Midterm 2 Preparation

DSC 152: Lectures 6–11 Discussion

Zida Jin

University of California, San Diego

Fall 2026

All studies and data in this discussion are fictional and used for practice.

Scenario: how do students learn statistics?

A teaching team considers several studies of practice methods and quiz scores.

- **Q1:** randomly assign students to coding or handwritten exercises
- **Q2–Q4:** study quiz scores, study hours, and practice methods
- **Q5:** randomly assign coding and handwritten exercises separately

Definitions

`gain` = post-test score − pre-test score (points)
`score` = quiz score (points), `hours` = weekly study hours

Each question uses a **separate study**. Unless stated otherwise, use $\alpha = 0.05$ and assume the conditions for the stated test hold.

Course references use the page numbers printed in the lecture PDFs.

Q1(a): What does random assignment give us?

2 minutes

Students are randomly assigned to **coding** or **handwritten** exercises. Each student contributes one `gain`. Students work independently, and everyone completes the study. ☐ **Select all that apply.**

A. Random assignment supports a causal comparison of the two methods for these students.

B. Random assignment guarantees that these students represent all UCSD students.

C. The default two-sample `t.test()` requires equal group variances.

D. We must use a paired t-test because `gain` uses pre- and post-test scores.

E. With a binary predictor, the usual `lm()` slope test gives the same p-value as the equal-variance two-sample t-test.

Q1(a): Solution

✓ **A** Assignment makes the method independent of students' pre-existing characteristics, apart from chance imbalance.

✗ **B** Generalizing to a population depends on *sampling*, not just assignment.

✗ **C** The default is **Welch's** test. Equal variances are not required.

✗ **D** We compare gains from *different students* in two independent groups.

✓ **E** A binary-predictor `lm()` uses the pooled variance assumption.

For H_0: $\mu_C = \mu_H$ versus H_A: $\mu_C \neq \mu_H$:

```
t.test(gain ~ group, data = trial)$p.value
```

Review: Lec06 pp. 4–5, 11–13; Lec08 pp. 11–13

Q1(b): A permutation test

3 minutes

A small randomised study in `trial` gives gains $H: 0, 2, 4, 6$ and $C: 4, 6, 8, 10$, so $\bar{x}_C - \bar{x}_H = 4$. Let `diff_mean(d)` return the coding mean minus the handwritten mean.

Fill in the blanks for a **two-sided** permutation test of no treatment effect.

```
null_diff <- numeric(10000)
for(b in 1:10000){
  shuffled <- trial
  shuffled$group <- sample(trial$group, replace = __(a)__)
  null_diff[b] <- diff_mean(shuffled)
}
pval <- mean(__(b)__)
```

If 1,430 shuffled statistics are at least as extreme as 4, give the estimated p-value and decision.

Q1(b): Solution

```r
diff_mean <- function(d){
  mean(d$gain[d$group == "coding"]) -
    mean(d$gain[d$group == "handwritten"])
}
# (a) FALSE: keep four students in each group
# (b) abs(null_diff) >= abs(4)
```

$p = 1430/10000 =$ **0.143**, so fail to reject H_0.

Why shuffle?

Under the null of no effect on any student's gain, the randomised labels are exchangeable. Shuffling preserves outcomes and group sizes.

Review: Lec06 pp. 14–19

Q2(a): A slope test and a residual

3 minutes

An **observational** sample of $n = 20$ students gives the fitted model

$$\widehat{\texttt{score}} = 50 + 2\,\texttt{hours}, \qquad SE(\hat{\beta}_1) = 1.$$

Weekly study hours range from 0 to 10.

1. State H_0 and H_A for a two-sided test of a linear association.
2. Compute t, give its degrees of freedom, and write one R expression for the p-value **without** using `lm()`.
3. A student studies 6 hours and scores 65. Find the fitted score and residual.

Reference: `pt(-2, df = 18)` $= 0.0304$.

