STATISTICS

SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II grade—25

6. Researchers want to see whether training increases the capability of people to correctly predict outcomes of coin tosses. Each of twenty people is asked to predict the outcome (heads or tails) of 100 independent tosses of a fair coin. After training, they are retested with a new set of 100 tosses. (All 40 sets of 100 tosses are independently generated.) Since the coin is fair, the probability of a correct guess by chance is 0.5 on each toss. The numbers correct for each of the 20 people were as follows.

Score Before Training (number correct)	Score After Training (number correct)
46	61
48	62
50	53
54	46
54	50
54	52
54	53
54	59
54	60
54	61
55	55
56	59
57	55
58	50
58	56
61	58
61	64
63	57
64	61
65	54
Sum 1,120	Sum 1,126

To answer the following questions, you may want to enter these data into your calculator. As a check that you have entered the data correctly, the sum of the first column is 1,120 and the sum of the second column is 1,126.

a. Do the data suggest that after training people can correctly predict coin toss outcomes better than the 50 percent expected by chance guessing alone?

   Give appropriate statistical evidence to support your conclusion.

GO ON TO THE NEXT PAGE

Question 6

Points:

Each part (a), (b), and (c) is essentially correct, partially correct, or incorrect.

4 Complete Response
Essentially correct on all 3 parts

3 Substantial Response
Essentially correct on 2 parts,
OR
Essentially correct on part (a) or (b) and partially correct on other 2 parts

2 Developing Response
Essentially correct on part (c) and partially correct on either part (a) or (b)
OR
Partially correct on at least 2 parts
OR
Essentially correct on part (a) or (b)

1 Minimal Response
Essentially correct on part (c), but incorrect on the other two parts.
OR
Partially correct on any 1 part
OR
Holistically partially correct overall (e.g. recognizes need for matched pairs in part (b))

Key Elements in Solutions:

a. This question focuses on just the score *after training* to assess whether performance is better than 50%.
b. This question focuses on assessing *change*.
c. This question focuses on correlation or regression slope to assess *association*.

Complete Solutions

a. Let MU_A = mean number of correct responses after training

H_0: $MU_A = 50$ (or <=) H_a: $MU_A > 50$

One-sample t test Test statistic: $$t = \frac{\bar{x}_A - \mu_0}{s_A / \sqrt{n}}$$

Question 6 (cont.)

Assumption: Normal population distribution.

Checking Assumption for t test: Boxplot, dot plot, stem and leaf plot or histogram does not show any outliers or extreme skewness. Or normal probability plot looks OK. (See attached plots.)

$\bar{x}_A = 56.3$

$s_A = 4.725$

n = 20

$$t = \frac{56.3 - 50}{4.725 / \sqrt{20}} = 5.963$$

df = 19

P-value = .0000049

For any reasonable choice of ALPHA, reject H_o. There is convincing evidence that the mean number of correct responses after training is higher than 50, the value expected by chance alone.

(For a rejection region approach, Rejection Region boundaries are at 1.33 for ALPHA = .1, 1.73 for ALPHA = .05, and 2.54 for ALPHA= .01)

Other possible approaches scored as Essentially correct:

- Sign test (1 below 50, 2 at 50, and 17 above 50. P-value = .0001). Doesn't require normality assumption.
- Binomial test or Z test based on the normal approximation to binomial (H_o: p = .5, H_a: p > .5, assumptions: large sample (e.g. np_o>=5, $n(1-p_o)$>=55), ok to combine across individuals, n = 2000, = 1126/2000 = .563, z = 5.63, P-value = .000000008782).
- A chi-square test can be a correct method (x^2=48.72, df=20) but has a minor error in that it assumes a two-sided alternative.

b. The analysis in part (a) does not provide evidence that the training is effective, because it does not compare before and after scores. Data is paired.

Question 6 (cont.)

H_o: MU_D = 0 Where MUD is the mean difference in the score before training and the score after training

H_a: MU_D > 0 (would be < if student defines differences as before-after)

OR

H_o: $MU_B - MU_A = 0$ H_a: $MU_B - MU_A > 0$ (< if student defines differences as before-after)

One-sample t test Test statistic:

$$t = \frac{\bar{x}_D - \mu_0}{s_D / \sqrt{n}}$$

Assumption: Normal difference population

Checking Assumption for t test: Boxplot, dot plot, stem and leaf plot or histogram of sample differences does not show any outliers or extreme skewness. Or normal probability plot of differences looks OK. (See attached graphs.)

$\bar{x}_D = .3$

$s_D = 6.837$

n = 20

$$t = \frac{.3 - 0}{6.837 / \sqrt{20}} = .1962$$

df = 19

P-value = .4233

For any reasonable choice of ALPHA, fail to reject Ho. There is not convincing evidence that training is effective in improving a person's ability to predict coin tossing outcomes.

