Stat C131A: Statistical Methods for Data Science

Lecture10: Hypothesis Testing III

Sep 30 2025

Alternate Final Exam

- Please email me the alternate date and times you can make it

Today

- A bit of a dive into R
  - With the context of hypothesis testing
- More hypothesis testing problems

Hypothesis Testing with R

R: **Distribution** Functions

- The “4-pack”
  - d*: density
    - height of curve
  - p*: probability
    - cumulative
  - q*: quantile
    - inverse cumulative
  - r*
    - random draws

- This pattern applies to all distributions: Normal, t, Z, χ^2, F, Binomial, …

- `pnorm(-2.77)`
- `qnorm(0.975)`
- `rnorm(5, mean=0, sd=1)`

Normal Distribution Functions

- dnorm(x): curve height at x
- pnorm(z): $P(Z \le z)$
  - pnorm(z) = the cumulative probability up to a Z-value
  - pnorm(1.96) = 0.975
    - → 97.5% of the standard normal distribution lies below Z = 1.96
- qnorm(p): z cutoff for probability p
  - qnorm(p) = the quantile (cutoff Z-value) for probability p.
  - qnorm(0.975) = 1.96
    - → the Z cutoff where 97.5% of the area is to the left.
  - inverse of pnorm: what value of z will give me this p ?
- - rnorm(n, mean, sd): random draws

```
pnorm(1.96)    # ~0.975
qnorm(0.025)    # ~-1.96
```

t Distribution Functions

- **dt(x, df)** → *density at x*
  - Example: dt(2, df=10) gives the height of the t-distribution curve with 10 degrees of freedom at x = 2.
  - Analogous to dnorm(2) for the normal (Z) distribution.
  - Used for plotting curves.
- **pt(t, df)** → *probability up to t (cumulative)*
  - Example: pt(2, df=10) gives the probability that a t random variable is ≤ 2.
  - Analogous to pnorm(2).
  - Used for computing p-values.
- **qt(p, df)** → *cutoff for probability p (inverse cumulative)*
  - Example: qt(0.975, df=10) gives the t-value that leaves 97.5% of probability below it.
  - Analogous to qnorm(0.975) = 1.96 for the Z distribution.
  - Used for finding critical values (cutoffs).
- **rt(n, df)** → *random draws*
  - Example: rt(5, df=10) generates 5 random samples from a t-distribution with df=10.
  - Analogous to rnorm(5).
  - Used for simulation or resampling.

Visualization Helpers in R

- Plot a distribution curve
- Shade tail region for p-values
- Helpful for visualizing hypothesis tests

```
shade_right_tail <- function(x0, dens_fun) {
  xs <- seq(x0, 4, length.out=200)
  ys <- dens_fun(xs)
  polygon(c(x0, xs, 4), c(0, ys, 0), col='lightblue')
}
```

- **Custom function definition**
  - shade_right_tail <- function(x0, dens_fun) { ... }
  - Defines a function with two arguments:
    - x0: the observed statistic (e.g., Z_obs or t_obs).
    - dens_fun: the density function to use (e.g., dnorm or function(z) dt(z, df)).
- **Create x-values for the tail**
  - xs <- seq(x0, 4, length.out=200)
  - Generates 200 evenly spaced values from x0 to 4 (the right tail of the axis).
  - Assumes the distribution is effectively zero beyond +4.
- **Evaluate the density function**
  - ys <- dens_fun(xs)
  - Applies the density function (dnorm, dt, etc.) to each x-value.
  - Produces the y-coordinates (heights of the curve).
- **Draw the shaded polygon**
  - polygon(c(x0, xs, 4), c(0, ys, 0), col='lightblue')
  - Creates a filled polygon to shade the area under the curve.
  - c(x0, xs, 4) → the x-coordinates: start at x0, trace the curve, end at 4.
  - c(0, ys, 0) → the y-coordinates: start at 0, follow densities, back to 0.
  - col='lightblue' → fill color.

