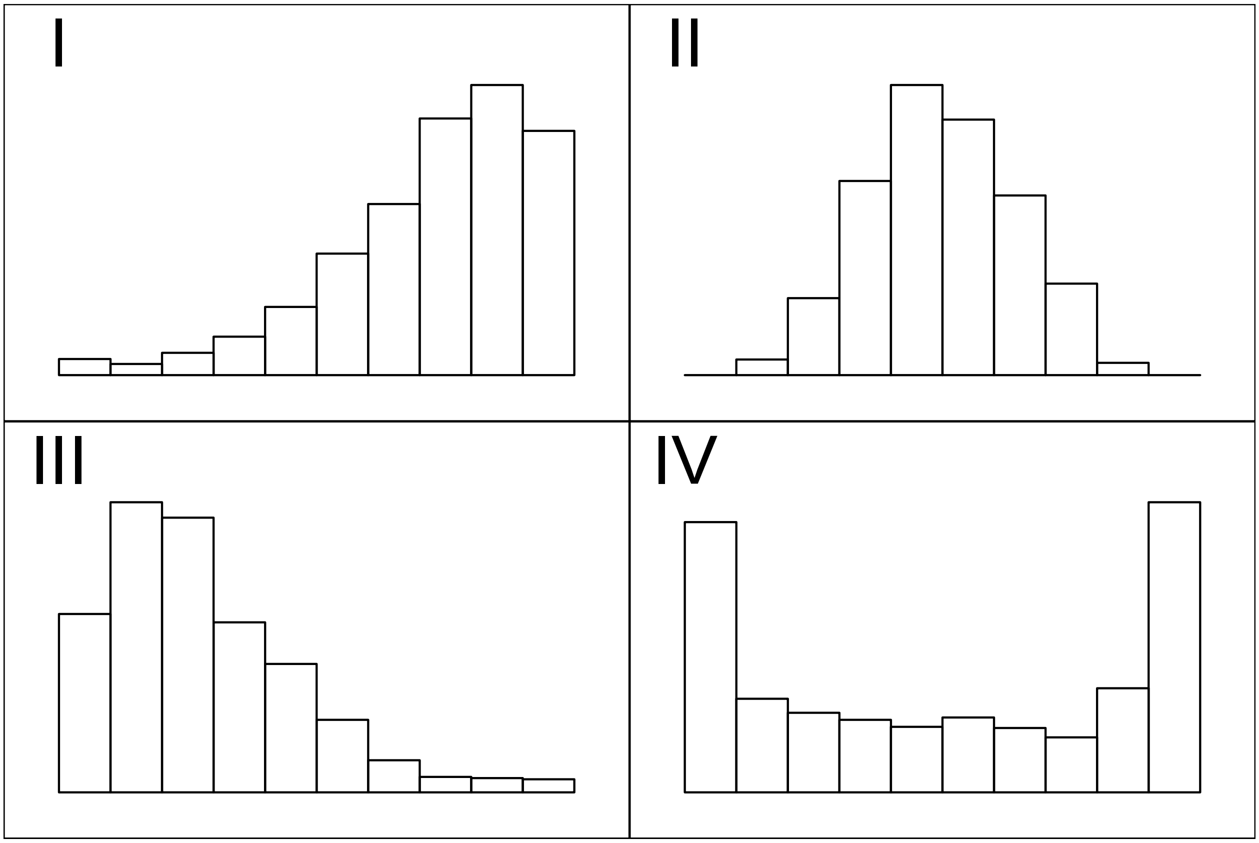
MAT-181 FINAL TAKE-HOME EXAM

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 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.

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

(d) The distribution of quiz scores on an easy quiz. Most students did very well, but a few did poorly.

Solution:

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

2. (15 Points)

In a deck of strange cards, there are 406 cards. Each card has an image and a color. The amounts are shown in the table below.

	gray	green	red	Total
dog	35	24	15	74
horn	36	26	46	108
kite	44	42	20	106
rug	49	28	41	118
Total	164	120	122	406

(a) Is a horn or a rug more likely to be red?

