MA22004

Seminar 4

Dr Eric Hall

29/10/2020

Announcements

Reminders

- It is week 4! You should have read the remainder of §3 of the notes on **Perusall**.
- Feedback for Lab 2 is posted.

Upcoming

- Lab 3 due **Thursday 5 Nov** at **17:00**: upload to **Gradescope**.
- Worksheet #4 (Estimating Variances) on **Blackboard**: do before next workshop.
- Investigation #3 on **Perusall**: do before next workshop.
- Reading assignment #5 (start of §4) on **Perusall**: do before next seminar.

Estimating Population Proportions p

Let X be the count of members with a given property based on a sample of size m from a population where a proportion p share the property.

Then $\widehat{p} = X/m$ is an estimator of p.

Assume $mp_0 \geq 10$ and $m(1 - p_0) \geq 10$.

Eg.

$p \in (0, 1)$ and consider Bernoulli trials $X_1, \ldots, X_m \sim$ Bernoulli(p).

$\widehat{p} = \frac{1}{m} \sum_{i=1}^{m} X_i \,, \qquad \mu_{\widehat{p}} = \mathbf{E}\, X_i = p \,, \qquad \sigma_{\widehat{p}} = \sqrt{\frac{\operatorname{Var} X_i}{m}} = \sqrt{\frac{p(1-p)}{m}}$

$100(1-\alpha)\%$ CI for p

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{m}}$$

> How is the standard error different from the standard error for estimate of mean?

Hypothesis test for p at level α:

Consider $H_0 : p = p_0$. The test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/m}} \ .$$

For a hypothesis test at level α, we use the following procedure:

If $H_a : p > p_0$, then P-value is the area under $\mathrm{N}(0, 1)$ to the right of z.

If $H_a : p < p_0$, then P-value is the area under $\mathrm{N}(0, 1)$ to the left of z.

If $H_a : p \neq p_0$, then P-value is twice the area under $\mathrm{N}(0, 1)$ to the right of $|z|$.

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.

Estimating Normal Variances σ^2

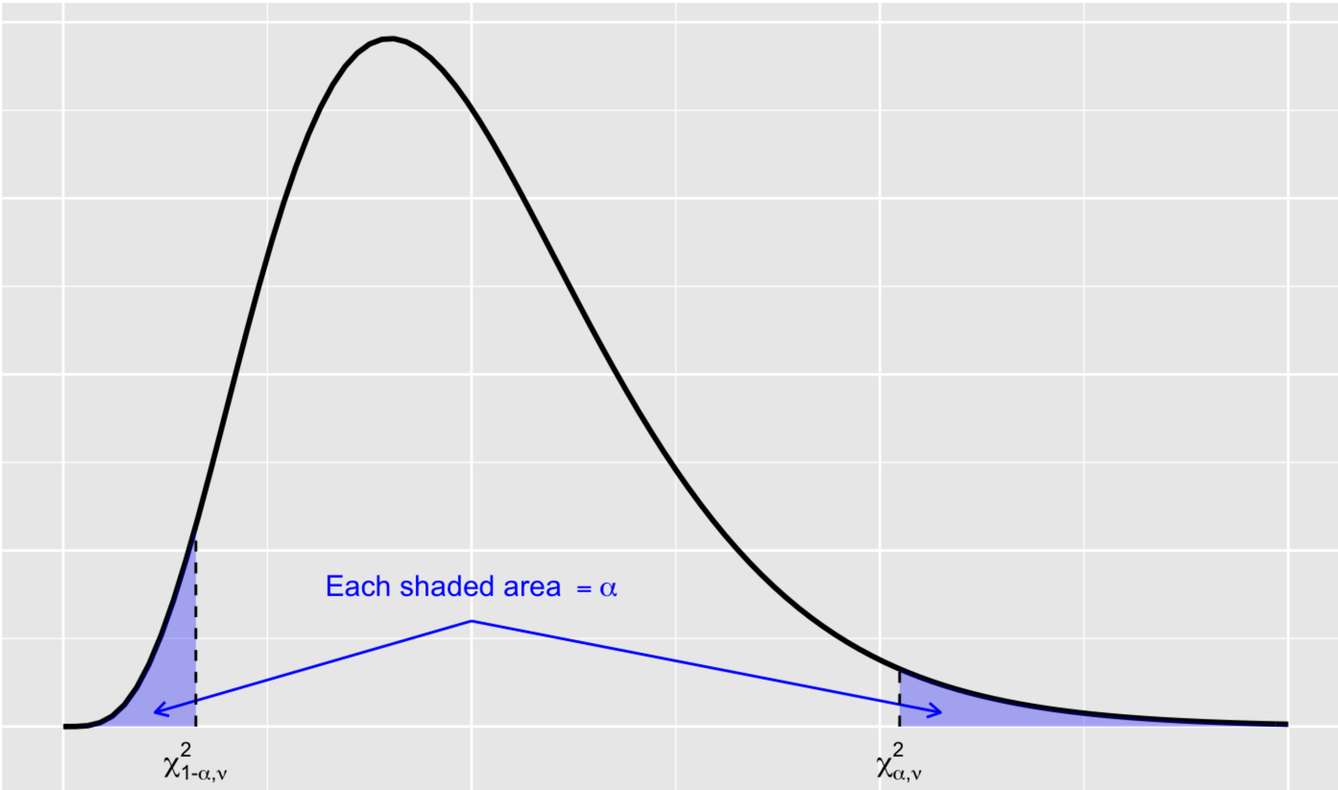
Consider iid $X_1, \dots, X_m \sim \mathsf{N}(\mu, \sigma^2)$.

$$V = \frac{(m-1)S^2}{\sigma^2} = \frac{\sum_i (X_i - \overline{X})^2}{\sigma^2} \sim \chi^2(m-1)$$

$100(1-\alpha)\%$ CI for σ^2 of normal population

$$\left((m-1)s^2/\chi^2_{\alpha/2,m-1}\, , (m-1)s^2/\chi^2_{1-\alpha/2,m-1}\right)$$

χ^2 critical values



As the χ^2 distribution is not symmetric, the upper and lower critical values will not be the same (the shaded areas are equal).

Example: Churchill 1/n

Churchill claims that he will receive half the votes for the House of Commons seat for the constituency of Dundee.

We are skeptical that he is as popular as he says. Suppose 116 out of 263 Dundonians polled claimed that they intended to vote for Churchill.

Can it be concluded at significance level 0.10 that more than half of all eligible Dundonains will vote for Churchill?

Example: Churhill 2/n

The parameter of interest is p, the proportion of votes for Churchill.

What are the null and alternative hypotheses?

Example: Churhill 3/n

The null hypothesis is $H_0 : p = 0.5$.

The alternative hypothesis is $H_a : p < 0.5$.

Since $263(0.5) = 131.5 > 10$, we satisfy the assumptions to carry out the hypothesis test using the statistic described earlier.

Example: Churhill 4/n

Based on the sample, $\widehat{p} = 116/263 = 0.4411$.

The test statistic value is

$$z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1 - p_0)/m}} = \frac{0.4411 - 0.5}{\sqrt{0.5(1 - 0.5)/263}} = -1.91\,.$$

The P-value for this lower-tailed z test is $P = \Phi(-1.91) = 0.028$.

Example: Churhill 5/n

Since $P < 0.10 = \alpha$, we reject the null hypothesis at the 0.05 level.

> What does this mean in words?

Example: Churhill 6/n (conclusion)

The evidence for concluding that the true proportion is different from $p_0 = 0.5$ at the 0.10 level is compelling.

Summary

Today we discussed CI and hypothesis tests for:

- population proportions
- population standard deviations