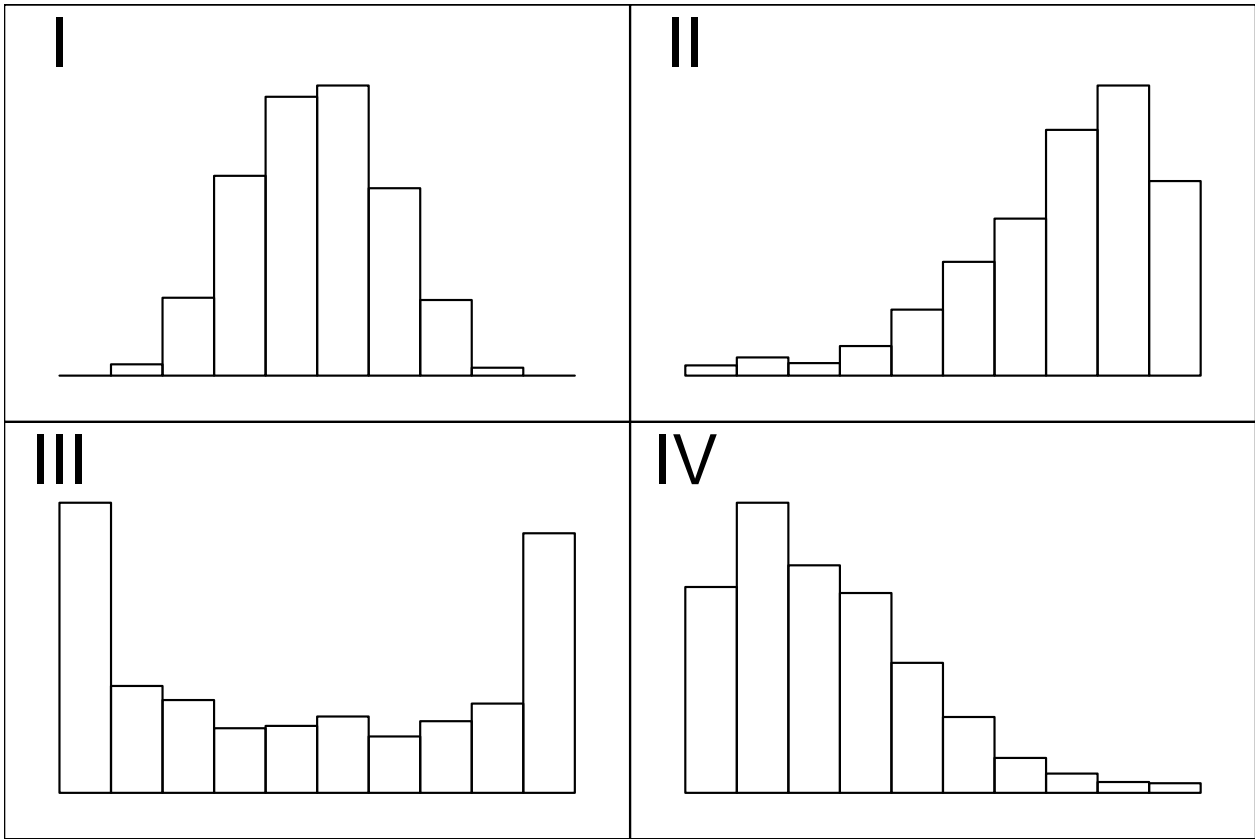# MAT-181 FINAL TAKE-HOME EXAM

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 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.

(b) The distribution of heights of adult women

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

(d) 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.

Solution:

(a) III

(b) I

(c) II

(d) IV

2. (15 Points)

In a deck of strange cards, there are 449 cards. Each card has an image and a color. The amounts are shown in the table below.

	blue	gray	orange	white	Total
dog	42	34	21	26	123
flower	45	29	18	50	142
gem	44	19	23	43	129
wheel	13	12	16	14	55
Total	144	94	78	133	449

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

(b) Is a gem or a wheel more likely to be orange?

(c) What is the probability a random card is orange given it is a gem?

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

(e) What is the probability a random card is both a dog and white?

