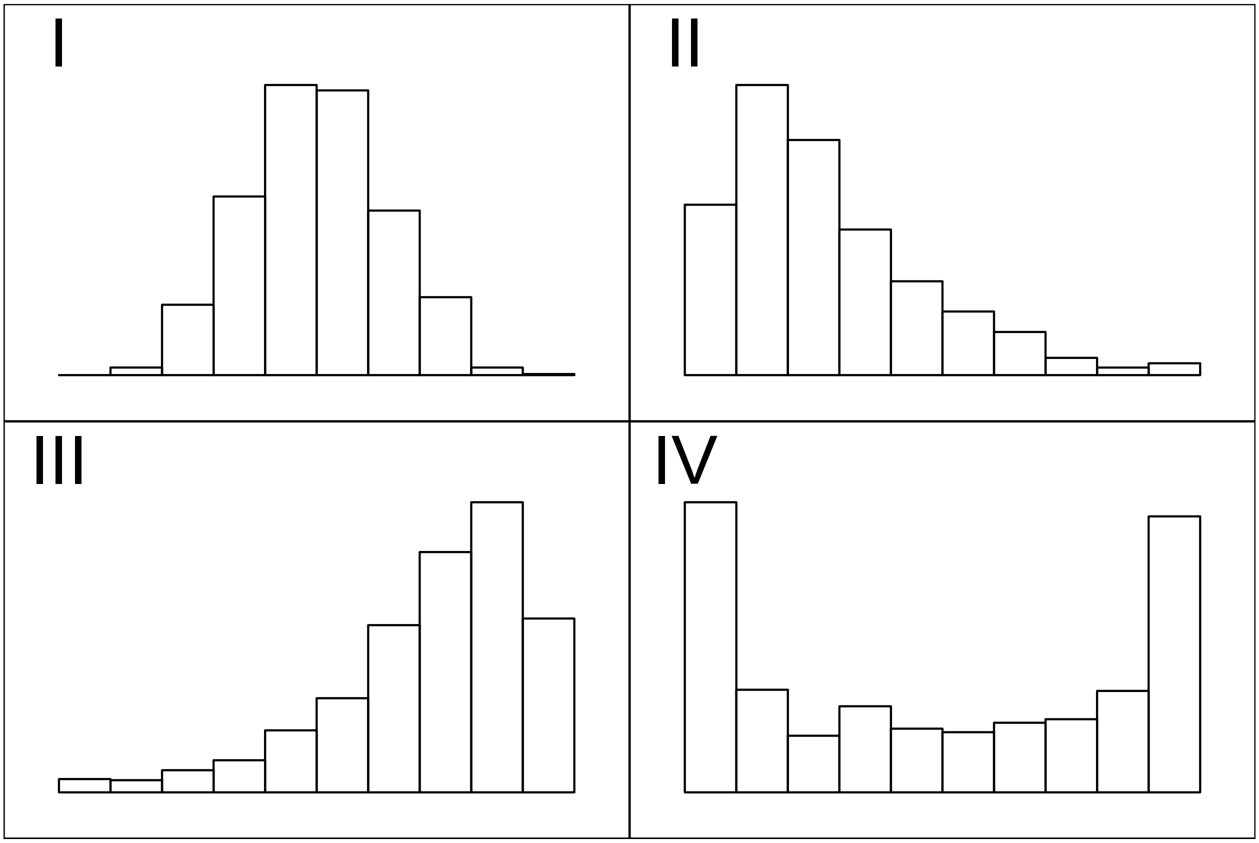
MAT-181 FINAL TAKE-HOME EXAM

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 heights of adult men

(b) The distribution of test scores on a very difficult exam, in which most students have poor to average scores, but a few did quite well.

(c) The distribution of hours spent per week reading by adults. In this distribution, many people do not read much, and a similar number of people read a lot.

