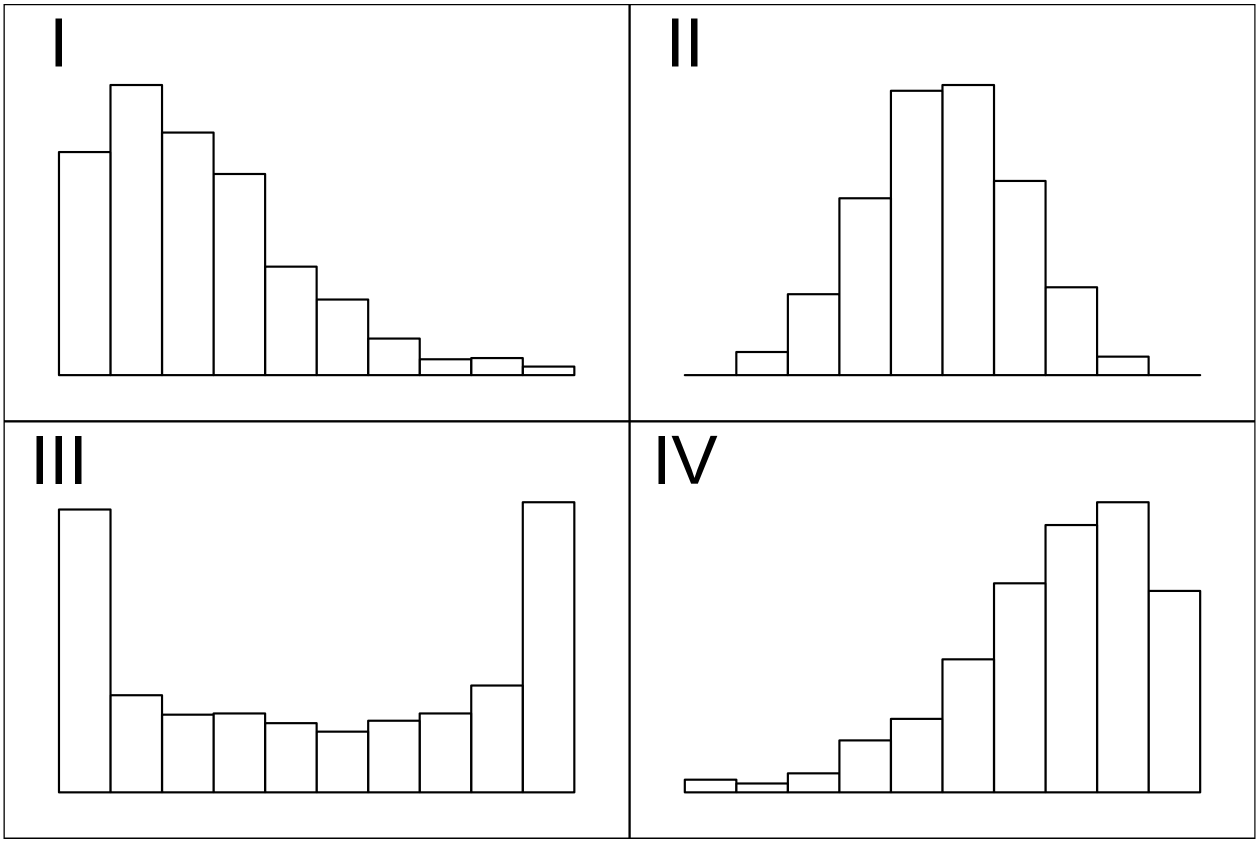
MAT-181 FINAL TAKE-HOME EXAM

This exam is to be taken without discussion or correspondance with any human. Please show work!

question	available points	earned points
1	10	
2	15	
3	10	
4	10	
5	10	
6	10	
7	15	
8	20	
EC	5	
EC	5	
Total	100	

1. (10 Points)



For each description below, choose which histogram best fits (I, II, III, or IV). Each histogram should be used once.