(f) What is the probability a random card is a wheel given it is white?

(g) What is the probability a random card is either a dog or white (or both)?

Solution:

   (a) $P(\text{gray}) = 0.209$

   (b) $P(\text{orange given gem}) = 0.178$ and $P(\text{orange given wheel}) = 0.291$, so a wheel is more likely to be orange than a gem is.

   (c) $P(\text{orange given gem}) = 0.178$

   (d) $P(\text{dog}) = 0.274$

   (e) $P(\text{dog and white}) = 0.0579$

   (f) $P(\text{wheel given white}) = 0.105$

   (g) $P(\text{dog or white}) = 0.512$

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	116	4
B	146	5
C	67	11
D	143	6

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

Type of fruit	Mass of specimen (g)
A	113
B	143.7
C	67
D	146.5

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	$z = \frac{113-116}{4}$	-0.76
B	$z = \frac{143.7-146}{5}$	-0.46
C	$z = \frac{67-67}{11}$	0
D	$z = \frac{146.5-143}{6}$	0.59

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

4. (10 points)

A tree's leaves were found to be normally distributed with a mean of 103 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 82.6 and 107.3 millimeters?

Solution:

$$\mu = 103$$
$$\sigma = 8.2$$
$$x_1 = 82.6$$
$$x_2 = 107.3$$

$$z_1 \; = \; \frac{x_1 - \mu}{\sigma} \; = \; \frac{82.6 - 103}{8.2} \; = \; -2.49$$

$$z_2 \; = \; \frac{x_2 - \mu}{\sigma} \; = \; \frac{107.3 - 103}{8.2} \; = \; 0.52$$

$$P(x_1 < X < x_2) \; = \; P(z_1 < Z < z_2) \; = \; 0.6985 - 0.0064 \; = \; 0.6921$$

5. (10 points)

A species of duck is known to have a mean weight of 226.1 grams and a standard deviation of 20 grams. A researcher plans to measure the weights of 64 of these ducks sampled randomly. What is the probability the **sample mean** will be between 223.6 and 227.6 grams?

Solution:

$$n = 64$$

$$\mu = 226.1$$

$$\sigma = 20$$

$$SE = \frac{20}{\sqrt{64}} = 2.5$$

$$x_1 = 223.6$$

$$x_2 = 227.6$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{223.6 - 226.1}{2.5} = -1$$

$$z_2 = \frac{x_2 - \mu}{SE} = \frac{227.6 - 226.1}{2.5} = 0.6$$

$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.7257 - 0.1587 = 0.567$$

6. (10 points)

An ornithologist wishes to characterize the average body mass of *Icterus spurius*. She randomly samples 35 adults of *Icterus spurius*, resulting in a sample mean of 21.67 grams and a sample standard deviation of 2.26 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.

$$n = 35$$
$$\bar{x} = 21.67$$
$$s = 2.26$$
$$\gamma = 0.95$$

Find the degrees of freedom.

$$df = n - 1$$
$$= 35 - 1$$
$$= 34$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$ and $df = 34$.

$$t^\star = 2.03$$

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

$$LB = \bar{x} - t^\star \frac{s}{\sqrt{n}} \qquad\qquad UB = \bar{x} + t^\star \frac{s}{\sqrt{n}}$$
$$= 21.67 - 2.03 \times \frac{2.26}{\sqrt{35}} \qquad\qquad = 21.67 + 2.03 \times \frac{2.26}{\sqrt{35}}$$
$$= 20.9 \qquad\qquad\qquad\qquad\qquad = 22.4$$

We are 95% confident that the population mean is between 20.9 and 22.4 grams.

$$CI = (20.9,\ 22.4)$$

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 156 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.

