### Question 1

Which of the following would be the scatterplot for the given data?

<table>
<thead>
<tr>
<th>x</th>
<th>4</th>
<th>8</th>
<th>8</th>
<th>13</th>
<th>14</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>21</td>
<td>29</td>
<td>28</td>
<td>40</td>
<td>31</td>
<td>41</td>
</tr>
</tbody>
</table>

![ Scatterplot Diagram ]
Question 2

What can be said about the relationship between the explanatory variable and the response variable in the following scatterplot?

- Strength -> strong
- Direction -> negative
- Form -> linear
- Outliers

- d)
- e) None of the above
a) There is a strong negative linear association.

b) There is a strong positive linear association.

c) There is a weak negative linear association.

d) The explanatory variable causes the responses.

e) None of the above

Question 3

Determine the correlation coefficient for the data shown in this table:

<table>
<thead>
<tr>
<th>x</th>
<th>3</th>
<th>5</th>
<th>10</th>
<th>11</th>
<th>17</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>23</td>
<td>29</td>
<td>27</td>
<td>41</td>
<td>33</td>
<td>44</td>
</tr>
</tbody>
</table>

a) 0.3651

b) 0.5332

c) -0.5332

d) 0.7302

e) -0.7302

f) None of the above

Question 4

Choose the best correlation coefficient for the data shown in this scatterplot:
Question 5

Which of the following is a true statement?

a) The variable that is being predicted in regression analysis is the independent variable.

b) If there is no correlation between the independent and dependent variables, then the value of the correlation coefficient must be -1.

c) The correlation coefficient $r$ is always between -1 and +1.

d) The coefficient of determination can assume negative values.

e) A negative correlation indicates that as values of $x$ decrease, values of $y$ will decrease.

f) None of the above
Question 6

The decline of salmon fisheries along the Columbia River in Oregon has caused great concern among commercial and recreational fishermen. The paper 'Feeding of Predaceous Fishes on Out-Migrating Juvenile Salmonids in John Day Reservoir, Columbia River' (Trans. Amer. Fisheries Soc. (1991: 405-420)) gave the accompanying data on \( y \) = maximum size of salmonids consumed by a northern squaw fish (the most abundant salmonid predator) and \( x \) = squawfish length, both in mm. Use the following statistics to give the equation of the least squares regression line.

\[
\bar{x} = 524.940, \quad \bar{y} = 519.050, \quad s_x = 9.329, \quad s_y = 11.200, \quad r = 0.9662
\]

a) \( \hat{y} = 1.160x - 89.880 \)

b) \( \hat{y} = 0.805x + 89.880 \)

c) \( \hat{y} = 1.160x + 89.880 \)

d) \( \hat{y} = 0.805x - 89.880 \)

e) \( \hat{y} = -89.880x + 1.160 \)

f) None of the above

Least Squares Regression Line (LSRL)

\[ \hat{y} = a + bx \]

\( a = y -intercept \)

\( b = slope \)

\[ b = r \frac{s_y}{s_x} = 0.9662 \left( \frac{11.2}{9.329} \right) = 1.16 \]

\[ a = \bar{y} - b \bar{x} = 519.05 - 1.16(524.94) = 89.88 \]

Eq: \( \hat{y} = 89.88 + 1.16x = 1.16x - 89.88 \)

Question 7

Suppose that you are given the following results. Find the correlation coefficient of the data.

\[ s_x = 1.791, \quad s_y = 13.400, \quad b = -6.480 \]

a) \( 0.155 \)

b) \( -0.433 \)

c) \( -0.866 \)

d) \( 0.866 \)

e) \( -0.155 \)

f) None of the above

Question 8

Suppose you find that the correlation coefficient for a set of data is 0.841. What is the coefficient of determination and what does it mean?

\[ r = 0.841 \]

Correlation Coefficient: \( r = 0.841 \)

Coefficient of determination: \( \begin{align*}
R^2 &= 0.841^2 \\
&= 0.707
\end{align*} \)

a) \( 0.841; \) This means that 84.1% of the variation of \( y \) is explained by the LSRL of \( y \) on \( x \).
b) 0.841; This means that we are 84.1% accurate with our prediction of the LSRL equation.