Q2(a): Solution

$$H_0 \colon \beta_1 = 0, \qquad H_A \colon \beta_1 \neq 0, \qquad t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{2}{1} = 2.$$

Two coefficients are estimated, so df $= 20 - 2 =$ **18**.

```
2 * pt(-abs(2), df = 18)      # 0.0608
```

Fail to reject at $\alpha = 0.05$.

$$\hat{y} = 50 + 2(6) = \mathbf{62}, \qquad e = y - \hat{y} = 65 - 62 = \mathbf{3}.$$

Review: Lec07 pp. 12–20, 25; Lec10 pp. 10–14

Q2(b): Simulating power for a slope

2 minutes

We simulate the test assuming $Y_i = 50 + 2X_i + \epsilon_i$, with independent $\epsilon_i \sim N(0, 10^2)$. Let $\alpha = 0.05$. Test $H_0 : \beta_1 = 0$ against $H_A : \beta_1 \neq 0$. Fill in the blanks.

```
count <- 0
for(i in 1:10000){
  x <- runif(40, min = 0, max = 10)
  y <- __(a)__
  fit <- __(b)__
  pval <- summary(fit)$coefficients[2, 4]      # why?
  if(__(c)__){ count <- count + 1 }
}
power <- count / 10000
```

Q2(b): Solution

```
count <- 0
for(i in 1:10000){
  x <- runif(40, min = 0, max = 10)
  y <- 50 + 2*x + rnorm(40, mean = 0, sd = 10)  # (a)
  fit <- lm(y ~ x)                              # (b)
  pval <- summary(fit)$coefficients[2, 4]       # row 2 is x, col 4 is p-value
  if(pval < 0.05){ count <- count + 1 }         # (c)
}
power <- count / 10000
```

Which truth are we simulating?

The true slope is 2, so rejections estimate **power**. Replacing `2*x` with `0*x` would instead estimate the **Type I error rate** of the same test.

Review: Lec07 pp. 26–33

Q2(c): What changes power?

1 minute

Starting from the planning model in Q2(b), which changes, applied **one at a time**, would increase power? ☐ **Select all that apply.**

A. Increase the sample size from 40 to 80, with the same distribution of X.

B. Reduce the error standard deviation from 10 to 5.

C. Change the intercept from 50 to 70.

D. Reduce the true slope from 2 to 1.

Q2(c): Solution

✓ **A** More independent observations give a more precise slope estimate.

✓ **B** Less noise makes the same slope easier to detect.

✗ **C** Adding a constant to every outcome changes the intercept, but leaves the slope estimate and its test unchanged.

✗ **D** The slope is closer to the null value of 0, so power falls.

Interpretation

Power is the probability of rejecting $\beta_1 = 0$ in repeated studies under the specified nonzero slope, sample size, predictor distribution, and error distribution.

Review: Lec07 pp. 23–33

Q3(a): Adjustment and centering

3 minutes

In an **observational** study, `coding` is 1 for coding exercises and 0 for handwritten exercises. Coding users tend to study longer.

$$\text{Unadjusted:} \quad \hat{y} = 55 + 6\,\texttt{coding}$$
$$\text{Adjusted:} \quad \hat{y} = 50 - 2\,\texttt{coding} + 3\,\texttt{hours}$$

Define `hours.c = hours - 4`. ☐ **Select all that apply.**

A. At the same study hours, coding users score an estimated 2 points lower on average.

B. The adjusted model proves that coding exercises lower scores.

C. With `hours.c`, the intercept becomes 62 and the coding coefficient stays -2.

D. A large p-value for `hours` would by itself justify dropping it from a planned adjustment model.

Q3(a): Solution

✓ **A** The coding coefficient compares the two methods *holding hours fixed*.

✗ **B** Adjustment for hours alone does not eliminate all possible confounding.

✓ **C** Substitute $\texttt{hours} = \texttt{hours.c} + 4$:

$$\hat{y} = 50 - 2\,\texttt{coding} + 3(\texttt{hours.c} + 4) = 62 - 2\,\texttt{coding} + 3\,\texttt{hours.c}.$$

✗ **D** Choose adjustment variables using the study question and design. Their individual p-values do not determine whether confounding matters.

```
obs$hours.c <- obs$hours - 4
fit <- lm(score ~ coding + hours.c, data = obs)
```

The new intercept is the fitted mean for **handwritten users studying 4 hours**.

Review: Lec08 pp. 14–24

Q3(b): A predictor with three categories

2 minutes

A different study includes **control, coding, and handwritten** groups. R uses control as the reference level. Suppose the fitted model is

$$\hat{y} = 50 + 4\, I_C + 2\, I_H + 3\, \texttt{hours},$$

where $I_C = 1$ for coding and $I_H = 1$ for handwritten, and 0 otherwise.

