|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Description** | **Model** | **Parameter of Interest** | **Point Estimate** | **Standard Error** | **Test Statistic** | **Sampling Distribution** |
| Single Proportion | *Assumptions: BINS* | , the true population proportion |  |  | , the observed number of successes | For confidence intervals: *.*  For hypothesis testing: where is the value of under the null hypothesis. |
| Difference in Proportions | *Assumptions: BINS within each group* | , the difference in true population proportions |  | For confidence intervals:  For hypothesis testing:  Where | where is the value of under the null hypothesis (often 0). |  |
| Single Mean | for  where D is an arbitrary distribution,  is the mean and is the std. dev. of D.  *Assumptions: Independence* | , the true population mean |  |  | where is the value of under the null hypothesis. |  |
| Difference in Means | See above for definitions.  *Assumptions: Independence* | , the difference in true population means |  |  | where is the value of under the null hypothesis (often 0). | W=  (provided in t.test() output) |
| Linear Regression | *Assumptions: Linear form (1), errors are normally distributed (2), around 0 (3), with constant variance (4)* | , the true population slope | (provided in lm() output) | (provided in lm() summary output) | Where is the value of under the null hypothesis (often 0). |  |

Created for STAT 240: Spring 2024, by Cameron Jones

**Hypothesis Testing**

* **Step 1: State your model.**
  + Define variables
  + State distribution (or equation for linear regression) which connects your *observed data* to the *parameter of interest.*
* **Step 2: State your hypotheses.**
  + usually captures the idea that there is no pattern, relationship, or difference of interest in the population.
  + or or captures the idea that there is a true pattern/difference/relationship in the population.
    - The decision to use a two-sided (≠) or one-sided (< or >) should be based on any background knowledge (or lack thereof) about the context. It will have consequences for how your p-value is calculated.
* **Step 3: Identify test statistic and null distribution.**
  + The test statistic is some function of the sample whose distribution we know.
    - This is different from the point estimate, which is used in confidence intervals to estimate the parameter of interest.
    - The test statistic measures how “extreme” your data sample is as a whole – we choose it because we know its distribution.
    - Many of them happen to take the “standardized” form .
  + A test statistic has an associated *sampling distribution* (which might be based on the unknown parameter of interest), which becomes the known *null distribution* when you plug in the statement.
* **Step 4: Identify relevant outcomes from data and alternative hypothesis.**
  + Calculate the value of your test statistic from your data, call this value T.
  + The set of relevant outcomes then depends on T and your alternative hypothesis:
    - If has ≠, the set of relevant outcomes is *all outcomes with probability less than or equal to T on the null distribution*. These outcomes are the two outer tails of the distribution, which are symmetric for all normal distributions, all t distributions, and binomial with p = 0.5.
    - If has > the set of relevant outcomes is *all outcomes greater than T*, no matter where T is.
    - If has < the set of relevant outcomes is *all outcomes less than T*, no matter where T is.
* **Step 5: Calculate p-value.**
  + The area/probability of the outcomes of the outcomes identified in step 4 is the p-value.
  + Keep the visual in your head of where T is on the null distribution and what outcomes are relevant; recall that p<dist> (such as pnorm, pt) gives you area to the left of including the given point.
* **Step 6: Interpret in context**
  + If the p-value from step 5 is large, i.e. greater than 0.05, we conclude “We fail to find strong evidence against < in context>. (p = <#>, <type of test>)”.
    - For example: “We fail to find strong evidence the psychic is not picking randomly (p = 0.25, single proportion binomial test)”.
  + If the p-value from step 5 is small, i.e. less than 0.05, we conclude “We find strong evidence for < in context>. (p = <#>, <type of test>)”.
    - For example: “We find strong evidence that the chimpanzees pick the prosocial choice more than half the time in the long run (p = 0.001, single proportion binomial test).”
  + We NEVER say that or is true or false, we only evaluate the strength of the evidence we have.

**Confidence Intervals**

**Point Estimate ± Quantile Confidence Score \* Standard Error**

* **Point Estimate:** The single, sample-based estimate of the parameter of interest. Often intuitive, such as the sample mean predicting the true population mean.
* **Quantile Confidence Score:** The quantile of the sampling distribution, where is your confidence level, such as 0.95.
* **Standard Error:** A known formula capturing the variability of the point estimate. Often inversely related with sample size (n), meaning more data results in smaller error, resulting in a narrower interval.

For example, for a 95% confidence interval for a single proportion, where X = 80 and n = 120,

Point estimate = .

Quantile confidence score = the 0.95 + (1-0.95)/2 = 0.975 quantile of the sampling distribution. The sampling distribution is N(0,1), so we find this with qnorm(0.975) = approximately 1.96.

Standard error = = .

Therefore, the final confidence interval is: = [0.578, 0.745]