c) 0.707; This means that 70.7% of the variation of \( y \) is explained by the LSRL of \( y \) on \( x \).

d) 0.707; This means that we are 70.7% accurate with our prediction of the LSRL equation.

e) None of the above

**Question 9**

A real estate agent would like to predict the selling price of a single-family house by predicting the price (in thousands of dollars) based on the square footage (in 100 square feet). If the LSRL for the data is \( \hat{y} = 3.8785x + 18.3538 \), predict the price of a 3000 square foot house (in thousands of dollars).

\[
\hat{y} = 3.8785x + 18.3538
\]

\[
\hat{y} = 3.8785(30) + 18.3538 = 134.709
\]

\( x = \frac{3000}{100} = 30 \)

a) 139.709

b) **134.709**

c) 248.709

d) 252.709

e) 942.962

f) None of the above

**Question 10**

Select the equation of the least squares line for the data: (27.2, 1.50), (26.0, 3.75), (28.0, .75), (24.8, 7.50), (24.0, 6.75), (22.0, 9.75), (23.2, 7.50).

\[
\hat{y} = 1.5206x - 43.434
\]

\[
\hat{y} = -43.434 - 1.5206x
\]

\[
\hat{y} = 43.434 - 1.6727x
\]

\[
\hat{y} = 43.434 - 1.5206x
\]

\[
\hat{y} = 47.777 - 1.6727x
\]

f) None of the above
Quiz 8

Question 1

Which of the following would be the LSRL for the given data?

<table>
<thead>
<tr>
<th>x</th>
<th>4</th>
<th>6</th>
<th>11</th>
<th>13</th>
<th>17</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>23</td>
<td>29</td>
<td>27</td>
<td>42</td>
<td>26</td>
<td>41</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\hat{y} &= 0.7795 - 22.24x \\
\hat{y} &= 22.24 - 0.7795x \\
\hat{y} &= 0.7795 + 22.24x \\
\hat{y} &= 22.24 + 0.7795x \\
\text{e)} & \text{ None of the above}
\end{align*}
\]

Question 2

Determine the correlation coefficient for the data shown in this table:

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>8</th>
<th>9</th>
<th>12</th>
<th>17</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>41</td>
<td>38</td>
<td>28</td>
<td>20</td>
<td>18</td>
<td>14</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
a) &= -0.9352 \\
b) &= 0.9352 \\
c) &= 0.8746 \\
d) &= -0.8746 \\
e) &= -0.4676 \\
f) &= \text{None of the above}
\end{align*}
\]

Question 3

Suppose you have the following data:
and the LSRL is $\hat{y} = 47.60 - 5.743x$. Find the residual value for $x = 1$.

\[
\text{residual} = \text{observed } y - \text{predicted } y \\
= 44 - \left[47.6 - 5.743(1)\right] \\
= 2.143
\]

a) 44
b) 41.857
c) 2.143
d) -2.143
e) None of the above

Question 4

Which of the following is the residual plot for the data in the given table?

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>13</td>
<td>19</td>
<td>17</td>
<td>31</td>
<td>17</td>
<td>34</td>
</tr>
</tbody>
</table>

\[
> x = c(1,2,3,4,5,6) \\
> y = c(13,19,17,31,17,34) \\
> \text{residual}(y-x) \\
> \begin{array}{cccccccc}
0.781942 & 2.005528 & 3.210478 & 7.5523810 \\
-0.781942 & 1.005528 & 2.210478 & 6.552381 \end{array}
\]

\[
> \text{residual}(y-x) \\
> \begin{array}{cccccccc}
0.781942 & 2.005528 & 3.210478 & 7.5523810 \\
-0.781942 & 1.005528 & 2.210478 & 6.552381 \end{array}
\]

a)
Question 5

Which of the following residual plots would indicate a good LSRL model?