1. Give (I_C, I_H) for each group.
2. Predict the score for a coding student who studies 4 hours.
3. The p-value for I_C is 0.0065. Which comparison does it test? Does it test whether *all three* adjusted group means are equal?

Q3(b): Solution

Group	I_C	I_H
Control	0	0
Coding	1	0
Handwritten	0	1

Coding at 4 hours:

$$\hat{y} = 50 + 4 + 3(4) = \mathbf{66}.$$

At the same hours, the fitted coding mean is 4 points above control.

The p-value 0.0065 tests **coding versus control**, adjusting for hours. The corresponding null is $\beta_C = 0$.

Testing the whole categorical variable

All three adjusted group means are equal when $\beta_C = \beta_H = 0$. This requires a **joint test**.

Review: Lec09 pp. 7–12, 21–24

Q3(c): A partial F-test

3 minutes

For Q3(b), suppose the two models use the **same 60 students**. Test whether `group` is associated with score **after adjusting for hours**. Suppose:

$$RSS_{\text{reduced}} = 1280, \qquad RSS_{\text{full}} = 1120.$$

1. State H_0 and fill in the reduced model.

```
full <- lm(score ~ group + hours, data = dat)
reduced <- lm(__(a)__, data = dat)
anova(reduced, full)
```

2. Compute F, give both degrees of freedom, and state the decision.

Reference: `pf(4, 2, 56, lower.tail = FALSE)` $= 0.0238$.

Q3(c): Solution

H_0: $\beta_C = \beta_H = 0$. The alternative is that at least one of these coefficients is nonzero.

```
reduced <- lm(score ~ hours, data = dat)    # keep hours
anova(reduced, full)
```

$$F = \frac{(RSS_{\text{reduced}} - RSS_{\text{full}})/2}{RSS_{\text{full}}/(60-4)} = \frac{(1280-1120)/2}{1120/56} = \frac{80}{20} = \mathbf{4}.$$

Degrees of freedom: **2 and 56**. Since $p = 0.0238 < 0.05$, reject H_0.

What can we conclude?

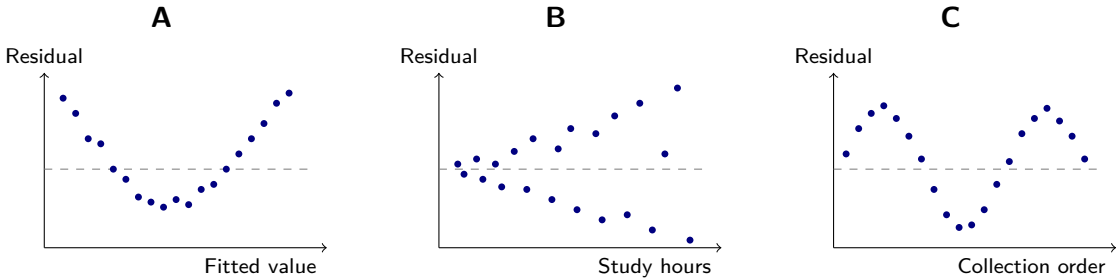
There is evidence of at least one adjusted group difference. This test does not establish that every pair of groups differs.

Review: Lec09 pp. 12–25

Q4(a): Which condition is in doubt?

3 minutes

The plots below come from **three separate regression fits**. Match each pattern to the LINE condition it calls into question.



For fit B, a residual Q-Q plot is roughly straight and the Shapiro-Wilk p-value is 0.40. Does this establish that **all** LINE conditions hold?

Schematic diagnostic patterns.

Q4(a): Solution

Plot	Pattern	Condition in doubt
A	Curved mean pattern	**L**: linear mean relationship
B	Increasing vertical spread	**E**: equal error variance
C	Runs or cycles over collection order	**I**: independence

- Check **N** with a residual histogram, Q-Q plot, and Shapiro-Wilk test. The two checks given here do not show strong evidence against normality.
- They do not check linearity, independence, or equal variance. Fit B still has evidence of unequal variance.

Review: Lec10 pp. 21–22, 26–29, 36–46; p. 48

Q4(b): Good estimates, bad p-values?

2 minutes

A simulation repeatedly generates independent errors with mean 0, a correct linear mean model, true slope $\beta_1 = 0$, and **error variance increasing with** X.

