INTRODUCTORY STATISTICS

Questions	290
Field	Statistics
Target Audience	Science Students
Target Level	First or Second-year Undergraduate
Topics	▪ Introduction to Statistics ▪ Descriptive Statistics ▪ Basic Probability ▪ Discrete Random Variables ▪ Continuous Random Variables ▪ Sampling Distributions ▪ Inference for Means ▪ Inference for Proportions ▪ Inference for Variances ▪ Chi-square Tests for Count Data ▪ One-Way ANOVA ▪ Simple Linear Regression and Correlation

Outline

The material in this module is designed to cover a single-semester course in statistics for science students at the second-year university level. The questions are designed to span the topics listed above, allowing for practice, homework or testing throughout the semester. The questions are mainly of an applied nature and do not delve very deeply into the underlying mathematical theory.

For each topic there is a variety of calculation questions with numeric responses, as well as True/False questions that test understanding of fundamental concepts. Calculation questions have randomly generated values and many conceptual questions are composed of True/False statements randomly selected from a pool of relevant statements.

Calculation questions have full solutions provided, including illustrative diagrams where appropriate. Information fields are included on all questions indicating topics and difficulty level.

The module was first implemented in Winter 2012 at the University of Guelph and has seen a subsequent round of updates. In the Winter 2012 semester, the questions were given as (short) weekly quizzes, worth a small portion of the overall grade. After the graded quiz deadline, they were also released as practice quizzes, giving the students an opportunity to master the material.

Sample quizzes are provided, spanning the course material.

Since most questions are numeric in nature, the included Maple TA Syntax Sheet may be of limited use, but it is provided in the event that you may wish to add more algebraic style questions.

Contact Info: Jeremy Balka, jbalka@uoguelph.ca, 519-824-4120 x54481.

MAPLE TA SYNTAX SHEET

Expression	Entry Syntax
$x \cdot y$	x*y
$\frac{x}{y}$	x/y
x^y	x^y
$\frac{a}{b \cdot c}$	a/(b*c) (although it will accept a/b/c)
$\sqrt{x}$	sqrt(x) or x^(1/2) (do not use x^0.5)
$x^{\frac{2}{3}} = \sqrt[3]{x}$	x^(2/3)
$\lvert x \rvert$	abs(x)
$\ln(x)$	ln(x)
$\log_n(x)$	log[n](x)
e^x	exp(x)
e	e or exp(1)
π	pi or Pi
∞	infinity
$\sin^2(x) = (\sin(x))^2$	sin(x)^2 or (sin(x))^2

Notes

Maple TA likes to make the following substitutions when displaying equations

Simple Form	Maple TA Will Show
sec(x)^2	1+tan(x)^2
csc(x)^2	1+cot(x)^2
sec(x)*tan(x)	sin(x)/cos(x)^2
csc(x)*cot(x)	cos(x)/sin(x)^2

Introductory Statistics
Sample Quizzes

Prepared by Lorna Deeth

Under the Supervision of Professor Jeremy Balka
(jbalka@uoguelph.ca)
Department of Mathematics & Statistics
University of Guelph
Summer 2011

Edited by Aron Pasieka

(aron@aron.ca)

Summer 2012

Table of Contents

Quiz 1: Introductory Concepts

Question 1: *Score 1/1*

Your response	Correct response
Identify each of the following statements as either TRUE or FALSE. a) **True** (20%) A survey is a type of observational study. b) **True** (20%) A response variable is the outcome of the experiment or study. c) **False** (20%) If an explanatory variable and a response variable are correlated, then it can safely be assumed that the explanatory variable causes the response variable. d) **False** (20%) In an experiment, researchers do not impose any conditions on the experimental units. e) **True** (20%) Assuming a large population, in simple random sampling each individual in the population has the same probability of being selected.	Identify each of the following statements as either TRUE or FALSE. a) **True** A survey is a type of observational study. b) **True** A response variable is the outcome of the experiment or study. c) **False** If an explanatory variable and a response variable are correlated, then it can safely be assumed that the explanatory variable causes the response variable. d) **False** In an experiment, researchers do not impose any conditions on the experimental units. e) **True** Assuming a large population, in simple random sampling each individual in the population has the same probability of being selected.

Correct

Comment:

Question 2: *Score 1/1*

All the items matched correctly.

Correct

Match	Your Choice	Correct Choice	
Statistic	1.7 kg	1.7 kg	
Sample	10 dogs	10 dogs	
Parameter	2.1 kg	2.1 kg	
Population	50 dogs	50 dogs	

Comment:

A *population* is the set of all experimental units of interest to the researcher; here, it is the 50 dogs on the low-calorie diet.

A *sample* is a subset of the population, which in this case is the 10 dogs the veterinarian randomly selected.

A *parameter* is a characteristic of the population. Since the population is the 50 dogs, the parameter is the mean weight of these 50 dogs, which is 2.1 kg.

A *statistic* is a characteristic of the sample. Since the sample consisted of 10 randomly selected dogs on the low-calorie diet, the statistic is the mean weight of the dogs in the sample, which is 1.7 kg.

Question 3: *Score 1/1*

Your response	Correct response	
A veterinary epidemiologist wishes to investigate the use of antibiotics on cattle farms in southern Alberta. However, cattle farms in Alberta can be classified as either small, medium, or large farms, depending on how many cattle are present. The epidemiologist decides to randomly sample farms within each classification level. This type of sampling method is best described as: **Stratified Random Sampling** (100%)	A veterinary epidemiologist wishes to investigate the use of antibiotics on cattle farms in southern Alberta. However, cattle farms in Alberta can be classified as either small, medium, or large farms, depending on how many cattle are present. The epidemiologist decides to randomly sample farms within each classification level. This type of sampling method is best described as: **Stratified Random Sampling**	Correct

Comment:

This is an example of *stratified random sampling*, in which the veterinary epidemiologist takes a random sample of experimental units from within each stratum (i.e. each classification level).

Question 4: *Score 1/1*

All the items matched correctly.

Correct

Match	Your Choice	Correct Choice	✔/✘
Statistic	Weight of 50 corn plants.	Weight of 50 corn plants.	✔
Population	All the corn plants.	All the corn plants.	✔
Parameter	Weight of all the corn plants.	Weight of all the corn plants.	✔
Sample	50 corn plants.	50 corn plants.	✔

Comment:

A *population* is the set of all experimental units of interest to the researcher; here, it is all the corn plants.

A *sample* is a subset of the population, which in this case is the 50 corn plants randomly selected by the farmer.

A *parameter* is a characteristic of the population. Since the population is all the corn plants, the parameter is the mean weight of all the corn plants.

A *statistic* is a characteristic of the sample. Since the sample consisted of 50 randomly selected corn plants, the statistic is the mean weight of the 50 corn plants in the sample.

Question 5: *Score 1/1*

Your response	Correct response	
Consider the following random sample of data: -10, -2, -3, 3, 2, -1, 0, -9, -7, 86 a) What is the mean of the sample data? Round your response to at least 2 decimal places. **5.9** (50%) b) If the outlier is removed, what is the mean of the remaining sample data? Round your response to at least 2 decimal places. **-3** (50%)	Consider the following random sample of data: -10, -2, -3, 3, 2, -1, 0, -9, -7, 86 a) What is the mean of the sample data? Round your response to at least 2 decimal places. **5.9** b) If the outlier is removed, what is the mean of the remaining sample data? Round your response to at least 2 decimal places. **-3**	Correct

Comment:

a) To calculate the mean, we use the formula $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. Since there are 10 observations, this becomes

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{(-10 + -2 + \ldots + 86)}{10} = 5.9$$

b) The outlier is the observation that is noticeably different from the other observations. In this case, the outlier is 86, as it is significantly larger than all the other observations. If this outlier is removed, and the mean of the remaining 9 observations is calculated, we get $\bar{x} = \frac{\sum_{i=1}^{9} x_i}{9} = \frac{(-10 + -2 + \ldots + -7)}{9} = -3$

Question 6: *Score 1/1*

Your response	Correct response	
Consider the following random sample of data: 13, 27, 30, 15, 21, 6, 10, 3, 107, 20 a) What is the median of the sample data? Round your response to 2 decimal places. **17.5** (50%) b) If the outlier is removed, what is the median of the remaining sample data? **15** (50%)	Consider the following random sample of data: 13, 27, 30, 15, 21, 6, 10, 3, 107, 20 a) What is the median of the sample data? Round your response to 2 decimal places. **17.5** b) If the outlier is removed, what is the median of the remaining sample data? **15**	Correct

Comment:

a) To find the median, we first need to list the numbers in order from smallest to largest:

3 6 10 13 15 20 21 27 30 107

Since there is an even number of observations, the median value will be the mean of the middle two numbers. Here, the median will be the average of observations 5 and 6, giving us $Median = \frac{15 + 20}{2} = 17.5$.

b) The outlier in this data set is 107, as it is significantly larger than the other observations. If this observation is removed, and the data set is ordered from smallest to largest, we get:

3 6 10 13 15 20 21 27 30

As there are now only 9 observations, the median is simply the middle value, which in this case is 15.

Quiz 2: Basic Probability and Transformations

Question 1: *Score 1/1*

Your response	Correct response	
Consider the following random sample of data: 7, 3, 9, 1, 7, 4, 7, -4, 7, 89 a) What is the variance of the sample data? Round your response to at least 3 decimal places. **727.778** (50%) b) If the outlier is removed, what is the variance of the remaining sample data? Round your response to at least 3 decimal places. **16.528** (50%)	Consider the following random sample of data: 7, 3, 9, 1, 7, 4, 7, -4, 7, 89 a) What is the variance of the sample data? Round your response to at least 3 decimal places. **727.778** b) If the outlier is removed, what is the variance of the remaining sample data? Round your response to at least 3 decimal places. **16.528**	Correct

Comment:

a) To calculate the variance of the data set, we can use the formula $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$. For this data set, the mean is $\bar{x} = 13$; substituting this into the equation for variance, along with the 10 observation values, we get

$$s^2 = \frac{(7-13)^2 + (3-13)^2 + \ldots + (89-13)^2}{10-1} = 727.777778$$

b) In this case, the outlier is 89, as this value is noticeably higher than the other values. Upon removing this value, the mean of the new data set is $\bar{x} = 4.555556$. Using this value, the remaining 9 observations, and the formula for variance from part (a), we get

$$s^2 = \frac{\left((7-4.555556)^2 + (3-4.555556)^2 + \ldots + (7-4.555556)^2\right)}{9-1} = 16.527778$$

Question 2: *Score 1/1*

Your response	Correct response	
Consider the following random sample of data: 6, 10, 4, 4, 1, 2, -6, 2, -6, 82 a) What is the mean of the sample data? Round your response to at least 2 decimal places. **9.90** (50%) b) If the outlier is removed, what is the mean of the remaining sample data? Round your response to at least 2 decimal places. **1.89** (50%)	Consider the following random sample of data: 6, 10, 4, 4, 1, 2, -6, 2, -6, 82 a) What is the mean of the sample data? Round your response to at least 2 decimal places. **9.90** b) If the outlier is removed, what is the mean of the remaining sample data? Round your response to at least 2 decimal places. **1.89**	**Correct**

Comment:

a) To calculate the mean, we use the formula $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$. Since there are 10 observations, this becomes

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{(6 + 10 + \ldots + 82)}{10} = 9.9$$

b) The outlier is the observation that is noticeably different from the other observations. In this case, the outlier is 82, as it is significantly larger than all the other observations. If this outlier is removed, and the mean of the remaining 9 observations is calculated, we get $\bar{x} = \frac{\sum_{i=1}^{9} x_i}{9} = \frac{(6 + 10 + \ldots + -6)}{9} = 1.888889$

Question 3: *Score 1/1*

Your response	Correct response	
Consider the following random sample of data: 6, -1, -1, 2, 2, 6, -10, -10, 1, 89 a) What is the standard deviation of the sample data? Round your response to at least 3 decimal places. **28.864** (50%) b) If the outlier is removed, what is the standard deviation of the remaining sample data? Round your response to at least 3 decimal places. **5.918** (50%)	Consider the following random sample of data: 6, -1, -1, 2, 2, 6, -10, -10, 1, 89 a) What is the standard deviation of the sample data? Round your response to at least 3 decimal places. **28.864** b) If the outlier is removed, what is the standard deviation of the remaining sample data? Round your response to at least 3 decimal places. **5.918**	Correct

Comment:

a) To calculate the standard deviation of the data set, we can use the formula $s = \sqrt{\dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}}$, where the mean of the data set is $\bar{x} = 8.4$. Substituting all 10 observations into the formula gives us

$$s = \sqrt{\frac{(6 - 8.4)^2 + (-1 - 8.4)^2 + \ldots + (89 - 8.4)^2}{10 - 1}} = 28.864434$$

b) In this case, the outlying value is 89, as it is noticeably larger than the other values. Upon removing this observation, we get a new mean of $\bar{x} = -0.555556$. Using this value in the formula for standard deviation seen in part (a), we get

$$s = \sqrt{\frac{(6 - -0.555556)^2 + (-1 - -0.555556)^2 + \ldots + (1 - -0.555556)^2}{9 - 1}} = 5.918427$$

Question 4: *Score 1/1*

Your response	Correct response	
Consider the following random sample of data: 11, 24, 29, 15, 23, 4, 9, 2, 102, 20 a) What is the median of the sample data? Round your response to 2 decimal places. **17.50** (50%) b) If the outlier is removed, what is the median of the remaining sample data? **15** (50%)	Consider the following random sample of data: 11, 24, 29, 15, 23, 4, 9, 2, 102, 20 a) What is the median of the sample data? Round your response to 2 decimal places. **17.50** b) If the outlier is removed, what is the median of the remaining sample data? **15**	✔ **Correct**

Comment:

a) To find the median, we first need to list the numbers in order from smallest to largest:

2 4 9 11 15 20 23 24 29 102

Since there is an even number of observations, the median value will be the mean of the middle two numbers. Here, the median will be the average of observations 5 and 6, giving us $Median = \frac{15 + 20}{2} = 17.5$.

b) The outlier in this data set is 102, as it is significantly larger than the other observations. If this observation is removed, and the data set is ordered from smallest to largest, we get:

2 4 9 11 15 20 23 24 29

As there are now only 9 observations, the median is simply the middle value, which in this case is 15.

Question 5: *Score 1/1*

Your response

The following observations represent the weights, in pounds (lbs), of a sample 6 NHL goalies:

ie 1	Goalie 2	Goalie 3	Goalie 4	Goalie 5	Goalie
	201	208	219	217	190

Some summary statistics are: $\bar{x} = 201.666667$, and $s = 16.872068$.

Before going on the ice, each goalie must put on approximately 12.76 kilograms (kg) of equipment.

If 1 lbs = 0.45359 kg, then:

a) What is the mean total weight (body weight plus equipment), in kilograms, of the goalies by the time they reach the ice?

Round your answer to at least 3 decimal places.

104.234 (50%)

b) What is the variance in total weight (in kilograms) of the goalies by the time they reach the ice?

Round your answer to at least 3 decimal places.

58.568 (50%)

Correct response

The following observations represent the weights, in pounds (lbs), of a sample 6 NHL goalies:

ie 1	Goalie 2	Goalie 3	Goalie 4	Goalie 5	Goalie
	201	208	219	217	190

Some summary statistics are: $\bar{x} = 201.666667$, and $s = 16.872068$.