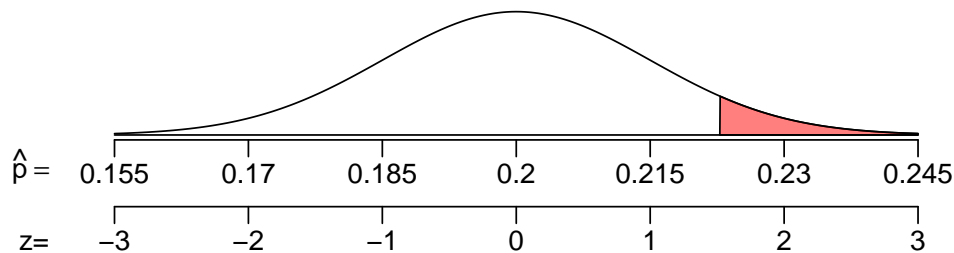$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1 - p_0)}{n}} = \sqrt{\frac{0.2(1 - 0.2)}{700}} = 0.0151$$

Determine the sample proportion.

$$\hat{p} = \frac{156}{700} = 0.223$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.223 - 0.2}{0.0151} = 1.52$$

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.

$$\begin{aligned}
p\text{-value} &= P\left(\hat{p} > 0.223\right) \\
&= P(Z > 1.52) \\
&= 1 - P(Z < 1.52) \\
&= 0.0643
\end{aligned}$$

Compare *p*-value to α (which is 0.05).

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

Make the conclusion: we retain the null hypothesis.

We think the student might just be guessing.

(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 0.0643

(d) We retain the null hypothesis.

(e) We think the student might just be guessing.

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

   You have collected the following data:

x	y	xy
8.3	51	
4.5	88	
2	98	
4.1	43	
1.9	94	
2.1	86	
9.1	76	
1.5	76	
5.4	39	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

(a) Complete the table.

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

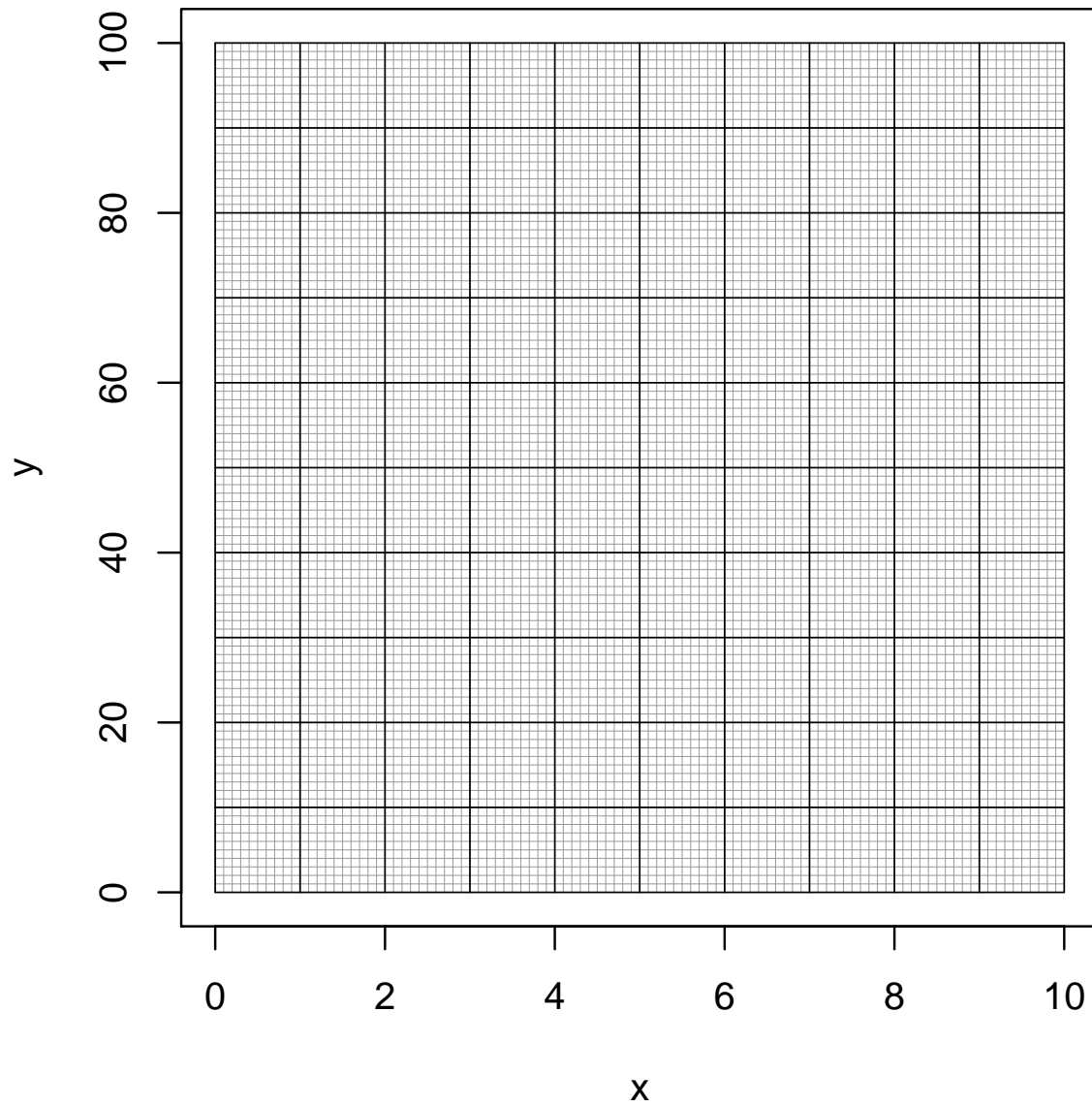$$r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

