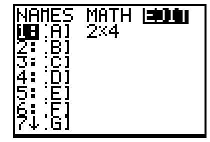



**Test for Independence**

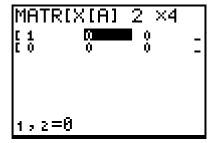
H0: <u>Variable 1</u> is independent of <u>Variable 2</u> (Change the variable names to fit the context of the question.)

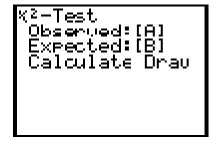
H1: <u>Variable 1</u> is dependent of <u>Variable 2</u>

Expected values are $\frac{(\text{Row Total})\cdot(\text{Column Total})}{\text{Grand Total}}$ for each cell & df = (r – 1)(c – 1)

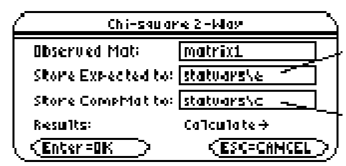



Test statistic: $\chi^2 = \sum \frac{(O-E)^2}{E}$ Critical Value: =CHI.INV.RT(α,df)




- For the TI-83 or TI-84: Press the [2nd] then [MATRX] key. Arrow over to the EDIT menu and 1:[A] should be highlighted, press the [ENTER] key. For a m X n contingency table, type in the number of rows(m) and the number of columns(n) at the top of the screen so that it looks like this MATRIX[A] m X n. For example a 2 X 3 contingency table, the top of the screen would look like this MATRIX[A] 2 X 4, as you hit [ENTER] the table will automatically widen to the size you put in. Now enter all of the observed values in there proper positions. Then press the [STAT] key, arrow over to the [TESTS] menu, arrow down to the option [ $\chi^2$ -Test] and press the [ENTER] key. Leave the default as Observed:[A] and Expected:[B], arrow down to [Calculate] and press the [ENTER] key. The calculator returns the $\chi^2$ -test statistic and the p-value. If you go back to the matrix menu [2nd] then [MATRX] key, arrow over to EDIT and choose 2:[B], you will see all of the expected values.
- TI-89: First you need to create the matrix for the observed values: Press: [Home] to return to the Home screen, press [Apps] and select **6:Data/Matrix Editor**. A menu is displayed, select **3:New**. The **New** dialog box is displayed. Press the right arrow key to highlight **2:Matrix,** and press [ENTER] to choose **Matrix** type. Press the down arrow key to highlight **1:main**, and press [ENTER], to choose **main** folder. Press the down arrow key, and then enter the name **o** in the **Variable** field. Enter **3** for **Row dimension** and **2** for **Column dimension**. Press [ENTER] to display the matrix editor. Enter **4** , **9** , **5** in **c1** and **7** , **2** , **3** in **c2**. Press ♦ [Apps] [ENTER] to close the matrix editor and return to the list editor. If you have more than one Application loaded, press ♦ [Apps], and then select **Stats/List Editor** .
  To display the **Chi-square 2-Way** dialog box, press 2nd then F6 [Tests], then select **8: Chi-2 2-way**. Enter in in the Observed Mat: **o** ; Store Expected to: **statvars\e** ; Store CompMat to: **statvars\c** . This will store the expected values in the matrix folder statvars with the name e, and the $(o\text{-}e)^2/e$ values in the matrix c. Press the [ENTER] key to calculate. The calculator returns the $\chi^2$ -test statistic and the p-value. If you go back to the matrix menu you will see all of the expected and $(o\text{-}e)^2/e$ values.




**Correlation and Regression**

$H_0: \rho = 0$

$H_1: \rho \neq 0$

Test statistic for correlation: $t = r\sqrt{\left(\frac{(n-2)}{(1-r^2)}\right)}$ with df= n – 2 $\hat{y} = a + bx$

a = y-intercept, b = slope, s = $s_{est}$ = standard error of estimate, $R^2$ = coefficient of determination, r = correlation coefficient.

**One-Factor ANOVA** table k=#of groups, N=total of all n's

$H_0$: $\mu_1=\mu_2=\mu_3=\ldots=\mu_k$

$H_1$: At least one mean is different CV: Always a right-tailed F, use Excel =F.INV.RT(α, $df_B$, $df_W$)

ANOVA Table

| Source | SS | df | MS | F (Test Statistic) |
|---|---|---|---|---|
| Between (Treatment or Factor) | $SS_B=\Sigma n(\bar{x}-\bar{x}_{GM})^2$ | k–1 | $MS_B = SS_B / df_B$ | $F = MS_B / MS_W$ |
| Within (Error) | $SS_W=\Sigma(n-1)s^2$ | N–k | $MS_W = SS_W / df_W$ | |
| Total | $SS_T$ | N–1 | | |