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

2. (15 Points)

In a deck of strange cards, there are 408 cards. Each card has an image and a color. The amounts are shown in the table below.

	green	orange	teal	violet	white	Total
bike	33	20	47	30	17	147
gem	11	13	24	38	41	127
jigsaw	50	28	31	15	10	134
Total	94	61	102	83	68	408

(a) What is the probability a random card is both a jigsaw and teal?

(b) What is the probability a random card is violet?

(c) What is the probability a random card is a bike given it is green?

(d) What is the probability a random card is a gem?

(e) What is the probability a random card is violet given it is a jigsaw?

(f) What is the probability a random card is either a bike or green (or both)?

(g) Is a bike or a jigsaw more likely to be green?

Solution:

(a) P(jigsaw and teal) = 0.076

(b) P(violet) = 0.203

(c) P(bike given green) = 0.351

(d) P(gem) = 0.311

(e) P(violet given jigsaw) = 0.112

(f) P(bike or green) = 0.51

(g) P(green given bike) = 0.224 and P(green given jigsaw) = 0.373, so a jigsaw is more likely to be green than a bike is.

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	90	4
B	147	15
C	72	14
D	93	6

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

Type of fruit	Mass of specimen (g)
A	90.2
B	136
C	61.22
D	87.36

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	$z = \frac{90.2-90}{4}$	0.05
B	$z = \frac{136-147}{15}$	-0.73
C	$z = \frac{61.22-72}{14}$	-0.77
D	$z = \frac{87.36-93}{6}$	-0.94

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 127.9 millimeters and a standard deviation of 7.6 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 119.5 and 132.4 millimeters?

Solution:

$$\mu = 127.9$$
$$\sigma = 7.6$$
$$x_1 = 119.5$$
$$x_2 = 132.4$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{119.5 - 127.9}{7.6} = -1.11$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{132.4 - 127.9}{7.6} = 0.59$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.7224 - 0.1335 = 0.5889$$

5. (10 points)

   A species of duck is known to have a mean weight of 216 grams and a standard deviation of 14 grams. A researcher plans to measure the weights of 49 of these ducks sampled randomly. What is the probability the **sample mean** will be between 217 and 218.5 grams?

Solution:

$$n = 49$$
$$\mu = 216$$
$$\sigma = 14$$
$$SE = \frac{14}{\sqrt{49}} = 2$$
$$x_1 = 217$$
$$x_2 = 218.5$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{217 - 216}{2} = 0.5$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{218.5 - 216}{2} = 1.25$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.8944 - 0.6915 = 0.2029$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Ammodramus maritimus*. She randomly samples 27 adults of *Ammodramus maritimus*, resulting in a sample mean of 21.67 grams and a sample standard deviation of 1.56 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.

$$\begin{aligned} n &= 27 \\ \bar{x} &= 21.67 \\ s &= 1.56 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 27 - 1 \\ &= 26 \end{aligned}$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$ and $df = 26$.

$$t^\star = 2.06$$

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

$$\begin{aligned} LB &= \bar{x} - t^\star \frac{s}{\sqrt{n}} \\ &= 21.67 - 2.06 \times \frac{1.56}{\sqrt{27}} \\ &= 21.1 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^\star \frac{s}{\sqrt{n}} \\ &= 21.67 + 2.06 \times \frac{1.56}{\sqrt{27}} \\ &= 22.3 \end{aligned}$$

We are 95% confident that the population mean is between 21.1 and 22.3 grams.

$$CI = (21.1,\ 22.3)$$

7. (15 points)

   A student is taking a multiple choice test with 800 questions. Each question has 3 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 292 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.

$$p_0 = \frac{1}{3} = 0.333$$

State the hypotheses.

$$H_0 \quad \text{claims} \quad p = 0.333$$
$$H_A \quad \text{claims} \quad p > 0.333$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.333(1-0.333)}{800}} = 0.0167$$