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

(c) What is the probability a random card is both a dog and gray?

(d) What is the probability a random card is a kite given it is gray?

(e) What is the probability a random card is green?

(f) What is the probability a random card is red given it is a dog?

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

Solution:

(a) P(red given horn) = 0.426 and P(red given rug) = 0.347, so a horn is more likely to be red than a rug is.

(b) P(rug) = 0.291

(c) P(dog and gray) = 0.0862

(d) P(kite given gray) = 0.268

(e) P(green) = 0.296

(f) P(red given dog) = 0.203

(g) P(horn or red) = 0.453

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	70	12
B	125	15
C	69	5
D	92	10

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

Type of fruit	Mass of specimen (g)
A	58.12
B	138.4
C	64.35
D	90.2

Which specimen is the most unusually far (in either direction) from average (relative to others of its type)?

Solution: We compare the absolute z-scores. The largest absolute z-score corresponds to the specimen that is most unusually far from average.

Type of fruit	formula	absolute z-score
A	$z = \frac{\lvert 58.12 - 70 \rvert}{12}$	0.99
B	$z = \frac{\lvert 138.4 - 125 \rvert}{15}$	0.89
C	$z = \frac{\lvert 64.35 - 69 \rvert}{5}$	0.93
D	$z = \frac{\lvert 90.2 - 92 \rvert}{10}$	0.18

Thus, the specimen of type A is the most unusually far from average.

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 35.8 millimeters and a standard deviation of 7.5 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 34.3 and 48.6 millimeters?

Solution:

$$\mu = 35.8$$
$$\sigma = 7.5$$
$$x_1 = 34.3$$
$$x_2 = 48.6$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{34.3 - 35.8}{7.5} = -0.2$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{48.6 - 35.8}{7.5} = 1.71$$

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.9564 - 0.4207 = 0.5357$$

5. (10 points)

   A species of duck is known to have a mean weight of 238.4 grams and a standard deviation of 55 grams. A researcher plans to measure the weights of 121 of these ducks sampled randomly. What is the probability the **sample mean** will be between 236.9 and 243.9 grams?

Solution:

$$n = 121$$
$$\mu = 238.4$$
$$\sigma = 55$$
$$SE = \frac{55}{\sqrt{121}} = 5$$
$$x_1 = 236.9$$
$$x_2 = 243.9$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{236.9 - 238.4}{5} = -0.3$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{243.9 - 238.4}{5} = 1.1$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.8643 - 0.3821 = 0.4822$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Catharus fuscescens*. She randomly samples 30 adults of *Catharus fuscescens*, resulting in a sample mean of 40.67 grams and a sample standard deviation of 4.76 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.

$$\begin{aligned} n &= 30 \\ \bar{x} &= 40.67 \\ s &= 4.76 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 30 - 1 \\ &= 29 \end{aligned}$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$ and $df = 29$.

$$t^\star = 2.05$$

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

$$\begin{aligned} LB &= \bar{x} - t^\star \frac{s}{\sqrt{n}} & UB &= \bar{x} + t^\star \frac{s}{\sqrt{n}} \\ &= 40.67 - 2.05 \times \frac{4.76}{\sqrt{30}} & &= 40.67 + 2.05 \times \frac{4.76}{\sqrt{30}} \\ &= 38.9 & &= 42.5 \end{aligned}$$

We are 95% confident that the population mean is between 38.9 and 42.5 grams.

$$CI = (38.9,\ 42.5)$$

7. (15 points)

   A student is taking a multiple choice test with 700 questions. Each question has 5 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 159 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.

$$p_0 = \frac{1}{5} = 0.2$$

State the hypotheses.

$$H_0 \quad \text{claims} \quad p = 0.2$$

$$H_A \quad \text{claims} \quad p > 0.2$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.2(1-0.2)}{700}} = 0.0151$$