(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
8.3	51	423.3
4.5	88	396
2	98	196
4.1	43	176.3
1.9	94	178.6
2.1	86	180.6
9.1	76	691.6
1.5	76	114
5.4	39	210.6
$\sum x = 38.9$	$\sum y = 651$	$\sum x_i y_i = 2567$
$\bar{x} = 4.322$	$\bar{y} = 72.33$	
$s_x = 2.83$	$s_y = 22.4$	

$$r = \frac{2567 - (9)(4.322)(72.33)}{(9-1)(2.83)(22.4)} = -0.486$$

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.

$$r_{\text{exact}} = -0.486653$$

The regression line has the form

$$y = a + bx$$

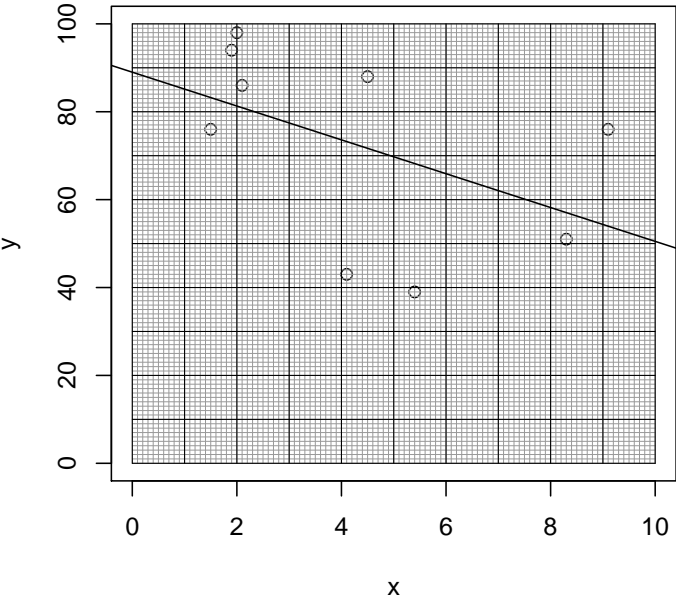So, a is the y-intercept and b is the slope. We have formulas to determine them:

$$b = r\frac{s_y}{s_x} = -0.486 \cdot \frac{22.4}{2.83} = -3.85$$

$$a = \bar{y} - b\bar{x} = 72.3 - (-3.85)(4.32) = 89$$

Our regression line:

$$y = 89 + (-3.85)x$$

Make a plot.

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.15$. If 162 trials occur, what is the probability of getting more than 16 but less than 35 successes?

   In other words, let $X \sim \text{Bin}(n = 162, p = 0.15)$ and find $P(16 < X < 35)$.