Before going on the ice, each goalie must put on approximately 12.76 kilograms (kg) of equipment.

If 1 lbs = 0.45359 kg, then:

a) What is the mean total weight (body weight plus equipment), in kilograms, of the goalies by the time they reach the ice?

Round your answer to at least 3 decimal places.

104.234

b) What is the variance in total weight (in kilograms) of the goalies by the time they reach the ice?

Round your answer to at least 3 decimal places.

58.568

Correct

Comment:

a) To find the mean total weight of the goalies in kilograms, we can use the linear transformation equation $\bar{x}^* = a + b\bar{x}$. Because the average weight is given in pounds, we first need to convert this to kilograms, and then add on the weight of the equipment (which is already given in kilograms). Using the conversion from pounds to kilograms, we get the linear transformation

$\bar{x}_{kg} = 12.76 + 0.45359 \cdot (201.666667) = 104.233983$.

b) Recall that the addition of a constant does not affect measures of variability. Therefore, to find the variance of the weight of the goalies, in kilograms, we only need to consider the multiplier of 0.45359. Using the formula $s^2_{x^*} = b^2 s^2_x$, we get

$s^2_{kg} = 0.45359^2 \cdot 16.872068^2 = 58.568426$

Question 6: *Score 1/1*

Consider a data set with variance σ^2.

What is the effect on the variance if each observation in the data set is multiplied by 6, then increased by 5?

Correct

Your Answer: The variance is multiplied by 36.

Correct Answer: The variance is multiplied by 36.

Comment: Recall that the addition of a constant does not affect measures of variability. Therefore, the variance is only affected by the multiplier, which in this case is 6. Using the formula $s^2_{x^*} = b^2 \cdot s^2_x$, we get $s^2_{x^*} = 6^2 \cdot s^2_x = 36 \cdot s^2_x$. Hence, the variance is multiplied by 36.

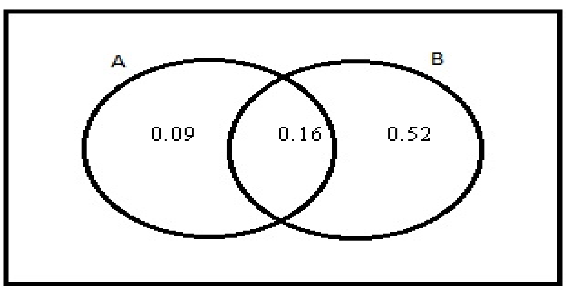
Question 7: *Score 1/1*

Your response	Correct response
Given that $P(A) = 0.25$, $P(B) = 0.68$ and the $P(A \cap B) = 0.16$, what is $P(A^c \cap B)$? Enter your answer to 2 decimal places. $P(A^c \cap B) =$ **0.52** (100%)	Given that $P(A) = 0.25$, $P(B) = 0.68$ and the $P(A \cap B) = 0.16$, what is $P(A^c \cap B)$? Enter your answer to 2 decimal places. $P(A^c \cap B) =$ **0.52**

Correct

Comment:

To find $P(A^c \cap B)$, it is easiest to display the given probabilities in a Venn diagram:



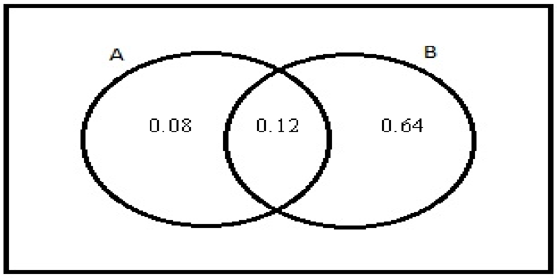
From the Venn diagram, it can be seen that $P(A^c \cap B)$, the area that is **not** in *A* but is in *B*, is 0.52.

Question 8: *Score 1/1*

Your response	Correct response	
Given that $P(A) = 0.20$, $P(B) = 0.76$ and the $P(A \cap B) = 0.12$, what is $P(A \cap B^c)$? Enter your answer to 2 decimal places. $P(A \cap B^c) =$ **0.08** (100%)	Given that $P(A) = 0.20$, $P(B) = 0.76$ and the $P(A \cap B) = 0.12$, what is $P(A \cap B^c)$? Enter your answer to 2 decimal places. $P(A \cap B^c) =$ **0.08**	Correct

Comment:

To find $P(A \cap B^c)$, it is easiest to display the given probabilities in a Venn diagram:

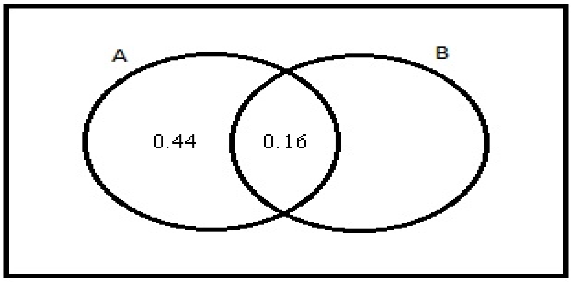


From the Venn diagram, it can be seen that $P(A \cap B^c)$, the area that is in *A* but **not** in *B*, is 0.08.

Question 9: *Score 1/1*

Your response

Consider the following Venn Diagram:



a) Given the additional information that $P(A \cup B)$ =0.89, what is $P(A \mid B)$?

Round your response to at least 3 decimal places.

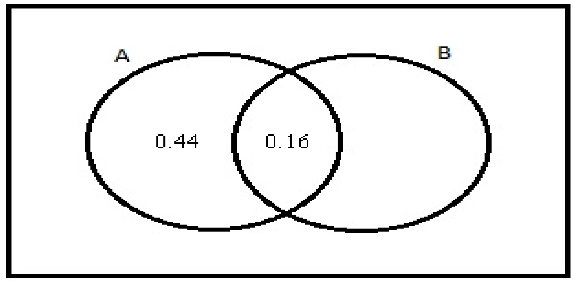
$P(A \mid B) =$ **0.356** (50%)

b) Are events *A* and *B* independent?

No (50%)

Correct response

Consider the following Venn Diagram:



a) Given the additional information that $P(A \cup B)$ =0.89, what is $P(A \mid B)$?

Round your response to at least 3 decimal places.

$P(A \mid B) =$ **0.356**

b) Are events *A* and *B* independent?

No

Correct

Comment:

a) To calculate $P(A \mid B)$, we can use the conditional probability formula $P\left(A \middle| B\right) = \frac{P(A \cap B)}{P(B)}$. The numerator is given to us in the question: $P(A \cap B) = 0.16$, but we need to determine the value of the denominator, $P(B)$. To do this, we can use the additional information, that $P(A \cup B) = 0.89$. Using the formula $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, we can rearrange it to get $P(B) = P(A \cup B) - P(A) + P(A \cap B) = 0.89 - 0.6 + 0.16 = 0.45$. Finally, we get $P\left(A \middle| B\right) = \frac{0.16}{0.45} = 0.355556$.

b) To determine if the events *A* and *B* are independent, we can use the check for independence $P(A \mid B) = P(A)$. Using the answer from part (a), we can conclude that the events *A* and *B* are not independent.

Question 10: *Score 1/1*

Your response

Given that $P(A) = 0.28$, $P(B) = 0.65$ and the $P(A \cap B) = 0.16$, what is $P(A \cup B)$?

Enter your answer to 2 decimal places.

$P(A \cup B) =$ **0.77** (100%)

Correct response

Given that $P(A) = 0.28$, $P(B) = 0.65$ and the $P(A \cap B) = 0.16$, what is $P(A \cup B)$?

Enter your answer to 2 decimal places.

$P(A \cup B) =$ **0.77**

Correct

Comment:

To find $P(A \cup B)$, we can use the formula $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Therefore, $P(A \cup B)$ = 0.28 + 0.65 -0.16 = 0.77

Question 11: *Score 1/1*

Your response	Correct response	
Given that $P(A) = 0.24$, $P(B) = 0.79$ and the $P(A \cap B) = 0.16$, what is $P(A \mid B)$? Round your answer to at least 3 decimal places. $P(A \mid B) =$ **0.203** (100%)	Given that $P(A) = 0.24$, $P(B) = 0.79$ and the $P(A \cap B) = 0.16$, what is $P(A \mid B)$? Round your answer to at least 3 decimal places. $P(A \mid B) =$ **0.203**	Correct

Comment:

To find $P(A \mid B)$, we can use the conditional probability equation $P\left(A \middle| B\right) = \dfrac{P(A \cap B)}{P(B)}$. Because these values are all given to us in the question, we can simply substitute them into the equation. Therefore, $P\left(A \middle| B\right) = \dfrac{0.16}{0.79} = 0.202532$.

Question 12: *Score 1/1*

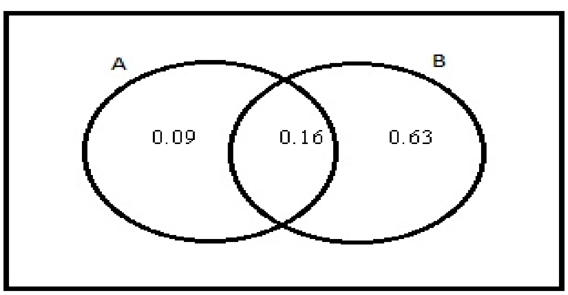
Your response	Correct response	
Given that $P(A) = 0.25$, $P(B) = 0.79$ and the $P(A \cap B) = 0.16$, what is $P(A \mid B^c)$? Round your answer to at least 3 decimal places. $P(A \mid B^c) =$ **0.429** (100%)	Given that $P(A) = 0.25$, $P(B) = 0.79$ and the $P(A \cap B) = 0.16$, what is $P(A \mid B^c)$? Round your answer to at least 3 decimal places. $P(A \mid B^c) =$ **0.429**	Correct

Comment:

To find $P(A \mid B^c)$, we need to make use of the general formula for conditional probability: $P\left(A \middle| B\right) = \dfrac{P(A \cap B)}{P(B)}$. Therefore,

$$P\left(A \middle| B^c\right) = \frac{P(A \cap B^c)}{P(B^c)}.$$

Since *P(B)* = 0.79, we get $P(B^c) = 1 - P(B) = 1 - 0.79 = 0.21$. To find $P(A \cap B^c)$, it is easiest to display the given probabilities in a Venn diagram:



It can be seen from the Venn diagram that $P(A \cap B^c)$ is 0.09. Therefore, $P\left(A \middle| B^c\right) = \dfrac{0.09}{0.21} = 0.428571$

Question 13: *Score 1/1*

Your response	Correct response	
Identify each of the following statements as either TRUE or FALSE: a) **False** (20%) If event A is a subset of event B, then the probability of their intersection is 1. b) **False** (20%) The union of events A and B is the event that either A occurs or B occurs, but not that both occur. c) **True** (20%) If two events (A and B) are mutually exclusive, then the conditional probability of A given B is 0. d) **False** (20%) Independent events cannot occur at the same time. e) **True** (20%) Mutually exclusive events are never independent, but dependent events are not always mutually exclusive.	Identify each of the following statements as either TRUE or FALSE: a) **False** If event A is a subset of event B, then the probability of their intersection is 1. b) **False** The union of events A and B is the event that either A occurs or B occurs, but not that both occur. c) **True** If two events (A and B) are mutually exclusive, then the conditional probability of A given B is 0. d) **False** Independent events cannot occur at the same time. e) **True** Mutually exclusive events are never independent, but dependent events are not always mutually exclusive.	Correct

Comment:

Question 14: *Score 1/1*

Your response	Correct response	
Identify each of the following statements as either TRUE or FALSE: a) **False** (20%) A and A^c are independent events. b) **True** (20%) If event A equals event B, then the conditional probability of event B, given event A, is 1. c) **True** (20%) If the probability of event A is influenced by the occurrence of event B, then the two events are dependent. d) **False** (20%) If event A is a subset of event B, then the conditional probability of event A, given event B, is 1. e) **False** (20%) If event A equals event B, then the probability of their intersection is 1.	Identify each of the following statements as either TRUE or FALSE: a) **False** A and A^c are independent events. b) **True** If event A equals event B, then the conditional probability of event B, given event A, is 1. c) **True** If the probability of event A is influenced by the occurrence of event B, then the two events are dependent. d) **False** If event A is a subset of event B, then the conditional probability of event A, given event B, is 1. e) **False** If event A equals event B, then the probability of their intersection is 1.	Correct

Comment:

Quiz 3: Discrete Probability Distributions, Binomial Distribution

Question 1: *Score 1/1*

Which of the following statements are true?

There may be more than one correct answer; select all that are true.

Choice	Selected	✔/✘	Points
A binomial random variable X can take on a total of n possible values.	No		
In a binomial distribution, the random variable X is a count of the number of successes.	Yes	✔	+1
If the probability of success is greater than zero, then the mean of a binomial random variable is greater than the variance.	Yes	✔	+1
The mean of a binomial random variable must be an integer, since a binomial random variable is discrete.	No		
The probability of success, p, in a binomial distribution increases as X, the number of successes, increases.	No		

Correct

Number of available correct choices: 2

Partial Grading Explained

Comment:

Question 2: *Score 1/1*

Your response	Correct response
Sweden is reported to have among the highest percentage of citizens aged 80 years and over, at approximately 5.4 % (i.e. 5.4 % of its population falls within this age category). Suppose 20 individuals are randomly selected from the population, and *X* is the number of individuals out of 20 that are 80 or more years old.	Sweden is reported to have among the highest percentage of citizens aged 80 years and over, at approximately 5.4 % (i.e. 5.4 % of its population falls within this age category). Suppose 20 individuals are randomly selected from the population, and *X* is the number of individuals out of 20 that are 80 or more years old.
a) On average, how many of the 20 individuals would be 80 years old or older?	a) On average, how many of the 20 individuals would be 80 years old or older?
Round your answer to at least 3 decimal places.	Round your answer to at least 3 decimal places.
1.080 (50%)	**1.080**
b) What is the probability that less than 2 of the 20 individuals are 80 years old or older?	b) What is the probability that less than 2 of the 20 individuals are 80 years old or older?
Round your answer to at least 3 decimal places.	Round your answer to at least 3 decimal places.
0.706 (50%)	**0.706**

Correct

Comment:

a) The random variable *X* follows a binomial distribution, with an expected value of $E[X] = \mu = np$, where *n* is the sample size, and *p* is the probability of success. Therefore, the expected value of *X* is $\mu = 20(0.054) = 1.08$.

b) The probability of less than 2 citizens being 80 years of age or older can be calculated using the binomial formula for *X* = 0 and *X* = 1. Therefore:

$$\begin{aligned} P(X < 2) &= P(X = 0) + P(X = 1) \\ &= C_0^{20} 0.054^0 (1 - 0.054)^{20-0} + C_1^{20} 0.054 (1 - 0.054)^{20-1} \\ &= .7056202450 \end{aligned}$$

Question 3: *Score 1/1*

Your response	Correct response	
According to the *Statistics Canada* website, the unemployment rate in Canada is approximately 7.6 % of the eligible workforce. If 25 members of the eligible workforce are randomly selected, and it is determined that at least 2 of the 25 are unemployed, what is the probability that exactly 4 of the 25 members are unemployed? Round your response to at least 3 decimal places. **0.139** (100%)	According to the *Statistics Canada* website, the unemployment rate in Canada is approximately 7.6 % of the eligible workforce. If 25 members of the eligible workforce are randomly selected, and it is determined that at least 2 of the 25 are unemployed, what is the probability that exactly 4 of the 25 members are unemployed? Round your response to at least 3 decimal places. **0.139**	Correct

Comment:

To calculate the probability of exactly 4 members being unemployed, given that 2 members of the labour force are unemployed, we need to calculate the conditional probability $P(X=4 \mid X \geq 2)$. This can be done using the general formula $P\left(A \middle| B\right) = \frac{P(A \cap B)}{P(B)}$, which in this case becomes $\frac{P(X=4 \cap X \geq 2)}{P(X \geq 2)}$.