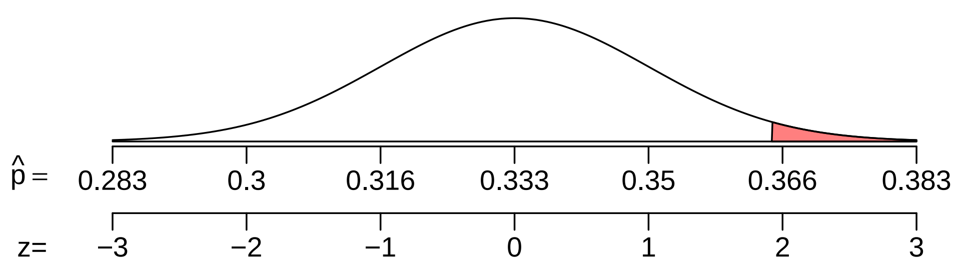
Determine the sample proportion.

$$\hat{p} = \frac{292}{800} = 0.365$$

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

$$z = \frac{\hat{p}-p_0}{\sigma_{\hat{p}}} = \frac{0.365-0.333}{0.0167} = 1.92$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.365) \\ &= P(Z > 1.92) \\ &= 1 - P(Z < 1.92) \\ &= 0.0274 \end{aligned}$$

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 $p = 0.333$ and H_A claims $p > 0.333$.

(c) The p-value is 0.0274

(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
7.6	30	
2	60	
3.3	66	
5.8	50	
5.2	56	
7.1	45	
3.6	61	
8.8	30	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

   (a) Complete the table.

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

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

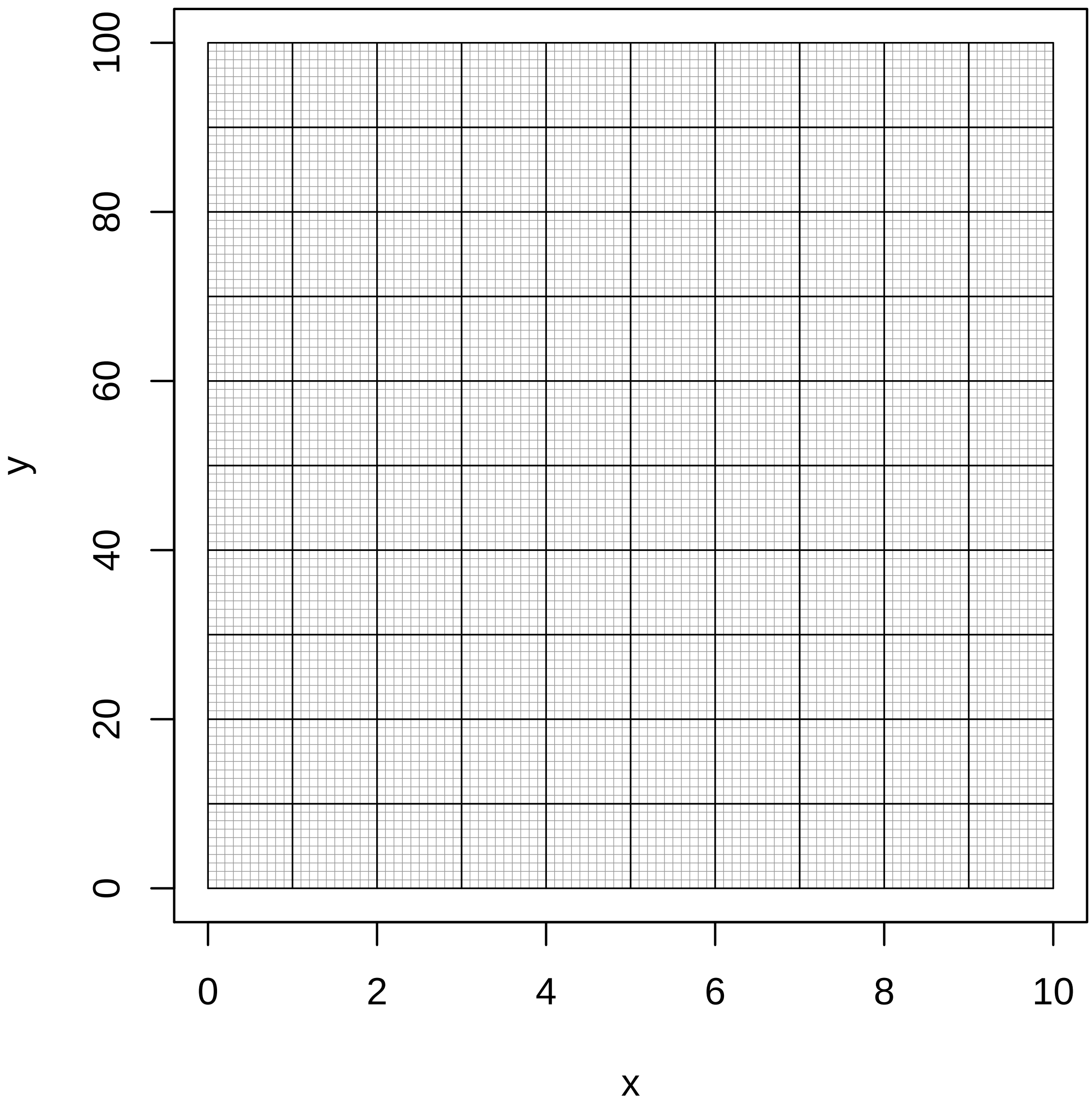
   (c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