(a) The distribution of lengths of newborn babies

(b) The distribution of annual income for NBA basketball players where only a few are high-paid superstars.

(c) The distribution of hours that students studied for an exam when about half of students studied a lot and a similar number of students studied very little.

(d) The distribution of ages at a skilled nursing facility, where most of the patients are elderly but a few are quite young.

Solution:

(a) II
(b) I
(c) III
(d) IV

2. (15 Points)

   In a deck of strange cards, there are 364 cards. Each card has an image and a color. The amounts are shown in the table below.

	pink	red	teal	white	Total
horn	38	50	40	32	160
pig	20	10	22	13	65
tree	45	35	43	16	139
Total	103	95	105	61	364

   (a) What is the probability a random card is a tree?

   (b) What is the probability a random card is pink given it is a horn?

   (c) What is the probability a random card is both a tree and red?

   (d) What is the probability a random card is a pig given it is red?

   (e) Is a horn or a tree more likely to be white?

   (f) What is the probability a random card is teal?

   (g) What is the probability a random card is either a pig or red (or both)?

Solution:

(a) $P(\text{tree}) = 0.382$

(b) $P(\text{pink given horn}) = 0.238$

(c) $P(\text{tree and red}) = 0.0962$

(d) $P(\text{pig given red}) = 0.105$

(e) $P(\text{white given horn}) = 0.2$ and $P(\text{white given tree}) = 0.115$, so a horn is more likely to be white than a tree is.

(f) $P(\text{teal}) = 0.288$

(g) $P(\text{pig or red}) = 0.412$

3. (10 points)

A farm produces 4 types of fruit: *A*, *B*, *C*, and *D*. The fruits' masses follow normal distributions, with parameters dependent on the type of fruit.

Type of fruit	Mean mass (g)	Standard deviation of mass (g)
A	93	13
B	115	14
C	123	15
D	125	5

One specimen of each type is weighed. The results are shown below.

Type of fruit	Mass of specimen (g)
A	107.3
B	105.6
C	110.8
D	120

Which specimen is the most unusually small (relative to others of its type)?

Solution: We compare the z-scores. The smallest z-score corresponds to the specimen that is most unusually small.

Type of fruit	formula	z-score
A	$z = \frac{107.3-93}{13}$	1.1
B	$z = \frac{105.6-115}{14}$	-0.67
C	$z = \frac{110.8-123}{15}$	-0.81
D	$z = \frac{120-125}{5}$	-0.99

Thus, the specimen of type D is the most unusually small.

4. (10 points)

A tree's leaves were found to be normally distributed with a mean of 61.5 millimeters and a standard deviation of 8.2 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 45.3 and 66.8 millimeters?

Solution:

$$\mu = 61.5$$
$$\sigma = 8.2$$
$$x_1 = 45.3$$
$$x_2 = 66.8$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{45.3 - 61.5}{8.2} = -1.98$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{66.8 - 61.5}{8.2} = 0.65$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.7422 - 0.0239 = 0.7183$$

5. (10 points)

   A species of duck is known to have a mean weight of 163 grams and a standard deviation of 18 grams. A researcher plans to measure the weights of 81 of these ducks sampled randomly. What is the probability the **sample mean** will be between 159.5 and 162 grams?

Solution:

$$n = 81$$
$$\mu = 163$$
$$\sigma = 18$$
$$SE = \frac{18}{\sqrt{81}} = 2$$
$$x_1 = 159.5$$
$$x_2 = 162$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{159.5 - 163}{2} = -1.75$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{162 - 163}{2} = -0.5$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.3085 - 0.0401 = 0.2684$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Zonotrichia albicollis*. She randomly samples 21 adults of *Zonotrichia albicollis*, resulting in a sample mean of 24.2 grams and a sample standard deviation of 1.91 grams. Determine a 95% confidence interval of the true population mean.