In the numerator, $P(X=4 \cap X \geq 2)$ simplifies to $P(X=4)$, as $X=4$ and $X \geq 2$ intersect only at the point $X=4$. To calculate $P(X=4)$, we can use the binomial formula, such that:

$$P(X=4) = C_4^{25} 0.076^4 (1 - 0.076)^{25-4} = 0.080252$$

In the denominator, $P(X \geq 2)$ can be calculated by using the binomial formula $P(X=x) = C_x^{25} 0.076^x (1 - 0.076)^{25-x}$ for *x* = 2, 3, 4...25, and summing together the resulting probabilities. Alternately, we can use the binomial formula on the complement of $P(X \geq 2)$, which is $P(X \leq 1)$. Since

$$P(X \geq 2) + P(X \leq 1) = 1 \Rightarrow P(X \geq 2) = 1 - P(X \leq 1) = 1 - (P(X=0) + P(X=1)) = 0.576365$$

Therefore, $P\left(X=4 \middle| X \geq 2\right) = \frac{0.080252}{0.576365} = 0.139238$

Question 4: *Score 1/1*

Your response	Correct response	
Consider the following discrete probability distribution: <table><tr><td>*X*</td><td>2</td><td>4</td><td>6</td><td>8</td></tr><tr><td>*P(X)*</td><td>0.11</td><td>0.16</td><td>0.27</td><td>0.46</td></tr></table>What is $P(X > 6.5 \mid X > 2.5)$? Round your answer to at least 3 decimal places. **0.517** (100%)	Consider the following discrete probability distribution: <table><tr><td>*X*</td><td>2</td><td>4</td><td>6</td><td>8</td></tr><tr><td>*P(X)*</td><td>0.11</td><td>0.16</td><td>0.27</td><td>0.46</td></tr></table>What is $P(X > 6.5 \mid X > 2.5)$? Round your answer to at least 3 decimal places. **0.517**	Correct

Comment:

To determine the conditional probability, use the general formula $P\left(A \middle| B\right) = \frac{P(A \cap B)}{P(B)}$.

This becomes $P\left(X > 6.5 \middle| X > 2.5\right) = \frac{P(X > 6.5 \cap X > 2.5)}{P(X > 2.5)}$, where $P(X > 6.5 \cap X > 2.5)$ is the sum of all the probabilities for *X* values greater than 6.5 **and** greater than 2.5, and $P(X > 2.5)$ is the sum of all the probabilities for *X* values greater than 2.5. This results in $P(X > 6.5 \cap X > 2.5)$ = 0.46, and $P(X > 2.5)$ = 0.89.

Therefore, $P\left(X > 6.5 \middle| X > 2.5\right) = \frac{0.46}{0.89} = 0.516854$

Question 5: *Score 1/1*

Your response	Correct response	
Consider the following discrete probability distribution:	Consider the following discrete probability distribution:	

Your response

Consider the following discrete probability distribution:

X	-2.25	0	0.25	1.5	2.5
P(X)	0.19	0.15	0.25	0.16	*C*

a) What is the missing value, *C*?

0.25 (33%)

b) What is E[*X*]? Your answer should be a numeric value that does not include *C*.

Round your answer to at least 3 decimal places.

0.500 (33%)

c) What is Var[*X*]? Your answer should be a numeric value that does not include *C*.

Round your answer to at least 3 decimal places.

2.650 (33%)

Correct response

Consider the following discrete probability distribution:

X	-2.25	0	0.25	1.5	2.5
P(X)	0.19	0.15	0.25	0.16	*C*

a) What is the missing value, *C*?

0.25

b) What is E[*X*]? Your answer should be a numeric value that does not include *C*.

Round your answer to at least 3 decimal places.

0.500

c) What is Var[*X*]? Your answer should be a numeric value that does not include *C*.

Round your answer to at least 3 decimal places.

2.650

Correct

Comment:

a) To determine the missing value, remember that the sum of the probabilities must equal 1, $\sum_{i=1}^{n} p(x_i) = 1$

Therefore, 0.19 + 0.15 + 0.25 + 0.16 + *C* = 1 $\Longrightarrow$ 1 - (0.19 + 0.15 + 0.25 + 0.16) = *C* = 0.25

b) To calculate E[*X*], use the formula $E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i)$. Since *n* = 5, we get

$$E[X] = \sum_{i=1}^{5} x_i \cdot p(x_i) = (-2.25 \cdot 0.19) + (0 \cdot 0.15) + (0.25 \cdot 0.25) + (1.5 \cdot 0.16) + (2.5 \cdot 0.25) = 0.5$$

c) To calculate Var[X], use the formula $Var[X] = \sum_{i=1}^{n} (x_i - \mu_X)^2 \cdot p(x_i)$. Here, *n* = 5, and $\mu_X = E[X] = 0.5$, so we get

$$\begin{aligned} Var[X] &= \sum_{i=1}^{5} (x_i - 0.5)^2 \cdot p(x_i) \\ &= (-2.25 - 0.5)^2 \cdot 0.19 + (0 - 0.5)^2 \cdot 0.15 + (0.25 - 0.5)^2 \cdot 0.25 + (1.5 - 0.5)^2 \cdot 0.16 + (2.5 \\ &\quad - 0.5)^2 \cdot 0.25 \\ &= 2.65 \end{aligned}$$

Question 6: *Score 1/1*

Your response	Correct response

Your response

Consider the following discrete probability distribution:

X	-1.3	0	2.1	5.4	6.2
$P(X)$	0.100	0.19	0.24	0.11	0.36

a) What is $P(X=0)$?

0.19 (50%)

b) What is $P(X=0 \mid X<5)$?

Round your answer to at least 3 decimal places.

0.358 (50%)

Correct response

Consider the following discrete probability distribution:

X	-1.3	0	2.1	5.4	6.2
$P(X)$	0.100	0.19	0.24	0.11	0.36

a) What is $P(X=0)$?

0.19

b) What is $P(X=0 \mid X<5)$?

Round your answer to at least 3 decimal places.

0.358

Correct

Comment:

a) The $P(X=0)$ can be read directly from the discrete probability distribution table, and is equal to 0.19.

b) To calculate $P(X=0 \mid X<5)$, use the general formula $P\left(A \middle| B\right)=\frac{P(A\cap B)}{P(B)}$. Therefore, we are calculating $\frac{P(X=0\cap X<5)}{P(X<5)}$, where $P(X=0\cap X<5)$ is the sum of the probabilities for X values that are equal to 0 **and** less than 5. In this case, X = 0 and X < 5 only intersect at the point X = 0, so $P(X=0\cap X<5)=P(X=0)=0.19$. To determine $P(X<5)$, it is the sum of the probabilities for X values that are less than 5, so $P(X<5)=0.100+0.19+0.24=0.53$.

Therefore, $P\left(X=0 \middle| X<5\right)=\frac{0.19}{0.53}=0.358491$

Quiz 4: Sampling Distributions

Question 1: *Score 1/1*

Your response	Correct response	
Determine whether the following statements are TRUE or FALSE: a) **True** (20%) The value of a parameter does not vary from sample to sample. b) **False** (20%) The standard deviation of the sampling distribution of the sample mean depends on the value of μ . c) **False** (20%) All else being equal, the standard deviation of the sampling distribution of the sample mean will be smaller for n = 10 than for n = 40. d) **False** (20%) The value of a statistic does not vary from sample to sample. e) **False** (20%) We cannot possibly determine any characteristics of a statistic's sampling distribution without repeatedly sampling from the population.	Determine whether the following statements are TRUE or FALSE: a) **True** The value of a parameter does not vary from sample to sample. b) **False** The standard deviation of the sampling distribution of the sample mean depends on the value of μ . c) **False** All else being equal, the standard deviation of the sampling distribution of the sample mean will be smaller for n = 10 than for n = 40. d) **False** The value of a statistic does not vary from sample to sample. e) **False** We cannot possibly determine any characteristics of a statistic's sampling distribution without repeatedly sampling from the population.	✔ Correct

Question 2: *Score 1/1*

Your response	Correct response	
Which one of the following statements is the best definition of the Central Limit Theorem? **The sampling distribution of the sample mean is approximately normal for large sample sizes, regardless of the distribution of the population.** (100%)	Which one of the following statements is the best definition of the Central Limit Theorem? **The sampling distribution of the sample mean is approximately normal for large sample sizes, regardless of the distribution of the population.**	✔ Correct

Question 3: *Score 1/1*

Your response	Correct response	
Which of the following statements is the best definition of a **biased** estimator? **An estimator whose expected value does not equal the true value.** (100%)	Which of the following statements is the best definition of a **biased** estimator? **An estimator whose expected value does not equal the true value.**	✔ Correct

Comment:

In order to be an unbiased estimator, the expected value of the estimator must equal the true value. Therefore, a biased estimator is one in which the expected value of the estimator does not equal the true value.

Question 4: *Score 1/1*

Your response	Correct response	
A random sample of size 123 was taken from a population with a population mean 28 and a population standard deviation 8. Determine each of the following about the sampling distribution of the sample mean. Round your answer to at least 3 decimal places where appropriate. a) $\mu_{\bar{x}}$ = **28** (33%) b) $\sigma_{\bar{x}}$ = **0.721** (33%) c) Can we conclude that the sampling distribution of the sample mean is approximately normal? **Yes** (33%)	A random sample of size 123 was taken from a population with a population mean 28 and a population standard deviation 8. Determine each of the following about the sampling distribution of the sample mean. Round your answer to at least 3 decimal places where appropriate. a) $\mu_{\bar{x}}$ = **28** b) $\sigma_{\bar{x}}$ = **0.721** c) Can we conclude that the sampling distribution of the sample mean is approximately normal? **Yes**	**Correct**

Comment:

a) The mean of the sampling distribution of the sample mean, $\mu_{\bar{x}}$, will be the same as the mean of the population from which the sample was taken. Therefore, $\mu_{\bar{x}} = \mu = 28$.

b) The standard deviation of the sampling distribution of the sample mean, $\sigma_{\bar{x}}$, is given by the formula $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Therefore,

$$\sigma_{\bar{x}} = \frac{8}{\sqrt{123}} = 0.721336 .$$

c) There is nothing to indicate whether or not the population from which we are sampling is normally distributed, however we can apply the Central Limit Theorem because the sample size is large. Therefore, we can conclude that the sampling distribution of the sample mean is approximately normal.

Question 5: *Score 1/1*

Your response	Correct response	
A random sample of size 12 was taken from a population with a population mean 26 and a population standard deviation 8. Determine each of the following about the sampling distribution of the sample mean. Round your answer to at least 3 decimal places where appropriate. a) $\mu_{\bar{x}} =$ **26** (33%) b) $\sigma_{\bar{x}} =$ **2.309** (33%) c) Can we conclude that the sampling distribution of the sample mean is approximately normal? **No** (33%)	A random sample of size 12 was taken from a population with a population mean 26 and a population standard deviation 8. Determine each of the following about the sampling distribution of the sample mean. Round your answer to at least 3 decimal places where appropriate. a) $\mu_{\bar{x}} =$ **26** b) $\sigma_{\bar{x}} =$ **2.309** c) Can we conclude that the sampling distribution of the sample mean is approximately normal? **No**	Correct

Comment:

a) The mean of the sampling distribution of the sample mean, $\mu_{\bar{x}}$, will be the same as the mean of the population from which the sample was taken. Therefore, $\mu_{\bar{x}} = \mu = 26$.

b) The standard deviation of the sampling distribution of the sample mean, $\sigma_{\bar{x}}$, is given by the formula $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$. Therefore,

$$\sigma_{\bar{x}} = \frac{8}{\sqrt{12}} = 2.309401 \ .$$

c) There is nothing to indicate whether or not the population from which we are sampling is normally distributed, and we cannot rely on the Central Limit Theorem because the sample size is small. Therefore, we cannot determine whether or not the sampling distribution is approximately normal.

Quiz 5: Inference for Single Population Mean

Question 1: *Score 1/1*

Your response	Correct response	
Suppose a random sample of size 31 was drawn from a normally distributed population, with a known population standard deviation of $\sigma = 8.9$. a) What is the margin of error for a 95% confidence level? Round your response to at least 3 decimal places. **3.133** (50%) b) What is the margin of error for a 90% confidence level? Round your response to at least 3 decimal places. **2.630** (50%)	Suppose a random sample of size 31 was drawn from a normally distributed population, with a known population standard deviation of $\sigma = 8.9$. a) What is the margin of error for a 95% confidence level? Round your response to at least 3 decimal places. **3.133** b) What is the margin of error for a 90% confidence level? Round your response to at least 3 decimal places. **2.630**	Correct

Comment:

a) The margin of error is calculated by the formula $ME = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$. For a 95% confidence level, the corresponding $z_{\frac{\alpha}{2}}$ value is 1.96. Therefore, the margin of error is $1.96 \cdot \left(\frac{8.9}{\sqrt{31}}\right) = 3.133035$.

b) Using the same formula for margin of error as above, but with $z_{\frac{\alpha}{2}}$ = 1.645 for a 90% confidence level, the calculated margin of error is then $1.645 \cdot \left(\frac{8.9}{\sqrt{31}}\right) = 2.629511$.

Question 2: *Score 1/1*

Your response	Correct response	
a) In order to estimate the population mean, μ, to within 3 at 95% confidence, what is the minimum sample size required? (Assume $\sigma = 6.1$). **16** (25%) b) If just the population standard deviation were to increase, then the minimum sample size required would: **increase** (25%) c) If just the confidence level were to decrease (i.e. go from 95% to 90% confidence), then the minimum sample size required would: **decrease** (25%) d) If just the bound within which μ was to be estimated were to increase, then the minimum sample size required would: **decrease** (25%)	a) In order to estimate the population mean, μ, to within 3 at 95% confidence, what is the minimum sample size required? (Assume $\sigma = 6.1$). **16** b) If just the population standard deviation were to increase, then the minimum sample size required would: **increase** c) If just the confidence level were to decrease (i.e. go from 95% to 90% confidence), then the minimum sample size required would: **decrease** d) If just the bound within which μ was to be estimated were to increase, then the minimum sample size required would: **decrease**	Correct

Comment:

a) To estimate the sample size, we can re-arrange the formula for margin of error to get $n = \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{ME}\right)^2$. Plugging in the given values, and the $z_{\frac{\alpha}{2}}$ value for a 95% confidence level, we get $n = \left(\frac{1.96 \cdot 6.1}{3}\right)^2 = 15.882882$. However, we cannot take a fraction of an individual, so our sample size must be rounded UP to the nearest whole number. Therefore, the minimum sample size required is $n = 16$.

b) If everything else were to remain constant, and only the population standard deviation were to increase, we can see from the formula above that this would result in an increase in the numerator, and consequently the minimum sample size would also increase. That is to say, if the variation in our population were to increase, then we would need a larger sample to be within the same margin of error of μ, with the same level of confidence.

c) If everything else were to remain constant, and only the confidence level were to decrease, then the value for $z_{\frac{\alpha}{2}}$ would decrease. This would result in a decrease in the numerator, and a subsequent decrease in the minimum sample size. That is to say, if we wanted to be within the same margin of error of μ, but with less confidence, then we could use a smaller sample.

d) If the bound in which we wanted to estimate μ (i.e. the margin of error) were to increase, then the denominator in our formula for sample size would increase, which results in an decrease in the minimum sample size required. That is to say, if we wanted to be within a larger margin of error of μ, but with the same confidence level, we could take a smaller sample.

