Stat 50 - Elementary Statistics — Spring 2019

Exam 4

Ch 8, 9

Dr. Jorge Basilio

Pasadena City College

June_4

gbasilio@pasadena.edu

Honesty Pledge

On my honor, by printing and signing my name below, I vow to neither receive nor given any unauthorized assistance on this examination:

NAME (PRINT): Solutions _____ SIGNATURE: _____

Directions

- YOU ARE ALLOWED TO USE A CALCULATOR ON THIS EXAM. (Ti83/Ti83+/Ti84/Ti84+/Ti84+CE-T, or scientific calculator)
- You have 80 minutes to complete this exam.
- The exam totals **108 points** but will be graded out of 100 only. (So it is possible to get 108% on this exam ☺).
- There are 7 problems, many of them with multiple parts.
- Place all of your belongings in the front of the classroom and I will assign you a seat. Bring with you your writing utensils.
- Cell phones must be turned off and put away in with your items in the front of the classroom.
- Handwriting should be neat and legible. If I cannot read your writing, zero points will be given.
- Make sure to ALWAYS SHOW YOUR WORK; you will not receive any partial credits unless work is clearly shown. *If in doubt, ask for clarification.*
- Leave answers in exact form (as simplified as possible), unless told otherwise.
- Put a box around your final answer where applicable.
- **PLEASE INCLUDE UNITS** where applicable
- **PLEASE CHECK YOUR WORK!!!**
- If you need extra space, there is extra space on the back of the cover page and clearly indicate that you are continuing your work there in the original location.
- If you finish early, you may take a break but you must come back to class by 2:45 and we will have class.
- I will take attendance at the end of class
- Some questions contain multiple-parts which you must do individually and the parts are denoted by (a), (b), (c), etc. Some questions are multiple-choice and the choices are denoted with (A), (B), (C), (D), and (E).

Score	Grade

This page is intentionally blank. It may be used for scratch paper. If you wish for me to grade your work on this page, please (i) label the problem you are working on, (ii) box your answer, (iii) indicate in the original problem's location that you will continue your work on this page.

Problem 1: 10 pts (2 pts each)

TRUE or **FALSE** (please spell out/write the entire word for credit). *(No work needed)*

(a) TRUE — In a t-distribution when the degrees of freedom df increases, distributions becomes more like the z- distribution.

(b) TRUE — The null hypothesis is a claim about a parameter assumed true until there is enough statistical evidence to reject it.

(c) FALSE — A type ~~I~~ II error is made by failing to reject a false null hypothesis.

(d) TRUE — All other things being equal, choosing a smaller value of α will increase the probability of making a type II error.

(e) TRUE — Two samples are said to be independent when the selection of the individuals in one sample has no bearing on the selection of those in the other sample.

Problem 2: 24 pts (2 pts each blank)

Fill in the blanks:

Eg $p > 0.5$

(a) In a right-tailed hypothesis test, the sign in the alternate hypothesis is greater than (>) .

(b) If we get a p-value of 0.015 in a hypothesis test with a significance level of $\alpha = 0.02$, then we REJECT the null hypothesis. $P < \alpha$ low, null go

(c) In a hypothesis test, the p-value is the probability of selecting a sample whose test statistic is at least as extreme as the observed test statistic that we got, assuming the null hypothesis is true .

(d) When converting two samples into one sample using the difference, d, of matched-pairs, we use the following notation:

Point Estimate: $\bar{d} = \bar{x}_1 - \bar{x}_2$

True mean difference: μ_d Sample mean standard deviation: s_d

Null Hypothesis: $\mu_d = 0$ Alternate Hypothesis for a left-tailed test: $\mu_d < 0$

differences are

Requirements are: SRS (simple random samples) and normal (t_y) distributed or $n > 30$

Problem 3: 4 pts (2 pts each)

Multiple-choice. Select the correct answer:

(a) A confidence interval is an interval that is used to estimate a:

A

(A) population parameter based on information from a sample.

(B) population parameter based on information from a population.

(C) sample statistic based on the information from a sample.

