Towards Inference

ST551 Lecture 7

Charlotte Wickham

2017-10-04

Review so far

Where have we been and where are we going?

(Charlotte's sketch)

Where have we been and where are we going?

We know a lot about the situation when our statistic is the sample mean:

- If population is exactly Normal, sampling distribution is exactly Normal.
- If population is anything else, sampling distribution is approximately Normal for large sample sizes.
- In both cases, mean and variance of sampling distribution relate directly to population mean and variance.

We'll walk through how this knowledge guides our inference for the situation where we are interested in the population mean.

Inference

Reminder about the types of inference

A population inference is making a statement about a population parameter based on a sample.

Types of statement:

- **Point Estimate:** the single best guess of the population parameter value
- **Interval Estimate:** a range of likely values for the population parameter
- **Hypothesis test:** is a specific value of the population parameter plausible?

Estimation

Example

Population: OSU freshman

Variable of interest: Hours spent preparing for classes

Parameter of interest: Population mean

Question of interest: What's our best guess for the mean hours spent preparing for classes OSU freshman?

Sample: Randomly sample 100 freshman from Fall 2017.

Inference: Estimation

We'll use the symbol θ to denote a general population parameter of interest.

We put a hat over a symbol to represent an **estimate** of that parameter:

$$\hat{\theta} \text{ is an estimate of } \theta$$

What's our best guess of the population mean? I.e for $\theta = \mu =$ population mean.

What makes an estimate good?

Some criteria for good estimates:

- The estimate is correct on average. A.k.a. Unbiased

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta = 0$$

- The estimate is close to true value. I.e. has a small mean squared error

$$MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right] = Bias(\hat{\theta})^2 + Var(\hat{\theta})$$

- The estimate gets closer to the true value as the sample size increases. A.k.a. Consistent: the estimate converges in probability to the true value

$$\hat{\theta} \rightarrow_p \theta$$

Your turn: Is the sample mean a good estimate of the population mean?

- Is it unbiased?

- What is its mean squared error?

- Is it consistent?

Notice, determining if an estimate is good, requires knowing **how its sampling distribution relates to the population distribution**.

Hypothesis Tests

Population: OSU freshman
Variable of interest: Hours spent preparing for classes
Parameter of interest: Population mean
Question of interest: Is mean hours spent preparing for classes OSU freshman equal to 30?
Sample: Randomly sample 100 freshman from Fall 2017.

Inference: Hypothesis Tests

Key Components

- Hypotheses
    - Null Hypothesis
    - Alternative Hypothesis
- Test Statistic
- Reference Distribution (Null Distribution)
    - Rejection Region

Hypotheses

Null Hypothesis H_0: specified value (or range of values) for the **parameter** of interest "uninteresting results"

Alternative Hypothesis H_A (or H_1): A different specified value or range of values for the parameter of interest (often NEGATION of null) "interesting results"

Two sided versus one-sided One sided $\theta > a$ Two sided $\theta \neq c$ or $\theta < a$ or $\theta > b$

Possible outcomes of a test

Reject null hypothesis Data are inconsistent with null hypothesis

Fail to reject null hypothesis Data are consistent with null hypothesis

We don't **accept the null** because the null hypothesis might not be the **only** plausible value for the parameter.

Possible outcomes

	H_0 is true	H_A is true
Reject Null		
Fail to reject Null		

Possible outcomes

	H_0 is true	H_A is true
Reject Null	Type I error	Correct decision
Fail to reject Null	Correct decision	Type II error

How do we design a hypothesis test?