Printing and Numeric Functions

- cat()
  - Concatenates text and numbers into a single output string.
  - Unlike print(), it does not add quotes or indexes.
  - Example: cat("Observed Z:", Z_obs, "\n") prints a clean message.

- round() and signif()
  - round(x, 2) rounds a number to 2 decimal places.
  - signif(x, 3) keeps 3 significant digits (better for very small p-values).

- pt() and pnorm()
  - Return probabilities from distributions.
  - pnorm(z) = probability under the normal curve up to z.
  - pt(t, df) = probability under the t curve up to t with given degrees of freedom.

Plotting functions

- curve()
  - Plots a function over an interval by sampling points.
  - Example: curve(dnorm(x), from=-4, to=4) plots the standard normal density.
  - Automatically evaluates the function at many x values.

- abline()
  - Adds straight lines to an existing plot.
  - Vertical line: abline(v = value)
  - Horizontal line: abline(h = value)
  - Customizable with color, width, line type.

- text()
  - Places annotations on the plot.
  - Requires x and y coordinates.
  - Example: text(2, 0.15, labels="p ≈ 0.04") writes text inside the plot.

- dnorm() and dt()
  - Return densities (heights of the curve).
  - Used inside curve() and polygon() to draw the shape of normal or t-distributions.

Z-Test: Lunch Packet Example

- Scenario: Packet labeled 300g, sample mean 298.2, s=3.9, n=36
- H0: mu = 300 vs H1: mu < 300

```
mu0 <- 300; xbar <- 298.2; s <- 3.9; n <- 36
SE <- s / sqrt(n)
Z_obs <- (xbar - mu0) / SE
p_value <- pnorm(Z_obs)
```

Plot: Lunch Packet Z-Test

- Visualize the Z distribution and shade the left tail region for p-value.

- `curve(dnorm(x), from=-4, to=4)`
- `shade_left_tail(Z_obs, dnorm)`
- `abline(v=Z_obs, col='red', lty=2)`

Quick p-value Utilities

- Once Z is known, compute p-values with one-liners:

```
Z_obs <- -2.77
p_left  <- pnorm(Z_obs)
p_right <- 1 - pnorm(Z_obs)
p_two   <- 2 * (1 - pnorm(abs(Z_obs)))
```

t-Test: Dealcoholized Wine Example

- Scenario: Testing polyphenols remain ≥ 450 mg/L
- Sample: n=8, mean=456.5, s=8

```
mu0 <- 450; xbar <- 456.5; s <- 8; n <- 8; df <- n-1
SE <- s / sqrt(n)
t_obs <- (xbar - mu0) / SE
p_value <- 1 - pt(t_obs, df)
```

Plot: Wine t-Test

- Plot the t distribution (df=7) and shade the right tail for p-value.

- `curve(dt(x, df), from=-4, to=4)`
- `shade_right_tail(t_obs, function(z) dt(z, df))`
- `abline(v=t_obs, col='red', lty=2)`

Comparing t and Z Distributions

- - Small df: heavy tails (wider variability)
- - Large df: t ≈ Normal

```
x_vals <- seq(-4,4, length=400)
plot(dt(x_vals,7), type='l', col='red')
lines(dnorm(x_vals), col='green')
lines(dt(x_vals,99), col='blue')
```

Tail Options in R

- By default `pnorm` and `pt` compute $P(X \leq x)$.
- Use `lower.tail=FALSE` for right-tail probabilities.
- Two-sided p-values require doubling the smaller tail.

```
Z_obs <- 2.1
pnorm(Z_obs)                    # Left tail
pnorm(Z_obs, lower.tail=FALSE)  # Right tail
2 * pnorm(-abs(Z_obs))          # Two-sided
```

Simulation with rnorm / rt

- Use random draws to simulate sampling distributions.
- Good for bootstrapping and intuition building.

```
set.seed(42)
# Draw 1000 sample means (n=36)
sim_means <- replicate(1000, mean(rnorm(36, mean=300, sd=3.9)))
hist(sim_means, breaks=30, col='lightblue')
```

Why learn R (now) ?