Solution: We are given the sample size, sample mean, sample standard deviation, and confidence level.

$$\begin{aligned} n &= 21 \\ \bar{x} &= 24.2 \\ s &= 1.91 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 21 - 1 \\ &= 20 \end{aligned}$$

Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 20$.

$$t^{\star} = 2.09$$

Use the formula for bounds (mean, σ unknown).

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} & UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 24.2 - 2.09 \times \frac{1.91}{\sqrt{21}} & &= 24.2 + 2.09 \times \frac{1.91}{\sqrt{21}} \\ &= 23.3 & &= 25.1 \end{aligned}$$

We are 95% confident that the population mean is between 23.3 and 25.1 grams.

$$CI = (23.3,\ 25.1)$$

7. (15 points)

A student is taking a multiple choice test with 400 questions. Each question has 2 choices. You want to detect whether the student does significantly better than random guessing, so you decide to run a hypothesis test with a significance level of 0.05.

Then, the student takes the test and gets 218 questions correct.

(a) What kind of hypothesis test is appropriate?

(b) State the hypotheses.

(c) Determine the test statistic (z or t), draw a sketch, and determine the p-value.

(d) Decide whether we reject or retain the null hypothesis.

(e) Did the student do significantly better than random guessing?

Solution: This is a right-tail (one-tail) proportion test because we only care whether the student does better than random.

Determine the null population proportion.

$$p_0 = \frac{1}{2} = 0.5$$

State the hypotheses.

$$H_0 \quad \text{claims} \quad p = 0.5$$
$$H_A \quad \text{claims} \quad p > 0.5$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{400}} = 0.025$$

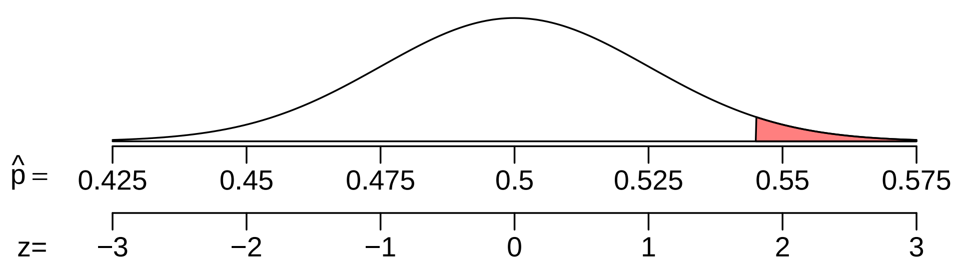
Determine the sample proportion.

$$\hat{p} = \frac{218}{400} = 0.545$$

Determine a z score. For simplicity, we ignore the continuity correction.

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.545 - 0.5}{0.025} = 1.8$$

Make a sketch of the null's sampling distribution. The p-value is a right area.



To determine that right area, we use the z table.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.545) \\ &= P(Z > 1.8) \\ &= 1 - P(Z < 1.8) \\ &= 0.0359 \end{aligned}$$

Compare p-value to α (which is 0.05).

$$p\text{-value} < \alpha$$

Make the conclusion: we reject the null hypothesis.

We think the student did better than random guessing typically allows.

(a) Right tail (one-tail) proportion test

(b) Hypotheses: H_0 claims $p = 0.5$ and H_A claims $p > 0.5$.

(c) The p-value is 0.0359

(d) We reject the null hypothesis.

(e) We think the student did better than random guessing typically allows.

8. (20 points) [Note: this question uses 2 pages.]

You have collected the following data:

x	y	xy
95	2.1	
41	3.5	
16	6.3	
97	1.4	
34	1.7	
98	1.8	
73	5.5	
54	3.7	
69	2.3	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

(a) Complete the table.

