MA22004

Seminar 3

Dr Eric Hall

22/10/2020

Announcements

Reminders

- It is week 3! You should have read §3.1 of the notes on **Perusall**.
- Feedback for Lab 1 is posted.

Upcoming

- Lab 2 due **TODAY** at **17:00**: upload to **Gradescope**. Late submissions will be accepted until Sat 24/10 at 13:00.
- Worksheet #3 (Estimating Proportions) on **Blackboard**: do before next workshop.
- Investigation #2 on **Perusall**: do before next workshop.
- Reading assignment remainder §3 on **Perusall**: do before next seminar.

Estimating means

Confidence intervals

$$\text{point estimate}\mu \pm (\text{critical value}) \cdot (\text{precision of point estimate})$$

Hypothesis tests

A hypothesis test asks if the available data provides sufficient evidence to reject a null hypothesis H_0 which is assume to be true.

Tests on equality of mean μ, i.e., $H_0 : \mu = \mu_0$ against

- $H_a : \mu \neq \mu_0$
- $H_a : \mu \geq \mu_0$
- $H_a : \mu \leq \mu_0$

Estimating means: consider the population

We consider inferences for three different parent populations.

1. Normal population with known σ
2. Any population with unknown σ, when m large
3. Normal population with unknown σ, m may be small

What changes? What stays the same?

Estimating means: $100(1-\alpha)\%$ CIs

1. Normal population with known σ:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{m}}$$

2. Any population with unknown σ, when m large

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{m}}$$

3. Normal population with unknown σ, m may be small

$$\bar{x} \pm t_{\alpha/2,m-1} \frac{s}{\sqrt{m}}$$

Estimating means: hypothesis tests

1. Normal population with known σ:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{m}} \sim \mathrm{N}(0, 1)$$

2. Any population with unknown σ, when m large

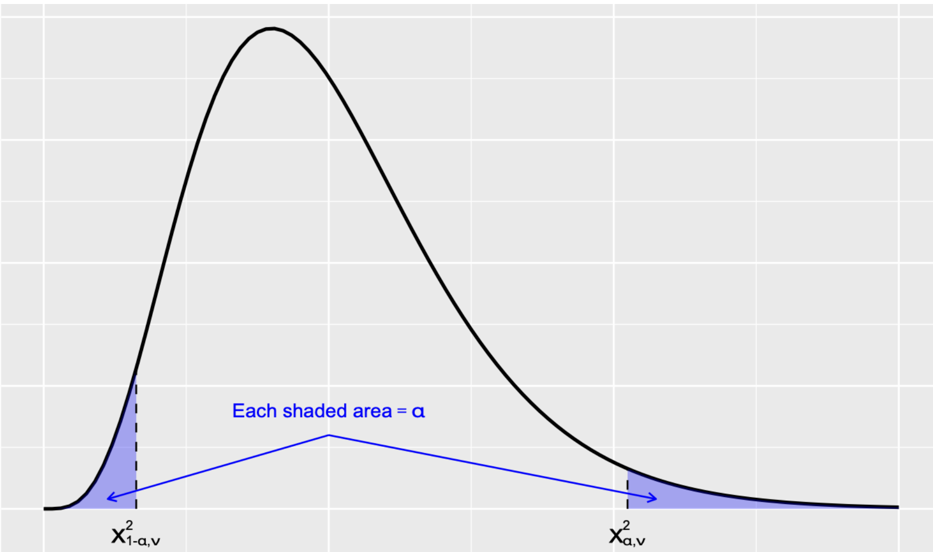
$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{m}} \sim \mathrm{N}(0, 1)$$

3. Normal population with unknown σ, m may be small

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{m}} \sim \mathrm{t}(m - 1)$$

Recall: critical values

Critical values are quantiles of the reference distribution.



- Using R: `qnorm`, `qt`, `qchisq`
- Using table

The skinny on hypothesis testing

1. Compute the value of an appropriate test statistic.
2. Determine the P-value, probability calculated assuming the null is true of observing a test statistic value at least as contradictory to H_0 as what was obtained from evidence.
3. For given α, reject H_0 if $P \leq \alpha$. If evidence is not strong enough, we fail to reject H_0.

Example (hypothesis testing 1/n)

Freshman 15 [Devore eg 8.2]

A common belief among the lay public is that body weight increases after entry into [university], and the phrase “freshman 15” has been coined to describe the 15 pounds that students presumably gain over their freshman [(i.e. first)] year.

Example (2/n)

Let μ denote the true average weight gain of women over the course of their first year at university. The foregoing quote suggests that we should test the hypotheses $H_0 : \mu = 15$ versus $H_a : \mu \neq 15$.

Suppose that a random sample of m such individuals is selected and the weight gain of each one is determined, resulting in a sample mean weight gain $\overline{x}$ and a sample standard deviation s.

Example (3/n)

Example (4/n)

Suppose $\bar{x} = 13.7$ and plugging in values for s and $\sqrt{n}$ yields $z = -2.8$.

Which values of the test statistic are at least as contradictory to H_0 as -2.80 itself?

Example (5/n)

$P(Z \leq -2.80 \quad \text{or} \quad Z \geq 2.80 \quad \text{assuming } H_0 \text{ true})$

Example (6/n)

The article *Freshman 15: Fact or Fiction* (Obesity, 2006: 1438– 1443), for $m = 137$ students, the $\bar{x} = 2.42$ lb with $s = 5.72$ lb.

This gives

$$z = (2.42 - 15)/(5.72/\sqrt{137}) = -25.7 \,.$$

The probability of observing a value at least this extreme in either direction is essentially zero!

The data very strongly contradicts the null hypothesis, and there is substantial evidence that true average weight gain is much closer to 0 than to 15.

Summary

Today we discussed CI and hypothesis tests for population means for three cases:

- normal populations with known σ
- population with unknown σ when sample size is large
- normal population with unknown σ when sample size is small