- **Interpretation > Automation**
  - AI can *write* code, but we need to know enough R to **interpret what it's doing**
- **Statistical Thinking Is in the Code**
  - Functions like pnorm, qnorm, pt, and qt *encode statistical concepts* (cumulative probability, quantiles, degrees of freedom).
  - Writing them by hand forces us to **internalize the logic of hypothesis testing**.
- **Transparency & Trust**
  - Not "that's what the system says"
- **Reproducible Research**
  - R scripts can be re-run, modified, peer-reviewed, and archived. This matters for scientific integrity, especially in disciplines like biostatistics, economics, or psychology.
- **Error-Checking Muscle**
- **AI as a Partner, Not a Crutch**
  - The future is "human + AI."
  - Students who know the fundamentals can **use AI more effectively**:
    - Prompt better, spot errors, and extend auto-generated code intelligently.

Proportion statistic example

Should a new product be introduced or not ?

Scenario: *A blind test of the new fitness app was conducted with 200 users. The company will only proceed if more than 30% of the test users indicate they would adopt the app. In the trial, 32% of users reported that they would use it.*

- Question: Should the new fitness app be launched, or not?

- P = proportion
- Null Hypothesis: $H_0 : P = 0.3$
- Alternate Hypothesis: $H_1 : P > 0.3$

Z-stat

- $$Z = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}}$$

Z-stat

- $$Z = \frac{0.32 - 0.30}{\sqrt{\frac{0.30 \times (1 - 0.30)}{200}}} \approx 0.62$$

R code

```
##############################################
# One-sample Z-test for a proportion
# Example: p = 0.32, P = 0.30, n = 200
##############################################

# Parameters
p_hat <- 0.32    # sample proportion
P0    <- 0.30    # hypothesized population proportion
n     <- 200     # sample size

# Compute Standard Error
SE <- sqrt(P0 * (1 - P0) / n)

# Observed Z statistic
Z_obs <- (p_hat - P0) / SE

# One-sided p-value (H1: p > P0)
```