Suppose the average fitted slope is close to 0, but the usual `lm()` test rejects $\beta_1 = 0$ in **12%** of repetitions. At $\alpha = 0.05$: ☐ **Select all that apply.**

A. The usual slope test has an inflated Type I error rate in this simulation.

B. An approximately unbiased slope estimate guarantees a valid p-value.

C. Increasing the number of simulation repetitions will by itself fix the test's Type I error rate.

D. These results are consistent with an approximately unbiased slope estimate and an incorrectly estimated standard error.

Q4(b): Solution

✓ **A** The null is true, and $0.12 > 0.05$.

✗ **B** Bias concerns the center of the slope estimator's sampling distribution. Inference also requires its uncertainty to be estimated correctly.

✗ **C** More repetitions estimate the rejection rate more precisely. They do not change the procedure being evaluated.

✓ **D** Mean-zero errors and a correct linear mean can give an unbiased slope even when unequal variances invalidate the usual standard error.

Two separate checks

$\text{Bias}(\hat{\beta}_1) = E(\hat{\beta}_1) - \beta_1$ concerns estimation. The rejection rate under H_0 concerns test validity.

Review: Lec10 pp. 30–34

Q5(a): When the effect depends on the other factor

3 minutes

A new 2^2 factorial experiment independently assigns **coding** ($C = 0/1$) and **handwritten** ($H = 0/1$) exercises. Each student receives one combination. The sample means are:

	$H = 0$	$H = 1$
$C = 0$	50	52
$C = 1$	54	62

For $\hat{y} = \hat{\beta}_0 + \hat{\beta}_C C + \hat{\beta}_H H + \hat{\beta}_{CH} CH$:

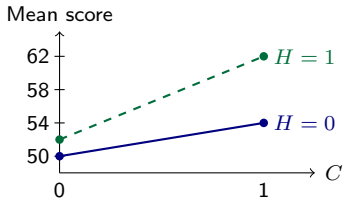
1. Find the four coefficient estimates.
2. What is the estimated coding effect when $H = 0$? When $H = 1$?
3. Interpret the interaction coefficient in one sentence.

Q5(a): Solution

$$\hat{\beta}_0 = 50, \qquad \hat{\beta}_C = 54 - 50 = 4, \qquad \hat{\beta}_H = 52 - 50 = 2,$$
$$\hat{\beta}_{CH} = (62 - 52) - (54 - 50) = 10 - 4 = \mathbf{6}.$$

- No handwritten exercises: coding effect $=$ **4** points.
- With handwritten exercises: coding effect $= 4 + 6 =$ **10** points.

The estimated coding effect is **6 points larger** when students also receive handwritten exercises.



Review: Lec11 pp. 3, 12–19

Q5(b): Which hypothesis does each p-value test?

3 minutes

The factorial study has 10 students per combination. The fitted model is:

```
full <- lm(score ~ coding * handwritten, data = fact)
```

The overall model p-value is **0.011**, the interaction p-value is **0.243**, and the `coding` coefficient p-value is **0.271**.

1. State H_0 for the overall test and for the interaction test. Give both decisions.
2. A colleague says $p = 0.271$ means coding has no effect at either level of handwritten exercises. Explain the error.
3. Write R code for a joint test of **all terms involving coding**. This test gives $p = 0.018$. State the conclusion.

Q5(b): Solution

- **Overall:** H_0: $\beta_C = \beta_H = \beta_{CH} = 0$. Reject ($p = 0.011$). The intercept is not part of this null.
- **Interaction:** H_0: $\beta_{CH} = 0$. Fail to reject ($p = 0.243$). A sample interaction of 6 does not by itself establish a population interaction.
- **Coding coefficient:** $p = 0.271$ tests $\beta_C = 0$, the coding effect *only when* $H = 0$. It does not test coding at both levels.

```
reduced <- lm(score ~ handwritten, data = fact)
anova(reduced, full)      # joint H0: beta_C = beta_CH = 0
```

Since $p = 0.018 < 0.05$, reject the joint null. There is evidence that coding affects the mean score at **at least one** level of handwritten exercises.

Review: Lec11 pp. 20–26

Three things to remember

1. **The study design determines the claim.**
   Random assignment supports causation. Random sampling supports generalization. Adjustment compares students with the same covariate values.

2. **Write the null before choosing the p-value.**
   One coefficient, an entire categorical predictor, and all terms involving a factor are different hypotheses.

3. **A fitted line is only part of the analysis.**
   Check LINE conditions. An interaction makes effects conditional. Failing to reject does not establish that an effect is zero.