Practice Midterm - DSC 152, Fall 2026

Instructions:

- This is a practice midterm. It provides an example of the midterm exam, but this does not constitute any guarantee for the midterm exam.
- This practice midterm consists of 8 questions. You have **50 minutes** to complete it.
- Please write clearly in the provided answer boxes; work elsewhere may not be graded. Completely fill in bubbles and square boxes; unclear selections may lose points.
  ◯ A bubble means select only one choice.
  ☐ A square box means select all that apply.
- Show all work and R code where requested. Partial credit may be awarded.
- You may use one double-sided handwritten sheet of notes. No calculators, computers, phones, or other devices.
- Assume all necessary `library()` calls in R have already been run.

Additional Practice Midterm notes:

- This practice midterm is intended to reflect the format and difficulty of the provided practice quiz while sampling important concepts from Lectures 1 through 5.
- It is not intended to cover every concept or course activity. Use it as practice, then check that you can explain why each answer is correct.

A city is evaluating trip durations for its shared e-scooter program. The operations team claims that the average trip duration across all trips is 18 minutes. To investigate this claim, we will use the `scoot` data frame, which contains one row per randomly sampled trip, with one trip per rider. Treat the observations as independent. The columns are:

- `trip_id` (double): unique identifier for each trip
- `zone` (character): service area (`"North"` or `"South"`)
- `day_type` (character): `"Weekday"` or `"Weekend"`
- `duration_min` (double): trip duration in minutes
- `helmet` (logical): `TRUE` if a helmet was recorded, `FALSE` otherwise

The first five rows of `scoot` are shown below. Not all columns will be needed on this midterm.

	trip_id	zone	day_type	duration_min	helmet
0	81001	North	Weekday	12.4	FALSE
1	81002	South	Weekend	21.1	TRUE
2	81003	North	Weekday	16.3	TRUE
3	81004	South	Weekday	24.7	FALSE
4	81005	North	Weekend	18.9	TRUE

The city claims that the average trip duration is 18 minutes. You have been asked to assess this claim and consider related statistical questions.

PID: _____

Question 1

You want to test whether the true average trip duration differs from 18 minutes.

a) Write the null and alternative hypotheses in terms of the appropriate quantity of interest.

H_0 : H_A :

b) Write one line of R code to run this two-sided t-test on the durations in `scoot` and output the p-value.

c) Suppose the p-value is 0.032. A teammate says: "This means there is a 3.2% chance that the true average trip duration is 18 minutes." Is this interpretation correct? Why or why not?

d) Name the two key conditions discussed in class that support the nominal Type I Error rate of a one-sample t-test, especially for a small sample. Briefly explain why each matters.

PID: _____

Question 2

Suppose the true mean trip duration is 30 minutes, the standard deviation is 4 minutes, and durations follow a normal distribution.

a) What is the name of the quantity that measures the probability of correctly rejecting $H_0 : \mu = 18$ in this scenario?

b) In a few words, describe what must happen every time the test is performed for this quantity to equal 1.0.

c) A colleague runs the following R code to calculate power for a two-sided one-sample t-test with $n = 10$, $\delta = 12$, $\sigma = 4$, and $\alpha = 0.05$:

```
power.t.test(n = 10, delta = 12, sd = 4,
             sig.level = 0.05,
             type = "one.sample",
             alternative = "two.sided")
## power = 0.9999
```

Briefly explain what this power value means in the e-scooter study.

PID: _____

Question 3

One condition for the t-test is that the data come from a normal distribution, especially for small samples. Suppose instead that trip durations follow a right-skewed distribution with true mean 18 minutes.

a) If we run the t-test at $\alpha = 0.05$ with $n = 6$ observations from this skewed distribution, what happens to the actual Type I Error rate compared with 0.05? Select the best answer.

◯ A. It will be exactly 0.05.

◯ B. It may differ from 0.05; it can be inflated or deflated.

◯ C. It will be exactly 0.