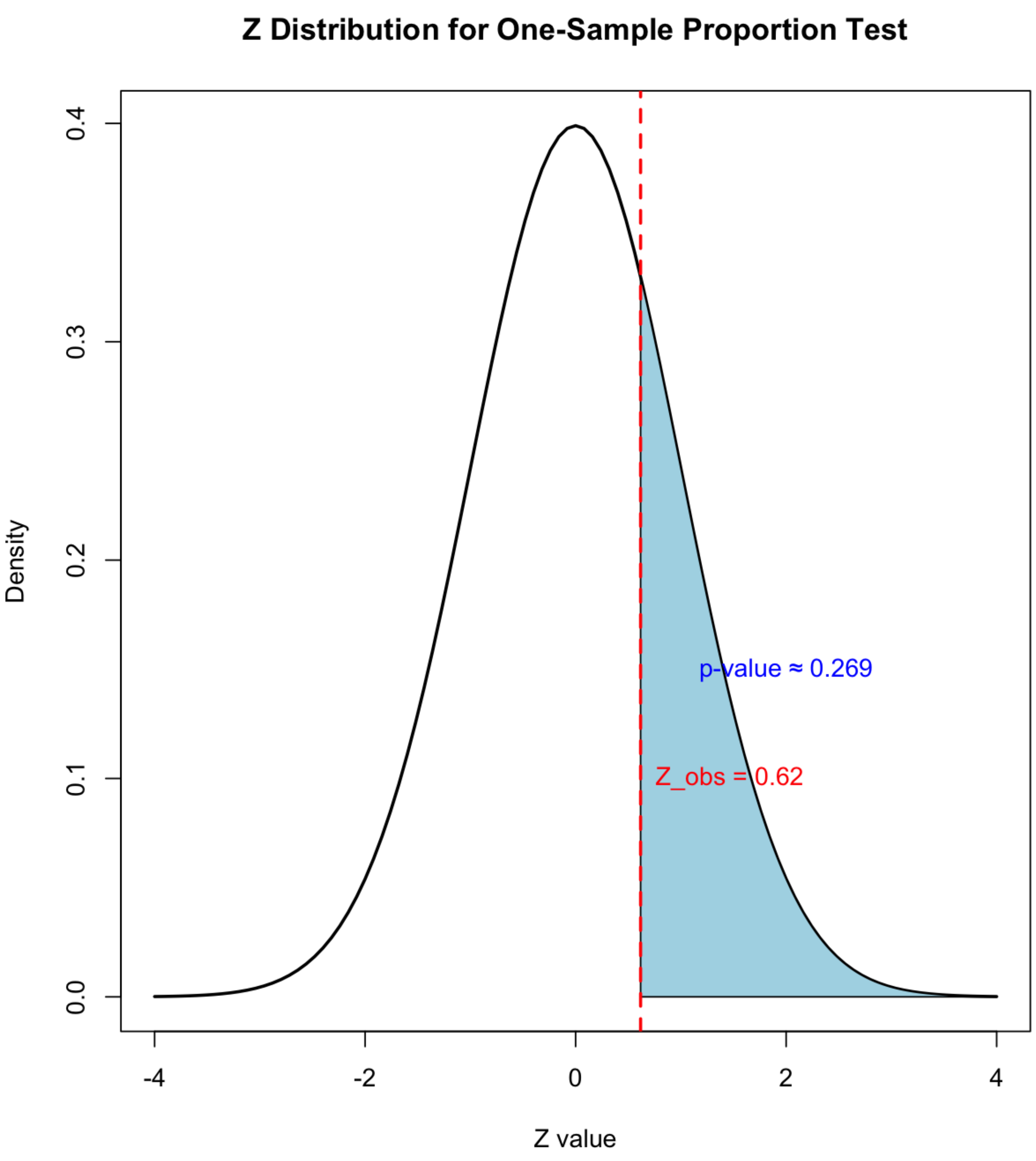
R code

```
############################################
# One-sample Z-test for a proportion
# Example: p = 0.32, P = 0.30, n = 200
############################################

# Parameters
p_hat <- 0.32    # sample proportion
P0    <- 0.30    # hypothesized population proportion
n     <- 200     # sample size

# Compute Standard Error
SE <- sqrt(P0 * (1 - P0) / n)

# Observed Z statistic
Z_obs <- (p_hat - P0) / SE
p_value <- 1 - pnorm(Z_obs)
```


Z Distribution for One-Sample Proportion Test
0.4
0.3
0.2
0.1
0.0
Density
p-value ≈ 0.269
Z_obs = 0.62
-4
-2
0
2
4
Z value

Decision

- **Result:** p-value = 0.269
- **Chosen alpha (significance level):** 0.05 (5%)
- **Decision:**
  - Since p-value (0.269) > alpha (0.05), we **fail to reject the null hypothesis**.
  - Evidence is insufficient to conclude that adoption is above 30%.
- **Implication:**
  - We cannot justify launching the app based on this test alone.
- **Type I error:**
  - If we had rejected the null here, the chance of making a Type I error would equal alpha = 5%.
  - In this case, because p-value is much larger than alpha, rejecting the null would have been very risky (about a 26.9% chance of seeing data this extreme under the null).