Question 3: *Score 1/1*

Your response	Correct response
Suppose a random sample of size 82 was drawn from a normally distributed population, with a known population standard deviation of $\sigma = 13.0$.	Suppose a random sample of size 82 was drawn from a normally distributed population, with a known population standard deviation of $\sigma = 13.0$.
a) What is the appropriate $z_{\frac{\alpha}{2}}$ value for a 50 % confidence interval?	a) What is the appropriate $z_{\frac{\alpha}{2}}$ value for a 50 % confidence interval?
Round your response to at least 2 decimal places.	Round your response to at least 2 decimal places.
0.674 (50%)	**0.674**
b) What is the margin of error for a 50 % confidence interval?	b) What is the margin of error for a 50 % confidence interval?
Round your response to at least 2 decimal places.	Round your response to at least 2 decimal places.
0.968 (50%)	**0.968**

Correct

Comment:

a) The $z_{\frac{\alpha}{2}}$ value for a 50 % confidence interval is .6744897502.

b) Using the formula for margin of error, with $z_{\frac{\alpha}{2}}$ = .6744897502 for a 50 % confidence level, we get

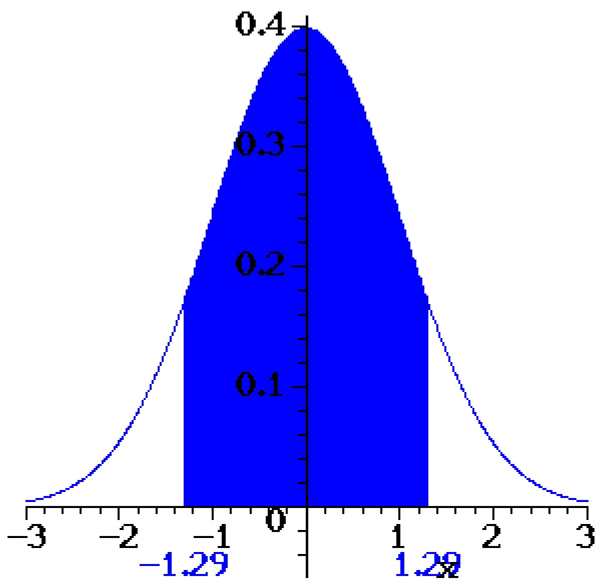
$$.6744897502 \cdot \left(\frac{13.0}{\sqrt{82}}\right) = 0.968304 .$$

Question 4: *Score 1/1*

Your response	Correct response	
For the confidence interval given by the formula $\bar{X} \pm 1.29 \cdot \frac{\sigma}{\sqrt{n}}$, what is the corresponding confidence level? For the purposes of this question, we can assume that *n* is sufficiently large, and σ is known. Express your answer as a percent, but do NOT include the percent sign (%) in your response. Round your response to at least 2 decimal places. **80.29** (100%)	For the confidence interval given by the formula $\bar{X} \pm 1.29 \cdot \frac{\sigma}{\sqrt{n}}$, what is the corresponding confidence level? For the purposes of this question, we can assume that *n* is sufficiently large, and σ is known. Express your answer as a percent, but do NOT include the percent sign (%) in your response. Round your response to at least 2 decimal places. **80.29**	Correct

Comment:

The confidence level is given as the area under the standard normal distribution in between the values of -1.29 and 1.29, which is graphically represented as:



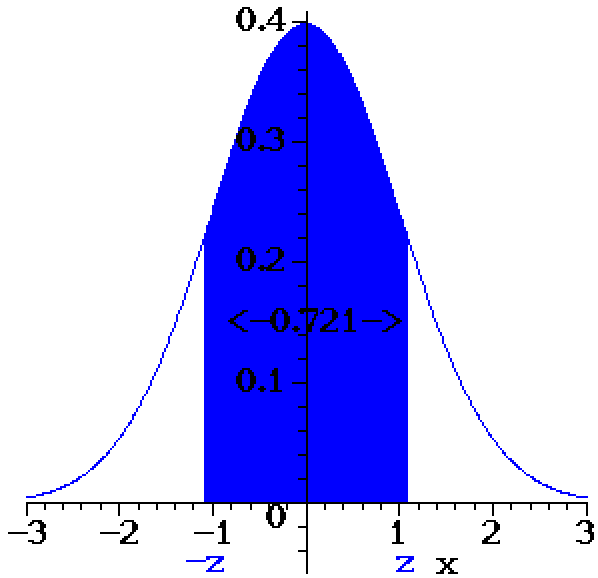
Using a standard normal table, we can find this area to be 0.802949. Therefore, the confidence level is 0.802949 X 100% = 80.294934.

Question 5: *Score 1/1*

Your response	Correct response	
What is the appropriate $z_{\frac{\alpha}{2}}$ value for a 72.1 % confidence interval for μ ? Round your response to at least 2 decimal places. **1.08** (100%)	What is the appropriate $z_{\frac{\alpha}{2}}$ value for a 72.1 % confidence interval for μ ? Round your response to at least 2 decimal places. **1.08**	Correct

Comment:

In order to determine the $z_{\frac{\alpha}{2}}$ values, we need to capture the middle 72.1 % of the standard normal distribution. That is, we need to find a *z* value such that the area between *-z* and *z* is 0.721:



Using a standard normal table, we can find the *z* value to be 1.082568490.

Question 6: *Score 1/1*

Your response	Correct response

Your response

Suppose in a popular fishing lake, 10 largemouth bass are caught, and each of their lengths (in inches) are recorded below:

Fish	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Length (inches)	13.6	16.1	15.8	16.7	15.0	16.8	17.5	13.3	12.0	16.1

Furthermore, suppose it is known that the population standard deviation for length is $\sigma = 1.22$, and that the length of largemouth bass is normally distributed.

a) What is a point estimate of the population mean length?

Round your response to at least 2 decimal places.

15.29 (50%)

b) What is a 90% confidence interval for the population mean length?

Enter your response in the interval notation: **(lower limit, upper limit)** . Include the brackets in your response.

Round your values for lower limit and upper limit to at least 3 decimal places.

(14.655,15.925) (50%)

Correct response

Suppose in a popular fishing lake, 10 largemouth bass are caught, and each of their lengths (in inches) are recorded below:

Fish	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Length (inches)	13.6	16.1	15.8	16.7	15.0	16.8	17.5	13.3	12.0	16.1

Furthermore, suppose it is known that the population standard deviation for length is $\sigma = 1.22$, and that the length of largemouth bass is normally distributed.

a) What is a point estimate of the population mean length?

Round your response to at least 2 decimal places.

15.29

b) What is a 90% confidence interval for the population mean length?

Enter your response in the interval notation: **(lower limit, upper limit)** . Include the brackets in your response.

Round your values for lower limit and upper limit to at least 3 decimal places.

(14.655,15.925)

Correct

Comment:

a) A point estimate for the population mean, μ , is the sample estimate, $\bar{x}$. Therefore, the point estimate is

$$\bar{x} = \frac{\sum_{i=10}^{10} x_i}{n} = \frac{(13.6 + 16.1 + \ldots + 16.1)}{10} = 15.29 .$$

b) The formula for a $(1 - \alpha) \cdot 100\ \%$ confidence interval is $\bar{x} \pm z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$. For a 90% confidence interval, $z_{\frac{\alpha}{2}}$ = 1.645.

Plugging all the values into the equation gives us $15.29 \pm 1.645 \cdot \frac{1.22}{\sqrt{10}}$. Written in interval notation, the 90% confidence interval for μ is (14.655362, 15.924638).

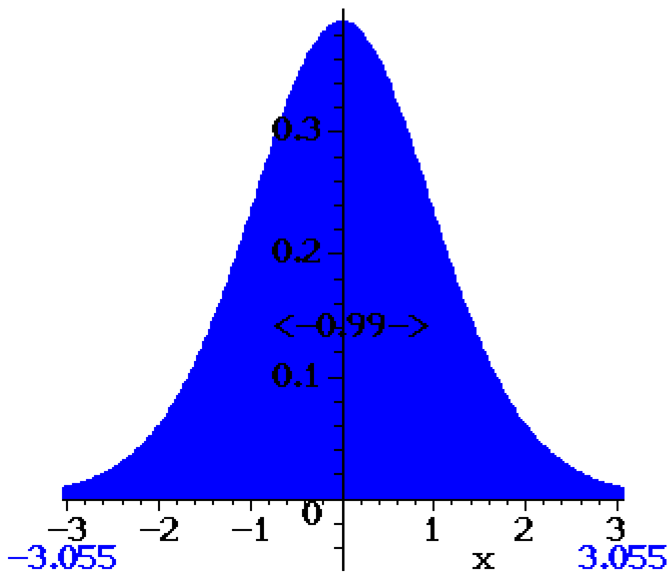
Question 7: *Score 1/1*

Your response	Correct response	
Suppose a random sample of size 13 is selected from a normally distributed population. If the confidence interval for the population mean is given by the formula $\bar{X} \pm 3.055 \cdot \left(\frac{s}{\sqrt{n}}\right)$, then what is the corresponding confidence level? Express your answer as a percent, but do NOT include the percent sign (%) in your response. **99** (100%)	Suppose a random sample of size 13 is selected from a normally distributed population. If the confidence interval for the population mean is given by the formula $\bar{X} \pm 3.055 \cdot \left(\frac{s}{\sqrt{n}}\right)$, then what is the corresponding confidence level? Express your answer as a percent, but do NOT include the percent sign (%) in your response. **99**	Correct

Comment:

The confidence level is determined as the area under the *t* distribution, with 12 degrees of freedom, between the values of -3.055 and 3.055. Using computer software, or approximating with a standard normal table, we can find this value to be 0.99.

Graphically, this is represented as:



Therefore, this is a 99 % confidence interval.

Question 8: *Score 1/1*

Your response	Correct response	
Suppose a random sample of size 13 was taken from a normally distributed population, and the sample standard deviation was calculated to be *s* = 6.5. a) Calculate the margin of error for a 95% confidence interval for the population mean. Round your response to at least 3 decimal places. **3.928** (50%) b) Calculate the margin of error for a 90% confidence interval for the population mean. Round your response to at least 3 decimal places. **3.213** (50%)	Suppose a random sample of size 13 was taken from a normally distributed population, and the sample standard deviation was calculated to be *s* = 6.5. a) Calculate the margin of error for a 95% confidence interval for the population mean. Round your response to at least 3 decimal places. **3.928** b) Calculate the margin of error for a 90% confidence interval for the population mean. Round your response to at least 3 decimal places. **3.213**	Correct

Comment:

a) To determine the margin of error, we first need to determine the $t_{\frac{\alpha}{2}}$ value. For a *t* distribution with 12 degrees of freedom, for a 95% confidence interval for the mean this value is $t_{\frac{\alpha}{2}} = 2.178809$. Therefore, the margin of error for a 95% confidence interval for the population mean is $ME = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 2.178809 \cdot \frac{6.5}{\sqrt{13}} = 3.927903$.

b) For a 90% confidence interval for the population mean, $t_{\frac{\alpha}{2}} = 1.782285$, and therefore

$$ME = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1.782285 \cdot \frac{6.5}{\sqrt{13}} = 3.213061.$$

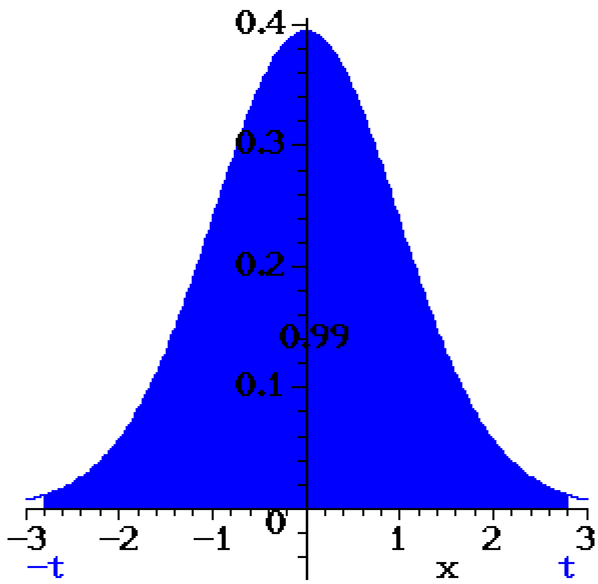
Question 9: *Score 1/1*

Your response	Correct response	
What is the appropriate $t_{\frac{\alpha}{2}}$ for a 99 % confidence interval with 25 degrees of freedom? Round your response to at least 3 decimal places. **2.787** (100%)	What is the appropriate $t_{\frac{\alpha}{2}}$ for a 99 % confidence interval with 25 degrees of freedom? Round your response to at least 3 decimal places. **2.787**	Correct

Comment:

The $t_{\frac{\alpha}{2}}$ value for a 99 % confidence interval is from a *t* distribution with 25 degrees of freedom, such that the area between +/- *t* is equal to 0.99.

Graphically, this is represented as:



Using computer software, or approximating with a *t* distribution table, we can find the value of *t* to be *t = 2.787*.

Quiz 6: Inference for Single Population Mean II

Question 1: *Score 1/1*

Your response	Correct response
A researcher wishes to test the null hypothesis that the mean of a normally distributed population is equal to 110, against the alternative hypothesis that the mean is greater than 110. She is not that familiar with hypothesis testing, so she decides to randomly select a sample of size 47, and if the *Z* test statistic is greater than 1.1, she will reject the null hypothesis. Assuming that the population standard deviation is known to be 25, then: a) What value of $\bar{x}$ would she need to obtain in order to get a *Z* test statistic of 1.1? Round your response to at least 3 decimal places. **114.011** (33%) b) Suppose that the *true* population mean was actually 120. What is the value of the power for the test? Round your responses to at least 3 decimal places. **0.950** (33%) c) What if, instead of 120, the true population mean was only 115. Would you expect the power of the test to be higher or lower than the value you found in part (b)? **Lower** (33%)	A researcher wishes to test the null hypothesis that the mean of a normally distributed population is equal to 110, against the alternative hypothesis that the mean is greater than 110. She is not that familiar with hypothesis testing, so she decides to randomly select a sample of size 47, and if the *Z* test statistic is greater than 1.1, she will reject the null hypothesis. Assuming that the population standard deviation is known to be 25, then: a) What value of $\bar{x}$ would she need to obtain in order to get a *Z* test statistic of 1.1? Round your response to at least 3 decimal places. **114.011** b) Suppose that the *true* population mean was actually 120. What is the value of the power for the test? Round your responses to at least 3 decimal places. **0.950** c) What if, instead of 120, the true population mean was only 115. Would you expect the power of the test to be higher or lower than the value you found in part (b)? **Lower**

Correct

Comment:

a) To determine the value of $\bar{x}$ that would result in a *z* test statistic of 1.1, we can rearrange the equation $Z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$ to get

$\bar{x} = Z \cdot \frac{\sigma}{\sqrt{n}} + \mu$. Substituting in values for *Z*, *n*, σ and μ results in $\bar{x} = 1.1 \cdot \left(\frac{25}{\sqrt{47}}\right) + 110 = 114.011287$.

b) In order to calculate the Power of the test, we need to calculate the probability of rejecting the null hypothesis when it is false. Since the researcher stated she would reject the null hypothesis if she calculated a *z* test statistic greater than 1.1, and in part (a) we calculated that this corresponded to a sample mean of 114.011287, it is equivalent to say that the researcher will reject the null hypothesis if she obtains a sample mean greater than 114.011287.

