Probability and Confidence Intervals; Hypothesis Tests

Statistics 800
Fall 2000
Dr. Nancy Pfenning

Conditions for Proportions

1. (a) A population has a fixed proportion falling into the category of interest; **or** (b) A repeatable situation has a long-run proportion of occurrences of a certain outcome; **and**

2. (a) A random sample of a given size is taken from the population; **or** (b) The situation is repeated independently a given number of times; **and**

3. (a) Sample size **or** (b) number of repetitions is large enough that there are at least 5 occurring both in and out of the category of interest.

Rules for Sample Proportions

If numerous samples or repetitions of the same size are taken,

1. (center) The mean of all sample proportions will be true proportion for the population.

2. (spread) Standard deviation of proportions = standard error =

$$\sqrt{\frac{\text{population proportion}(1 - \text{population proportion})}{\text{sample size}}}$$

3. (shape) The frequency curve made from proportions from the various samples will be approximately bell-shaped/normal.

Probability Intervals based on Empirical Rule

1. Probability is 68% that sample proportion falls within 1 standard error of population proportion.

2. Probability is 95% that sample proportion falls within 2 standard errors of population proportion.

3. Probability is 99.7% that sample proportion falls within 3 standard errors of population proportion.

Confidence Interval for Proportion

A 95% confidence interval for unknown population proportion is sample proportion plus or minus 2 standard errors, which is approximately

$$\text{sample proportion} \pm 2\sqrt{\frac{\text{sample proportion}(1 - \text{sample proportion})}{\text{sample size}}}$$

Conditions for Means

1. Samples must be random; **and**
2. Either (a) the population of measurements is bell-shaped; then any sample size is OK; **or** (b) if the population is not bell-shaped, then the sample size must be large enough

Rules for Sample Means

If numerous samples or repetitions of the same size are taken from a population of values for a measurement variable, and sample means are found,

1. (center) The mean of all sample means will equal population mean
2. (spread) Standard deviation of sample means = standard error =

$$\frac{\text{population standard deviation}}{\sqrt{\text{sample size}}}$$

3. (shape) The frequency curve made from means from the various samples will be approximately bell-shaped/normal.

Probability Intervals based on Empirical Rule

1. Probability is 68% that sample mean falls within 1 standard error of population mean.
2. Probability is 95% that sample mean falls within 2 standard errors of population mean.
3. Probability is 99.7% that sample mean falls within 3 standard errors of population mean.

Confidence Interval for Mean

A 95% confidence interval for unknown population mean is sample mean plus or minus 2 standard errors, which is approximately

$$\text{sample mean} \pm 2\frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$$

Confidence Interval for Difference in Means

1. Collect large samples of observations independently for two groups. Compute sample mean and standard deviation for each.
2. Compute each standard error of the mean $SEM = \dfrac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$
3. Compute SE of the difference $= \sqrt{(SEM_1)^2 + (SEM_2)^2}$
4. A 95% confidence interval for difference in two population means is

$$\text{difference in sample means} \pm 2\sqrt{(SEM_1)^2 + (SEM_2)^2}$$

Steps for Testing Hypotheses about Proportions

1. Determine null and alternative hypotheses:

   null hypothesis: population proportion = ____________

   alternative hypothesis: population proportion $\left\{ \begin{array}{c} > \\ < \\ \neq \end{array} \right\}$ ____________

2. Collect and summarize data, including sample size, sample proportion,

$$\text{standard error} = \sqrt{\frac{\text{population proportion}(1 - \text{population proportion})}{\text{sample size}}}$$

$$\text{test statistic} \quad z = \frac{\text{sample proportion} - \text{population proportion}}{\text{standard error}}$$

   [Where population proportion is as stated in the null hypothesis.]

3. Determine p-value = unlikelihood of the test statistic, assuming the null hypothesis is true
4. Make a decision: If the p-value is “small”, reject the null hypothesis and conclude the alternative is true. (Results are **statistically significant**.) If the p-value is not so small, we cannot reject the null hypothesis.

Steps for Testing Hypotheses about Means

1. Determine null and alternative hypotheses:
   null hypothesis: population mean = ____________

   alternative hypothesis: population mean $\left\{ \begin{array}{c} > \\ < \\ \neq \end{array} \right\}$ ____________

2. Collect and summarize data, including the test statistic

$$z = \frac{\text{sample mean} - \text{population mean}}{\text{standard error}} = \frac{\text{sample mean} - \text{population mean}}{\frac{\text{standard deviation}}{\sqrt{\text{sample size}}}}$$

[Where population mean is as stated in the null hypothesis.]

3. Determine p-value = unlikelihood of the test statistic, assuming the null hypothesis is true

4. Make a decision: If the p-value is "small", reject the null hypothesis and conclude the alternative is true. (Results are **statistically significant**.) If the p-value is not so small, we cannot reject the null hypothesis.