Bivariate t-test: 2 independent populations

Bivariate t-test: Independent populations

- **Central concept**
  - Compares the means of two **separate, unrelated groups**.
  - Example: undergraduates vs faculty, treatment group vs control group.
  - Each observation in one group has no natural pairing with an observation in the other.
- **What "independent" means**
  - Different subjects in each group.
  - No overlap or repeated measurement.
  - Group membership is mutually exclusive.
- **Null hypothesis**
  - The two population means are equal.
- **Variance handling**
  - If population variances are assumed equal → use **pooled t-test** (common variance estimate).
  - If variances may differ → use **Welch's t-test** (adjusts denominator and degrees of freedom).
  - Always safer to check for unequal variance.

Example: Effectiveness of fitness app in undergrad student versus faculty populations

- ***Scenario:*** *A university fitness program wants to evaluate the impact of a new step-tracking app. They collect data on the increase in average daily steps after installing the app from two independent groups: undergraduates and faculty members (seniors). Since undergrads and faculty are different populations, we want to test whether the mean improvement in steps is the same across the two groups, or whether one group benefits more than the other.*

  - **Question:** Is there a difference in the effectiveness of the fitness app in one population versus the other ?

Undergrads (U) data: $\{1500, 1700, 1600, 1550, 1800, 1400, 1650, 1750, 1580\}$

Faculty (F) data: $\{1200, 1250, 1300, 1100, 1400, 1350, 1280\}$

$n_U = 9, \quad n_F = 7$

$\overline{X}_U = 1603.3, \quad \overline{X}_F = 1268.6$

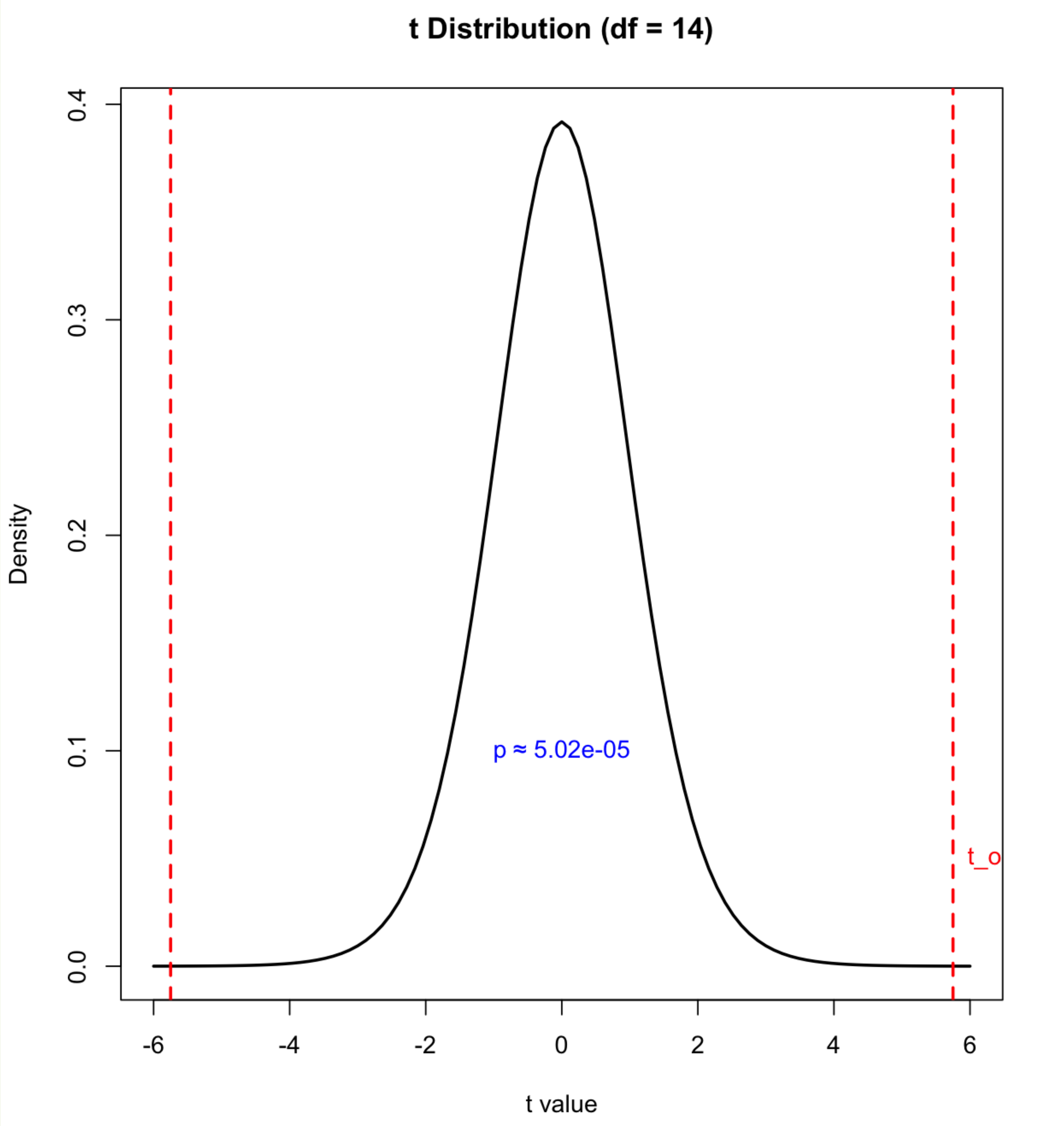
Null hypothesis (H0): $\mu_U = \mu_F$

Alternative hypothesis (H1): $\mu_U \neq \mu_F$

t-stat: Welch’s formula, Pooled

- $$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}}$$
- $$s_p{}^2 = \frac{(n_1 - 1)s_1{}^2 + (n_2 - 1)s_2{}^2}{n_1 + n_2 - 2}$$

- $$t = \frac{334.7}{\sqrt{\frac{15800}{9} + \frac{9900}{7}}} = \frac{334.7}{57.9} \approx 5.78$$