(D) sample statistic based on the information from a population.

(E) sample parameter based on the information from a population statistic.

(b) When constructing a 95% confidence interval of the average yearly salary difference between males and females at a specific company (average female salary minus average male salary), our calculator gives us the interval $(-18240, -1115)$. This can be interpreted as $\bar{x}_1$ $\bar{x}_2$ $\bar{d} = \bar{x}_1 - \bar{x}_2 < 0 \rightarrow \bar{x}_1 < \bar{x}_2 \rightarrow$ female make less males

B

(A) we are 95% confident that the females make between \$1,115 and \$18,240 more than males per year at this company.

(B) we are 95% confident that the females make between \$1,115 and \$18,240 less than males per year at this company.

(C) we are 95% confident that the females make between -\$1,115 and -\$18,240 per year at this company.

(D) we are 95% confident that the males make between \$1,115 and \$18,240 per year at this company.

(E) we are 95% confident that the males make between \$1,115 and \$18,240 per year at this company.

Problem 4: 10 pts (2 pts each)

Match the test with the scenario:
Write the letter of the hypothesis test or confidence interval from the list on the right that would be used for the situation on the left.

E Cole wants to test → HT if there is a difference in moisture content among turkeys that have been baked in the oven compared to those that have been deep fried. He takes two independent samples and measures their moisture level. (difference, two μ, HT)

I Nadine wants to estimate → CI the average speed difference of cars on Highway 1 (between Santa Cruz and San Francisco) and Highway 17. She selects two simple random, independent samples and measures their speeds. (difference, two μ, CI)

B Irene wants to test → HT if the percentage of students that own a smartphone is different than the percentage of faculty that own a smartphone at Pasadena City College. (two p, HT)

C Ethan wants to test → HT if the average height of female volleyball players in the U.S. is more than 78 inches. (single μ, HT)

F Jesus wants to estimate → CI the percentage of statistics students at Pasadena City College who took Math 150. (single p, CI)

(A) Test about p

(B) Test about $p_1 - p_2$

(C) Test about μ

(D) Test about μ_d

(E) Test about $\mu_1 - \mu_2$

(F) Confidence Interval about p

(G) Confidence Interval about $p_1 - p_2$

(H) Confidence Interval about μ

(I) Confidence Interval about $\mu_1 - \mu_2$

(J) Confidence Interval about μ_d

HT

n = 310
x = 170

Problem 5: 20 pts

→ population proportion $p_0 = 0.5$

Students conducted an experiment to determine whether the Belgium-minted Euro coin was equally likely to land heads up or tails up. Coins were spun on a smooth surface, and in 310 spins, 170 landed with the heads side up. → proportion of sample $\hat{p}$

claim:

Should the students interpret this result as convincing evidence that the proportion of the time the coin would land heads up is not 0.5? Test the relevant hypotheses using a significance level of $\alpha = 0.01$. *Round final answers to three decimal places where appropriate*

(3 pt) (a) Are the **requirements** met for the hypothesis test? Why or why not?

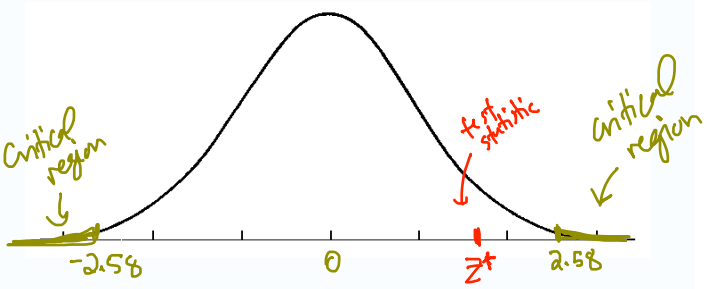
single proportion HT — here: spining coin fixed 310 times.

