0.0001248 0.01117139

Lesson 16: Describing Categorical Data: Proportions; Sampling Dist. Of a Sample Proportion

Graphing categorical data:

Pie chart - Used when you want to represent observations as part of a whole.

- Bar chart - Typically used when data represents counts
- Pareto chart - A bar chart where bars are in desending order $\hat{p}=\frac{x}{n}$

Population ~~mean~~ proportion: p, Sample proportion: $\hat{p}$ → $\hat{p}$ is a point estimator for true proportion p.

We can apply the CLT to a sample proportion if:

- $np \geq 10$
- $n(1-p) \geq 10$

Mean of $\hat{p}$ = mean of p

Standard deviation of $\hat{p} = \sqrt{\frac{p\cdot(1-p)}{n}}$

If our sample size is large we can use the Normal Probability Applet to make probability calculations for proportions.

1. Find the z-score: $z=\frac{\hat{p}-p}{\sqrt{\frac{p\cdot(1-p)}{n}}}$ → Enter z-score into the Applet and shade accordingly

Lesson 17: Inference for One Proportion

The formula for the ~~confinec~~ confidence interval for 1 proportion is: $\left(\hat{p}-z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, + z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$ where $\hat{p}=\frac{x}{n}$

- Use the normal probability applet to compute z^* by shading the middle only and putting the confidence interval into the upper box. (or use the toolbox)
- Formula for margin of error is $m=z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- To solve the margin of error equation for n, we get: $n=\left(\frac{z^*}{m}\right)^2 p^*(1-p^*)$
- If no prior estimate for p is available, use the formula $n=\left(\frac{z^*}{2m}\right)^2$

☆ Important: The requirements for a confidence interval are $n\hat{p}\geq 10$ and $n(1-\hat{p})\geq 10$. The requirements for a hypothesis test involving 1 proportion are $np \geq 10$ and $n(1-p)\geq 10$.

$n=\left(\frac{1.96}{0.02}\right)^2(0.15)(1-0.15)$

4268.4 9,604 .1275

1225

$0.02=1.96\sqrt{\frac{0.1875}{0.0009375}}$

$n=\left(\frac{1.96}{0.02}\right)0.15(1-0.15)$ $0.02=1.96\sqrt{\frac{0.25(1-0.25)}{200}}$

$n=(98)(0.1275)$

$n=\left(\frac{1.96}{0.02}\right)^2 0.2(1-0.2)$ ~~0.02=1.96~~$\sqrt{\ }$

$n=9604\ (0.16)$ $n=$

$0.02=1.96(0.03061862)$

$0.02=0.060012$

$0.02=$

$0.02=1.96\sqrt{\frac{0.1875}{n}}$

0.09044586

Direction

Lesson 18: Inference for Two Proportions

- When conducting a hypothesis test for two proportions, $\hat{p}_1=\frac{x_1}{n_1}$ and $\hat{p}_2=\frac{x_2}{n_2}$
- The null and alternative hypotheses are $H_0: p_1=p_2$ and $H_a: p_1 >, <, \neq p_2$
- $Z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ → (Use toolbox, haha)
- n_1 = Sample size for group 1
- n_2 = Sample size for group 2
- $\hat{p}_1$ = Sample proportion for group 1 → $\hat{p}_1=\frac{x_1}{n_1}$
- $\hat{p}_2$ = Sample proportion for group 2 → $\hat{p}_2=\frac{x_2}{n_2}$
- $\hat{p}$ = Overall sample proportion

→ overall sample proportion: $\hat{p}=\frac{x_1+x_2}{n_1+n_2}$

- Hypothesis Test Requirements:
  - $n_1\hat{p}_1\geq 10$
  - $n_2\hat{p}_2\geq 10$
  - $n_1(1-\hat{p}_1)\geq 10$
  - $n_2(1-\hat{p}_2)\geq 10$
- The equation of the confidence interval for the difference between 2 proportions is: $\left((\hat{p}_1-\hat{p}_2)\pm z^*\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$ → Use the toolbox
- Confidence Interval Requirements: Same as Hypothesis Test

☆ Whenever zero is contained in the confidence interval of the difference of the true proportions, we conclude that there is no significant difference between the proportions.

Lesson 19: Inference for Independence of Categorical Data (Chi-Squared)

χ^2 → Test statistic that compares the observed counts to expected counts (The counts you should expect to get if the null hypothesis is true.)

- Different χ^2 distributions are distinguished by the number of degrees of freedom, which are determined by the number of rows and columns in the table.
- df = (number of rows − 1)(number of columns − 1)

Ho: The [row variable] and the [column variable] are independent

Ha: The [row variable] and the [column variable] are not independent

- Requirements: 1. You must use a SRS to obtain a sample from a single population.

2. Each expected count must be greater than or equal to 5

☆ No Confidence Intervals