Therefore, to calculate the probability of rejecting the null hypothesis when it is false, we need to calculate

$P(\bar{X} > 114.011287, \textit{when } \mu = 120) = P\left(Z > \frac{(114.011287 - 120)}{\frac{25}{\sqrt{47}}}\right) = P(Z > -1.642262)$. Using a standard normal table, we can calculate this probability to be 0.949732, which is the Power of our test.

c) The Power of a test increases as the difference between the hypothesized mean and true mean increases. Therefore, if the difference between the hypothesized mean and true mean decreases, the Power will also decrease, as it becomes more difficult for the test to distinguish between the two values.

Question 2: *Score 1/1*

Your response	Correct response	
A researcher wishes to test the null hypothesis that the mean of a normally distributed population is equal to 110, against the alternative hypothesis that the mean is not equal to 110. He is not that familiar with hypothesis testing, so he decides to randomly select a sample of size 46, and if the *Z* test statistic is greater than 1.1 or less than -1.1, he will reject the null hypothesis. Assuming that the population standard deviation is known to be 24, then:	A researcher wishes to test the null hypothesis that the mean of a normally distributed population is equal to 110, against the alternative hypothesis that the mean is not equal to 110. He is not that familiar with hypothesis testing, so he decides to randomly select a sample of size 46, and if the *Z* test statistic is greater than 1.1 or less than -1.1, he will reject the null hypothesis. Assuming that the population standard deviation is known to be 24, then:	
a) What is the probability he will make a Type I error?	a) What is the probability he will make a Type I error?	
Round your response to at least 3 decimal places.	Round your response to at least 3 decimal places.	**Correct**
0.271 (50%)	**0.271**	
b) What are the corresponding $\bar{x}$ values associated with his arbitrarily selected rejection region? That is, what values for $\bar{x}$ will result in a *Z* test statistic of +/- 1.1?	b) What are the corresponding $\bar{x}$ values associated with his arbitrarily selected rejection region? That is, what values for $\bar{x}$ will result in a *Z* test statistic of +/- 1.1?	
Round your responses to at least 3 decimal places.	Round your responses to at least 3 decimal places.	
Enter your responses in the format: ***Smaller Value, Larger Value*** (include the ',' between the values).	Enter your responses in the format: ***Smaller Value, Larger Value*** (include the ',' between the values).	
106.108,113.892 (50%)	**106.108,113.892**	

Comment:

a) Type I error occurs when the null hypothesis is rejected when it is actually true. Since the researcher will reject the null hypothesis if she calculates a *z* test statistic greater than 1.1 or less than -1.1, the probability of a Type I error is calculated as $P(|Z| > 1.1)$. Using a standard normal table, we can find this area to be 0.271332. Therefore, *P(Type I error) = 0.271332.*

b) To find the values of $\bar{x}$ that correspond to a *z* test statistic value of +/- 1.1, we can rearrange the formula $Z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$ to get

$\bar{x} = Z \cdot \left(\frac{\sigma}{\sqrt{n}}\right) + \mu$. Because there are two possible *z* test statistic values, we need to calculate two $\bar{x}$ values:

$\bar{x} = 1.1\left(\frac{24}{\sqrt{46}}\right) + 110 = 113.892468$, and $\bar{x} = -1.1 \cdot \left(\frac{24}{\sqrt{46}}\right) + 110 = 106.107532$. Therefore, corresponding sample means are 106.107532 and 113.892468.

Question 3: *Score 1/1*

Your response	Correct response	
Test the null hypothesis $H_0 : \mu = 3.4$ against the alternative hypothesis $H_A : \mu < 3.4$, based on a random sample of 25 observations drawn from a normally distributed population with $\bar{x} = 3.2$ and $\sigma = 0.68$. a) What is the value of the test statistic? Round your response to at least 3 decimal places. **-1.471** (25%) b) What is the appropriate p-value? Round your response to at least 3 decimal places. **0.071** (25%) c) Is the null hypothesis rejected at: i) the 10% level of significance? **Yes** (25%) ii) the 5% level of significance? **No** (25%)	Test the null hypothesis $H_0 : \mu = 3.4$ against the alternative hypothesis $H_A : \mu < 3.4$, based on a random sample of 25 observations drawn from a normally distributed population with $\bar{x} = 3.2$ and $\sigma = 0.68$. a) What is the value of the test statistic? Round your response to at least 3 decimal places. **-1.471** b) What is the appropriate p-value? Round your response to at least 3 decimal places. **0.071** c) Is the null hypothesis rejected at: i) the 10% level of significance? **Yes** ii) the 5% level of significance? **No**	Correct

Comment:

a) The test statistic can be calculated using the formula: $Z = \dfrac{(\bar{x} - \mu)}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{(3.2 - 3.4)}{\dfrac{0.68}{\sqrt{n}}} = -1.470588$.

b) Since the alternative hypothesis indicates that the test is a one-sided, lower tailed test, the p-value is calculated as the area under the standard normal curve to the *left* of the test statistic. Therefore,
$p - value = P(Z < -1.470588) = .707012531065958627E\text{-}1$.

c) i) Since the p-value calculated in part (b) is less than $\alpha = 0.10$, the null hypothesis *is* rejected at the 10% level of significance.

ii) At the 5% level of significance, the p-value is greater than $\alpha = 0.05$, and therefore the null hypothesis is *not* rejected.

Question 4: *Score 1/1*

Your response	Correct response
A local newscaster reports that the average rainfall in the month of June is approximately 90 mm. However, a meteorologist wishes to test this claim, believing that the average rainfall in June is actually higher than 90 mm. He collects data on the average June rainfall for 10 randomly selected years, and computes a mean of 92 mm. Assuming that the population standard deviation is known to be 16.0, and that rainfall is normally distributed, determine each of the following:	A local newscaster reports that the average rainfall in the month of June is approximately 90 mm. However, a meteorologist wishes to test this claim, believing that the average rainfall in June is actually higher than 90 mm. He collects data on the average June rainfall for 10 randomly selected years, and computes a mean of 92 mm. Assuming that the population standard deviation is known to be 16.0, and that rainfall is normally distributed, determine each of the following:
a) What are the appropriate hypotheses:	a) What are the appropriate hypotheses:
$H_0 : \mu = 90,\ H_A : \mu > 90$ (33%)	$H_0 : \mu = 90,\ H_A : \mu > 90$
b) Calculate the appropriate test statistic.	b) Calculate the appropriate test statistic.
Round your answer to at least 3 decimal places.	Round your answer to at least 3 decimal places.
0.395 (33%)	**0.395**
c) What is the appropriate conclusion that can be made, at the 5% level of significance?	c) What is the appropriate conclusion that can be made, at the 5% level of significance?
There is insufficient evidence to reject the null hypothesis, and therefore no significant evidence that the mean rainfall in June is different from 90 mm. (33%)	**There is insufficient evidence to reject the null hypothesis, and therefore no significant evidence that the mean rainfall in June is different from 90 mm.**

Correct

Comment:

a) Hypothesis testing is always carried out on the population parameter, which is this case is the population mean μ . The null hypothesis is what the current belief is; here, it is that the mean rainfall is 90 mm. The alternative hypothesis is the new idea that the researcher believes, which in this case is that the mean rainfall is actually greater than 90 mm. Therefore, the appropriate null and alternative hypotheses are $H_0 : \mu = 90,\ H_A : \mu > 90$.

b) To calculate the test statistic, we use the formula $Z = \dfrac{\left(\bar{x} - \mu_0\right)}{\dfrac{\sigma}{\sqrt{n}}}$. Substituting in the corresponding values, we get

$$Z = \frac{(92 - 90)}{\frac{16.0}{\sqrt{10}}} = 0.395285 \ .$$

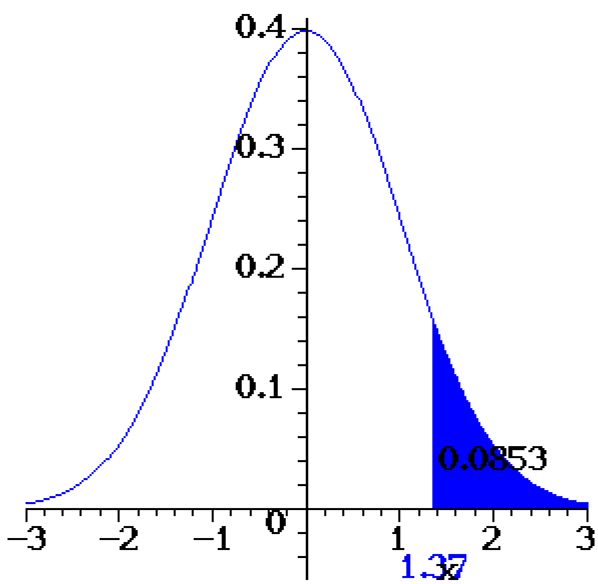
c) To determine what conclusion can be made, we need to calculate the p-value for the test statistic calculated in part (b). Because the alternative hypothesis indicates that we are carrying out a one-sided, upper tailed test, the p-value is the area under the standard normal curve to the right of our calculated test statistic. Using computer software, or a standard normal table, we can find this area to be 0.346316. Comparing the p-value to $\alpha = 0.05$, we see that our p-value is greater than α, indicating that there is **insufficient evidence** to reject the null hypothesis.

Question 5: *Score 1/1*

Your response	Correct response	
Suppose a hypothesis test of $H_0 : \mu = \mu_0$ is tested against the alternative hypothesis $H_A : \mu > \mu_0$, and the resulting *Z* test statistic is $Z = 1.37$. a) What is the appropriate p-value for the hypothesis test. Round your response to at least 4 decimal places. **0.0853** (50%) b) Based on the p-value, would you reject the null hypothesis at the 10% level of significance? **Yes** (50%)	Suppose a hypothesis test of $H_0 : \mu = \mu_0$ is tested against the alternative hypothesis $H_A : \mu > \mu_0$, and the resulting *Z* test statistic is $Z = 1.37$. a) What is the appropriate p-value for the hypothesis test. Round your response to at least 4 decimal places. **0.0853** b) Based on the p-value, would you reject the null hypothesis at the 10% level of significance? **Yes**	Correct

Comment:

a) Since the alternative hypothesis indicates a one-sided, upper tailed test, the p-value is the area under the standard normal curve to the right of the test statistic. Using computer software, or approximating with a standard normal table, we can find this area to be *p-value = 0.085343*. Graphically, this is represented as:



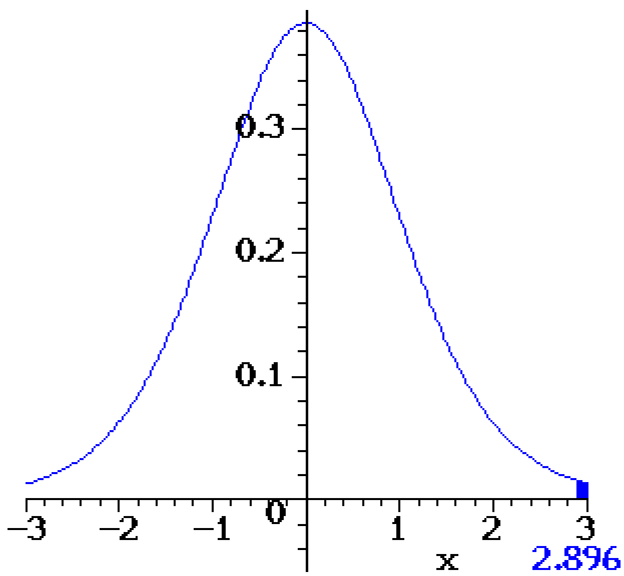
b) Since the p-value is less than $\alpha = 0.10$, there is sufficient evidence to reject the null hypothesis, at the 10% level of significance.

Question 6: *Score 1/1*

Your response	Correct response	
Suppose the null hypothesis of $H_0 : \mu = \mu_0$ is being tested against the one-sided alternative hypothesis $H_A : \mu > \mu_0$, at the 1% level of significance. If a random sample of size 9 is taken, and the population is assumed to be normally distributed, what is the appropriate rejection region? $t \geq 2.896$ (100%)	Suppose the null hypothesis of $H_0 : \mu = \mu_0$ is being tested against the one-sided alternative hypothesis $H_A : \mu > \mu_0$, at the 1% level of significance. If a random sample of size 9 is taken, and the population is assumed to be normally distributed, what is the appropriate rejection region? $t \geq 2.896$	Correct

Comment:

A one-sided, upper tailed hypothesis test indicates that the rejection region is determined by a value of *T* such that $P(t \geq T) = 0.01$, under a *t* distribution with 8 degrees of freedom. Using computer software, or approximating with a *t* distribution table, we can find this value to be *T = 2.896459*. Therefore, the rejection region is approximately $t \geq 2.896$, which is represented graphically as:

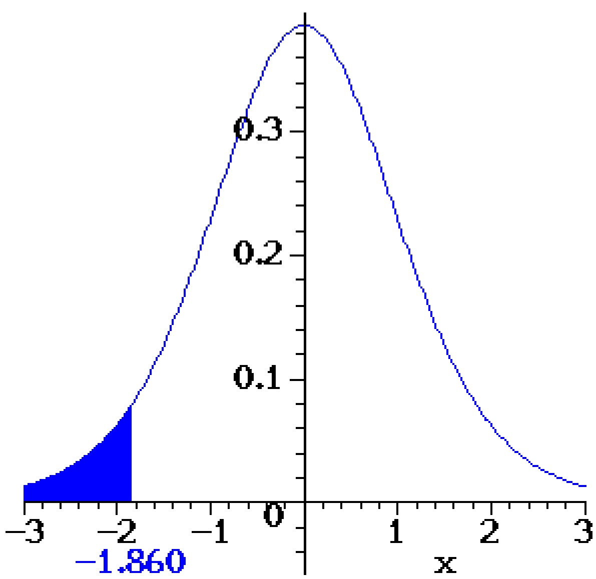


Question 7: *Score 1/1*

Your response	Correct response	
Suppose the null hypothesis of $H_0 : \mu = \mu_0$ is being tested against the one-sided alternative hypothesis $H_A : \mu < \mu_0$, at the 5% level of significance. If a random sample of size 9 is taken, and the population is assumed to be normally distributed, what is the appropriate rejection region? $t \leq -1.860$ (100%)	Suppose the null hypothesis of $H_0 : \mu = \mu_0$ is being tested against the one-sided alternative hypothesis $H_A : \mu < \mu_0$, at the 5% level of significance. If a random sample of size 9 is taken, and the population is assumed to be normally distributed, what is the appropriate rejection region? $t \leq -1.860$	Correct