◯ D. It will always be greater than 0.05.

b) Fill in the blanks in the R code below to estimate the Type I Error rate by simulation. Each sample is drawn under H_0 from a Gamma distribution with mean 18.

```
count <- 0
for (i in 1:10000) {
  skewed_data <- rgamma(n = 6, shape = 2,
                        scale = 9)
  p.val <- t.test(skewed_data,
                  mu = __(a)__)$__(b)__
  if (p.val < __(c)__) {
    count <- count + 1
  }
}
TypeI <- count / 10000
```

(a): _____ (b): _____

(c): _____

c) In one sentence, explain what the resulting value of `TypeI` estimates.

PID: _____

Question 4

A colleague suggests using the sign test because trip durations may be skewed.

a) Suppose a sample of 8 trip durations, in minutes, is:

$$12,\ 21,\ 16,\ 24,\ 18,\ 27,\ 14,\ 20$$

Using the sign-test reference value $\tilde{\mu}_0 = 18$, write the sign sequence. What do you do with an observation equal to the reference value?

b) Using your sequence, state the number of positive signs, negative signs, and nonzero signs.

c) Let X be the number of positive signs. Give the two-sided p-value using a Binomial$(n, 0.5)$ probability or a `pbinom()` expression. You do not need to evaluate it.

d) The sign test is valid here, but it tests a different quantity than the t-test. Explain the difference.

PID: _____

Question 5

Assume six independent observations are tested with a two-sided sign test and there are no ties. The test rejects when its p-value is strictly less than $\alpha = 0.05$.

a) What is the smallest possible two-sided p-value?

b) What actual Type I Error rate does this test have at $\alpha = 0.05$? Briefly explain using the rejection outcomes.

PID: _____

Question 6

Researchers want to estimate power for the one-sample t-test of $H_0 : \mu = 18$ versus $H_A : \mu \neq 18$ using simulation. Assume data are normally distributed with mean $\mu_0 + \delta$ and standard deviation `sd`.

a) Complete the function below to return a simulated estimate of power. The null value is $\mu_0 = 18$.

```
sim_power <- function(n, delta, sd, alpha = 0.05,
                      reps = 10000) {
  count <- 0
  for (i in 1:reps) {
    samp <- rnorm(n = __(a)__,
                  mean = __(b)__,
                  sd = __(c)__)
    pval <- t.test(samp, mu = 18)$p.value
    if (__(d)__) count <- count + 1
  }
  return(__(e)__)
}
```

(a): _____ (b): _____

(c): _____ (d): _____

(e): _____

b) If `sim_power(n=12, delta=3, sd=6)` returns approximately 0.35, what does this number represent in the e-scooter context?

PID: _____

Question 7

The city collects a large random sample of $n = 40{,}000$ trips and runs a t-test of $H_0 : \mu = 18$ versus $H_A : \mu \neq 18$. The sample mean is $\bar{x} = 18.12$ minutes, the sample standard deviation is $s = 8$ minutes, and the p-value is 0.003.

a) At $\alpha = 0.05$, what is the decision?

◯ Reject H_0. ◯ Fail to reject H_0.

b) Despite the highly significant p-value, a city manager argues that a 0.12-minute difference in average trip duration is operationally unimportant. In 2–3 sentences, explain what concept this illustrates and why the p-value alone may be misleading with a very large sample.

Question 8

A bootstrap analysis reports a 95% confidence interval of [17.4, 20.8] minutes for the population mean trip duration.

a) Based on this interval, what is the decision for testing $H_0 : \mu = 18$ at $\alpha = 0.05$?

◯ A. Reject H_0, because 18 is inside the interval.

◯ B. Fail to reject H_0, because 18 is inside the interval.

◯ C. Reject H_0, because 18 is outside the interval.

◯ D. Fail to reject H_0, because 18 is outside the interval.

b) A colleague says bootstrap confidence intervals are guaranteed to have a Type I Error rate of exactly 0.05 for any population distribution because bootstrap is nonparametric. Is this claim correct? Select the best answer.

◯ A. Yes, the claim is correct.

◯ B. No. The bootstrap can have an inflated or deflated Type I Error rate, especially for small samples.

◯ C. No. The bootstrap is a parametric procedure.

◯ D. Yes. The bootstrap is nonparametric, so it is valid for the median but not the mean.

c) Briefly describe how you could estimate the actual Type I Error rate of the bootstrap test by simulation under a specified population distribution with mean 18.