Exam 2 — Review Handout

Stat 515 (Fall 2026)

Instructor: Yen-Yi Ho

Each question is reprinted exactly as it appeared on the exam, followed by a full solution and — where points were commonly lost — a note on the mistake to avoid. Point values show how each question was weighted. Probabilities are given to four decimal places; values you read from the z and t tables may differ from software by about ±0.001, and either is correct.

---

1. (18 pts) Short questions.

1) (3 pts) *Define what is meant by a Type I error of a test.*

Solution. A **Type I error** is rejecting the null hypothesis H_0 when H_0 is in fact true.

2) (3 pts) *How is the Type I error of a test related to the significance level (α)?*

Solution. The significance level α **is** the probability of a Type I error: $\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$. It is the largest Type I error rate you are willing to tolerate, and it is chosen **before** the data are examined. Because we reject whenever the p-value $\leq \alpha$, in repeated sampling from a population where H_0 is true we would falsely reject H_0 in about $100\alpha\%$ of samples — so testing at $\alpha = 0.05$ accepts a 5% false-positive rate.

3) (3 pts) *Define what is meant by the p-value (observed significance level) of a test.*

Solution. The **p-value** is the probability of obtaining a sample result at least as extreme as the one actually observed, computed **assuming H_0 is true**. A small p-value means the observed data would be unusual if H_0 were true, and therefore counts as evidence against H_0.

Common mistake. Describing the p-value as "the probability that H_0 is true." The p-value is computed *assuming* H_0 is true, so it cannot be a probability about H_0. It is P(data this extreme | H_0 true), not P(H_0 true | data).

4) (3 pts) *What would happen to the Type II error of a test, if we decrease α?*

Solution. Decreasing α **increases** the probability of a Type II error, β, and therefore **decreases** the power, $1 - \beta$, with the sample size and the true parameter value held fixed. A smaller α shrinks the rejection region, so stronger evidence is demanded before H_0 is rejected, and genuine departures from H_0 are missed more often. The two error probabilities trade against each other — the only way to reduce both at once is to increase the sample size n.

5) (2 pts) *(Circle the correct answers) In general, a $(1-\alpha)100\%$ confidence interval for μ becomes wider / narrower if the sample size n is increased, and wider / narrower if the confidence level $(1-\alpha)$ is increased.*

Solution. narrower if the sample size n is increased; **wider** if the confidence level $(1 - \alpha)$ is increased.

6) (2 pts) *(Circle the correct answers) A test of hypotheses results in a p-value of 0.0372. If the α-level is 0.01 you would fail to reject / reject the null hypothesis. If the α-level is 0.05 you would fail to reject / reject the null hypothesis.*

Solution. $\alpha = 0.01$: **fail to reject** $(0.0372 > 0.01)$. $\alpha = 0.05$: **reject** $(0.0372 < 0.05)$.

7) (2 pts) *(Circle the correct answer) Suppose that for a test of H_0: $\mu = \mu_0$ versus H_a: $\mu \neq \mu_0$ the calculated test statistic is $z = -1.84$. The p-value for this test is $P(Z < -1.84)$ / $P(Z > -1.84)$ / $2P(Z < -1.84)$*

Solution. $2 \cdot P(Z < -1.84)$. The test is two-sided, so both tails contribute; the observed statistic is negative, so the lower-tail area is doubled. Numerically, $2(0.0329) = 0.0658$.

2. (12 pts) Textbook spending.

A national survey reports that the mean annual amount undergraduates spend on textbooks and course materials is $\mu = \$1{,}240$, with standard deviation $\sigma = \$400$. A random sample of 100 undergraduates is selected, and $\bar{X}$ denotes the sample mean amount spent.

a) (3 pts) *Based on the central limit theorem (CLT), find the mean of $\bar{X}$.*

Solution. By the CLT the sampling distribution of $\bar{X}$ is centered at the population mean: $\mu(\bar{X}) = \mu =$ **$1,240**.

b) (3 pts) *Based on the CLT, find the standard deviation of $\bar{X}$.*

Solution. $\sigma(\bar{X}) = \sigma / \sqrt{n} = 400 / \sqrt{100} = 400 / 10 =$ **$40**.

