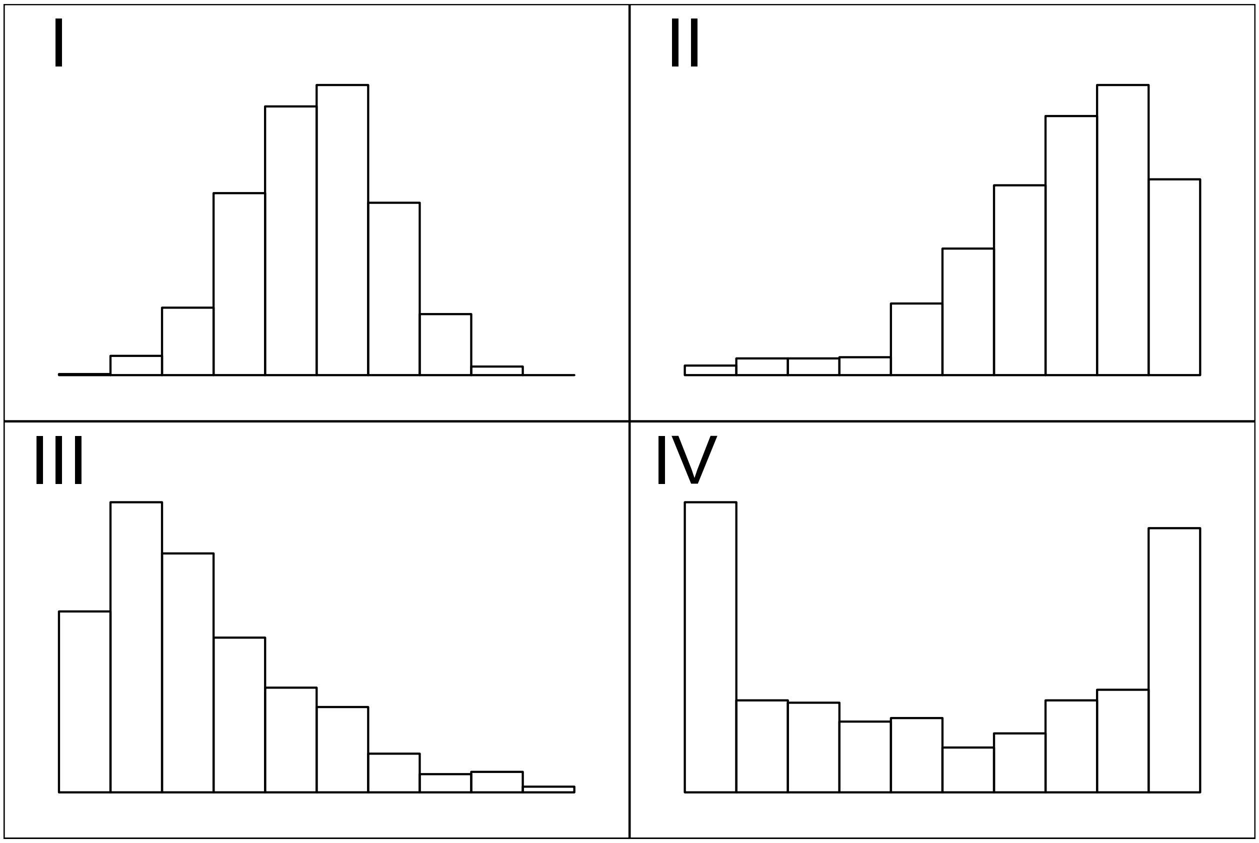
MAT-181 FINAL TAKE-HOME EXAM

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 ages at a skilled nursing facility, where most of the patients are elderly but a few are quite young.

(b) The distribution of weights of newborn babies

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

(d) 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.

Solution:

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

2. (15 Points)

   In a deck of strange cards, there are 352 cards. Each card has an image and a color. The amounts are shown in the table below.

	black	blue	green	Total
kite	25	19	37	81
quilt	40	15	12	67
shovel	36	20	45	101
wheel	30	24	49	103
Total	131	78	143	352

   (a) What is the probability a random card is a shovel given it is black?

   (b) Is a quilt or a wheel more likely to be green?

   (c) What is the probability a random card is blue given it is a wheel?

   (d) What is the probability a random card is both a wheel and blue?

   (e) What is the probability a random card is either a quilt or black (or both)?

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

   (g) What is the probability a random card is a kite?

Solution:

(a) P(shovel given black) = 0.275

(b) P(green given quilt) = 0.179 and P(green given wheel) = 0.476, so a wheel is more likely to be green than a quilt is.

(c) P(blue given wheel) = 0.233

(d) P(wheel and blue) = 0.0682

(e) P(quilt or black) = 0.449

(f) P(blue) = 0.222

(g) P(kite) = 0.23

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	119	15
B	60	14
C	104	5
D	131	11

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

Type of fruit	Mass of specimen (g)
A	128.3
B	67
C	103.8
D	128.6

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

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

Type of fruit	formula	z-score
A	$z = \frac{128.3-119}{15}$	0.62
B	$z = \frac{67-60}{14}$	0.5
C	$z = \frac{103.8-104}{5}$	-0.03
D	$z = \frac{128.6-131}{11}$	-0.22

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

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 97.6 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 90.2 and 107.3 millimeters?