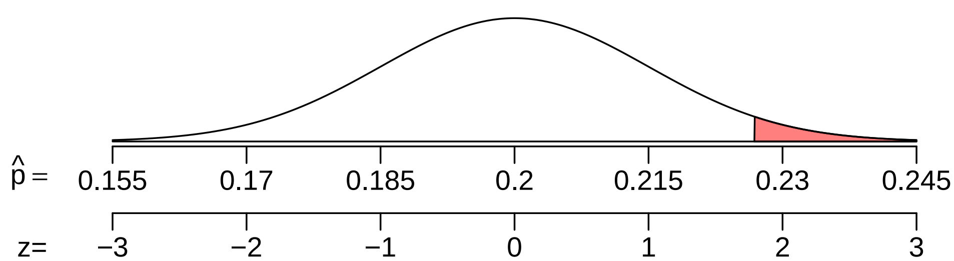
Determine the sample proportion.

$$\hat{p} = \frac{159}{700} = 0.227$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.227 - 0.2}{0.0151} = 1.79$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.227) \\ &= P(Z > 1.79) \\ &= 1 - P(Z < 1.79) \\ &= 0.0367 \end{aligned}$$

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 $p = 0.2$ and H_A claims $p > 0.2$.

(c) The p-value is 0.0367

(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
26	18	
22	52	
15	30	
93	80	
44	55	
32	42	
81	63	
21	39	
13	19	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

   (a) Complete the table.

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

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

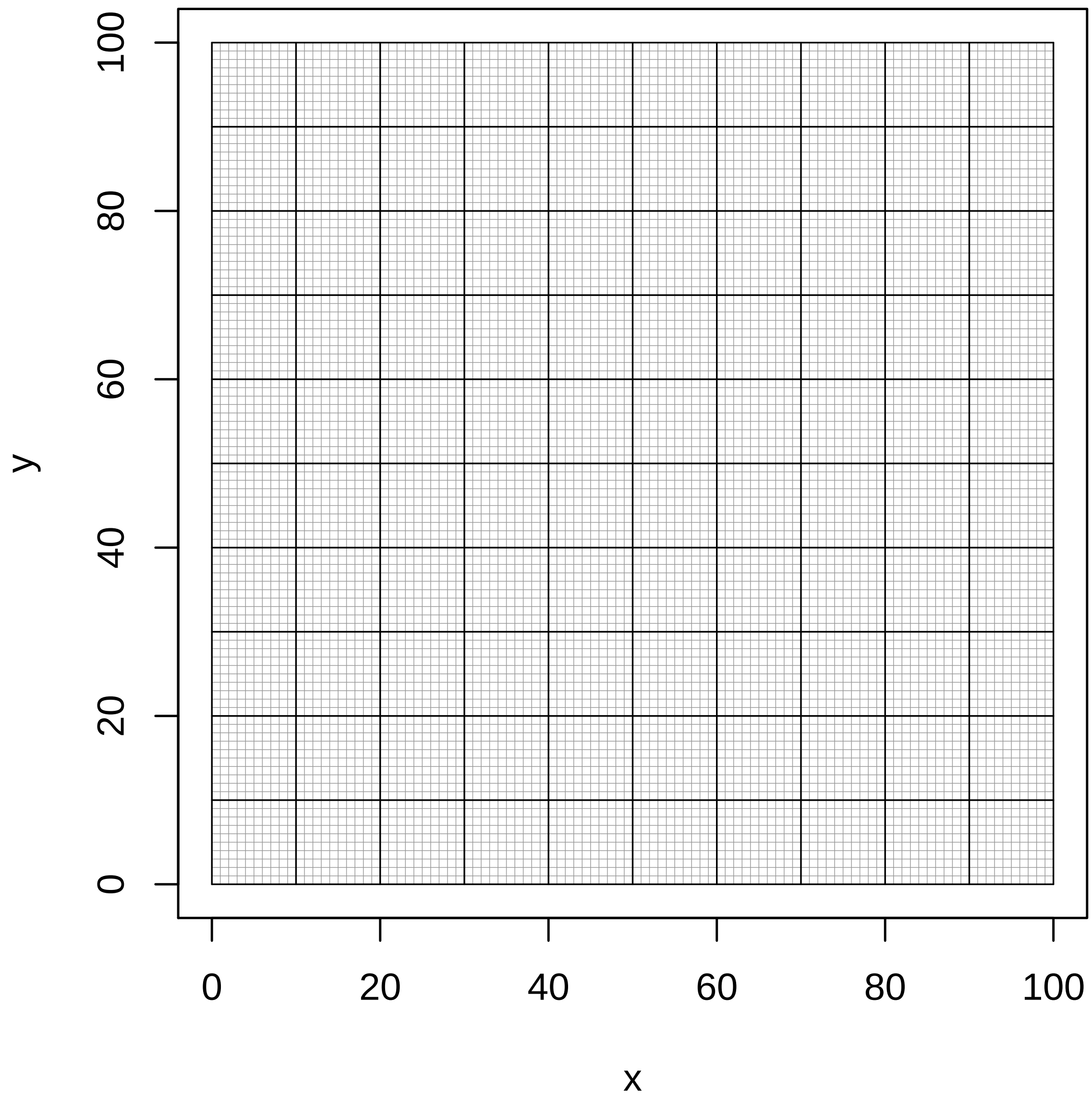
   (c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

