See this page in the course material.

Let's Summarize

Overview of Statistical Inference

- Inference is based on probability.
- A parameter is a number that describes a population. A statistic is a number that describes a sample. In inference, we use a statistic to draw a conclusion about a parameter. These conclusions include a probability statement that describes the strength of the evidence or our certainty.
- For a categorical variable, the parameter and statistics are proportions. For a quantitative variable, the parameter and statistics are means.
- For a given situation, we assume the parameter is fixed. It does not change. In contrast, statistics always vary. When we take random samples, the fluctuation in statistics is due to chance. We create simulations and mathematical models to describe the variability we expect to see in sample statistics.

Sampling Distribution for a Sample Proportion

- Larger samples have less variability.
- For a categorical variable we assume that population has a proportion *p* of successes. When we select random samples from this population, the sample proportions have a pattern in the long run. We can describe this pattern with a mathematical model of the sampling distribution. The model has the following center, spread, and shape.

Center: Mean of the sample proportions is *p*, the population proportion.

Spread: Standard deviation of the sample proportions is [latex]\sqrt{\frac{p(1-p)}{n}}[/latex]

Shape: A normal model is a good fit if the expected number of successes and failures is at least 10. We can translate these conditions into formulas:
[latex]np≥10\text{}\mathrm{and}\text{}n(1-p)≥10[/latex].

- When a normal is a good fit for the sampling distribution, we can calculate a *z*-score, which allows us to use the standard normal model to find probabilities associated with the sampling distribution.

[latex]\begin{array}{l}\mathrm{standard}\text{}\mathrm{error}=\sqrt{\frac{p(1-p)}{n}}\\
Z=\frac{\mathrm{statistic}-\mathrm{parameter}}{\mathrm{standard}\text{}\mathrm{error}}=\frac{\st

ackrel{^}{p}-p}{\mathrm{standard}\text{}\mathrm{error}}\end{array}[/latex]

We can also write this as one formula:

[latex]Z=\frac{\stackrel{^}{p}-p}{\sqrt{\frac{p(1-p)}{n}}}[/latex]

Introduction to Statistical Inference

This course presents two types of inference procedures: confidence intervals and hypothesis tests. The goal of a confidence interval is to estimate a parameter value. The goal of a hypothesis test is to test a claim about a parameter. Both types of inference are based on the sampling distribution of sample statistics. For both, we report probabilities that state what would happen if we used the inference method many times.

Confidence Intervals

The purpose of a confidence interval is to estimate a population parameter on the basis of a sample statistic. Sample statistics vary, so there is always error in our estimate, but we never know how much. We therefore use the standard error, which is the average error in our sample estimates, to create a margin of error. The margin of error is related to our confidence that the interval contains the population parameter.

We investigated the 95% confidence interval for a population proportion in depth. When a normal model is a good fit for the sampling distribution, the 95% confidence interval has a margin of error equal to 2 standard errors.

[latex]\begin{array}{l}\mathrm{sample}\text{}\mathrm{statistic}\text{}±\text{}\mathrm{margin}\text{}\mathrm{of}\text{}\mathrm{error}\\ \mathrm{sample}\text{}\mathrm{proportion}\text{}±\text{}2(\mathrm{standard}\text{}\mathrm{errors})\\ \stackrel{^}{p}\text{}±\text{}2\sqrt{\frac{p(1-p)}{n}}\end{array}[/latex]

We say we are 95% confident that the calculated interval contains the population proportion. This means that 95% of the time, these intervals will actually contain the population proportion, and we will be right. Five percent of the time, we will be wrong. We can never tell if a confidence interval does or does not contain the population proportion we are trying to estimate.

Hypothesis Tests

The purpose of a hypothesis test is to use sample data to test a claim about a population parameter. We investigated testing a claim about a population proportion informally.

We make a claim about a population proportion. From the claim, we state an assumption about the value of the population proportion. Could the data have come from this population? Or is the sample proportion too far off? It depends on how much random samples from this population vary. We construct a simulation or a normal model to represent the sampling distribution that occurs when sampling from a population with this assumed value. We make a judgment about whether the sample proportion is likely or unlikely to occur in the sampling distribution. If the data supports our claim and is unlikely, then we doubt our assumption about the population proportion.

Likely or unlikely? It depends on how much the sample proportions vary. If the normal model is a good fit for the sampling distribution, we can find a *z*-score and use a simulation to associate a probability with our “likely” or “unlikely” statement.

Licenses and Attributions

CC licensed content, Shared previously

- Concepts in Statistics. **Provided by**: Open Learning Initiative. **Located at**: http://oli.cmu.edu. **License**: *CC BY: Attribution*

</div