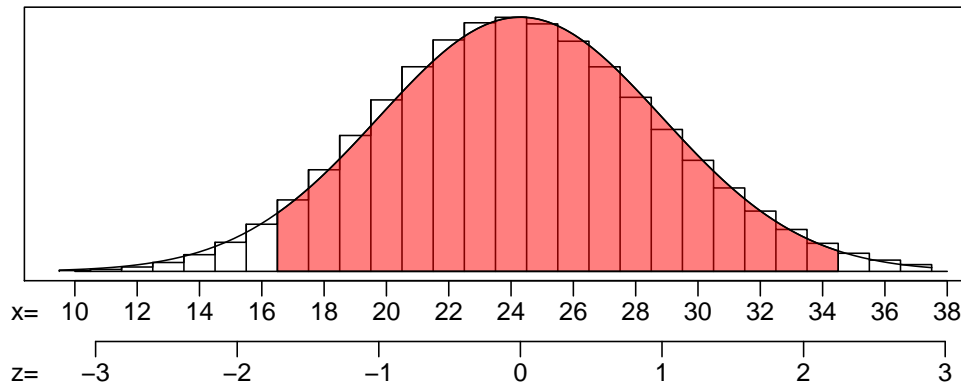   Use a normal approximation along with the continuity correction.

Solution: Find the mean.

$$\mu = np = (162)(0.15) = 24.3$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(162)(0.15)(1-0.15)} = 4.5448$$

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

Find the z scores.

$$z_1 = \frac{16.5 - 24.3}{4.5448} = -1.72$$

$$z_2 = \frac{34.5 - 24.3}{4.5448} = 2.24$$

Find the percentiles (from z-table).

$$\ell_1 = 0.0427$$

$$\ell_2 = 0.9875$$

Calculate the probability.

$$P(16 < X < 35) = 0.9875 - 0.0427 = 0.944$$

10. (Extra credit: 5 points)

A null hypothesis claims a population has a mean μ = 90. You decide to run two-tail test on a sample of size n = 10 using a significance level α = 0.1.

You then collect the sample:

108.3	117	91.5	83.2	113.8
69.3	123.6	127.4	120.2	68.5

(a) Determine the p-value.

(b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 90$$

$$H_A \quad \text{claims} \quad \mu \neq 90$$

Find the mean and standard deviation of the sample.
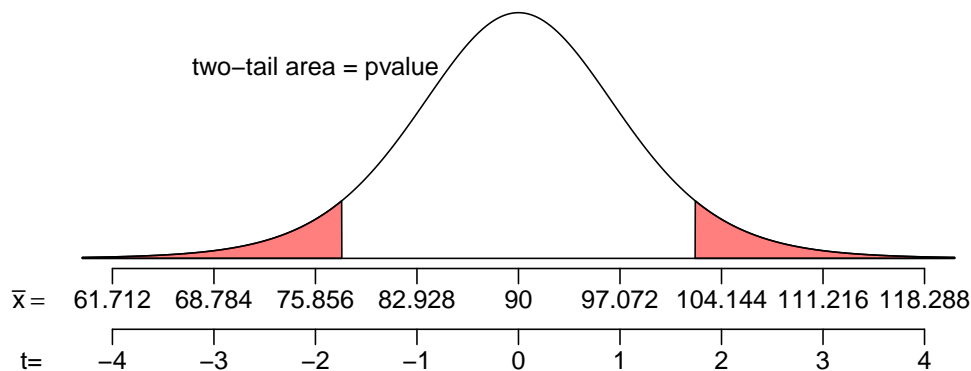
$$\bar{x} = 102.28$$

$$s = 22.364$$

Determine the degrees of freedom.

$$df = 10 - 1 = 9$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{22.364}{\sqrt{10}} = 7.072$$

Make a sketch of the null's sampling distribution.

$\bar{x} =$	61.712	68.784	75.856	82.928	90	97.072	104.144	111.216	118.288
$t =$	−4	−3	−2	−1	0	1	2	3	4

Find the *t* score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{102.28 - 90}{7.072} = 1.74$$

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* = 9.)

$$P(|T| > 1.83) = 0.1$$

$$P(|T| > 1.38) = 0.2$$

Basically, because *t* is between 1.83 and 1.38, 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) $0.1 < p$-value < 0.2

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