Common mistake. Reporting \$400. The standard deviation of $\bar{X}$ is the **standard error** $\sigma/\sqrt{n}$, not the population standard deviation σ. $\bar{X}$ averages 100 observations, so it varies far less than a single student's spending does.

c) (3 pts) *Describe the shape of the sampling distribution of $\bar{X}$. (Justify your answer clearly.)*

Solution. Approximately **normal**. The shape of the population is not given, but $n = 100$ is well above the rule of thumb of $n \geq 30$, so by the **Central Limit Theorem** the sampling distribution of $\bar{X}$ is approximately normal.

Common mistake. Justifying this by saying the population is normal. The problem never says that — and it does not need to. The CLT is what does the work here, and citing it (together with $n = 100 \geq 30$) is the justification the question asks for.

d) (3 pts) *How likely is it that the mean amount spent by these 100 students is greater than \$1,290?*

Solution.

$$P(\bar{X} > 1290) = P(\, Z > (1290 - 1240) / 40\,) = P(Z > 1.25) = 1 - 0.8944 = \mathbf{0.1056}$$

Common mistake. Standardizing with 400 instead of 40. This is the same error as in part (b), and it changes z from 1.25 to 0.125 — a completely different probability. Whenever the question is about a *sample mean*, divide by the standard error.

3. (10 pts) Returned orders.

An online retailer knows that 20% of all orders placed on its site are eventually returned. A random sample of 400 orders is selected; let X denote the number of orders in the sample that are returned. Use the normal approximation to the binomial, and do not apply a continuity correction.

a) (4 pts) *Find the mean and standard deviation of X, and explain why the normal approximation is appropriate here.*

Solution.

$\mu = np = 400(0.20) = \mathbf{80}$

$\sigma = \sqrt{(np(1-p))} = \sqrt{(400 \times 0.20 \times 0.80)} = \sqrt{64} = \mathbf{8}$

The approximation is appropriate because $np = 80$ and $n(1-p) = 320$ are both far larger than 15, so the binomial distribution is close to bell-shaped. Hence X is approximately $N(\mu = 80, \sigma = 8)$.

Common mistake. Checking only one of the two conditions. Both $np \geq 15$ **and** $n(1-p) \geq 15$ have to hold — a sample can be large while p sits so close to 0 or 1 that the distribution is still badly skewed.

b) (3 pts) *Approximate $P(X \geq 92)$.*

Solution.

$P(X \geq 92) \approx P(Z \geq (92 - 80)/8) = P(Z \geq 1.50) = 1 - 0.9332 = \mathbf{0.0668}$

c) (3 pts) *Approximate $P(68 \leq X \leq 92)$.*

Solution.

$P(68 \leq X \leq 92) \approx P(-1.50 \leq Z \leq 1.50) = 0.9332 - 0.0668 = \mathbf{0.8664}$

For reference, the exact binomial values are 0.0770 and 0.8822 — close enough that either would lead to the same practical conclusion. The small gap is the price of using a smooth normal curve to describe a discrete count.

4. (6 pts) Setting up hypotheses.

For each situation, state the appropriate null and alternative hypotheses. Be sure to define any parameter you use in terms of the problem.

a) (3 pts) *A tire manufacturer advertises that its all-season tires last, on average, at least 60,000 miles. A consumer group suspects the true mean tread life is less than advertised and plans to test the claim.*

Solution. Let μ = the true mean tread life (miles) of the manufacturer's all-season tires.

H_0: $\mu = 60{,}000$ vs. H_a: $\mu < 60{,}000$

One-sided (lower-tailed): the consumer group's suspicion is the research claim, so it goes in H_a. *(Writing H_0: $\mu \geq 60{,}000$ is also acceptable; the test is carried out at the boundary either way.)*

b) (3 pts) *A school district claims that more than 60% of its graduates enroll in a four-year college. A researcher collects data hoping to find evidence supporting this claim.*

Solution. Let p = the true proportion of the district's graduates who enroll in a four-year college.

H_0: $p = 0.60$ vs. H_a: $p > 0.60$

One-sided (upper-tailed): the district's claim is what the researcher hopes to support, so it goes in H_a.

Common mistake. Writing the hypotheses about $\hat{p}$. Hypotheses are always statements about a **parameter** — the unknown population quantity — never about a statistic you computed from the sample. You already know $\hat{p}$; there is nothing to hypothesize about it.

