MATH 3339
Statistics for the Sciences
Live Lecture Help

James West
jdwest@uh.edu

University of Houston

Session 13

Office Hours: see schedule in the "Office Hours" channel on Teams
Course webpage: www.casa.uh.edu

Email policy

When you email me you **MUST** include the following

- MATH 3339 Section 20024 and a description of your issue in the **Subject Line**
- Your name and ID# in the **Body**
- Complete sentences, punctuation, and paragraph breaks
- Email messages to the class will be sent to your Exchange account (user@cougarnet.uh.edu)

Using R and RStudio

1. Download R from https://cran.r-project.org/
2. Download RStudio from https://www.rstudio.com/

Outline

Updates and Announcements

- Test 2 begins tomorrow.
  Check your time!
- Practice Test 2 (PT2) closes tonight
  I will add 5% of PT2 score to
  Test 2.
- Final Exam scheduling will open Monday
  11/22 at 12 AM.
- Lecture 23 is in Teams. It will close
  Sunday.

pdf - $f(x)$ - cts. function

cdf - $F(x) = \int_{-\infty}^{x} f(t)\,dt \Rightarrow F'(x) = f(x)$

- Density functions
- Know how to calculate expected values and quantiles for continuous distributions.

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\,dx$$

- Named distributions
  - Uniform punif(x, min, max)
  - Exponential pexp(x, λ)
  - Gamma pgamma(x, α, $\lambda = \frac{1}{\beta}$)
  - Normal pnorm(x, μ, σ)

- Expected values and variances of X + Y
- Applying the Central Limit Theorem
- The sampling distribution of the sample means, $\bar{X}$.
- The sampling distribution of the proportions, $\hat{p}$.

$\overline{X} \sim N(\mu, \sigma/\sqrt{n})$ if $n \geq 30$

$\hat{p} \sim N\left(\mu_{\hat{p}} = p, \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\right)$ if $np, n(1-p) \geq 10$

Confidence Intervals

Know how to calculate the confidence intervals for mean (μ), proportions (p) and standard deviation (σ).

- For mean if σ, population standard deviation is given: $\bar{x} \pm z^* \left(\frac{\sigma}{\sqrt{n}}\right)$.
- For mean if σ is not given: $\bar{x} \pm t^*_{df} \left(\frac{s}{\sqrt{n}}\right)$. $df = \nu = n-1$
- For proportions $\hat{p} = \frac{X}{n}$: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
- For standard deviation: $lcl = \sqrt{\frac{(n-1)s^2}{qchisq(1-\alpha/2, n-1)}}$ and $ucl = \sqrt{\frac{(n-1)s^2}{qchisq(\alpha/2, n-1)}}$
- For paired - t: $\bar{x}_D \pm t^*_{df} \left(\frac{s_D}{\sqrt{n}}\right)$
- For two-sample t for means: $(\bar{x}_1 - \bar{x}_2) \pm t^*_{\nu} \sqrt{\frac{s_1^2}{n_2} + \frac{s_s^2}{n_2}}$

Degrees of Freedom for Two-Sample T

$$df = \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$$

by hand use $df \approx \nu = \min(n_1 - 1, n_2 - 1)$

Confidence Intervals

- Know how the confidence interval changes as the confidence level C changes and as the sample size changes.
- Know how to interpret confidence intervals.
- Know how to determine a sample size given confidence level and margin of error.
  - For means $n = \left(\frac{z_{\alpha/2} \times \sigma}{E}\right)^2$, where E = margin of error.
  - For proportions $n = p^*(1 - p^*)(\frac{z_{\alpha/2}}{E})^2$, where p^* is some previous knowledge of the proportion if not known we use 0.5.

Hypothesis Tests

- Know how to set up null and alternative hypotheses.
- Be able to determine a rejection region, given the level of significance (α).
- Calculate a test statistic.
  - For one - sample

	$H_0 : \mu = \mu_0$ if σ is given	$H_0 : \mu = \mu_0$ if σ is unknown	$H_0 : p = p_0$
Test Statistic	$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

  - For two-sample

	$H_0 : \mu_1 = \mu_2$	$H_0 : p_1 = p_2$	$H_0 : \mu_D = 0$
Test Statistic	$t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	$z = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$	$t = \frac{\bar{x}_D-\mu_d}{s_D/\sqrt{n}}$

- Be able to calculate the p-value and know to reject (RH0) or fail to reject (FRH0).
- Know the difference between Type I and Type II errors.
- Be able to make a conclusion in context of the problem.

Strength of Evidence

If the P-value for testing H_0 is less than:

- 0.1 we have **some evidence** that H_0 is false.
- 0.05 we have **strong evidence** that H_0 is false.
- 0.01 we have **very strong evidence** that H_0 is false.
- 0.001 we have **extremely strong evidence** that H_0 is false.

If the P-value is greater than 0.1, we **do not have any evidence** that H_0 is false.

Review Questions

We wish to test the hypotheses $H_0 : \mu = 13.5$ versus $H_a : \mu > 13.5$ at $\alpha = 0.02$ significance level. From a random sample of 40 the sample mean is 15.9. Assume that the population standard deviation is 2.7.