- $$t = \frac{334.7}{115.2 \cdot \sqrt{\frac{1}{9} + \frac{1}{7}}} = \frac{334.7}{58.2} \approx 5.75$$


t Distribution (df = 14)
Density
0.0
0.1
0.2
0.3
0.4
-6
-4
-2
0
2
4
6
t value
p ≈ 5.02e-05
t_o

Decision

- **Observed test statistic:** $t = 5.75$
- **Degrees of freedom:** 14
- **p-value:** 0.0000502
- **Chosen alpha (significance level):** 0.05 (5%)
- **Decision:**
  - Since p-value ≪ alpha, we **reject the null hypothesis**.
  - Evidence strongly supports that the mean step increase differs between undergraduates and faculty.
- **Interpretation:**
  - The probability of seeing a difference this extreme by chance (if the means were equal) is ~0.005%.
  - This is far below our tolerance for Type I error (5%).
  - We conclude that undergraduates had significantly higher step increases than faculty.

p-value: Interpreting, Reporting

Definition: The p-value is the probability of observing a test statistic as extreme as (or more extreme than) the one we obtained, if the null hypothesis were true.

Not confidence:

- A p-value does not measure the probability that the null hypothesis is true.
- It is not a "confidence level" in the result.

What it reflects

- The **significance of the evidence** against the null hypothesis.
- It quantifies how surprising our data are under the assumption of "no effect" or "no difference."

Preset threshold (alpha)

- Researchers decide a significance level before testing (for example alpha = 0.05).
  - If p-value is less than alpha, we reject the null hypothesis.

Interpretation nuances

- Very small p-value (for example 0.001): **strong evidence** against the null.
- Borderline p-value (for example 0.048 when alpha = 0.05): **marginal evidence.**
- Larger p-value (for example greater than 0.10): **little to no evidence** against the null.

Key takeaways

- The p-value measures the strength of evidence against the null hypothesis.
- It does not tell us the size or practical importance of the effect.
- It does not give the probability that the alternative is true.

Next lecture

- **Paired t-test**: Bivariate **related** populations