- SRR ✓
- requirements for binomial: ① fixed trials ✓ ② independent ✓ ③ "bi" two outcomes ✓ ④ probability of trial success constant ✓

(2 pt) (b) State the **null** and **alternative hypotheses**:

$H_0: p = 0.5$ (null is coin is fair)

$H_1: p \neq 0.5$ (Two tail test) (alt: coin is not fair)



(1 pt) (c) State the **level of significance**: $\alpha = 0.01$

(3 pt) (d) Find the **test statistic**:

$$z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \qquad z^* = \frac{(0.548387 - 0.5)}{\sqrt{\frac{0.5 * 0.5}{310}}} = 1.70388\ldots$$

$z^* = 1.70$

z-scores round to 2 decimal places

$\hat{p} = \frac{x}{n} = \frac{170}{310} = 0.548387$. $p_0 = 0.5$ $q_0 = 0.5$ $n = 310$

(4 pt) (e) Use the **Critical Value Method** to find the critical value:

Label the critical value and the shade the critical region in the distribution provided. Also, label the test statistic in your graph. z^*

Critical Value $z_{\frac{\alpha}{2}} = \text{invNorm}(\frac{\alpha}{2}, 0, 1, \text{Right})$ $z_{\frac{\alpha}{2}} = \text{invNorm}(0.005, 0, 1, \text{Right}) = 2.575829\ldots$

$z_{\frac{\alpha}{2}} = 2.58$

$\alpha = 0.01$

$\frac{\alpha}{2} = 0.005$

(2 pt) (f) Make a **decision**:

z^* is outside critical region → Fail to Reject H_0

(5 pt) (g) $(M \to E)$ State your **conclusion**:

"There is not enough statistical evidence to support the claim that the true proportion of the time the Belgian-minted Euro coin is not 0.5"

Problem 6: 20 pts

A researcher wanted to estimate the effect a new drug would have on systolic blood pressure. The following table gives the systolic blood pressure (in $mm\ Hg$) of seven adults before taking this drug, and after having taken this drug for 2 months.

x_1	**Before**	210	180	195	220	231	199	224
x_2	**After**	195	178	186	223	218	195	224
$d = x_1 - x_2$	difference	15	2	9	-3	13	4	0

1-VAR stats

$\bar{d} = 5.714\,28\ldots$ $\bar{d} = 5.7143$

$s_d = 6.7753\ldots$ $s_d = 6.7753$

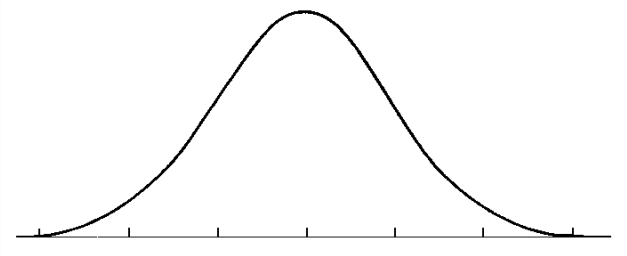
We will assume that the population of paired differences is normally distributed.

Construct a 90% confidence interval for the difference in systolic blood pressure before and after taking this drug. *Round final answers to three decimal places where appropriate*

(4 pt) (a) Identify the **point estimate**:

Matched-Pairs HT

$\bar{d} = \bar{x}_1 - \bar{x}_2 =$ 5.714 mm Hg per person



(4 pt) (b) Determine the **critical value**:

$CL = 0.9$

$\alpha = 1 - CL = 0.1$

$\frac{\alpha}{2} = 0.05$

$1 - \frac{\alpha}{2}$ df $\quad df = 6$

$t_{\frac{\alpha}{2}} = \text{invT}(1 - 0.05, 6)$

$t_{\frac{\alpha}{2}} = 1.943\,18\ldots \rightarrow t_{\frac{\alpha}{2}} = 1.94$

(3 pt) (c) Find the **margin of error**:

$E = t_{\frac{\alpha}{2}} \cdot \frac{s_d}{\sqrt{n}} \qquad E = \frac{(1.94) * 6.7753}{\sqrt{7}} = 4.967\,996\,0\ldots$

$E =$ 4.968 mg Hg per person

(4 pt) (d) Contsruct the **confidence interval**:

$CI: (\bar{d} - E, \bar{d} + E)$

$\bar{d} - E = 0.746$

$\bar{d} + E = 10.682$

$CI:$ (0.746 mg Hg per person, 10.682 mg Hg per person)

(5 pt) (e) $(M \rightarrow E)$ **Interpretation of CI**:

"We are 90% confident that the true difference in systolic blood pressure before and after taking this drug is between 0.746 mg Hg and 10.682 mg Hg."