5. (10 pts) Interpreting a confidence interval.

A city utility studied monthly household water use (in gallons). A random sample of 45 households produced a 90% confidence interval of (3,180 gallons, 4,020 gallons).

a) (4 pts) *Identify the error in the following interpretation and briefly explain why it is an error: "We are 90% confident that the average monthly water use of the 45 households in the sample is between 3,180 and 4,020 gallons."*

Solution. The interval estimates the **population** mean, not the mean of the 45 sampled households. The mean of those 45 households is a known, fixed number — it is the point estimate at the center of the interval, (3,180 + 4,020)/2 = 3,600 gallons — so there is nothing to be "confident" about for it.

b) (4 pts) *Identify the error in the following interpretation and briefly explain why it is an error: "There is a 90% probability that the mean monthly water use of all households in the city is between 3,180 and 4,020 gallons."*

Solution. Once the sample is drawn and the interval computed, the population mean μ is a **fixed, unknown constant**, not a random variable. This particular interval either does or does not contain μ — the probability is 0 or 1, we simply do not know which. The randomness lives in the sampling procedure, not in μ.

c) (2 pts) *Give a correct interpretation of this interval.*

Solution. "We are 90% confident that the true mean monthly water use of **all households in the city** is between 3,180 and 4,020 gallons." Equivalently: if this sampling and interval-construction procedure were repeated many times, about 90% of the resulting intervals would contain μ.

Worth remembering. The two errors in (a) and (b) are the two most common ways to misread any confidence interval: shifting it from the population to the sample, and treating a fixed parameter as if it were random. The "90%" is a property of the *method*, not of the one interval in front of you.

6. (9 pts) Battery charging time.

An engineer tests a new fast-charge battery. A random sample of 16 batteries gave a mean charging time of $\bar{x} = 42.5$ minutes with standard deviation $s = 6.0$ minutes.

a) (6 pts) *Construct and interpret a 95% confidence interval for the true mean charging time.*

Solution. σ is unknown and n is small, so use a t interval with df = n − 1 = 15.

$SE = s / \sqrt{n} = 6.0 / \sqrt{16} = 6.0 / 4 = 1.5$

$t^*(0.025, 15) = 2.131$

$42.5 \pm 2.131(1.5) = 42.5 \pm 3.1965 \Longrightarrow$ **(39.30, 45.70) minutes**

Interpretation: we are 95% confident that the true mean charging time of this battery is between 39.30 and 45.70 minutes.

Common mistake. Using $z^* = 1.96$ instead of $t^* = 2.131$. σ is unknown here — only s is available — so the t distribution with df = 15 applies. Using z gives a margin of error that is too small and an interval that is too narrow, which overstates how precisely you know μ.

b) (3 pts) *What must be true about the distribution of charging times in the population for the procedure in part (a) to be valid, and why does that condition matter here in particular?*

Solution. The population of charging times must be **approximately normal**. This matters here because n = 16 is small: the Central Limit Theorem has not had a chance to make $\bar{X}$ approximately normal on its own, so the t procedure must rely on the population itself being normal. With $n \geq 30$ the condition could be relaxed.

7. (9 pts) Sample size determination.

a) (5 pts) *A researcher wants to estimate the mean daily recreational screen time (in hours) of adults to within ±0.25 hours with 95% confidence. A pilot study suggests $\sigma \approx 1.4$ hours. What sample size is needed?*

Solution. 95% confidence $\Longrightarrow z(0.025) = 1.96$; $\sigma \approx 1.4$; SE (half-width) = 0.25.

$n = (z(\alpha/2))^2 \sigma^2 / (SE)^2 = (1.96 \times 1.4 / 0.25)^2 = (10.976)^2 = 120.47 \Longrightarrow$ **n = 121**

b) (4 pts) *A pollster wants to estimate the proportion of registered voters who support a ballot measure to within ±0.03 with 95% confidence. No prior estimate of p is available. What sample size is needed?*

Solution. With no prior estimate, use the most conservative value p = 0.5, which maximizes pq:

$n = (z(\alpha/2))^2 pq / (SE)^2 = (1.96 / 0.03)^2 (0.5)(0.5) = (65.333)^2(0.25) = 1{,}067.11 \Longrightarrow$ **n = 1,068**

Common mistake. Rounding down. A sample size always rounds **up**, no matter what the decimal is: n = 120.47 becomes 121, not 120. Rounding down leaves you just short of the precision you asked for, and you cannot sample 0.47 of a person.