Comment:

For a one-sided, lower-tailed test, the rejection region is determined by the *t* value that has an area to the left, under a *t* distribution with 8 degrees of freedom, of 0.05. Using computer software or a *t* distribution table, we can find this value to be -1.860. Graphically, the rejection region is seen as:

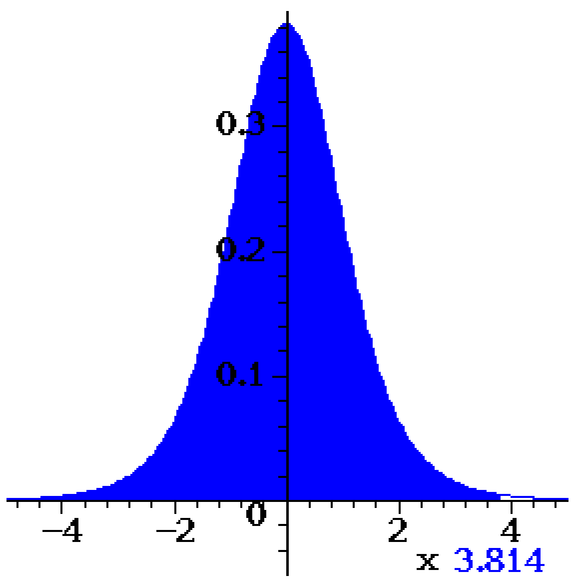


Question 8: *Score 1/1*

Your response	Correct response	
Suppose the null hypothesis of $H_0 : \mu = \mu_0$ is being tested against the one-sided alternative hypothesis $H_A : \mu < \mu_0$ in a population that is assumed to be normally distributed. If a random sample of size 7 is taken, and the *t* test statistic is calculated to be *t* = 3.814, then: a) The p-value falls within which one of the following ranges: **p-value > 0.50** (50%) b) Is there a significant amount of evidence against the null hypothesis at each of the 10%, 5% and 1% levels of significance? **No** (50%)	Suppose the null hypothesis of $H_0 : \mu = \mu_0$ is being tested against the one-sided alternative hypothesis $H_A : \mu < \mu_0$ in a population that is assumed to be normally distributed. If a random sample of size 7 is taken, and the *t* test statistic is calculated to be *t* = 3.814, then: a) The p-value falls within which one of the following ranges: **p-value > 0.50** b) Is there a significant amount of evidence against the null hypothesis at each of the 10%, 5% and 1% levels of significance? **No**	**Correct**

Comment:

a) The alternative hypothesis indicates a one-sided, lower-tailed test. Therefore, the p-value is the area to the left of the test statistic, under a *t* distribution with *6* degrees of freedom. Graphically, this becomes:



Using computer software or a *t* distribution table, we can find this area to be *0.995589*.

b) Because the p-value is extremely large (close to 1.0), there is no significant evidence at any of the 10%, 5% or 1% levels of significance against the null hypothesis.

Question 9: *Score 1/1*

Your response

Suppose a random sample of size 21 is taken from a normally distributed population, and the sample mean and variance are calculated to be $\bar{x} = 5.31$ and $s^2 = 0.5$ respectively.

Use this information to test the null hypothesis $H_0 : \mu = 5$ versus the alternative hypothesis $H_A : \mu > 5$.

a) What is the value of the test statistic, for testing the null hypothesis that the population mean is equal to 5?

Round your response to at least 3 decimal places.

2.009 (25%)

b) The p-value falls within which one of the following ranges:

0.025 < p-value < 0.05 (25%)

c) i) Is the null hypothesis rejected at the 5% level of significance?

Yes (25%)

ii) Is the null hypothesis rejected at the 1% level of significance?

No (25%)

Correct response

Suppose a random sample of size 21 is taken from a normally distributed population, and the sample mean and variance are calculated to be $\bar{x} = 5.31$ and $s^2 = 0.5$ respectively.

Use this information to test the null hypothesis $H_0 : \mu = 5$ versus the alternative hypothesis $H_A : \mu > 5$.

a) What is the value of the test statistic, for testing the null hypothesis that the population mean is equal to 5?

Round your response to at least 3 decimal places.

2.009

b) The p-value falls within which one of the following ranges:

0.025 < p-value < 0.05

c) i) Is the null hypothesis rejected at the 5% level of significance?

Yes

ii) Is the null hypothesis rejected at the 1% level of significance?

No

Correct

Comment:

a) To calculate the *t* test statistic, we use the formula $t = \frac{(\bar{x} - \mu)}{\frac{s}{\sqrt{n}}}$. Substituting in the appropriate values, we get

$$t = \frac{(5.31 - 5)}{\frac{\sqrt{0.5}}{\sqrt{21}}} = 2.00903 .$$

b) The alternative hypothesis indicates we are conducting a one-sided, upper-tailed test. Therefore, the p-value is the area to the right of the test statistic, under the *t* distribution with *21 - 1 =20*. Using computer software or a *t* distribution table, we can find the p-value to be *0.029112*.

c) i) The p-value = 0.029112 is less than $\alpha = 0.05$, therefore there is sufficient evidence to reject the null hypothesis at the 5% level of significance.

ii) The p-value = 0.029112 is greater than $\alpha = 0.01$, therefore there is insufficient evidence to reject the null hypothesis at the 1% level of significance.

Question 10: *Score 1/1*

Your response	Correct response	
Suppose a random sample of size 16 was taken from a normally distributed population, and the sample standard deviation was calculated to be *s* = 6.3. a) Calculate the margin of error for a 95% confidence interval for the population mean. Round your response to at least 3 decimal places. **3.357** (50%) b) Calculate the margin of error for a 90% confidence interval for the population mean. Round your response to at least 3 decimal places. **2.761** (50%)	Suppose a random sample of size 16 was taken from a normally distributed population, and the sample standard deviation was calculated to be *s* = 6.3. a) Calculate the margin of error for a 95% confidence interval for the population mean. Round your response to at least 3 decimal places. **3.357** b) Calculate the margin of error for a 90% confidence interval for the population mean. Round your response to at least 3 decimal places. **2.761**	Correct

Comment:

a) To determine the margin of error, we first need to determine the $t_{\frac{\alpha}{2}}$ value. For a *t* distribution with 15 degrees of freedom, for a 95% confidence interval for the mean this value is $t_{\frac{\alpha}{2}} = 2.131449$. Therefore, the margin of error for a 95% confidence interval for the population mean is $ME = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 2.131449 \cdot \frac{6.3}{\sqrt{16}} = 3.357032$.

b) For a 90% confidence interval for the population mean, $t_{\frac{\alpha}{2}} = 1.75305$, and therefore

$$ME = t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1.75305 \cdot \frac{6.3}{\sqrt{16}} = 2.761054.$$

Quiz 7: Inference for Two Population Means

Question 1: *Score 1/1*

Which of the following statements are true?

Note that there may be more than one correct answer; select all that are true.

Choice	Selected	✔/✘	Points
If the sample sizes are similar, the pooled-variance t procedure will still work relatively well, even if the population variances are not quite the same.	Yes	✔	+1
The pooled-variance t procedure is most appropriate when the observations between the two groups are dependent.	No		
The pooled sample variance used in a pooled-variance t procedure is a weighted average of the sample variances, and tends to be closer to the sample variance with the higher number of observations.	Yes	✔	+1
When applying the paired difference procedure, it is assumed that the differences between pairs of observations constitute a simple random sample from the population of differences.	Yes	✔	+1
The pooled-variance t procedure requires that the two populations be normally distributed, however because the Welch procedure is only an approximate procedure, it does not require this assumption.	No		

✔ Correct

Number of available correct choices: 3

Partial Grading Explained

Question 2: *Score 1/1*

Your response	Correct response
Consider the following summary statistics that were calculated on the ***difference*** between two dependent random samples, obtained from two normally distributed populations:	Consider the following summary statistics that were calculated on the ***difference*** between two dependent random samples, obtained from two normally distributed populations:

Summary Statistic
$n = 15$
$\bar{x}_D = -10.1$
$s_D = 41.9$

Test the null hypothesis $H_0 : \mu_D = 0$ against the alternative hypothesis $H_A : \mu_D < 0$.

a) Calculate the value of the *t* test statistic.

Round your response to at least 3 decimal places.

-0.934 (33%)

b) What is the range in which the p-value falls?

p-value > 0.10 (33%)

c) Is there sufficient evidence, at the 10% level of significance, to reject the null hypothesis, in favour of the alternative hypothesis?

No (33%)

Summary Statistic
$n = 15$
$\bar{x}_D = -10.1$
$s_D = 41.9$

Test the null hypothesis $H_0 : \mu_D = 0$ against the alternative hypothesis $H_A : \mu_D < 0$.

a) Calculate the value of the *t* test statistic.

Round your response to at least 3 decimal places.

-0.934

b) What is the range in which the p-value falls?

p-value > 0.10

c) Is there sufficient evidence, at the 10% level of significance, to reject the null hypothesis, in favour of the alternative hypothesis?

No

Correct

Comment:

a) The formula for the test statistic is given by the formula $t = \dfrac{(\bar{x}_D - \mu_D)}{\dfrac{s_D}{\sqrt{n}}}$. Substituting in the appropriate values, we get

$$t = \frac{(-10.1 - 0)}{\frac{41.9}{\sqrt{15}}} = -0.933583 .$$

b) The alternative hypothesis indicates that we are performing a one-sided, lower tailed test, which means that the p-value is the area to the left of the test statistic, under a *t* distribution with *15 - 1 = 14* degrees of freedom. Using computer software, we can find this value to be exactly *p-value = 0.183171*.

c) Since the p-value found in part (b) is greater than $\alpha = 0.10$, there is insufficient evidence to reject the null hypothesis.

Question 3: *Score 1/1*

Your response	Correct response	
Consider the following set of random measurements, taken from a normally distributed population before and after a treatment was applied.	Consider the following set of random measurements, taken from a normally distributed population before and after a treatment was applied.	

ent	[57.34, 56.91, 50.49, 52.82, 52.81, 54.57, 56.99, 50.52	**ent**	[57.34, 56.91, 50.49, 52.82, 52.81, 54.57, 56.99, 50.52
t	[61.25, 59.14, 58.4, 55.56, 64.57, 56.3, 63.05, 56.84, 5	**t**	[61.25, 59.14, 58.4, 55.56, 64.57, 56.3, 63.05, 56.84, 5
	[-3.91, -2.23, -7.91, -2.74, -11.76, -1.73, -6.06, -6.32, -3		[-3.91, -2.23, -7.91, -2.74, -11.76, -1.73, -6.06, -6.32, -3

Your response	Correct response	
a) Determine the point estimate for the mean difference. Round your response to at least 3 decimal places. **-5.168** (33%) b) Calculate the standard error of the sample mean difference. Round your response to at least 3 decimal places. **1.072** (33%) c) What is the margin of error for a 95% confidence interval for the mean difference? Round your response to at least 3 decimal places. **2.472** (33%)	a) Determine the point estimate for the mean difference. Round your response to at least 3 decimal places. **-5.168** b) Calculate the standard error of the sample mean difference. Round your response to at least 3 decimal places. **1.072** c) What is the margin of error for a 95% confidence interval for the mean difference? Round your response to at least 3 decimal places. **2.472**	✔ **Correct**

Comment:

a) A point estimate for the mean difference is found by using the formula $\bar{x}_D = \frac{\sum_{i=1}^{9} x_i}{9}$, where x_i is value for the difference.

Substituting in the appropriate values, we get $\bar{x}_D = -5.167778$.

b) The standard error for the mean difference is found by dividing the standard deviation of the differences by the number of observations, such as $SE(\bar{x}_D) = \frac{s_D}{\sqrt{n}}$. The standard deviation of the differences can be found by taking the square root of the variance, $s_D = \sqrt{\frac{\sum_{i=1}^{9}(x_i - \bar{x}_D)^2}{8}}$. Using the value found in part (a) for $\bar{x}_D$, we get $s_D = 3.21641$. Therefore, the standard error is $SE(\bar{x}_D) = \frac{3.21641}{\sqrt{9}} = 1.072137$.

c) The margin of error for a 95% confidence interval for the mean difference is given by $ME = t_{\frac{\alpha}{2}} \cdot SE(\bar{x}_D)$. The degrees of freedom for the *t* distribution is *n - 1 = 9 - 1 = 8,* and therefore for a 95% confidence interval, $t_{\frac{\alpha}{2}} = 2.306004$. Finally, the margin of error can be calculated as $ME = 2.306004 \cdot 1.072137 = 2.472352$.

Question 4: *Score 1/1*

Your response	Correct response	
Consider two independent random samples of sizes $n_1 = 15$ and $n_2 = 14$, taken from normally distributed populations, with sample standard deviations of $s_1 = 5.2$ and $s_2 = 4.1$, respectively. a) What is s_p^2, the estimate of the pooled variance? Round your response to at least 3 decimal places. **22.114** (50%) b) What are the appropriate degrees of freedom for the *t* value? **27** (50%)	Consider two independent random samples of sizes $n_1 = 15$ and $n_2 = 14$, taken from normally distributed populations, with sample standard deviations of $s_1 = 5.2$ and $s_2 = 4.1$, respectively. a) What is s_p^2, the estimate of the pooled variance? Round your response to at least 3 decimal places. **22.114** b) What are the appropriate degrees of freedom for the *t* value? **27**	Correct

Comment:

a) The pooled variance, s_p^2, is calculated by the formula $s_p^2 = \frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2}$. Substituting in the appropriate values, we get $s_p^2 = \frac{(15 - 1)\cdot 5.2^2 + (14 - 1)\cdot 4.1^2}{15 + 14 - 2} = 22.114444$.

b) The degrees of freedom for the pooled-variance *t* procedure are $n_1 + n_2 - 2$. Therefore, in this case the degrees of freedom are $15 + 14 - 2 = 27$.

Question 5: *Score 1/1*

Your response	Correct response
Consider the following summary statistics, calculated from two independent random samples taken from normally distributed populations.	Consider the following summary statistics, calculated from two independent random samples taken from normally distributed populations.