- d) None of the above
- e) None of the above

Diagram:

- a) [Diagram showing a residual plot]
- d) [Diagram showing a residual plot]
Question 6

For children between the ages of 18 months and 29 months, there is approximately a linear relationship between height and age. The relationship can be represented by \( \hat{y} = 62.72 + 0.65x \) where \( y \) represents height (in centimeters) and \( x \) represents age (in months). Joseph is 23.5 months old and is 82 centimeters tall. What is Joseph's residual?

a) -34.017
b) 116.017
c) 8.008
d) 77.992
e) 4.008
f) None of the above

Question 7

If the LSRL relating the independent variable \( x \) and the dependent variable \( y \) for a given problem is \( \hat{y} = 3x + 5 \), then an increase of 1 unit in \( x \) is associated with an increase of how many units in \( y \)?

a) 5
b) 0
c) 8

d) 1

Question 8

If the correlation between number of children in a family and annual income was close to -1, we could conclude that:

a) low incomes cause people to have more children.

b) low-income families tend to have more children.

c) high incomes cause sterility.

d) high-income families tend to have more children.

e) high incomes cause people to have more children.

Question 9

The following two-way table describes the preferences in movies and fast food restaurants for a random sample of 100 people.

<table>
<thead>
<tr>
<th></th>
<th>McDonalds</th>
<th>Taco Bell</th>
<th>Wendy's</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron Man</td>
<td>20</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Dispicable Me</td>
<td>11</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Harry Potter</td>
<td>5</td>
<td>15</td>
<td>12</td>
</tr>
</tbody>
</table>

What percent of people in the sample like the movie Dispicable Me?

a) 36%

b) 35%

c) 29%

d) 32%

e) 28%

Question 10

The following two-way table describes the preferences in movies and fast food restaurants for a random sample of 100 people.
What percent of the Iron Man lovers also like Wendy's?

a) 21%

b) 20%

c) 9%

d) 19%

e) 40%

\[
\frac{\text{Wendy} \cap \text{Iron Man}}{\text{Iron Man}} = \frac{8}{40} = 0.2 \ (20\%)
\]
## Quiz 9

### Question 1

Identify the population of interest in the following situation:
A radio talk show wanted to know whether Houstonians think Texas should continue to use the death penalty. The station asked for listeners to call in and give their opinion.

- a) Prisoners.
- b) Houstonians.
- c) The population of the United States.
- d) Texans.
- e) Listeners of the radio station.

### Question 2

Define: voluntary response

- a) population is divided into similar groups and a SRS is chosen from each group.
- b) the explanatory variable(s) in an experiment.
- c) people who choose themselves for a sample by responding to a general appeal.
- d) choosing the individuals easiest to reach.
- e) directly holding extraneous factors constant.
- f) gives each member of the population a known chance to be selected.
- g) using extraneous factors to create similar groups.
- h) successively smaller groups are selected within the population in stages.
- i) every possible sample of a given size has the same chance to be selected.

### Question 3

Define: stratified random sample

- a) population is divided into similar groups and a SRS is chosen from each group.
b) gives each member of the population a known chance to be selected.

c) people who choose themselves for a sample by responding to a general appeal.

d) the explanatory variable(s) in an experiment.

e) directly holding extraneous factors constant.

f) every possible sample of a given size has the same chance to be selected.

g) using extraneous factors to create similar groups.

h) successively smaller groups are selected within the population in stages.

i) choosing the individuals easiest to reach.

**Question 4**

Define: blocking

a) choosing the individuals easiest to reach.

b) the explanatory variable(s) in an experiment.

c) using extraneous factors to create similar groups.

d) directly holding extraneous factors constant.

e) successively smaller groups are selected within the population in stages.

f) population is divided into similar groups and a SRS is chosen from each group.

g) gives each member of the population a known chance to be selected.

h) people who choose themselves for a sample by responding to a general appeal.

i) every possible sample of a given size has the same chance to be selected.