Solution:

$$\mu = 97.6$$
$$\sigma = 7.6$$
$$x_1 = 90.2$$
$$x_2 = 107.3$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{90.2 - 97.6}{7.6} = -0.97$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{107.3 - 97.6}{7.6} = 1.28$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.8997 - 0.166 = 0.7337$$

5. (10 points)

   A species of duck is known to have a mean weight of 108.3 grams and a standard deviation of 60 grams. A researcher plans to measure the weights of 144 of these ducks sampled randomly. What is the probability the **sample mean** will be between 114.8 and 116.8 grams?

Solution:

$$n = 144$$
$$\mu = 108.3$$
$$\sigma = 60$$
$$SE = \frac{60}{\sqrt{144}} = 5$$
$$x_1 = 114.8$$
$$x_2 = 116.8$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{114.8 - 108.3}{5} = 1.3$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{116.8 - 108.3}{5} = 1.7$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.9554 - 0.9032 = 0.0522$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Catharus guttatus*. She randomly samples 20 adults of *Catharus guttatus*, resulting in a sample mean of 28.89 grams and a sample standard deviation of 1.44 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.

$$\begin{aligned} n &= 20 \\ \bar{x} &= 28.89 \\ s &= 1.44 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 20 - 1 \\ &= 19 \end{aligned}$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$ and $df = 19$.

$$t^\star = 2.09$$

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

$$\begin{aligned} LB &= \bar{x} - t^\star \frac{s}{\sqrt{n}} \\ &= 28.89 - 2.09 \times \frac{1.44}{\sqrt{20}} \\ &= 28.2 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^\star \frac{s}{\sqrt{n}} \\ &= 28.89 + 2.09 \times \frac{1.44}{\sqrt{20}} \\ &= 29.6 \end{aligned}$$

We are 95% confident that the population mean is between 28.2 and 29.6 grams.

$$CI = (28.2,\ 29.6)$$

7. (15 points)

   A student is taking a multiple choice test with 300 questions. Each question has 4 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 86 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.

$$p_0 = \frac{1}{4} = 0.25$$

State the hypotheses.

$$\begin{array}{lll} H_0 & \text{claims} & p = 0.25 \\ H_A & \text{claims} & p > 0.25 \end{array}$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.25(1-0.25)}{300}} = 0.025$$

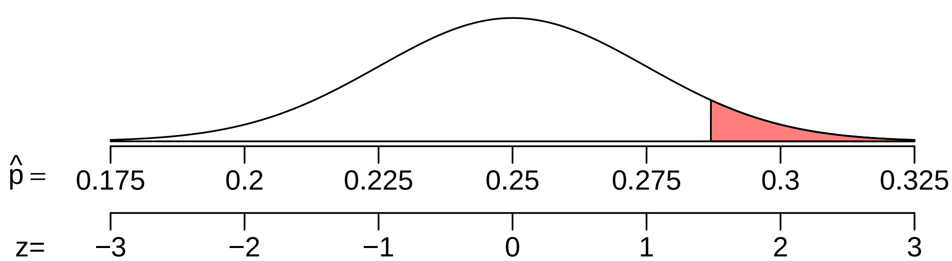
Determine the sample proportion.

$$\hat{p} = \frac{86}{300} = 0.287$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.287 - 0.25}{0.025} = 1.48$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.287) \\ &= P(Z > 1.48) \\ &= 1 - P(Z < 1.48) \\ &= 0.0694 \end{aligned}$$

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 $p = 0.25$ and H_A claims $p > 0.25$.

(c) The p-value is 0.0694

(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
240	7.1	
560	5.1	
180	7.3	
550	5.9	
850	4.4	
130	7.7	
680	4.5	
510	6.4	
160	7.1	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

(a) Complete the table.