   (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
7.6	30	228
2	60	120
3.3	66	217.8
5.8	50	290
5.2	56	291.2
7.1	45	319.5
3.6	61	219.6
8.8	30	264
$\sum x = 43.4$	$\sum y = 398$	$\sum x_i y_i = 1950.1$
$\bar{x} = 5.425$	$\bar{y} = 49.75$	
$s_x = 2.351$	$s_y = 13.82$	

$$r = \frac{1950.1 - (8)(5.425)(49.75)}{(8-1)(2.351)(13.82)} = -0.919$$

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.

$$r_{\text{exact}} = -0.9189161$$

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:

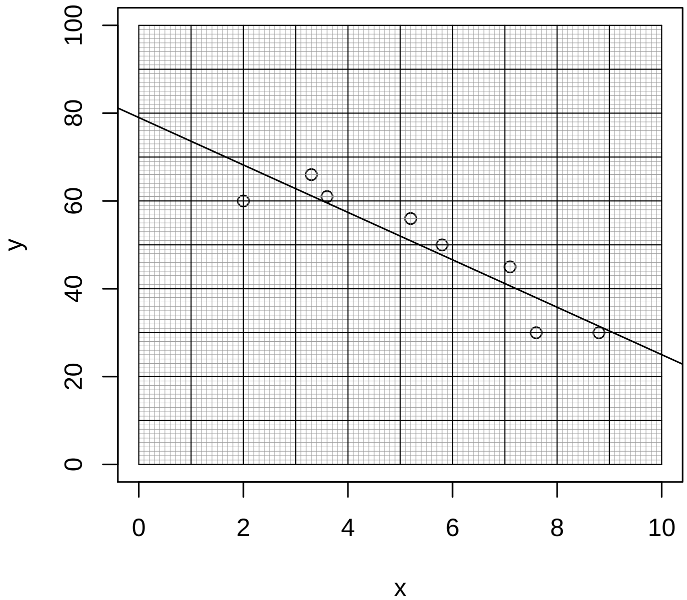
$$b = r\frac{s_y}{s_x} = -0.919 \cdot \frac{13.82}{2.351} = -5.4$$

$$a = \bar{y} - b\bar{x} = 49.8 - (-5.4)(5.42) = 79$$

Our regression line:

$$y = 79 + (-5.4)x$$

Make a plot.



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

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.1$. If 186 trials occur, what is the probability of getting at least 16 but less than 28 successes?

   In other words, let $X \sim \text{Bin}(n = 186,\ p = 0.1)$ and find $P(16 \leq X < 28)$.

   Use a normal approximation along with the continuity correction.