(b) Calculate the correlation coefficient (r) using the formula below.

$$r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

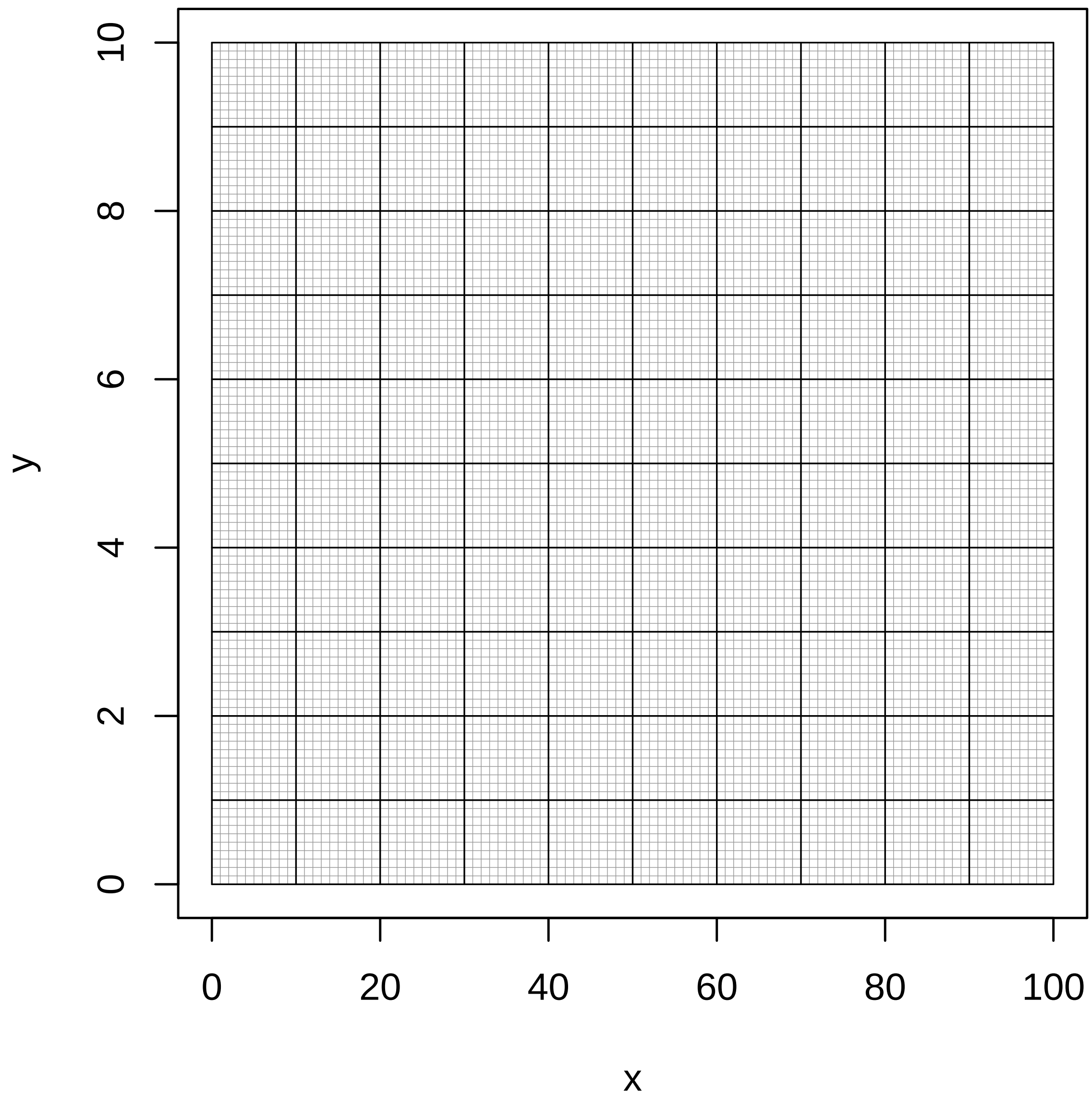
(c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

$$b = r\frac{s_y}{s_x}$$

$$a = \bar{y} - b\bar{x}$$

(d) Write the equation of the regression line (using the calculated values of a and b.)