   (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
26	18	468
22	52	1144
15	30	450
93	80	7440
44	55	2420
32	42	1344
81	63	5103
21	39	819
13	19	247
$\sum x = 347$	$\sum y = 398$	$\sum x_i y_i = 19435$
$\bar{x} = 38.56$	$\bar{y} = 44.22$	
$s_x = 29.13$	$s_y = 20.52$	

$$r = \frac{19435 - (9)(38.56)(44.22)}{(9-1)(29.13)(20.52)} = 0.855$$

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.

$$r_{\text{exact}} = 0.8555395$$

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:

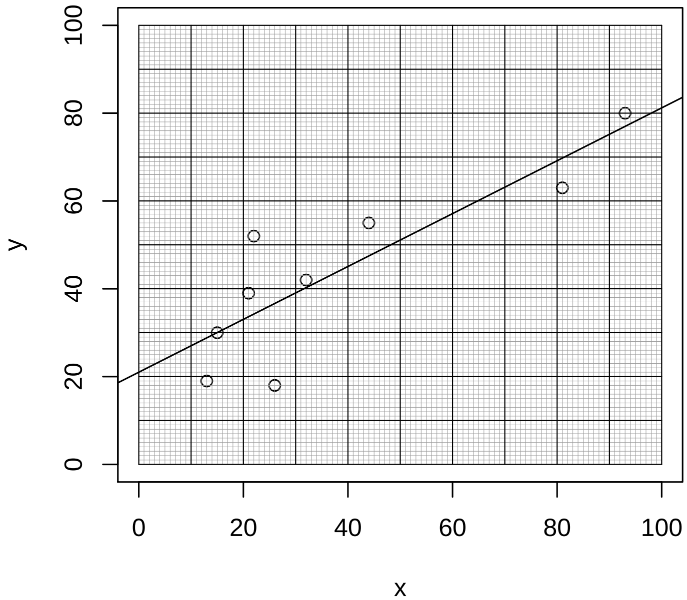
$$b = r\frac{s_y}{s_x} = 0.855 \cdot \frac{20.52}{29.13} = 0.602$$

$$a = \bar{y} - b\bar{x} = 44.2 - (0.602)(38.6) = 21$$

Our regression line:

$$y = 21 + (0.602)x$$

Make a plot.



100
80
60
40
20
0
y
0
20
40
60
80
100
x

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.67$. If 168 trials occur, what is the probability of getting at least 104 but at most 128 successes?

   In other words, let $X \sim \text{Bin}(n = 168,\ p = 0.67)$ and find $P(104 \leq X \leq 128)$.

   Use a normal approximation along with the continuity correction.