Problem 7: 20 pts

A Fair Isaac Corporation (FICO) score is used by credit agencies (such as mortgage companies and banks) to assess the creditworthiness of individuals. Values range from 300 to 850, with a FICO score over 700 considered to be a quality credit risk. According to Fair Isaac Corporation, the mean FICO score is 703.

A credit analyst wondered whether high-income individuals (incomes in excess of $100,000 per year) had higher credit scores. He obtained a random sample of 40 high-income individuals and found the sample mean credit score to be 735 with a standard deviation of 80.

Test the claim that high-income individuals have higher FICO scores. *Round final answers to three decimal places where appropriate*

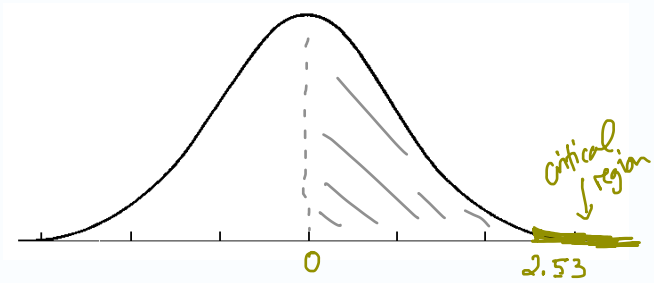
(3 pt) (a) Are the **requirements** met for the hypothesis test? Why or why not?

(2 pt) (b) State the **null** and **alternative hypotheses**:

$H_0: \mu = 703$ average FICO score

$H_1: \mu > 703$ average FICO score

(RIGHT-tailed test)



(1 pt) (c) State the **level of significance**:

none given, assume $\alpha = 0.05$

(3 pt) (d) Find the **test statistic**:

$$t^* = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \qquad t^* = \frac{(735 - 703)}{\left(\frac{80}{\sqrt{40}}\right)} = 2.52982\ldots$$

$\bar{x} = 735$ $\mu_0 = 703$ $n = 40$ $s = 80$

$t^* = 2.53$

(4 pt) (e) Use the P-**Value Method** to find the P-value:

Label and shade the critical region in the distribution provided.

$P = P(t > 2.53) = 1 - P(0 < t < 2.53)$ (a = 0, b = 2.53, df = 39)

$= 1 - \text{tcdf}(0, 2.53, 39)$

$P = 0.5077778\ldots \rightarrow P = 0.508$

(2 pt) (f) Make a **decision**:

$P = 0.508$ $\alpha = 0.05$ $P > \alpha$

P high, null will fly → Fail to Reject H_0

(5 pt) (g) $(M \rightarrow E)$ State your **conclusion**:

"There is not enough statistical evidence to support the claim that high-income individuals have higher FICO scores."

Post Exam Survey

Now that you have finished the exam, please take a few minutes to reflect on how you prepared for the exam and how you think you did. Then answer these questions.

1. When taking the exam I felt
   (a) Rushed. I wanted more time.
   (b) Relaxed. I had enough time.
   (c) Amazed. I had tons of extra time.
2. The week before the test I did all my homework on time: YES NO
3. The week before the test, in addition to the homework I followed a study plan. YES NO
   (a) I think this helped: YES NO
4. The day before the test I spend _____ hours studying and reviewing.
   (a) I think that was enough time: YES NO
5. The night before the test:
   (a) I stayed up very late cramming for the test
   (b) I stayed up very late, but I wasn't doing math
   (c) I didn't need to cram because I was prepared
   (d) I got a good night's sleep so my brain would function well.
6. I think I got the following grade on this test: _____
7. Strategies that worked well for me were (please elaborate):
8. Next time I will do an even better job preparing for the test by: