

Exam 2 Practice

Stat 515 Name: _____

Instructor: Ho, Yen-Yi

This is a 75-minute, closed book, closed note exam. You may use a non-programmable, non-graphing calculator. A formula sheet and normal, t, and chi-square tables should accompany this exam. Show all work for full credit.

Answer all twelve questions.

1. (8 pts) Definitions.

a) Define what is meant by "the significance level (α -level)" of a hypothesis test.

b) Define what is meant by "Type I error."

2. (10 pts) Quality control.

A manufacturer knows that 8% of its light bulbs are defective. Use the normal approximation to the binomial to approximate the probability that more than 25 bulbs are defective in a shipment of 250. Do not apply a continuity correction. You may leave your answer in the form of a probability statement about a standard normal (for example $P(Z > 1.5)$).

3. (10 pts) Normal probability (Q–Q) plot.

A quality engineer records the fill volume of 30 randomly selected bottles and constructs a normal probability (Q–Q) plot of the data before carrying out any inference.

a) (3 pts) What is plotted on each axis of a normal probability plot, and what pattern should the points follow if the sample is consistent with a normally distributed population?

b) (4 pts) Instead, the plot bends away from a straight line: the smallest observations fall slightly above the line and the largest observations fall far above it, producing an upward curve. What does this pattern tell you about the shape of the underlying distribution?

c) (3 pts) Based on the plot described in (b), the engineer still wants a confidence interval for the population variance σ^2 . Is that procedure appropriate here? Explain in one or two sentences.

4. (8 pts) Sampling concepts.

a) State the Central Limit Theorem. Under what conditions is n generally large enough for the CLT to apply?

5. (10 pts) Confidence interval for variance.

a) What assumption is required to construct a valid confidence interval for a population variance σ^2 ?

b) Is the confidence interval for the population variance robust against violations of that assumption? Explain briefly.

c) In a few sentences, explain what it means for a statistical procedure to be "robust against non-normality." Contrast this with a procedure that is NOT robust.

6. (10 pts) Interpreting a confidence interval.

A nutritionist studied the average daily sodium intake (in mg) of adults in a city. A sample of 40 adults yielded a 95% confidence interval of (2,104 mg, 2,896 mg). Identify the error in each of the following interpretations, and briefly explain why it is an error.

a) "We are 95% confident that the average daily sodium intake of the 40 adults sampled is between 2,104 mg and 2,896 mg."

b) "There is a 95% probability that the true population mean daily sodium intake is between 2,104 mg and 2,896 mg."

7. (10 pts) Sleep habits survey.

A university health center surveyed 980 undergraduate students and found that 412 reported getting fewer than 7 hours of sleep on a typical night. Construct a 95% confidence interval for the true proportion of all undergraduates at this university who get fewer than 7 hours of sleep per night.

8. (10 pts) Sample size determination.

a) A researcher wants to estimate the mean commute time (in minutes) for employees at a large company to within ± 2 minutes with 95% confidence. A pilot study suggests the standard deviation is approximately 10 minutes. What sample size is needed?

b) If instead the researcher had no prior estimate of σ , and wanted to estimate the proportion of employees who commute more than 30 minutes to within ± 0.04 with 90% confidence, what sample size would be needed?

9. (12 pts) Setting up hypotheses.

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

a) A coffee company claims its espresso pods contain an average of at least 75 mg of caffeine per pod. A consumer watchdog group suspects the true mean is less than 75 mg and plans to test the claim.

b) A hospital administrator wants to determine whether the average length of stay (in days) differs between two units — a standard care unit and an enhanced recovery unit. Patients are randomly assigned to one of the two units.

c) A packaging machine is set to fill bottles with exactly 500 mL. A quality inspector suspects the machine is not filling correctly (either over- or under-filling) and takes a random sample of bottles to test.

10. (14 pts) Hypothesis test.

A physical therapist believes that the mean resting heart rate of patients enrolled in a cardiac rehabilitation program is different from the national average of 72 bpm. A random sample of 18 patients gave a mean of 76.4 bpm and a standard deviation of 9.2 bpm. Assume the population is approximately normal.

- a) State the null and alternate hypotheses.
- b) Compute the test statistic.
- c) Find the rejection region using $\alpha = 0.05$.
- d) State your conclusion in the context of the problem.

11. (8 pts) p-value and decision.

Consider testing $H_0: \mu = 50$ vs. $H_a: \mu \neq 50$. A sample of size 22 gave $\bar{x} = 47.3$ and $s = 6.8$. The resulting two-sided p-value is 0.0761.

- a) (Circle the correct response.) For an α -level of 0.05 we would
fail to reject / reject the null hypothesis.
- b) (Circle the correct response.) For an α -level of 0.10 we would
fail to reject / reject the null hypothesis.
- c) What would the p-value have been if we were instead testing $H_0: \mu = 50$ vs. $H_a: \mu < 50$?
Briefly explain your reasoning.

12. (10 pts)

Answer the following questions (each counts as 5 pts).

- a) A colleague constructs a confidence interval for a proportion using $\hat{p} = X/n$ directly, where $n = 12$ and $X = 1$. Give one reason why the standard formula $\hat{p} \pm z^*(\sqrt{(\hat{p}(1-\hat{p})/n)})$ may be unreliable here, even though they calculated it correctly.

b) Explain why Wilson's adjusted interval (which adds two successes and two failures to the data before computing the interval) tends to perform better than the standard formula when $\hat{p}$ is close to 0 or 1, or when n is small.