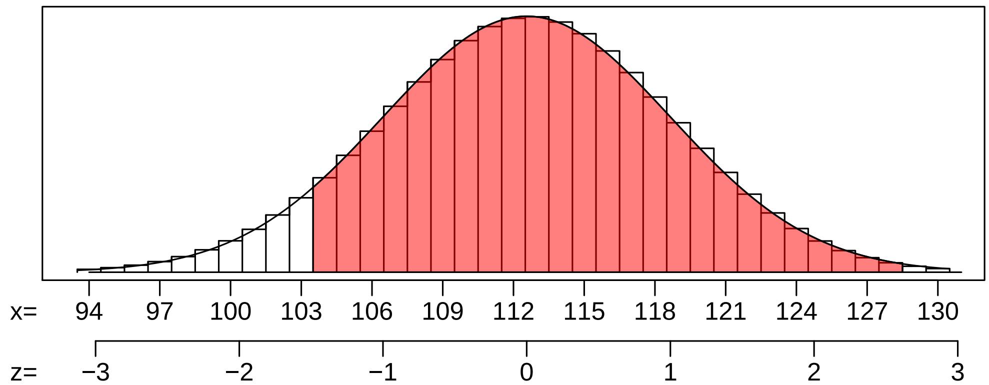
Solution: Find the mean.

$$\mu = np = (168)(0.67) = 112.56$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(168)(0.67)(1-0.67)} = 6.0947$$

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



Find the z scores.

$$z_1 = \frac{103.5 - 112.56}{6.0947} = -1.49$$

$$z_2 = \frac{128.5 - 112.56}{6.0947} = 2.62$$

Find the percentiles (from z-table).

$$\ell_1 = 0.0681$$

$$\ell_2 = 0.9956$$

Calculate the probability.

$$P(104 \leq X \leq 128) = 0.9956 - 0.0681 = 0.9275$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 60$. You decide to run two-tail test on a sample of size $n = 12$ using a significance level $\alpha = 0.1$.

    You then collect the sample:

81.3	70.6	55.6	64.9	69.8
78.3	50.3	66	69	68.5
46.1	65.1			

    (a) Determine the p-value.

    (b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \text{ claims } \mu = 60$$

$$H_A \text{ claims } \mu \neq 60$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 65.458$$

$$s = 10.381$$

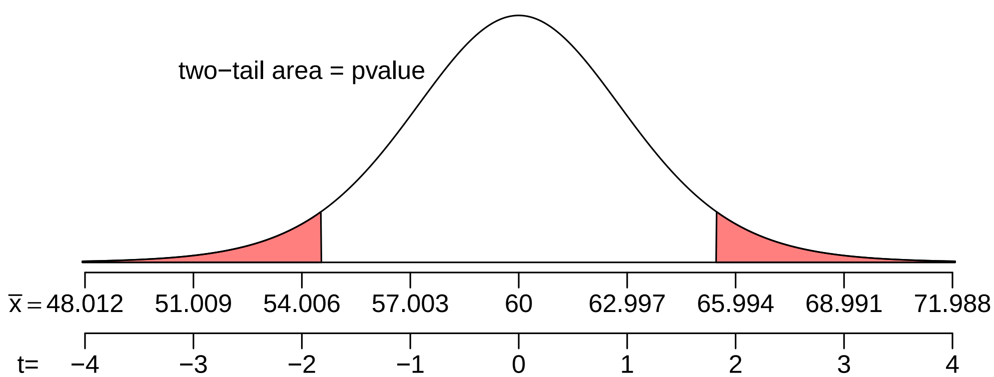
Determine the degrees of freedom.

$$df = 12 - 1 = 11$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{10.381}{\sqrt{12}} = 2.997$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{65.458 - 60}{2.997} = 1.82$$

Find the p-value.

$$p\text{-value} = P(|T| > 1.82)$$

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 $df = 11$.)

$$P(|T| > 2.2) = 0.05$$

$$P(|T| > 1.8) = 0.1$$

Basically, because t is between 2.2 and 1.8, 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.