8. (14 pts) Knee rehabilitation.

Historically, the mean recovery time following a certain knee surgery is 10 weeks. A hospital introduces a new rehabilitation protocol and wants to know whether the mean recovery time under the new protocol differs from 10 weeks. A random sample of 16 patients on the new protocol had a mean recovery time of 8.5 weeks with a standard deviation of 2.4 weeks. Assume recovery times are approximately normally distributed. Use $\alpha = 0.05$.

a) (3 pts) *State the null and alternative hypotheses, defining the parameter in terms of the problem.*

Solution. Let μ = the true mean recovery time (weeks) following knee surgery under the new rehabilitation protocol.

H_0: $\mu = 10$ vs. H_a: $\mu \neq 10$ (two-sided)

b) (3 pts) *Compute the appropriate test statistic and report its degrees of freedom.*

Solution.

$SE = s / \sqrt{n} = 2.4 / \sqrt{16} = 2.4 / 4 = 0.6$

$t = (\bar{x} - \mu_0) / (s/\sqrt{n}) = (8.5 - 10) / 0.6 = -1.5 / 0.6 = \mathbf{-2.50}$, df = **15**

c) (3 pts) *Find the rejection region.*

Solution. Two-sided test at $\alpha = 0.05$ with df = 15 $\Longrightarrow$ $t^*(0.025, 15) = 2.131$.

Rejection region: **reject H_0 if $|t| > 2.131$** (that is, $t < -2.131$ or $t > 2.131$)

d) (3 pts) *State your conclusion clearly, in the context of the problem.*

Solution. Since $|t| = 2.50 > 2.131$, we **reject H_0** at the 5% level. There is sufficient evidence to conclude that the mean recovery time under the new protocol differs from 10 weeks; because $\bar{x} = 8.5$, the evidence points to a **shorter** mean recovery time than the historical 10 weeks.

Common mistake. Writing that the test "proves the protocol works." A hypothesis test provides *evidence against* H_0; it never proves anything, and failing to reject is never proof that H_0 is true. State the conclusion in terms of the quantity the study was about — here, mean recovery time.

e) (2 pts) *Using the t table, find the p-value of this test. Report a single number, rounded to four decimal places.*

Solution. Use the **upper-tail probability table**. In the row $t = 2.50$, the df = 15 column gives the one-tail area

$P(T_{15} > 2.50) = 0.0123$

The test is two-sided, so double it:

p-value = $2(0.0123) = \mathbf{0.0246}$

Software gives 0.0245; either is correct. Since $0.0246 < 0.05$, this agrees with the rejection in part (d) — as it always must.

Common mistake. Reporting 0.0123 and stopping. The table gives a **one-tail** area, but H_a: $\mu \neq 10$ puts rejection region in *both* tails, so the p-value is twice the table entry. (The table also has no negative t values — use $|t| = 2.50$ and read the same entry, since the t distribution is symmetric.)

9. (12 pts) Public transit use.

A city transit authority claims that 40% of city residents use public transit at least once a week. In a random sample of 600 residents, 216 reported using public transit at least once a week. Test, at the $\alpha = 0.05$ level, whether the true proportion differs from 0.40.

a) (2 pts) *State the null and alternative hypotheses in terms of p, the true proportion of all city residents who use public transit at least once a week.*

Solution. Let p = the true proportion of all city residents who use public transit at least once a week.

H_0: $p = 0.40$ vs. H_a: $p \neq 0.40$

b) (3 pts) *State the conditions required for this test to be valid, verify them, and compute the appropriate test statistic.*

Solution. Conditions: the sample must be random (stated in the problem) and large enough for the normal approximation. Under H_0, $np_0 = 600(0.40) = 240 \geq 15$ and $n(1 - p_0) = 600(0.60) = 360 \geq 15$, so both conditions are satisfied.

