Useful Formulas for POLI 30

Taylor N. Carlson

Mean

\[ \bar{X} = \frac{\sum X_i}{n} \]

- \( \bar{X} \) = mean (sometimes denoted by \( \mu \), usually in reference to the population mean)
- \( X_i \) = the value for each observation \( i \)
- \( n \) = the number of observations

For which type of data would you use the mean as a measure of central tendency?

Standard Deviation

\[ s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}} \]

- \( s \) = standard deviation (sometimes denoted by \( \sigma \), usually in reference to the population standard deviation)
- \( \bar{X} \) = mean
- \( X_i \) = the value for each observation \( i \)
- \( n \) = the number of observations

For which type of data would you use the standard deviation as a measure of dispersion?

Standard deviation is a measure of spread (variability) of the values on a given variable. Note that we use \( n - 1 \) for the sample standard deviation. Note also that the standard deviation is the square root of the variance.

Variance

\[ s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} \]

- \( s^2 \) = variance
- \( \bar{X} \) = mean
- \( X_i \) = the value for each observation \( i \)
- \( n \) = the number of observations

For which type of data would you use the variance as a measure of dispersion? Note that variance is the standard deviation squared.
Margin of Error

Margin of Error for a Mean

\[ 2 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \]

- 2 (because we’re using a 95% level; it would be 3 if we were using a 99% level)
- \( \hat{\sigma} \) = standard deviation (sometimes denoted as \( s \))
- \( n \) = sample size

Margin of Error for a Proportion

\[ 2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \]

- \( \hat{p} \) = the proportion of interest (for example, the proportion of participants who reported voting for Candidate X)
- \( n \) = sample size

95% Confidence Intervals

When we calculate a confidence interval, we’re essentially going to take either the mean or the proportion of interest in our sample plus or minus the margin of error.

95% Confidence Interval for a Mean

\( \bar{X} \pm 2 \cdot \frac{\hat{\sigma}}{\sqrt{n}} \)

- \( \bar{X} \) = mean
- \( \pm \) plus or minus
- 2 (because we’re using a 95% confidence interval; it would be 3 if we were using a 99% confidence interval)
- \( \hat{\sigma} \) = standard deviation (sometimes denoted as \( s \))
- \( n \) = sample size

For what kind of data would you calculate a confidence interval for a mean?

95% Confidence Interval for a Proportion

\( \hat{p} \pm 2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)

- \( \hat{p} \) = the proportion of interest (for example, the proportion of participants who reported voting for Candidate X)
- \( \pm \) plus or minus
- \( n \) = sample size

For what kind of data would you calculate a confidence interval for a proportion?
Standard Error

Standard Error of a Mean
\[ \frac{\sqrt{n}}{s} \]
- \( s \) = standard deviation
- \( n \) = number of observations

How would you write this in terms of the variance instead of the standard deviation?

Standard Error of a Proportion
\[ \frac{\sqrt{(\hat{p}(1-\hat{p}))}}{\sqrt{n}} \]

95% Confidence Interval for the Difference of Proportions
\[ (\hat{p}_2 - \hat{p}_1) \pm 2 \times \sqrt{(\text{standard error}_2)^2 + (\text{standard error}_1)^2} \]
- \( \hat{p}_1 \) = proportion of interest from group 1 (e.g., proportion of Republicans who voted for Candidate X; proportion of women who voted for Candidate X)
- \( \hat{p}_2 \) = proportion of interest from group 2 (e.g., proportion of Democrats who voted for Candidate X; proportion of men who voted for Candidate X)
- \( \text{standard error}_1 \) = standard error for group 1
- \( \text{standard error}_2 \) = standard error for group 2
- \( \pm 2 \) = plus or minus 2 because we are using a 95% confidence interval. We’d use \( \pm 3 \) instead if it was a 99% confidence interval.

For what kind of data would you use a difference of proportions test? How would you write a null hypothesis for a difference of proportions test?

95% Confidence Interval for the Difference of Means
\[ (\bar{X}_2 - \bar{X}_1) \pm 2 \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \]
- \( \bar{X}_1 \) = mean from group 1 (e.g., mean contributions from Republicans)
- \( \bar{X}_2 \) = mean from group 2 (e.g., mean contributions from Democrats)
- \( s_1^2 \) = variance from group 1
- \( s_2^2 \) = variance from group 2
- \( n_1 \) = number of observations in group 1
- \( n_2 \) = number of observations in group 2
- \( \pm 2 \) = plus or minus 2 because we’re using a 95% confidence interval – if we wanted a 99% confidence interval, we’d use \( \pm 3 \)

For what kind of data would you use a difference of means test? How would you write a null hypothesis for a difference of means test?