**Question 5**

True or false: Convenience samples often lead to under coverage bias.

a) True

b) False

**Question 6**

True or false: The entire group of individuals we want information about is called the population of interest.
a) True
b) False

Question 7
Identify the type of sampling used in the following situation:
To judge the appeal to American adults of a proposed television sitcom, a sample of 10 people from each of three different age groups was selected and those chosen were asked to rate a pilot show.

a) Voluntary Response
b) Stratified Random Sample
c) Simple Random Sample
d) Convenience Sample

Question 8
A personnel director at a large company studied the eating habits of employees by watching the movements of a selected group of employees at lunchtime. The purpose of the study was to determine the proportion of employees who buy lunch in the cafeteria, bring their own lunches, or go out to lunch. If the director includes only the employees in the department nearest her office in her study, she is performing a(n)

a) convenience sample
b) simple random sample
c) quota sample
d) stratified random sample
e) None of the above

Question 9
You are going to test two new varieties of fish food vs. a commonly used fish food. You set up an experiment as follows: 60 fish are randomly assigned to each of three different tanks. One tank is randomly selected to receive one of the new foods and another to receive the other new food. The remaining tank will receive the commonly used food. Fish growth is measured over time. This is an example of:

a) a randomized block design.
b) a completely randomized design with no control group.
c) a completely randomized design with a control group.
d) a double blind matched pairs design.
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>None of the above</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question 10**

A game of chance is based on rolling a die two times in succession. The player wins if the larger of the two numbers is greater than 3. Which of the following situations would simulate 9 plays of this game?

- **a)** Choosing 9 digits from the random number table (discarding 0, 7, 8, and 9).
- **b)** Choosing 9 pairs of single digits from the random number table (discarding 0, 7, 8, and 9).
- **c)** Choosing 54 digits from the random number table (discarding 0, 7, 8, and 9).
- **d)** Choosing 2 single digits from the random number table (discarding 0, 7, 8, and 9).
- **e)** None of the above
Question 1

Think about a density curve that consists of two line segments. The first goes from the point (0, 1) to the point (0.7, 1). The second goes from (0.7, 1) to (0.9, 2) in the xy-plane. What percent of observations fall below 0.40?

a) 0.30
b) 0.20
c) 1.00
d) 0.40
e) 0.60
f) None of the above

Question 2

The heights of students in a class are normally distributed with mean 66 inches and standard deviation 5 inches. Use the Empirical Rule to determine the interval that contains the middle 95% of the heights.

a) [61, 71]
b) [53, 79]
c) [51, 71]
d) [51, 81]
e) [56, 76]
f) None of the above

Question 3

Suppose that \( x \) is normally distributed with a mean of 50 and a standard deviation of 10. What is \( P(x \leq 66.50) \)?

a) 0.951
b) 0.549
c) 0.455
d) 0.451  

e) 0.049  

f) None of the above

**Question 4**

Find a value of \( c \) so that \( P(Z \geq c) = 0.62 \).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
a) 0.31  
b) -0.31  
c) 0.51  
d) 1.31  
e) -0.61  
f) None of the above

**Question 5**

Suppose a random sample of 80 measurements is selected from a population with a mean of 25 and a variance of 200. Select the pair that is the mean and standard error of \( \bar{x} \).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
a) [25, 2.081]  
b) [25, 1.981]  
c) [25, 1.681]  
d) [25, 1.581]  
e) [80, 1.681]  
f) None of the above

**Question 6**

Current research indicates that the distribution of the life expectancies of a certain protozoan is normal with a mean of 47 days and a standard deviation of 10.2 days. Find the probability that a simple random sample of 25 protozoa will have a mean life expectancy of 53 or more days.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
a) 0.0033  
b) 0.2016
Question 7

In a large population, 75% of the households have cable tv. A simple random sample of 64 households is to be contacted and the sample proportion computed. What is the probability that the sampling distribution of sample proportions is less than 87%?

a) 0.0133
b) 0.1338
c) 0.4933
d) 0.8662
e) 0.9867
f) None of the above

Question 8

Determine the correlation coefficient for the data shown in this table:

<table>
<thead>
<tr>
<th>x</th>
<th>3</th>
<th>6</th>
<th>11</th>
<th>13</th>
<th>16</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>22</td>
<td>28</td>
<td>27</td>
<td>40</td>
<td>27</td>
<td>43</td>
</tr>
</tbody>
</table>

a) -0.7019
b) 0.7019
c) 0.3509
d) -0.4926
e) 0.4926
f) None of the above
Which of the following would be the LSRL for the given data?