Sample 1	Sample 2
$\bar{x}_1 = 5.308$	$\bar{x}_2 = 7.648$
$s_1^2 = 2.998$	$s_2^2 = 0.7692$
$n_1 = 7$	$n_2 = 10$

Sample 1	Sample 2
$\bar{x}_1 = 5.308$	$\bar{x}_2 = 7.648$
$s_1^2 = 2.998$	$s_2^2 = 0.7692$
$n_1 = 7$	$n_2 = 10$

Your response	Correct response
You are interested in determining if there is sufficient evidence of a difference between the population means.	You are interested in determining if there is sufficient evidence of a difference between the population means.
a) What are the appropriate null and alternative hypotheses?	a) What are the appropriate null and alternative hypotheses?
$H_0 : \mu_1 = \mu_2,\ H_A : \mu_1 \neq \mu_2$ (33%)	$H_0 : \mu_1 = \mu_2,\ H_A : \mu_1 \neq \mu_2$
b) Calculate the test statistic for the pooled-variance *t* procedure.	b) Calculate the test statistic for the pooled-variance *t* procedure.
Round your response to at least 3 decimal places.	Round your response to at least 3 decimal places.
-3.685 (33%)	**-3.685**
c) Determine the range in which the p-value falls:	c) Determine the range in which the p-value falls:
p-value < 0.025 (33%)	**p-value < 0.025**

Correct

Comment:

a) Since we are only interested in determining if there is a difference between the population means, the implication is that we are conducting a two-sided hypothesis test. Therefore, the appropriate null and alternative hypotheses are $H_0 : \mu_1 = \mu_2$ and $H_A : \mu_1 \neq \mu_2$.

b) The *t* test statistic is given by the formula $t = \dfrac{(\bar{X}_1 - \bar{X}_2)}{SE(\bar{X}_1 - \bar{X}_2)}$, where $SE(\bar{X}_1 - \bar{X}_2) = s_p \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$. To calculate the pooled standard deviation, we can first calculate the pooled variance:

$$s_p^2 = \frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2} = \frac{(7 - 1)\cdot 2.998 + (10 - 1)\cdot 0.7692}{7 + 10 - 2} = 1.66072$$

, and therefore $s_p = \sqrt{1.66072} = 1.288689$. Substituting these values into the equation for the test statistic, along with the other given information, results in

$$t = \frac{(5.308 - 7.648)}{1.288689 \cdot \sqrt{\dfrac{1}{7} + \dfrac{1}{10}}} = -3.684616 .$$

c) Since the alternative hypothesis indicates that we are conducting a two-sided hypothesis test, the p-value is determined as $2 \cdot P(t > |-3.684616|)$, where *t* follows a *t* distribution with *7 + 10 - 2 = 15* degrees of freedom. Using computer software, we can find the area in the tail to be exactly *0.001104,* and therefore the p-value is *2 X 0.001104 = 0.002208*.

Question 6: *Score 1/1*

Your response	Correct response
Consider two independent random samples of sizes $n_1 = 11$ and $n_2 = 9$, taken from normally distributed populations with sample standard deviations of $s_1 = 3.5$ and $s_2 = 2.4$, respectively.	Consider two independent random samples of sizes $n_1 = 11$ and $n_2 = 9$, taken from normally distributed populations with sample standard deviations of $s_1 = 3.5$ and $s_2 = 2.4$, respectively.
a) What is s_p^2, the estimate of the pooled variance?	a) What is s_p^2, the estimate of the pooled variance?
Round your response to at least 3 decimal places.	Round your response to at least 3 decimal places.
9.366 (33%)	**9.366**
b) What is the appropriate $t_{\frac{\alpha}{2}}$ value for a 95% confidence interval for $\mu_1 - \mu_2$?	b) What is the appropriate $t_{\frac{\alpha}{2}}$ value for a 95% confidence interval for $\mu_1 - \mu_2$?
Round your response to at least 3 decimal places.	Round your response to at least 3 decimal places.
2.101 (33%)	**2.101**
c) Using your responses from parts (a) and (b), what is the value for the margin of error?	c) Using your responses from parts (a) and (b), what is the value for the margin of error?
Round your response to at least 3 decimal places.	Round your response to at least 3 decimal places.
2.890 (33%)	**2.890**

Correct

Comment:

a) The formula for the pooled variance is $s_p^2 = \dfrac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2}$. Substituting in the appropriate values, we can calculate the pooled variance to be $s_p^2 = \dfrac{(11 - 1)\cdot 3.5^2 + (9 - 1)\cdot 2.4^2}{11 + 9 - 2} = 9.365556$.

b) The appropriate degrees of freedom for the *t* distribution to be used is $n_1 + n_2 - 2 = 11 + 9 - 2 = 18$. Therefore, for a 95% confidence interval $t_{\frac{\alpha}{2}} = 2.100922$.

c) The margin of error for a 95% confidence interval for the difference between two means is given as $ME = t_{\frac{\alpha}{2}} \cdot SE\left(\bar{X}_1 - \bar{X}_2\right)$, where $SE\left(\bar{X}_1 - \bar{X}_2\right) = s_p \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$. Substituting in the values from parts (a) and (b), we can calculate the margin of error to be $ME = 2.100922 \cdot \sqrt{9.365556} \cdot \sqrt{\dfrac{1}{11} + \dfrac{1}{9}} = 2.889842$.

Question 7: *Score 1/1*

Your response	Correct response	
Independent random samples of approximately the same size ($n_1 = 33$ and $n_2 = 35$, respectively) are drawn from two normally distributed populations with known variances of $\sigma_1^2 = 2.3$ and $\sigma_2^2 = 2.0$, respectively. The sample means are calculated to be $\bar{x}_1 = 35.0$ and $\bar{x}_2 = 33.5$. Use this information to test the null hypothesis $H_0 : \mu_1 = \mu_2$ against the alternative hypothesis $H_A : \mu_1 \neq \mu_2$. a) What is the value of the test statistic? Round your answer to at least 3 decimal places. **4.212** (33%) b) At the 5% level of significance, what is the appropriate rejection region for the test? $\lvert Z \rvert \geq 1.96$ (33%) c) Is there sufficient evidence to reject the null hypothesis at the 5% level of significance? **Yes** (33%)	Independent random samples of approximately the same size ($n_1 = 33$ and $n_2 = 35$, respectively) are drawn from two normally distributed populations with known variances of $\sigma_1^2 = 2.3$ and $\sigma_2^2 = 2.0$, respectively. The sample means are calculated to be $\bar{x}_1 = 35.0$ and $\bar{x}_2 = 33.5$. Use this information to test the null hypothesis $H_0 : \mu_1 = \mu_2$ against the alternative hypothesis $H_A : \mu_1 \neq \mu_2$. a) What is the value of the test statistic? Round your answer to at least 3 decimal places. **4.212** b) At the 5% level of significance, what is the appropriate rejection region for the test? $\lvert Z \rvert \geq 1.96$ c) Is there sufficient evidence to reject the null hypothesis at the 5% level of significance? **Yes**	**Correct**

Comment:

a) The z test statistic is given by the formula $z = \dfrac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$. Substituting in the appropriate values, we can calculate the test statistic to be $z = \dfrac{(35.0 - 33.5)}{\sqrt{\dfrac{2.3}{33} + \dfrac{2.0}{35}}} = 4.211758$.

b) Since the alternative hypothesis indicates we are conducting a two-sided hypothesis test, the rejection region is determined by a z value such that the areas above z and below $-z$ are each equal to 0.025. Using computer software, or approximating with a standard normal table, this z value is equal to +/- 1.96. Therefore, the rejection region is $\lvert Z \rvert \geq 1.96$.

c) Since the z test statistic is greater than 1.96, there is sufficient evidence to reject the null hypothesis at the 5% level of significance.

Question 8: *Score 1/1*

Your response	Correct response
Consider the following summary statistics, obtained from independent samples drawn from two normally distributed populations:	Consider the following summary statistics, obtained from independent samples drawn from two normally distributed populations:

SAMPLE 1	SAMPLE 2
$n_1 = 54$	$n_2 = 13$
$s_1 = 4.55$	$s_2 = 6.68$

a) Calculate $SE(\bar{X}_1 - \bar{X}_2)$, using the Welch approximate *t* procedure.
Round your response to at least 3 decimal places.

Your response: **1.953** (33%)
Correct response: **1.953**

Correct

b) Calculate the pooled standard deviation, s_p.
Round your response to at least 3 decimal places.

Your response: **5.012** (33%)
Correct response: **5.012**

c) Which method, the Welch approximate *t* procedure or the pooled-variance *t* procedure, is the most appropriate to use in this situation?

Your response: **The Welch Approximate t Procedure** (33%)
Correct response: **The Welch Approximate t Procedure**

Comment:

a) Using the Welch Approximate *t* procedure, $SE(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$. Substituting in the appropriate values leads to the calculation of the standard error of the difference in sample means to be

$$SE(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{4.55^2}{54} + \frac{6.68^2}{13}} = 1.953426.$$

b) The estimate of the pooled variance is given by the formula $s_p^2 = \frac{(n_1 - 1)\cdot s_1^2 + (n_2 - 1)\cdot s_2^2}{n_1 + n_2 - 2}$. Substituting in the appropriate values result in $s_p^2 = \frac{(54 - 1)\cdot 4.55^2 + (13 - 1)\cdot 6.68^2}{54 + 13 - 2} = 25.118482$, and therefore the estimate of the pooled standard deviation is $s_p = \sqrt{25.118482} = 5.011834$.

c) The more appropriate method to use in this situation is the Welch Approximate *t* procedure, since it is not known if the population variances are equal, and the sample sizes are very different from each other.

Question 9: *Score 1/1*

Your response	Correct response
Consider the following summary statistics, taken on two independent random samples drawn from normally distributed populations:	Consider the following summary statistics, taken on two independent random samples drawn from normally distributed populations:

SAMPLE 1	SAMPLE 2
$n_1 = 24$	$n_2 = 12$
$s_1^2 = 6.0516$	$s_2^2 = 26.2144$
$\bar{x}_1 = 3.95$	$\bar{x}_2 = 4.42$

SAMPLE 1	SAMPLE 2
$n_1 = 24$	$n_2 = 12$
$s_1^2 = 6.0516$	$s_2^2 = 26.2144$
$\bar{x}_1 = 3.95$	$\bar{x}_2 = 4.42$

Your response	Correct response
Using the Welch approximate *t* procedure, test the null hypothesis $H_0 : \mu_1 = \mu_2$ against the one-sided alternative $H_A : \mu_1 > \mu_2$.	Using the Welch approximate *t* procedure, test the null hypothesis $H_0 : \mu_1 = \mu_2$ against the one-sided alternative $H_A : \mu_1 > \mu_2$.
a) Calculate the value for the *t* test statistic. Round your response to at least 3 decimal places.	a) Calculate the value for the *t* test statistic. Round your response to at least 3 decimal places.
-0.301 (33%)	**-0.301**
b) Using a conservative estimate of the degrees of freedom (i.e. the lesser of n_1 - 1 and n_2 - 1), the p-value is within which one of the following ranges?	b) Using a conservative estimate of the degrees of freedom (i.e. the lesser of n_1 - 1 and n_2 - 1), the p-value is within which one of the following ranges?
p-value > 0.10 (33%)	**p-value > 0.10**
c) Is the null hypothesis rejected at the 10% level of significance? **No** (33%)	c) Is the null hypothesis rejected at the 10% level of significance? **No**

Correct

Comment:

a) Using the Welch Approximate *t* procedure, the *t* test statistic is given by the equation $t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$. Substituting in the appropriate values, we can calculate the test statistic to be $t = \frac{(3.95 - 4.42)}{\sqrt{\frac{6.0516}{24} + \frac{26.2144}{12}}} = -0.301091$.

b) The conservative estimate of the degrees of freedom is the lesser of $n_1 - 1 = 24 - 1$ and $n_2 - 1 = 12 - 1$, which in this case is 11. The alternative hypothesis indicates that we are performing a one-sided, upper tailed test, which means that the p-value is determined as the area to the right of the test statistic, under a *t* distribution with 11 degrees of freedom. Using computer software, we can find this area to be *p-value = 0.61552.*

c) Since the p-value is greater than $\alpha = 0.10$, there is insufficient evidence to reject the null hypothesis at the 10% level of significance.

Question 10: *Score 1/1*

Your response	Correct response	
Consider two independent random samples of sizes $n_1 = 15$ and $n_2 = 14$, taken from two normally distributed populations. The sample standard deviations are calculated to be $s_1 = 5.8$ and $s_2 = 9.5$, and the sample means are $\bar{x}_1 = 71.06$ and $\bar{x}_2 = 76.16$, respectively. a) Calculate a point estimate for the difference in population means, $\mu_1 - \mu_2$. Round your response to at least 3 decimal places. **-5.100** (33%) b) Using the Welch approximate *t* procedure and a conservative estimate of the degrees of freedom (i.e. the lesser of $n_1 - 1$ and $n_2 - 1$), what is the appropriate $t_{\frac{\alpha}{2}}$ value for a 99% confidence interval? Round your response to at least 3 decimal places. **3.012** (33%) c) Calculate the margin of error for a 99% confidence interval. Round your response to at least 3 decimal places. **8.879** (33%)	Consider two independent random samples of sizes $n_1 = 15$ and $n_2 = 14$, taken from two normally distributed populations. The sample standard deviations are calculated to be $s_1 = 5.8$ and $s_2 = 9.5$, and the sample means are $\bar{x}_1 = 71.06$ and $\bar{x}_2 = 76.16$, respectively. a) Calculate a point estimate for the difference in population means, $\mu_1 - \mu_2$. Round your response to at least 3 decimal places. **-5.100** b) Using the Welch approximate *t* procedure and a conservative estimate of the degrees of freedom (i.e. the lesser of $n_1 - 1$ and $n_2 - 1$), what is the appropriate $t_{\frac{\alpha}{2}}$ value for a 99% confidence interval? Round your response to at least 3 decimal places. **3.012** c) Calculate the margin of error for a 99% confidence interval. Round your response to at least 3 decimal places. **8.879**	✔ **Correct**

Comment:

a) A point estimate for $\mu_1 - \mu_2$ is given by $\bar{X}_1 - \bar{X}_2 = 71.06 - 76.16 = -5.1$.

b) The conservative estimate of the degrees of freedom is the smaller of $n_1 - 1 = 15 - 1$ and $n_2 - 1 = 14 - 1$, which in this case is 13. Therefore, for a 99% confidence interval for the difference in population means, $t_{\frac{\alpha}{2}} = 3.012266$.

c) Using the Welch Approximate *t* procedure, the $SE\left(\bar{X}_1 - \bar{X}_2\right) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{5.8^2}{15} + \frac{9.5^2}{14}} = 2.947727$.

Therefore, the margin of error for a 99% confidence interval for the difference in population means is

$ME = t_{\frac{\alpha}{2}} \cdot SE\left(\bar{X}_1 - \bar{X}_2\right) = 3.012266 \cdot 2.947727 = 8.87934$.

Quiz 8: ANOVA and Linear Regression

Question 1: *Score 1/1*

Your response

Researchers are interested in investigating the effect of four different treatments on a certain characteristic of interest. Random samples of individuals are drawn and assigned to the various treatment groups. The table below displays the summary statistics obtained upon the conclusion of the experiment.

Group 1	$s_1^2 = 19.3$	$\bar{x}_1 = 28.35$	$n_1 = 8$
Group 2	$s_2^2 = 7.24$	$\bar{x}_2 = 43.32$	$n_2 = 7$
Group 3	$s_3^2 = 26.2$	$\bar{x}_3 = 54.14$	$n_3 = 9$
Group 4	$s_4^2 = 22.9$	$\bar{x}_4 = 27.64$	$n_4 = 6$

A one-way ANOVA was carried out to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against the alternative hypothesis $H_A : \text{At least one } \mu_i \text{ is different}$, with the resulting p-value being small enough to reject the null hypothesis.

Using Fisher's Least Significant Difference (LSD) method, determine if there is a significant difference between Group 2 and Group 3, based on the calculation of a 95% confidence interval.

Correct

a) Calculate the value of s_p^2.

Round your response to at least 3 decimal places.

19.332 (33%)

b) Calculate the margin of error for a 95% confidence interval for $\mu_2 - \mu_3$, using Fisher's LSD method.

Round your response to at least 3 decimal places.

4.555 (33%)

c) Based on the confidence interval calculated in part (b), is there significant evidence of a difference between μ_2 and μ_3?