1. Which test would be appropriate to use?

   a) z-test for proportions b) z-tests for means
   c) t-test for means d) t-tests for proportions

2. Calculate the test statistic.

   a) 2.4 b) -5.62 c) 5.62 d) 0.89 e) 2.0537

3. Determine the p-value, approximately.

   a) 0 b) 0.02 c) 1.434 d) 4.68

4. What is our decision of the test?

   a) Reject H_0 b) Fail to reject H_0 c) Accept H_0

Review Questions

For each of the following scenarios, determine if it is **a) paired t-test** or **b) two sample t-test**

5. The weight of 14 patients before and after open-heart surgery. A
6. The smoking rates of 14 men measured before and after a stroke. A
7. The number of cigarettes smoked per day by 14 men who have had stokes compared with the number smoked by 14 men who have not had strokes. B
8. The photosynthetic rates of 10 randomly chosen Douglas-fir trees compared with 10 randomly chosen western red cedar trees. B
9. The photosynthetic rate measured on 10 randomly chosen Sitka spruce trees compared with the rate measured on the western red cedar growing next to each of the Sitka spruce trees. A

Study Habits Review Question

A study was conducted to determine the study habits of men versus women. Scores for this study ranged from 0 to 200. Here are the results: `https://www.math.uh.edu/~cathy/Math3339/data/ssha.csv`.

a) Give a side-by-side box plot of the scores by gender.
b) Test if the mean scores for males is lower than the mean scores for the women.
c) Give a 90% confidence interval for the mean difference between the SSHA scores of male and females.

Lloyd's Cereal company packages cereal in 1 pound boxes (16 ounces). A sample of 81 boxes is selected at random from the production line every hour, and if the average weight is less than 15 ounces, the machine is adjusted to increase the amount of cereal dispensed. If the mean for 1 hour is 1 pound and the standard deviation is 0.2 pound, what is the probability that the amount dispensed per box will have to be increased?

$$\bar{X} \sim N\left(\mu = 16, \sigma_{\bar{x}} = \sigma/\sqrt{n} = \frac{16}{45}\right)$$

$$\bar{X} \sim N\left(16, \frac{16}{45}\right)$$

$\sigma = 0.2$ lb
$= \frac{1}{5} \cdot 16$ oz

Want $P(\bar{X} < 15)$

An experimenter flips a coin 100 times and gets 58 heads. Test the claim that the coin is fair against the two-sided claim that it is not fair at the level α=.01.

$$H_0: p = \frac{1}{2}$$

$$\hat{p} = \frac{x}{n} \sim N\left(\frac{1}{2}, \sqrt{\frac{(\frac{1}{2})(\frac{1}{2})}{n}}\right)$$

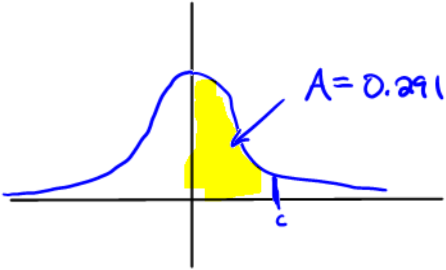
In a large population, 92% of the households have cable tv. A simple random sample of 64 households is to be contacted and the sample proportion computed. What is the probability that the sampling distribution of sample porportions is less than 90%?

$$\hat{p} \sim N\left(0.92, \sqrt{\frac{0.92(0.08)}{64}}\right)$$

But ! $64 \cdot 0.08 < 10$

So our methods won't work there

b. $P(0 \le Z \le c) = 0.291$



$$P(Z \le c) = P(Z \le 0) + P(0 \le Z \le c)$$
$$= 0.5 + 0.291$$
$$= 0.791$$

$$\Phi(c) = 0.791$$
$$c = \Phi^{-1}(0.791)$$

Let X be a random variable with Normal distribution, mean = 16, sd = 2. Find the value of x such that P(X > x) = 0.05.

$X \sim N(16, 2)$

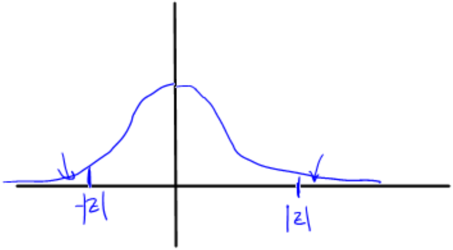
Want $P(X \geq x) = 0.05$

$\Rightarrow P(X \leq x) = 1 - 0.05 = 0.95$

$x = qnorm(0.95, 16, 2)$

$$2\cdot(1-\mathrm{pnorm}(z))$$

$$1-\mathrm{pnorm}(z) = 1 - P(Z \le z)$$
$$= P(Z > z)$$



$$2\cdot(1-\mathrm{pnorm}(|z|)$$
$$= 2\cdot\mathrm{pnorm}(-|z|)$$

Using R and RStudio

1. Download R from https://cran.r-project.org/
2. Download RStudio from https://www.rstudio.com/