<table>
<thead>
<tr>
<th>x</th>
<th>4</th>
<th>6</th>
<th>11</th>
<th>13</th>
<th>17</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>23</td>
<td>29</td>
<td>27</td>
<td>42</td>
<td>26</td>
<td>41</td>
</tr>
</tbody>
</table>

- a) $\hat{y} = 0.7795 - 22.24x$
- b) $\hat{y} = 22.24 - 0.7795x$
- c) $\hat{y} = 0.7795 + 22.24x$
- d) $\hat{y} = 22.24 + 0.7795x$
- e) None of the above

**Question 10**

For children between the ages of 18 months and 29 months, there is approximately a linear relationship between height and age. The relationship can be represented by $\hat{y} = 62.72 + 0.65x$ where $y$ represents height (in centimeters) and $x$ represents age (in months). Joseph is 23.5 months old and is 82 centimeters tall. What is Joseph's residual?

- a) -34.017
- b) 116.017
- c) 8.008
- d) 77.992
- e) 4.008
- f) None of the above

**Question 11**

Which of the following is the residual plot for the data in the given table?

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>33</td>
<td>27</td>
<td>19</td>
<td>11</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>
Question 12

The following two-way table describes the preferences in movies and fast food restaurants for a random sample of 100 people.
<table>
<thead>
<tr>
<th>Iron Man</th>
<th>McDonald's</th>
<th>Taco Bell</th>
<th>Wendy's</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Dispicable Me</td>
<td>11</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Harry Potter</td>
<td>6</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>

What percent of the Harry Potter lovers also like Taco Bell?

a) 13%

b) 40%

c) 22%

d) 44%

e) 17%

**Question 13**

Define: multistage sample

a) choosing the individuals easiest to reach.

b) people who choose themselves for a sample by responding to a general appeal.

c) gives each member of the population a known chance to be selected.

d) successively smaller groups are selected within the population in stages.

e) population is divided into similar groups and a SRS is chosen from each group.

f) using extraneous factors to create similar groups.

g) the explanatory variable(s) in an experiment.

h) every possible sample of a given size has the same chance to be selected.

i) directly holding extraneous factors constant.

**Question 14**

A personnel director at a large company studied the eating habits of employees by watching the movements of a selected group of employees at lunchtime. The purpose of the study was to determine the proportion of employees who buy lunch in the cafeteria, bring their own lunches, or go out to lunch.

If the director selects 50 employees at random from throughout the company and categorizes their lunchtime practices by gender, she is:

a) testing for bias
b) testing for a lurking variable

c) promoting sexual harassment

d) blocking for gender

e) None of the above

Question 15

You will need to do free response like the following:

From chapter 4, be able to find probabilities for a normal distribution. Be able to sketch the distribution and shade the appropriate probability. Find z-scores. Find the value of X corresponding to a particular probability.

From chapter 5, be able to draw a scatter plot, find the LSRL, interpret the slope, find the correlation coefficient, coefficient of determination (and interpret those values), find a residual value, show the residual plot and determine if the model is a good fit or not based on all observations of values found.

From chapter 6, given a section from the random digit table be able to simulate an experiment (like 6.3 number 7).

a) I will study these types of problems.

b) I won't study these types of problems.

Assume that the percentage of women in the labor force of a large metropolitan area is 40%. A company hires ten workers, two of whom are women. We want to see if this is likely.

a. Assign the digits 0 through 9, to represent the men and women in this situation. Describe how you will run the simulation using those digits and the random digit table.

b. Start on line 130 of the random digit table and carry out the simulation with 3 runs.

c. What is the expected number of women that should be hired, based on your simulation?

d. Repeat the 3 runs using line 135 of the random digit table. Is the expected number the same? Why or why not?

Let digits 0, 1, 2, 3 represent women

Let digits 4, 5, 6, 7, 8, 9 represent men

Select 10 numbers from the line "hire 10"

1 run: 8 0 0 5 0 4 5 6 9 8

2 run: 2 0 0 2 1 9 9 6 4 8

3 run: 2 3 9 0 6 4 9 3 7 5

https://assessment.casa.uh.edu/Assessment/PrintTest.htm