Section Summary | Introduction to Statistics
Confidence Intervals

What Are Interval Estimates?

1) Interval Estimates
   a) Previously, we approximated population parameters with point estimates (like using $\overline{x}$ and s to estimate μ and σ)
   b) We can instead use intervals to estimate our population parameters
   c) Intervals are more likely to contain the exact population parameter
2) Confidence Intervals
   a) A confidence interval provides a range of values that may contain the exact population parameter
   b) Confidence intervals are usually labeled by the probability of the exact population parameter falling within your interval
      i) A 95% confidence interval has a 95% chance of containing the exact μ
      ii) A 50% confidence interval has a 50% chance of containing the exact μ
      iii) Stated another way, if you have an x% confidence interval, and you construct 100 intervals from 100 different random samples, x of those intervals will contain the exact value of μ
   c) α is the complement to the confidence probability, giving the probability that an interval does not contain the exact population parameter
      i) Unlike the confidence interval, α is usually given as a decimal, not a percent
      ii) For example, a 95% confidence interval has α = 0.05
      iii) A 50% confidence interval has α = 0.5
      iv) Given a percent confidence, x, $\alpha = 1 - \frac{x}{100\%}$
      v) Given α, the percent confidence is given as $x = (1 - \alpha) \times 100\%$
3) Useless Confidence Intervals
   a) We want confidence intervals that are small enough to narrow down the search for μ, but wide enough to likely contain μ
   b) A 100% confidence interval is useless because it is usually too large
      i) If I make an interval that contains all real numbers, the confidence interval will definitely be 100%

    ii) An interval that contains all real numbers does not give us any information
  c) A 0% confidence interval is useless because it is too narrow
  d) A point estimate is an example of a 0% confidence interval
  e) Though it gives us an exact estimate for μ, it has a 0% chance of being the exact value of μ

4) Calculating Confidence Intervals
  a) Confidence interval = point estimate $\pm$MOE
  b) MOE = margin of error = critical value $\times$SE
  c) SE = standard error = $\frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$

Interval Estimates Using the Normal Distribution

1) When Can We Use a Normal Distribution to Estimate Confidence Intervals?
   a) According to the Central Limit Theorem: $\overline{X} \sim N(\mu_X, \frac{\sigma}{\sqrt{n}})$ for large n
   b) If n isn't that large but the underlying data is itself normally distributed, the sample means will still be normally distributed
   c) Other assumptions:
      i) Data are independent and identically distributed (iid)
      ii) σ^2 is known
2) Constructing the Confidence Interval
   a) $\overline{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
   b) $Z_{\frac{\alpha}{2}}$ is the critical value, which tells us how many standard errors $(\frac{\sigma}{\sqrt{n}})$ to go above and below the mean to construct the confidence interval
   c) As the percent confidence increases, $Z_{\frac{\alpha}{2}}$ increases
   d) More specifically, $Z_{\frac{\alpha}{2}}$ is the Z-score such that the top $\frac{\alpha}{2}$ of our data is excluded and $-Z_{\frac{\alpha}{2}}$ is the Z-score such that the bottom $\frac{\alpha}{2}$ of our data is excluded
      i) For a 95% confidence interval, $\frac{\alpha}{2}$= 0.025, and $Z_{\frac{\alpha}{2}} = 1.96$
      ii) 2.5% of the data is above 1.96 and 2.5% of the data is below -1.96 on the standard normal distribution
      iii) In other words, 95% of the data is contained within $Z_{\frac{\alpha}{2}} = 1.96$ and $-Z_{\frac{\alpha}{2}} = -1.96$ of the standard normal distribution
3) To find $Z_{\frac{\alpha}{2}}$:
   a) You are looking for the $Z_{\frac{\alpha}{2}}$of an x% confidence interval
   b) Get your standard normal distribution look-up table
   c) Locate the x% + $\frac{\alpha}{2} \times 100\%$ percentile (i.e. for a 95% confidence interval, look up the 95% + 2.5% = 97.5th percentile)
   d) The Z-score for this percentile is your $Z_{\frac{\alpha}{2}}$

4) Now with $Z_{\frac{\alpha}{2}}$, σ, n and $\overline{x}$, you can construct your confidence interval

Outlier

Section Summary | Introduction to Statistics
Confidence Intervals

Interval Estimates Using the "T" Distribution

1) Using a "T" distribution is useful when you have a small sample size and therefore a weak estimate of σ
2) T distributions are similar to normal distributions
   a) Unimodal
   b) Symmetric
   c) Bell-shaped
3) A T distribution is defined by the number of degrees of freedom
   a) For our purposes, the number of degrees of freedom is given by n - 1, where n is the sample size
   b) A T distribution has heavier tails, i.e. there is more density away from the mean and thus more outliers
   c) The less degrees of freedom, the heavier the tails
   d) As the number of degrees of freedom increases to infinity, the T distribution approaches the normal distribution
4) Assumptions necessary to use a T distribution
   a) Data are independent and identically distributed (iid)
   b) The underlying data are normally distributed (this is a necessary assumption because T distributions are usually used when the sample size is small and thus the Central Limit Theorem does not apply)
5) Confidence intervals from a T distribution
   a) $\overline{x} \pm t_{df,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$
   b) s is the sample standard deviation and n is the sample size
   c) $t_{df,\frac{\alpha}{2}}$ is the critical value for a confidence with a given value of $\frac{\alpha}{2}$ found from a T distribution with df number of degrees of freedom
   d) For example, $t_{5,\frac{0.05}{2}}$ is the critical value for a 95% confidence interval ($\alpha = 0.05$) from a T distribution with 5 degrees of freedom (sample size n = 6)
   e) Given your confidence interval, calculate α
   f) Like in the last section, use $\frac{\alpha}{2}$ to find the appropriate percentile where your critical value is located

g) Look up the Z-score/critical value of this percentile in a table for the T distribution with the appropriate number of degrees of freedom

h) For example, if I want to find $t_{5,\frac{0.05}{2}}$, I find a look-up table corresponding to a T distribution with 5 degrees of freedom and locate the value corresponding to the 97.5th percentile

i) As your sample size decreases, $t_{df,\frac{\alpha}{2}}$ increases, creating wider confidence intervals

j) This is because as your sample size decreases, you will find more variance in your sample and thus need to create larger intervals to have the same level of confidence

Interval Estimates for Population Proportions

1) This section looks at predicting the fraction (or proportion) of successes within a sample
2) Proportion intervals
   a) $\hat{p} \pm Z_{\frac{\alpha}{2}}$
   b) $\hat{p} = \frac{number\ of\ successes}{sample\ size}$, i.e. the proportion of trials that were a success
   c) $Z_{\frac{\alpha}{2}}$ is the margin of error, which is the critical value times the standard error of the sample
3) Standard error of proportions$\approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
4) Assumptions for this interval
   a) Data are independent and identically distributed (iid)
   b) n is large (at least 30)
   c) Because n is large, we can use the standard normal distribution to calculate our critical value
   d) $\hat{p}$ is not near 0 or 1
5) Once you calculate the standard error from $\hat{p}$ and n, calculate the critical value from the standard normal distribution given $\frac{\alpha}{2}$ and then calculate the margin of error to set up your interval