

OVERVIEW OF STATISTICAL INFERENCE PROCEDURES

Selecting the inferential procedure: You should start with these questions:

1. Does the question need a confidence interval (goal is to estimate parameter) or a test of significance (goal is to assess strength of evidence against some claim about the parameter)?
2. Is the question dealing with means (quantitative data) or proportions (categorical data)?
3. Do I have one mean or one proportion? Am I comparing the means or proportions between two groups? Or are there two categorical variables that might be associated or two quantitative variables to correlate?

Standardized Statistic:

$$z = \frac{\text{statistic} - \text{mean of null distribution}}{SD \text{ of null distribution}}$$

2SD Method 95% Confidence Interval:

$$\text{statistic} \pm 2 (\text{SD of statistic})$$

**The SD of statistic can be read off the simulated distribution on the applet or estimated with the SE formula.*

One Sample	One Proportion (Categorical)	One Mean (Quantitative)
<i>Descriptive Statistics</i>		
Graphical Summary	Bar graph	Histogram, dotplot
Numerical Summary	\hat{p}, n	\bar{x}, s, n
<i>Null Hypothesis</i>	$H_0 : \pi = \text{---}$ (hypothesized population proportion or long-run probability of a process)	$H_0 : \mu = \text{---}$ (hypothesized population mean)
<i>Null Distribution</i>		
Mean	Mean of \hat{p} 's = π	Mean of \bar{x} 's = μ
Standard Deviation	$SD(\hat{p}) = \sqrt{\pi(1-\pi)/n}$	$SD(\bar{x}) = \sigma/\sqrt{n}$
Standard Error	$SE(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}$	$SE(\bar{x}) = s/\sqrt{n}$
<i>Simulation</i>	One Proportion applet	—
<i>Theory Based</i>	<i>One proportion Z test</i>	<i>One sample T test</i>
Validity Conditions	At least 10 successes and at least 10 failures in sample	Symmetric population or $n \geq 20$ and sample not too skewed
Applet	TBI applet, "One Proportion"	TBI applet, "One Mean"
Two Samples	Two Proportions (Categorical)	Two Means (Quantitative)
<i>Descriptive Statistics</i>		
Graphical Summary	Segmented Bar Graph or Mosaic Plot	Boxplots, histograms, or dot plots for each group
Numerical Summary	$\hat{p}_1, \hat{p}_2, n_1, n_2$	$\bar{x}_1, \bar{x}_2, s_1, s_2, n_1, n_2$
<i>Null Hypothesis</i>	$H_0 : \pi_1 - \pi_2 = 0$	$H_0 : \mu_1 - \mu_2 = 0$
<i>Simulation</i>	<i>Two Proportions applet</i>	<i>Multiple Means applet</i>
Standardized Statistic	$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SD \text{ of null distribution}}$	$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{SD \text{ of null distribution}}$
<i>Theory Based</i>	<i>Two proportions Z test</i>	<i>Two sample T test</i>
Validity Conditions	At least 10 successes and at least 10 failures in each sample (The count in each cell of the 2 way table is at least 10)	Symmetric populations or $n_1, n_2 \geq 20$ and samples not too skewed
Applet	TBI applet, "Two Proportions"	TBI applet, "Two Means"

*TBI applet: Theory Based Inference applet

Decide if two variables are associated: You need to decide whether each variable is quantitative or categorical. Questions that might apply here include:

- Is there an association between gender and favorite Olympic sport (chi-square)?
- Is there an association between height and how far someone long-jumps (regression)?

Two Variables	Both Categorical	Both Quantitative
<i>Descriptive Statistics</i> Graphical Summary Numerical Summary	Segmented Bar Graph / Mosaic Plot Conditional Proportions \hat{p}_i 's and n_i 's	Scatterplot Correlation Coefficient r
<i>Inferential Statistics</i> Null Hypothesis Alternative Hypothesis	<i>Chi Square Test</i> H_0 : no association between variables H_A : is an association	<i>Regression</i> H_0 : no association between variables ($\rho = 0$) H_A : is a positive / a negative / an association ($\rho > 0, \rho < 0, \rho \neq 0$)
<i>Simulation</i> Statistic	<i>Multiple Proportions applet</i> Chi Square χ^2	<i>Corr/Regression applet</i> Correlation Coefficient r
<i>Theory-Based</i> Validity Conditions Standardized Statistic P-value Confidence Intervals	<i>Chi Square Test</i> At least 10 observations in each cell of the table χ^2 Applet click "show χ^2 output" 95% pairwise confidence intervals for all the possible pairs	
		<i>Regression Line</i> $\hat{y} = \text{intercept} + \text{slope}(x)$ residual= actual y – predicted \hat{y}