Testing Categorical (Qualitative) Variables

Type of Test	Purpose	Example	Hypotheses	Test Statistic and Formulas	Conditions* (Specific to this Test)
Chi Square Test for Independence Sections 10.1, 10.3	Test whether there is an association between 2 categorical variables. Data may be summarized in a Two-Way Table.	Is there an association between sugar consumption and ability to focus in children?	H_O : Var 1 and Var 2 are independent. H_a : Var 1 and Var 2 are associated.	$\chi^2 = \sum \frac{(O-E)^2}{E}$ **Df = (rows – 1)(columns – 1)** O = Observed counts E = Expected counts	Expected counts must be at least 5 in all cells
One Proportion Z Test Section 8.1 – 8.3	Test whether the proportion of successes in a single population is different than a specific percentage. Note: Proportions are often expressed as **percentages.**	Is the proportion of kids with asthma in the Central Valley higher than 8.4% (the national percentage)?	$H_O: \ p = p_O$ $H_a: \ p \neq p_O$ **(or >, or <, depending on the wording of the problem)**	$z = \frac{\hat{p} - p_O}{SE}$ $SE = \sqrt{\frac{p_O(1-p_O)}{n}}$	Expected counts in the Success and Failure groups must be at least 10: $E = np_O \geq 10$ $E = n(1-p_O) \geq 10$
One Proportion Confidence Interval Section 7.4		**Estimate $\pm$ Margin of Err.** $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	Z* is based on the confidence level 90% $\Rightarrow z^* = 1.645$ 95% $\Rightarrow z^* = 1.960$ 99% $\Rightarrow z^* = 2.576$	Notes: (1) For SE in the hypothesis test, you use p_O, since the entire test is based on the assumption that $p = p_O$. (2) For the confidence interval, you have to approximate the SE value by using $\hat{p}$.	

Testing Numerical (Quantitative) Variables

Type of Test	Purpose	Example	Hypotheses	Test Statistic and Formulas	Conditions* (Specific to this Test)
1 Sample T-Test for Mean	Test whether the average value of some numerical variable in a single population is different than a specific value.	Is the average height of male college students greater than 6.0 feet?	$H_O: \ \mu = \mu_O$ $H_a: \ \mu \neq \mu_O$ **(or >, or <, depending on the wording of the problem)**	$t = \frac{\bar{x} - \mu_O}{SE}$ $SE = \frac{s}{\sqrt{n}}$ $df = n - 1$	**Either** the sample size has to be 25 or more (Large Sample) **or**, if the sample is small, then the underlying population distribution of the variable has to be normal.
One Mean Confidence Interval Section 7.4		**Estimate $\pm$ Margin of Err.** $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$	t* is based on the confidence level and the degrees of freedom.	Notes: If the CI captures the null, then we do not reject the null. If the CI does not capture the null, then have a significant result; i.e., we do reject the null and accept the alternative.	

Testing Numerical (Quantitative) Variables: Two Groups					
Paired T-Test Section 9.5	Test whether the average of the <u>differences</u> between paired or dependent samples is zero or not.	Weigh a set of people. Put them on a diet plan. Weigh them after. Was the weight loss (the difference between Before and After) significant?	$H_O: \mu_D = 0$ $H_a: \mu_D \neq 0$ **(or >, or <, depending on the wording of the**	$t = \frac{\bar{D}}{SE}$ $SE = \frac{s_D}{\sqrt{n}}$ $df = n-1$	**Observations <u>between</u> groups are <u>Dependent</u>!** **Either** the sample size has to be 25 or more (Large Sample) **or**, if the sample is small, then the underlying population distribution of the variable has to be normal.
2 Sample T-Test For Means Section 9.5	Test whether there is a difference between the means of two groups.	Is the average speed of cyclists during rush hour greater than the average speed of drivers	$H_O: \mu_1 - \mu_2 = 0$ $H_a: \mu_1 - \mu_2 \neq 0$ **(or >, or <, depending on the wording of the problem)**	$t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$ $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	**Observations <u>between</u> groups are <u>Independent</u>!** **Either** each the sample for each group is 25 or more (Large Samples) **or**, if the samples are small, then the underlying population distributions of the variable has to be normal.
One-Way ANOVA Chapter 11	Test whether if there is a difference between the means of more than two groups.	In a diet study where there are three groups, with each group following a different diet, are there significant differences in average weight loss between the groups?	$H_O: \mu_1 = \mu_2 = \mu_3$ H_a : At least one mean is different from the others.	$F = \frac{MS_{between}}{MS_{within}}$	Same as Conditions for **2 Sample T-Test for means** **Also, population variance must be equal**

ANOVA Table

	DF	SS	MS	F	P-value
BETWEEN *SIGNAL*	***M* – 1**	$\sum n_i(\bar{x}_i - \bar{\bar{x}})^2$	$\frac{SS_{Between}}{DF_{Between}}$	$\frac{MS_{Between}}{MS_{Within}}$	$P(F \geq Test\ F)$
WITHIN *NOISE*	***N* – *M***	$\sum (n_i - 1) \cdot s_i^2$	$\frac{SS_{Within}}{DF_{Within}}$		
Total	***N* – 1**				

Confidence Intervals For Differences in Means (including Tukey Tests) Section 9.5, Section 11.4	**Estimated Difference** ± **Margin of Err.**	Notes: If the CI captures zero, there is not a significant difference between the groups, on average. If the CI does not capture zero, then there is a significant difference.

***All tests require that you have a random sample from the population of interest and independent observations within the sample group(s) (except for the Paired t-Test).**
All tests require a Large Population (Pop is at least 10 times the sample size).