Useful Formulas for POLI 30

Mean

$\bar{X} = \frac{\Sigma X_i}{n}$

- $\bar{X}$ = mean (sometimes denoted by μ, 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

$s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{n-1}}$

- s = standard deviation (sometimes denoted by σ, 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

$s^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1}$

- 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

$2 * \frac{\hat{\sigma}}{\sqrt{n}}$

- 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

$2 * \sqrt{\frac{(\hat{p})(1-\hat{p})}{n}}$

- $\hat{p}$ = 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

$\bar{X} \pm 2 * \frac{\hat{\sigma}}{\sqrt{n}}$

- $\bar{X}$ = 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

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

- $\hat{p}$ = 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

$\frac{s}{\sqrt{n}}$

- 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

$\frac{\sqrt{(\hat{p})(1-\hat{p})}}{\sqrt{n}}$

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

95% Confidence Interval for the Difference of Proportions

$(\hat{p}_2 - \hat{p}_1) \pm 2 * \sqrt{(standarderror_2)^2 + (standarderror_1)^2}$

- $\hat{p}_1$ = 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)
- $standarderror_1$ = standard error for group 1
- $standarderror_2$ = standard error for group 2
- ± 2 = plus or minus 2 because we are using a 95% confidence interval. We'd use ± 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

$(\bar{X}_2 - \bar{X}_1) \pm 2 * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

- $\bar{X}_1$ = 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
- ± 2 = plus or minus 2 because we're using a 95% confidence interval – if we wanted a 99% confidence interval, we'd use ± 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?

Chi Square Statistic

$\chi^2 = \Sigma_{k=1}^{k} \frac{(f_o - f_e)^2}{f_e}$

- K = the number of table cells
- f_o = frequencies observed in each category
- f_e = frequencies expected in each category if the variables were statistically independent

Expected Count

$\frac{RowTotal * ColumnTotal}{N}$

- Row total = The count of observations in a row
- Column total = The count of observations in a column
- N = The total number of observations in the table

Degrees of Freedom

df=(Rows-1)*(Columns-1)

- df = Degrees of freedom. Note that this tells us what row of the-chi square table to use for our significance test
- Rows = number of rows
- Columns = number of columns

For what kind of data would you calculate a chi-square statistic? How would you write a null hypothesis for a chi-square test? What would you compare your chi-square statistic to confirm statistic?

Useful Interpretations for POLI30

We calculate statistics and conduct hypothesis tests to better understand the phenomena and relationships that we care about. This means, that we also care about how we interpret the results.

Margin of Error for a Mean

Assuming a 95% confidence interval:

- You can conclude with 95% confidence that the true population mean is within ... of the sample mean.

Margin of Error for a Proportion

Assuming a 95% confidence interval:

- You can conclude with 95% confidence that the true population proportion is within ... of the sample proportion.

Confidence Interval for a Mean

Assuming a 95% confidence interval:

- You can be 95% confident that the true population mean is between ... and ...

Confidence Interval for a Proportion

Assuming a 95% confidence interval:

- You can be 95% confident that the true population proportion is between ... and ...

Confidence Interval for Difference of Means

Assuming a 95% confidence interval:

- You can be 95% confident that the true population value for the difference in means between groups is between ... and ...

Confidence Interval for Difference of Proportions

Assuming a 95% confidence interval:

- You can be 95% confident that the true population value for the difference in proportions between groups is between ... and ...

Hypothesis Test for Difference of Means

Where the null hypothesis is that there is no difference in the mean of the DV between two groups (in symbols, that $\bar{X}_1 - \bar{X}_2 = 0$)

- If the confidence interval does not cross zero
  - You reject the null hypothesis.
  - There is a significant difference between the group means.
  - You can be 95% confident that the true difference in means for the groups is different than 0.
- If the confidence interval crosses zero
  - You fail to reject the null hypothesis.
  - There is no significant difference between the group means.
  - You cannot be certain that the true value of the difference in means between groups is different than 0.

Hypothesis Test for Difference of Proportions

Where the null hypothesis is that there is no difference in the proportion of the DV between groups (in symbols, $\hat{p_1} - \hat{p_2} = 0$)

- If the confidence interval does not cross zero
  - You reject the null hypothesis.
  - There is a significant difference between the group proportions.
  - You can be 95% confident that the true difference in proportions between groups is different than 0.
- If the confidence interval crosses zero.
  - The fail to reject the null hypothesis.
  - There is no significant difference between the group proportions.
  - You cannot be certain that the true value of the difference in proportions between groups is different than 0.

Hypothesis Test for Chi-Squared Statistic

Where the null hypothesis is that there is no association between the IV and DV

- If $\chi^2 >$ threshold value
  - You reject the null hypothesis
  - The relationship is statistically significant
  - The IV has a significantly relationship to the DV.
- if $\chi^2 <$ threshold value
  - You fail to reject the null hypothesis.
  - The relationship is insignificant.
  - There is no significant relationship between the IV and DV.

Bivariate Regression Coefficients

Intercept

- When the IV=0, the DV= ...

Coefficient of a continuous variable

- A one unit increase in the value of the IV, leads to a ... unit increase/decrease in the DV
- If the coefficient is positive, it will be an increase
- If the coefficient is negative, it will be a decrease

Coefficient of a nominal variable

- Having your IV characteristic increases/decreases the DV by ..., compared to the absence of the characteristic

Multivariate Regression Coefficients

Intercept

- When *all the IVs* are=0, the DV= ...

Coefficient of a continuous variable

- *Holding all other variables constant*, a one unit increase in the value of the IV, leads to a ... unit increase/decrease in the DV
- If the coefficient is positive, it will be an increase
- If the coefficient is negative, it will be a decrease

Coefficient of a nominal variable

- *Holding all other variables constant*, having your IV characteristic increases/decreases the DV by ..., compared to the absence of the characteristic

R^2

- ....% of the variation in the DV is explained by the IV(s)
- Explains the variation in the DV that is explained by the whole model

Hypothesis Test for Regression Coefficient

Where the null hypothesis is that the regression coefficient for an IV = 0 (in symbols, that b=0)

- If the confidence interval does not cross zero (or the t-value is > 2 or < -2)...
  - You reject the null hypothesis.

  - The relationship is statistically significant
  - The IV has a significantly positive/negative impact on the DV

- If the confidence interval crosses zero (or the t-value is < 2 and > -2)
  - You fail to reject the null hypothesis.
  - The relationship is insignificant.
  - There is no significant relationship between the IV and DV.