Yes, there is significant evidence of a difference. (33%)

Comment:

a) The pooled variance is calculated with the equation $s_p^2 = MSE = \frac{SSE}{df_{\text{error}}} = \frac{\sum_{i=1}^{4}(n_i - 1)\cdot s_i^2}{N - k}$. Substituting in the appropriate values results in a pooled variance of

$$s_p^2 = \frac{((8-1)\cdot 19.3 + (7-1)\cdot 7.24 + (9-1)\cdot 26.2 + (6-1)\cdot 22.9)}{30 - 4} = 19.332308.$$

b) The margin of error for a 95% confidence interval for $\mu_2 - \mu_3$ is given by $ME = t_{\frac{\alpha}{2}} \cdot SE\left(\bar{X}_2 - \bar{X}_3\right)$, where

$SE\left(\bar{X}_2 - \bar{X}_3\right) = s_p \cdot \sqrt{\frac{1}{n_2} + \frac{1}{n_3}}$. Substituting in the appropriate values results in a standard error of

$SE\left(\bar{X}_2 - \bar{X}_3\right) = \sqrt{19.332308} \cdot \sqrt{\frac{1}{7} + \frac{1}{9}} = 2.215805$. For a *t* distribution with 26 degrees of freedom, the corresponding

$t_{\frac{\alpha}{2}}$ value is 2.055529, and therefore the margin of error is calculated as $ME = 2.055529 \cdot 2.215805 = 4.554653$.

c) In order to determine if there is evidence of a significant difference between μ_2 and μ_3 , we need to calculate the upper and lower bounds of the 95% confidence interval for $\mu_2 - \mu_3$, which is calculated by

$\left(\bar{X}_2 - \bar{X}_3\right) \pm t_{\frac{\alpha}{2}} \cdot SE\left(\bar{X}_2 - \bar{X}_3\right) \Rightarrow (43.32 - 54.14) \pm 4.554653$. Therefore, the 95% confidence interval for the difference between population means is *(-15.374653, -6.265347)*. Since this confidence interval does not contain 0, there is significant evidence of a difference between the means for Group 2 and Group 3.

Question 2: *Score 1/1*

Your response

A random sample of 30 individuals is selected from a population, and these individuals are then randomly assigned to 3 treatment groups (such that each group has an equal number of individuals).

The following ANOVA table summarizes the results of measurements taken from each individual, within each group, on a characteristic of interest. Due to the small p-value, the null hypothesis is rejected, and at least one of the treatments means is different.

Source	DF	SS	MS	F	p-value
Groups	2	2,195.71	1,097.855	8.113472	0.001741
Error	27	3,653.44	135.312593	---	
Total	29	5,849.15	---	---	

The sample means for each of the treatment groups were found to be $\bar{x}_1 = 22.60,\ \bar{x}_2 = 16.82$, and $\bar{x}_3 = 10.18$.

Use Fisher's Least Significant Difference (LSD) method to carryout pairwise comparisons between the treatment means, and determine which means are different.

Correct

a) What is the value of the standard error of the difference in treatment means, $SE\left(\bar{X}_i - \bar{X}_j\right)$?

Round your response to at least 3 decimal places.

5.202 (20%)

b) What are the appropriate degrees of freedom for the *t* distribution used to find $t_{\frac{\alpha}{2}}$?

27 (20%)

c) Determine whether or not there is significant evidence of a difference between each pair of treatment means, at the 5% level of significance:

Comparison	Significant Difference
Group 1 - Group 2	**No** (20%)
Group 1 - Group 3	**Yes** (20%)
Group 2 - Group 3	**No** (20%)

Comment:

a) To calculate the standard error for 95% confidence intervals using the LSD method, we can use the formula $SE\left(\bar{x}_i - \bar{x}_j\right) = s_p\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$, where $s_p = \sqrt{MSE}$, which can be obtained from the given ANOVA table. Note that because there are the same number of individuals in each group, $n_i = n_j$ for all values of *i* and *j*. Substituting in the appropriate values, we get

$$SE\left(\bar{x}_i - \bar{x}_j\right) = \sqrt{135.312593} \cdot \sqrt{\frac{1}{10} + \frac{1}{10}} = 5.202165$$

b) The appropriate degrees of freedom for $t_{\frac{\alpha}{2}}$ for the 95% confidence intervals using the LSD method is *30 - 3 = 27*, the degrees of freedom for error in the ANOVA table.

c) The 95% confidence intervals for pairwise comparisons, using the LSD method, are given by the formula $\left(\bar{X}_i - \bar{X}_j\right) \pm t_{\frac{\alpha}{2}} \cdot SE\left(\bar{X}_i - \bar{X}_j\right)$. Using the value for the standard error found in part (a), and $t_{\frac{\alpha}{2}} = 2.051831$, we can find the lower and upper limits of the confidence intervals:

Group 1 - Group 2: $(22.60 - 16.82) \pm 2.051831 \cdot 5.202165 \Rightarrow (-4.893961, 16.453961)$

Group 1 - Group 3: $(22.60 - 10.18) \pm 2.051831 \cdot 5.202165 \Rightarrow (1.746039, 23.093961)$

Group 2 - Group 3: $(16.82 - 10.18) \pm 2.051831 \cdot 5.202165 \Rightarrow (-4.033961, 17.313961)$

The confidence interval for Group 1 - Group 2, and Group 2 - Group 3 contains 0, therefore there is no significant difference between Groups 1 and 2, and between Groups 2 and 3. However, the confidence interval for Group 1 - Group 3 does not contain 0, therefore there is evidence of a significant difference between Group 1 and Group 3.

Question 3: *Score 1/1*

Your response

Researchers are interested in investigating the effect of four different treatments on a particular characteristic. Random samples of individuals are drawn, and assigned to various treatment groups. The table below displays the summary statistics obtained upon the conclusion of the experiment.

Group 1	$s_1^2 = 17.4$	$\bar{x}_1 = 25.63$	$n_1 = 8$
Group 2	$s_2^2 = 5.13$	$\bar{x}_2 = 42.38$	$n_2 = 7$
Group 3	$s_3^2 = 27.3$	$\bar{x}_3 = 53.64$	$n_3 = 10$
Group 4	$s_4^2 = 22.2$	$\bar{x}_4 = 30.70$	$n_4 = 8$
SSTotal = 1,269.68			

Use this information to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against the alternative hypothesis $H_A : \textit{At least one } \mu_i \textit{ is different}$.

Correct

a) Calculate the value of MSE.

Round your response to at least 3 decimal places.

19.092 (33%)

b) What is the value of the *F* test statistic?

Round your response to at least 3 decimal places.

12.501 (33%)

c) What conclusion can be made, at the 5% level of significance?

There is sufficient evidence to reject the null hypothesis, in favour of the alternative hypothesis that at least one population mean is different. (33%)

Comment:

a) The mean square error (MSE) is calculated with the equation $MSE = \frac{SSE}{df_{\mathbf{error}}} = \frac{\sum_{i=1}^{4}(n_i - 1)\cdot s_i^2}{N - k}$. Substituting in the appropriate values results in an MSE of

$$MSE = \frac{((8-1)\cdot 17.4 + (7-1)\cdot 5.13 + (10-1)\cdot 27.3 + (8-1)\cdot 22.2)}{33-4} = 19.092414.$$

b) Using the given SST value, along with the MSE value found in part (a), we can complete the ANOVA table as follows:

Source	DF	SS	MS	F
Groups	3	716	238.666667	12.500602
Error	29	553.68	19.092414	----
Total	32	1,269.68	----	----

Recall that if $MSE = \frac{SSE}{df_{\mathbf{error}}}$, then $SSE = MSE \cdot df_{\mathbf{error}}$. Therefore, we could find $SSE = 19.092414 \cdot 29 = 553.68$; using this, we can then find $SSG = SST - SSE = 1,269.68 - 553.68 = 716$. From this point, the remainder of the ANOVA table can be easily computed.

c) To determine whether or not there is sufficient evidence to reject the null hypothesis, we need to determine the p-value. In ANOVA, the p-value is the area under an *F* distribution, with 3 and 29 degrees of freedom, to the right of the test statistic. Using computer software, or approximating with an *F* distribution table, the p-value is found to be 2.013179E-5. As this value is less than $\alpha = 0.05$, there is sufficient evidence to reject the null hypothesis at the 5% level of significance.

Question 4: *Score 1/1*

Your response

Identify each of the following statements as either true or false.

a) **False** (20%) ANOVA is only valid when there are more than two populations. When there are only two groups being examined, you must use a t test.

b) **False** (20%) The null hypothesis in ANOVA is that the population means are all the same. Therefore, the appropriate alternative hypothesis is that all the population means are different.

c) **False** (20%) If the null hypothesis is true, the F test statistic follows an F distribution with k and n degrees of freedom, where k is the number of populations and n is the number of observations.

d) **False** (20%) An ANOVA table can contain negative values.

e) **False** (20%) If the population means are very different, it is likely that the MSE is much greater than the MST.

Correct

Question 5: *Score 1/1*

Your response

Measurements on two variables, *X* and *Y*, were taken from 17 individuals, with the following summary statistics:

$$\bar{X} = 29.87,\ \bar{Y} = 38.55,\ SS_{xy} = 142.7,\ SS_{xx} = 200.9,\ s^2 = 25.6$$

Use this information to make inferences on the slope for the least squares regression line, $Y = \beta_0 + \beta_1 \cdot X + \epsilon$.

a) What is the point estimate of the true slope, β_1 ?

Round your response to at least 3 decimal places.

0.710 (50%)

b) What is the margin of error for a 95% confidence interval for the true slope?

Round your response to at least 3 decimal places.

0.761 (50%)

Correct

Comment:

a) To estimate the slope, β_1 , we use the equation $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$. Substituting in the appropriate values gives us

$\hat{\beta}_1 = \frac{142.7}{200.9} = 0.710304$.

b) To estimate the margin of error for a 95% confidence interval for the true slope, we need to first estimate the standard error for our estimate of the slope, given as $SE\left(\hat{\beta}_1\right) = \frac{s}{\sqrt{SS_{xx}}}$. Once the appropriate values have been substituted in, we get

$SE\left(\hat{\beta}_1\right) = \frac{\sqrt{25.6}}{\sqrt{200.9}} = 0.356969$. This value is then used to estimate the margin of error, $ME = t_{\frac{\alpha}{2}} \cdot SE\left(\hat{\beta}_1\right)$, where for a *t* distribution with *n - 2 = 15* degrees of freedom, $t_{\frac{\alpha}{2}} = 2.131449$. Finally, we can calculate the margin of error as

$ME = 2.131449 \cdot 0.356969$.

Question 6: *Score 1/1*

Your response

A random sample of 10 individuals is selected from a population, and measurements on two variables (*X* and *Y*) are obtained, as seen in the table below.

Individual	*X*	*Y*
1	9.71	5.85
2	4.31	3.94
3	6.23	4.78
4	10.1	5.86
5	8.80	4.29
6	10.1	5.32
7	4.64	3.98
8	8.04	5.41
9	7.68	4.18
10	9.90	5.75

The fitted regression line for the linear relationship between *X* and *Y* was found to be

$\hat{Y} = 2.64 + 0.29 \cdot X$.

Assuming all the model assumptions are met, and the inference procedures are valid, then:

a) What is the predicted value of *Y* for the third individual?

Round your response to at least 3 decimal places.

4.447 (50%)

b) What is the value of the residual for the third individual?

Round your response to at least 3 decimal places.

0.333 (50%)

Correct

Comment:

a) The fitted regression line is given as $\hat{Y} = 2.64 + 0.29 \cdot X$, and therefore to find the predicted value of *Y* at the third observation, when *X* = *6.23*, we simply substitute this value into the equation. This results in $\hat{Y} = 2.64 + 0.29 \cdot 6.23 = 4.4467$.

b) The residual for the third observation is given by $e_3 = Y_3 - \hat{Y}_3$, where Y_3 is the observed value of *Y,* and $\hat{Y}_3$ is the predicted value of *Y*, found in part (a). Therefore, the residual for the third observation (i.e. when *X* = *6.23*) is

$e_3 = 4.78 - 4.4467 = 0.3333$.

Question 7: *Score 1/1*

Which of the following statements are true?

Note that there may be more than one correct answer; select all that are true.

Choice	Selected	✔/✘
A 'residual' is the difference between an observed and predicted value of Y, for a particular value of X.	Yes	[answer withheld]
The observed values of Y will fall on the estimated regression line, while the predicted values of Y will vary around the regression line.	No	[answer withheld]
If the purpose of our regression model is prediction, it does not matter which variables we define as the explanatory and response variable.	No	[answer withheld]
The purpose of linear regression is to investigate if there exists a linear relationship between a response variable and one or more explanatory variables.	Yes	[answer withheld]
The sum of the residuals must be 1.	No	[answer withheld]

✔ **Correct**

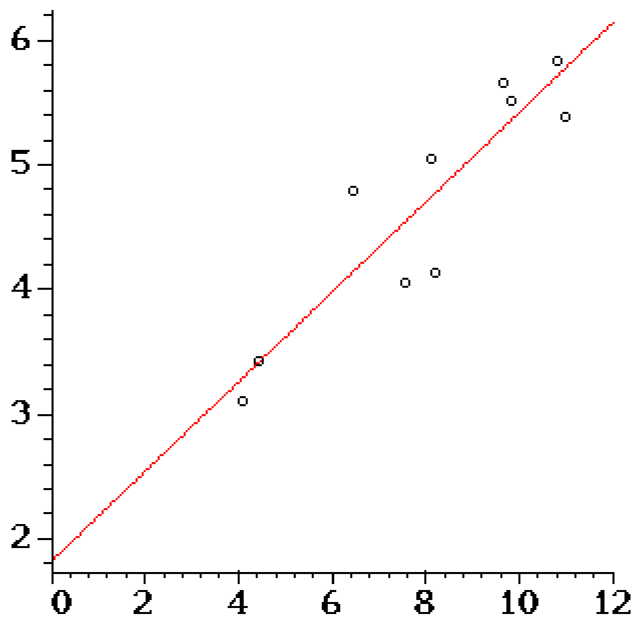
Number of available correct choices: 2

Partial Grading Explained

Question 8: *Score 1/1*

Your response

Researchers are interested in investigating whether or not there is a relationship between two variables, *X* and *Y*. To do so, they obtained a random sample of 10 individuals, measured values of *X* and *Y* on each individual, and fit a linear regression model to the resulting data, as seen in the plot below.



Correct

The fitted regression line was found to be $\hat{Y} = 1.82 + 0.36 \cdot X$.

Assuming all the model assumptions are met, and the inference procedures are valid, then:

a) What is the predicted value of *Y*, at *X = 35*?

Round your response to at least 2 decimal places.

14.42 (50%)

b) Is it reasonable to use the regression line to predict *Y* at *X = 35*?

Select one of the following options that offers the best explanation.

No. This is an example of extrapolation, and should be avoided as it can have misleading results. (50%)

Comment:

a) The fitted regression line is given as $\hat{Y} = 1.82 + 0.36 \cdot X$, and therefore to find the predicted value of *Y* when *X = 35*, we simply substitute this value into the equation. This results in $\hat{Y} = 1.82 + 0.36 \cdot 35 = 14.42$.

b) From the plot of the data, it is clear that an *X* value of 35 is well outside the range of observed data. To predict a value of *Y* at *X = 35* is an example of extrapolation, which should be avoided as it can lead to misleading results. Therefore, it is not reasonable to try and predict *Y* at *X = 35.*