$\hat{p} = 216 / 600 = 0.36$

$z = (\hat{p} - p_0) / \sqrt{(p_0(1 - p_0) / n)} = (0.36 - 0.40) / \sqrt{((0.40)(0.60) / 600)} = -0.04 / 0.02 =$ **-2.00**

Common mistake. Putting $\hat{p}$ inside the standard error instead of p_0. This was the most frequent slip on the exam. In a **test**, everything is computed *assuming H_0 is true*, so the standard error uses the hypothesized $p_0 = 0.40$. Using $\hat{p}$ gives $z = -2.04$ instead of -2.00. Compare this with part (e) below, where the opposite rule applies.

c) (2 pts) *Find the p-value of the test.*

Solution. Two-sided: p-value $= 2 \cdot P(Z < -2.00) = 2(0.0228) =$ **0.0456**. Software gives 0.0455; either is correct.

d) (2 pts) *State your conclusion clearly, in the context of the problem.*

Solution. Since the p-value $0.0456 < 0.05 = \alpha$, we **reject H_0**. There is sufficient evidence at the 5% level to conclude that the true proportion of city residents who use public transit at least once a week **differs from** 0.40 — and since $\hat{p} = 0.36$, the data suggest it is lower than the authority claims.

e) (3 pts) *Construct a 95% confidence interval for p. Does the interval agree with your test decision in part (d)? Explain.*

Solution. For a confidence interval the standard error uses $\hat{p}$ rather than p_0:

$SE = \sqrt{(\hat{p}\hat{q} / n)} = \sqrt{((0.36)(0.64) / 600)} = \sqrt{0.000384} = 0.0196$

$0.36 \pm 1.96(0.0196) = 0.36 \pm 0.0384 \Longrightarrow$ **(0.3216, 0.3984)**

Yes, the interval agrees with the test. The hypothesized value $p_0 = 0.40$ falls **outside** the 95% confidence interval, and a two-sided test at $\alpha = 0.05$ rejects H_0 exactly when p_0 falls outside the 95% interval. Both procedures say that 0.40 is not a plausible value for p.

Worth remembering. Parts (b) and (e) use *different* standard errors on purpose. A test assumes H_0 and uses p_0; an interval assumes nothing and uses $\hat{p}$. Knowing which one you are doing tells you which to use.

Carrying this into the final

Five ideas from this exam reappear on the final. If you review nothing else, review these.

1. A test and a confidence interval always agree. A two-sided test at level α rejects H_0 exactly when the hypothesized value falls outside the $(1 - \alpha)$ confidence interval. You saw this in Q9: the test rejected $p_0 = 0.40$, and 0.40 sat outside (0.3216, 0.3984). If your test and your interval ever disagree, you have made an arithmetic error — this is a free way to check your own work.

2. One tail or two. The z and t tables give **one-tail** areas. For a two-sided alternative ($\neq$), double the table entry; for a one-sided alternative ($<$ or $>$), do not. This decided Q1(7), Q8(e) and Q9(c), and forgetting to double was among the most common errors on the exam.

3. When the CLT rescues you — and when it does not. With n = 100 (Q2) the CLT makes $\bar{X}$ approximately normal no matter what the population looks like. With n = 16 (Q6) it does not, so the procedure needs the *population* to be approximately normal. The dividing line in this course is roughly $n \geq 30$. Always ask which situation you are in before choosing a procedure.

4. Which standard error. This trips people up more than any formula does:

- Mean, σ known $\rightarrow$ z, with $SE = \sigma/\sqrt{n}$
- Mean, σ unknown $\rightarrow$ t with $df = n - 1$, and $SE = s/\sqrt{n}$
- Proportion, **test** $\rightarrow$ SE uses the hypothesized p_0
- Proportion, **interval** $\rightarrow$ SE uses the observed $\hat{p}$

5. Hypotheses, and what a conclusion may claim. Hypotheses are statements about a *parameter* (μ, p), never about a statistic ($\bar{x}$, $\hat{p}$). The researcher's claim goes in H_a, because the logic of a test is one of disproof: we assume H_0 and look for evidence against it. So rejecting H_0 is evidence *for* the research claim, while failing to reject is never proof that H_0 is true — and no test ever "proves" anything.

One more habit worth keeping: sample sizes always round **up**, and when no prior estimate of p is available, p = 0.5 is the conservative choice because it maximizes pq.