(e) Please plot the data and a corresponding regression line.



Solution: Remember the formula for the correlation coefficient.

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

We calculate the necessary values.

x	y	xy
95	2.1	199.5
41	3.5	143.5
16	6.3	100.8
97	1.4	135.8
34	1.7	57.8
98	1.8	176.4
73	5.5	401.5
54	3.7	199.8
69	2.3	158.7
$\sum x = 577$	$\sum y = 28.3$	$\sum x_i y_i = 1573.8$
$\bar{x} = 64.11$	$\bar{y} = 3.144$	
$s_x = 29.88$	$s_y = 1.756$	

$$r = \frac{1573.8 - (9)(64.11)(3.144)}{(9-1)(29.88)(1.756)} = -0.572$$

If you didn't round any of the steps up to here, you'd get an exact value which is pretty close to our value.

$$r_{\text{exact}} = -0.5728032$$

The regression line has the form

$$y = a + bx$$

So, a is the y-intercept and b is the slope. We have formulas to determine them:

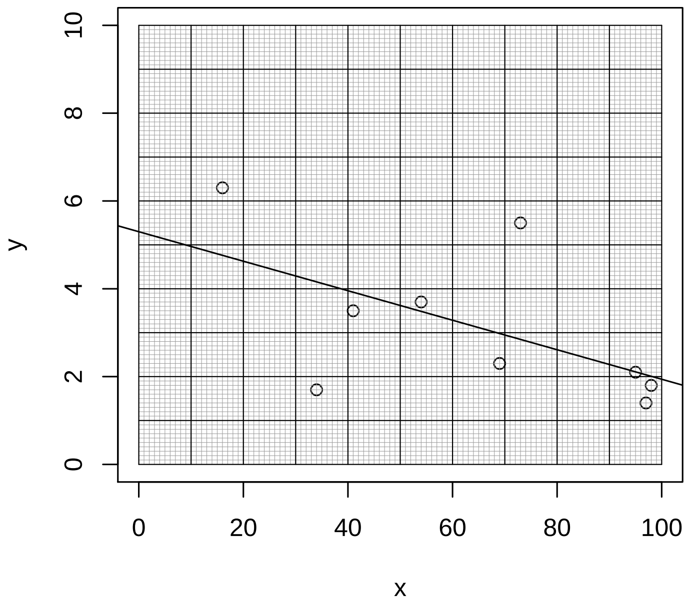
$$b = r\frac{s_y}{s_x} = -0.572 \cdot \frac{1.756}{29.88} = -0.0336$$

$$a = \bar{y} - b\bar{x} = 3.14 - (-0.0336)(64.1) = 5.3$$

Our regression line:

$$y = 5.3 + (-0.0336)x$$

Make a plot.



0
2
4
6
8
10
y
0
20
40
60
80
100
x

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.61$. If 132 trials occur, what is the probability of getting more than 77 but less than 88 successes?

   In other words, let $X \sim \text{Bin}(n = 132,\ p = 0.61)$ and find $P(77 < X < 88)$.

   Use a normal approximation along with the continuity correction.

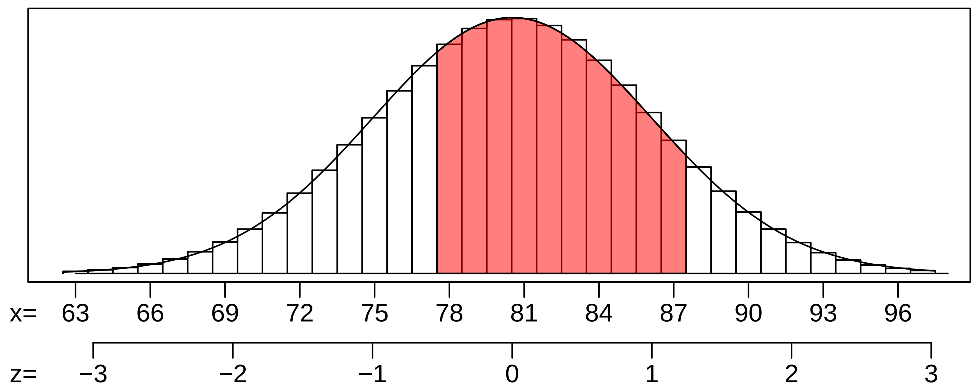
Solution: Find the mean.

