5. A polling agency showed the following two statements to a random sample of 1,048 adults in the United States.

> Environment statement: Protection of the environment should be given priority over economic growth.
>
> Economy statement: Economic growth should be given priority over protection of the environment.

The order in which the statements were shown was randomly selected for each person in the sample. After reading the statements, each person was asked to choose the statement that was most consistent with his or her opinion. The results are shown in the table.

	Environment Statement	Economy Statement	No Preference
Percent of sample	58%	37%	5%

(a) Assume the conditions for inference have been met. Construct and interpret a 95 percent confidence interval for the proportion of all adults in the United States who would have chosen the economy statement.

(b) One of the conditions for inference that was met is that the number who chose the economy statement and the number who did not choose the economy statement are both greater than 10. Explain why it is necessary to satisfy that condition.

(c) A suggestion was made to use a two-sample z-interval for a difference between proportions to investigate whether the difference in proportions between adults in the United States who would have chosen the environment statement and adults in the United States who would have chosen the economy statement is statistically significant. Is the two-sample z-interval for a difference between proportions an appropriate procedure to investigate the difference? Justify your answer.

GO ON TO THE NEXT PAGE.

STATISTICS
SECTION II
Part B
Question 6
Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. A newspaper in Germany reported that the more semesters needed to complete an academic program at the university, the greater the starting salary in the first year of a job. The report was based on a study that used a random sample of 24 people who had recently completed an academic program. Information was collected on the number of semesters each person in the sample needed to complete the program and the starting salary, in thousands of euros, for the first year of a job. The data are shown in the scatterplot below.
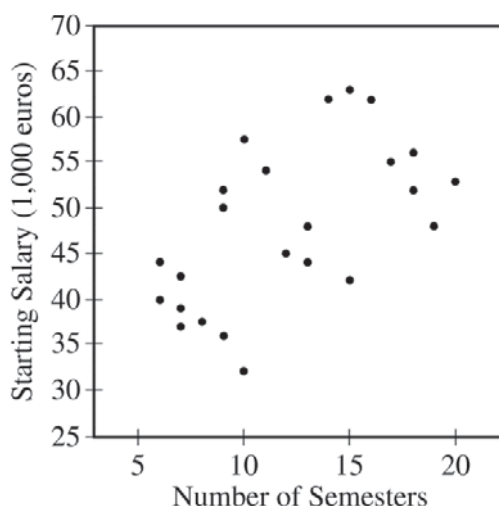
(a) Does the scatterplot support the newspaper report about number of semesters and starting salary? Justify your answer.

The table below shows computer output from a linear regression analysis on the data.

Predictor	Coef	SE Coef	T	P
Constant	34.018	4.455	7.64	0.000
Semesters	1.1594	0.3482	3.33	0.003
S = 7.37702	R-Sq = 33.5%		R-Sq(adj) = 30.5%	

(b) Identify the slope of the least-squares regression line, and interpret the slope in context.

GO ON TO THE NEXT PAGE.

Question 5

<u>Intent of Question</u>

The primary goals of this question were to assess a student's ability to (1) construct and interpret a confidence interval for a population proportion; (2) explain why one of the conditions for inference is necessary; and (3) explain why a suggested procedure for constructing a confidence interval is incorrect.

<u>Solution</u>

Part (a):

The appropriate procedure is a one-sample z-interval for a population proportion. The problem stated the conditions for inference have been met, so they do not need to be checked. A 95 percent confidence interval for the population proportion is given as $\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1 - \hat{p})}{n}}$, which is

$$0.37 \pm 1.96 \sqrt{\dfrac{(0.37)(0.63)}{1,048}} \approx 0.37 \pm 0.03 = (0.34, 0.40).$$

We are 95 percent confident that the population proportion of all adults in the U.S. who would have chosen the economy statement is between 0.34 and 0.40.

Part (b):

The condition is necessary because the formula for the confidence interval relies on the fact that the binomial distribution can be approximated by a normal distribution which then results in the sampling distribution of $\hat{p}$ being approximately normal. The approximation does not work well unless both $n\hat{p}$ and $n(1 - \hat{p})$ are at least 10.

Part (c):

The suggested procedure is not appropriate because one of the requirements for using a two-sample z-interval for a difference between proportions is that the two proportions are based on two independent samples. In the situation described the two proportions come from a single sample and thus are not independent.

<u>Scoring</u>

This question is scored in three sections. Section 1 consists of the mechanics of the confidence interval in part (a), section 2 consists of interpreting the confidence interval in part (a), and section 3 consists of parts (b) and (c).

Section 1 is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. States the correct procedure by name or formula
2. Calculates the confidence interval

Partially correct (P) if the response includes only one of the two components.

Incorrect (I) if the response does not meet the criteria for E or P.

Notes:
- A formula with correct values is sufficient for component 2.
- Component 2 can never be satisfied if the response contains an incorrect formula.

Section 2 is scored as follows:

Essentially correct (E) if the response includes the following four components for the interval interpretation:
1. Estimates a proportion
2. Infers about the population
3. States 95 percent confidence
4. Includes context

Partially correct (P) if the response satisfies components 1 and 2 *AND* satisfies only one of components 3 or 4;
    OR
if the response gives a correct interpretation of the confidence *level* in context without interpreting the specific interval.

Incorrect (I) if the response does not meet the criteria for E or P.

Notes:
- Any indication of an inference to the adults sampled rather than the population does not satisfy component 2 and is scored as I.
- Stating values that are unrealistic as proportions (including blanks) in the interpretation lowers the score by one level (from E to P or from P to I).
- When both the interpretation of the interval and the level are given, only the interpretation of the interval is scored.

Section 3 is scored as follows:

Essentially correct (E) if the response includes the following two components:
1. Part (b) states that the condition implies the sampling distribution of $\hat{p}$ is approximately normal *OR* the normal approximation to the binomial distribution is appropriate.
2. Part (c) indicates the procedure is not appropriate because the two proportions come from a single sample (dependent) rather than two (independent) samples.

Partially correct (P) if the response includes only one of the two components.

Incorrect (I) if the response does not meet the criteria for E or P.

Notes:
- Referring to the sampling distribution needing to be normal without explicitly stating $\hat{p}$ satisfies component 1.
- Referring to normal approximation must specify "to binomial" to satisfy component 1.
- Discussing the two-sample z-interval versus the two-proportion z-interval is extraneous information.

- In part (c) component 2 can never be satisfied if a procedure that is not possible for one sample of categorical data is suggested.

4   Complete Response

      Three parts essentially correct

3   Substantial Response

      Two parts essentially correct and one part partially correct

2   Developing Response

      Two parts essentially correct and no parts partially correct

OR

      One part essentially correct and one or two parts partially correct

OR

      Three parts partially correct

1   Minimal Response

      One part essentially correct

OR

      No parts essentially correct and two parts partially correct