(For a rejection region approach, Rejection Region boundaries are at 1.33 for ALPHA = .1, 1.73 for ALPHA= .05, and 2.54 for ALPHA = .01)

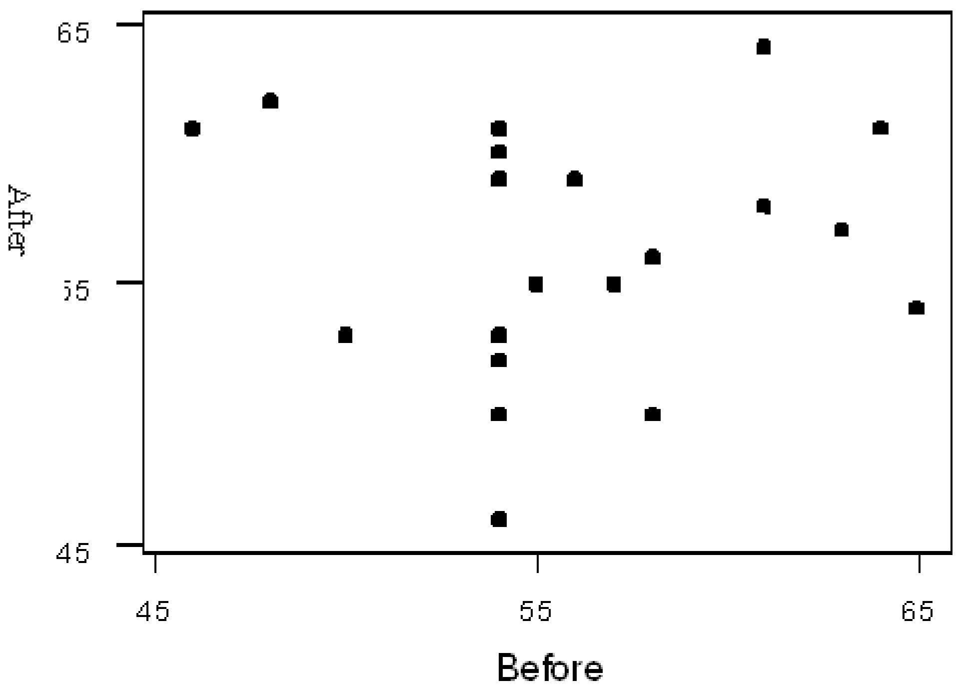
An alternate correct approach is to use the sign test (11 negative differences, 1 zero difference, 8 positive differences, P-value = .6762).

Question 6 (cont.)

Alternate incorrect approaches include using a two independent sample t test (t = .1943, P-value = .4235 for either pooled or unpooled. ff = 38 for pooled, df = 37 for unpooled) or a two independent sample z test for proportions (z = .1912, p-value = .4242). (See definition of partially correct.)

c. The response indicates that knowledge of a person's score before training would not be helpful in predicting after training score and justifies this response in at least one of the following ways:

- A scatter plot of the after scores vs. the before scores shows no pattern indicating a relationship between before and after scores
- The correlation coefficient for before and after scores is very close to 0 (r = .020 or R^2 = .0396%)
- The test for slope in a regression of y = "after score" on x = "before score" does not indicate that the slope is significantly different from 0 (t = .08, P-value = .934).



A complete solution in (a) and (b) contains the following 5 steps

1. Hypotheses are correct.
2. Assumptions are stated and checked.
3. The correct test is used and is either named or formula for the test statistic is given.
4. Test statistic and p-value computed correctly.
5. A correct conclusion is given in context.

Question 6 (cont.)

Parts (a) and (b) are essentially correct if student correctly completes 4 or 5 of the above steps.

Parts (a) and (b) are partially correct if student correctly completes 3 of the above steps.

Parts (a) and (b) are incorrect if only 1 or 2 steps are completely correct.

UNLESS:

- Part (a) is considered Incorrect if the student uses both columns of data in the analysis (e.g. matched pairs, two independent samples, chi-square).
- In part (b), a two using a two independent sample procedure (means or proportions) receives at most Partial credit in part (b). Student must get 3 of remaining 4 steps to receive the Partial credit in (b).
- In parts (a) or (b), if the students manufacture an invalid formula (e.g. mixes t and binomial numerator and denominator) they receive at most Partial credit.

If an error similar to the following is made, student does not receive credit for that step.

- Inconsistent or undefined nonstandard notation.
- Incorrect hypotheses (but if H_a is 2-sided, don't count **both** this & 2-sided p-values as errors).
- States but doesn't check assumptions.
- Plugs wrong numbers into correct formula (e.g. forgetting , uses n = 100, uses n = 20 in one-sample z-test for proportions).
- Uses z instead of t when SIGMA is unknown.
- Calculates two-sided p-value or one-sided in the wrong direction.
- Uses a two-sided confidence interval approach.
- Gives incorrect conclusion or conclusion not stated in context of problem.

Part (c) is essentially correct if student examines the association between before and after scores, instead of the differences.

Part (c) is partially correct if student merely recognizes the need to examine the association (e.g. say no pattern but gives no evidence has examined the plot or correlation coefficient).

Question 6 (cont.)

Part (c) is incorrect if

- Student only refers to results of (a) or (b):
  - Response indicates that since there was no significant difference between before and after, the before score is a useful predictor of the after score (should be about the same.)
  - Response indicates that since there was no significant difference, training doesn't matter, so the before score is not a useful predictor.
- Response is based on the coin flips being fair and independent, rather than on the guesses of the outcomes.
- Response examines association through a chi-square test using data given as two-way table.