$$\mu = np = (132)(0.61) = 80.52$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(132)(0.61)(1-0.61)} = 5.6038$$

Make a sketch, specifically try to picture whether you need to add or subtract 0.5 for the continuity correction.



Find the z scores.

$$z_1 = \frac{77.5 - 80.52}{5.6038} = -0.54$$

$$z_2 = \frac{87.5 - 80.52}{5.6038} = 1.25$$

Find the percentiles (from z-table).

$$\ell_1 = 0.2946$$

$$\ell_2 = 0.8944$$

Calculate the probability.

$$P(77 < X < 88) = 0.8944 - 0.2946 = 0.599$$

10. (Extra credit: 5 points)

A null hypothesis claims a population has a mean $\mu = 80$. You decide to run two-tail test on a sample of size $n = 11$ using a significance level $\alpha = 0.1$.

You then collect the sample:

112.7	73.7	99.6	112.3	93.6
73.4	81.1	85.7	93.5	68
76				

(a) Determine the p-value.

(b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \text{ claims } \mu = 80$$

$$H_A \text{ claims } \mu \neq 80$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 88.145$$

$$s = 15.528$$

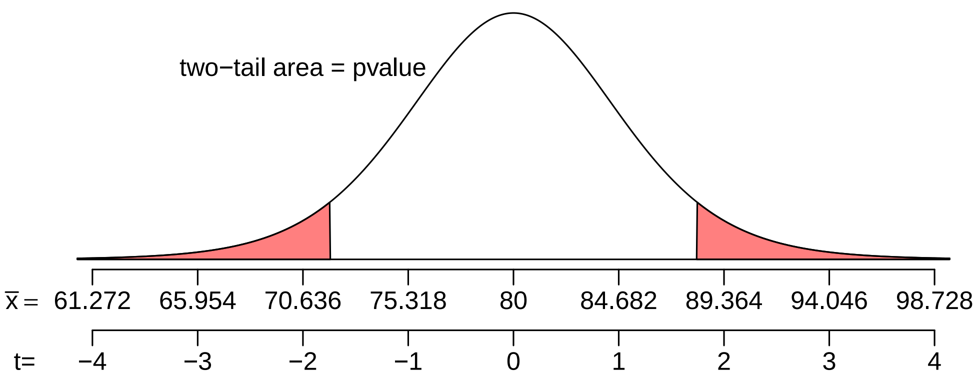
Determine the degrees of freedom.

$$df = 11 - 1 = 10$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{15.528}{\sqrt{11}} = 4.682$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{88.145 - 80}{4.682} = 1.74$$

Find the p-value.

$$p\text{-value} = P(|T| > 1.74)$$

We can't get an exact value with our table, but we can determine an interval that contains the p-value. (Look at row with $df = 10$.)

$$P(|T| > 1.81) = 0.1$$

$$P(|T| > 1.37) = 0.2$$

Basically, because t is between 1.81 and 1.37, we know the p-value is between 0.1 and 0.2.

$$0.1 < p\text{-value} < 0.2$$

Compare the p-value and the significance level ($\alpha = 0.1$).

$$p\text{-value} > \alpha$$

No, we do not reject the null hypothesis.

(a) $0.1 < p\text{-value} < 0.2$

(b) No, we do not reject the null hypothesis.