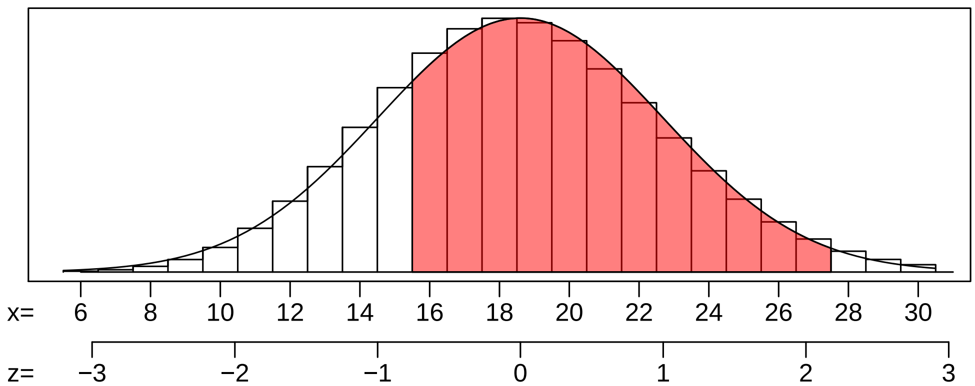
Solution: Find the mean.

$$\mu = np = (186)(0.1) = 18.6$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(186)(0.1)(1-0.1)} = 4.0915$$

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



Find the z scores.

$$z_1 = \frac{15.5 - 18.6}{4.0915} = -0.76$$

$$z_2 = \frac{27.5 - 18.6}{4.0915} = 2.18$$

Find the percentiles (from z-table).

$$\ell_1 = 0.2236$$

$$\ell_2 = 0.9854$$

Calculate the probability.

$$P(16 \leq X < 28) = 0.9854 - 0.2236 = 0.761$$

10. (Extra credit: 5 points)

A null hypothesis claims a population has a mean $\mu = 100$. 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:

125.7	66.5	70.5	86.2	123.4
143.1	107.7	150.4	129.7	132.2
163.8				

(a) Determine the p-value.

(b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 100$$

$$H_A \quad \text{claims} \quad \mu \neq 100$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 118.109$$

$$s = 31.985$$

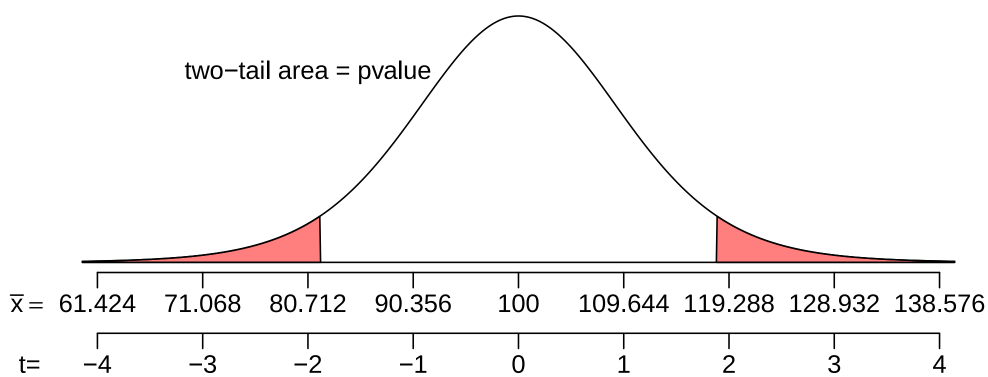
Determine the degrees of freedom.

$$df = 11 - 1 = 10$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{31.985}{\sqrt{11}} = 9.644$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{118.109 - 100}{9.644} = 1.88$$

Find the p-value.

$$p\text{-value} = P(|T| > 1.88)$$

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 $df = 10$.)

$$P(|T| > 2.23) = 0.05$$

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

Basically, because t is between 2.23 and 1.81, we know the p-value is between 0.05 and 0.1.

$$0.05 < p\text{-value} < 0.1$$

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

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

Yes, we reject the null hypothesis.

(a) $0.05 < p\text{-value} < 0.1$

(b) Yes, we reject the null hypothesis.