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

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

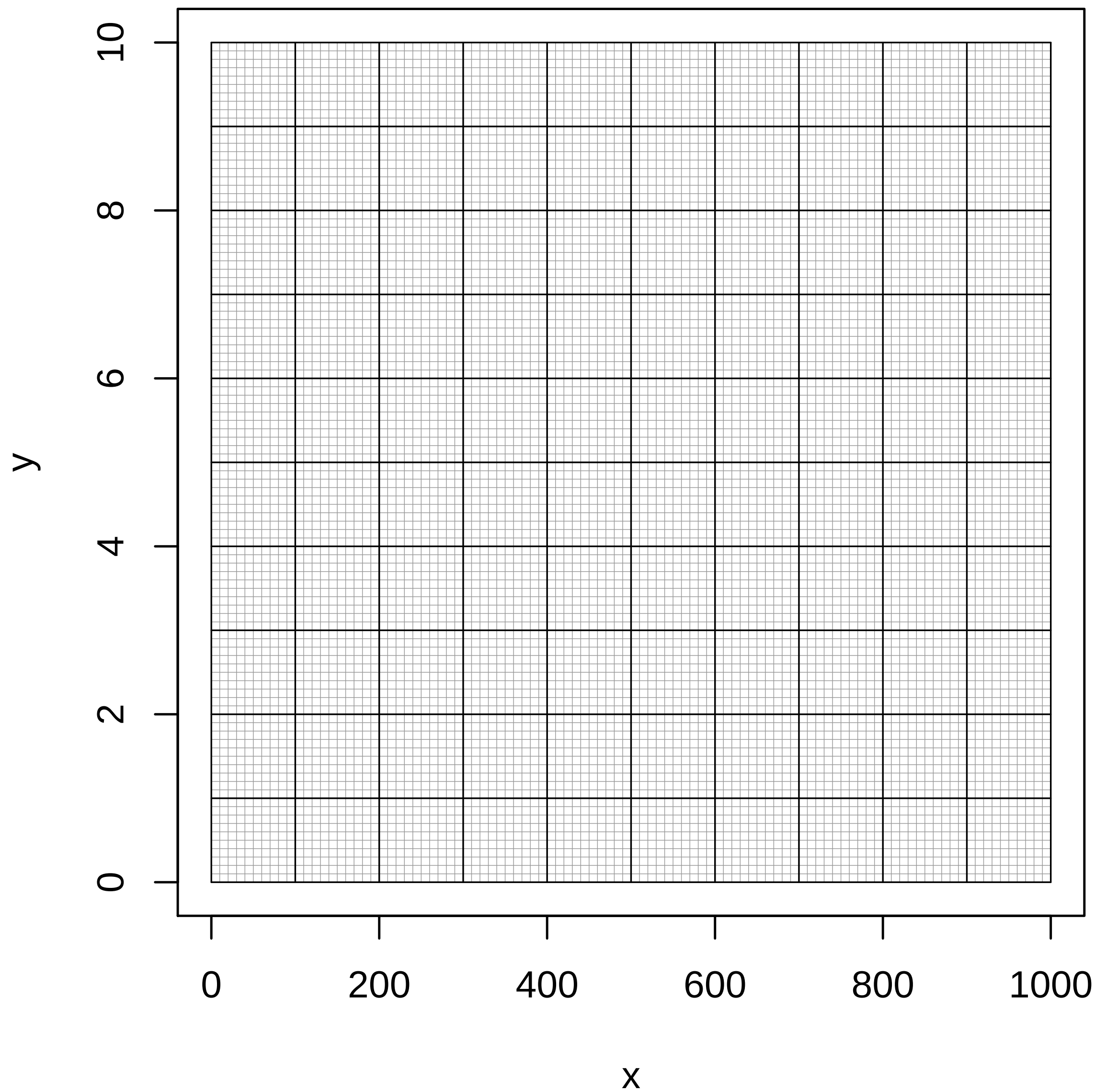
(c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

(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
240	7.1	1704
560	5.1	2856
180	7.3	1314
550	5.9	3245
850	4.4	3740
130	7.7	1001
680	4.5	3060
510	6.4	3264
160	7.1	1136
$\sum x = 3860$	$\sum y = 55.5$	$\sum x_i y_i = 21320$
$\bar{x} = 428.9$	$\bar{y} = 6.167$	
$s_x = 259.3$	$s_y = 1.252$	

$$r = \frac{21320 - (9)(428.9)(6.167)}{(9-1)(259.3)(1.252)} = -0.957$$

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.

$$r_{\text{exact}} = -0.956003$$

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:

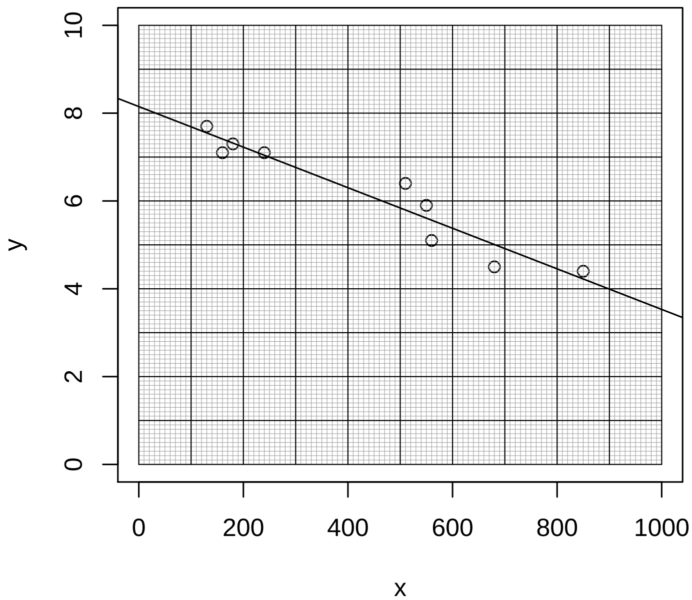
$$b = r\frac{s_y}{s_x} = -0.957 \cdot \frac{1.252}{259.3} = -0.00462$$

$$a = \bar{y} - b\bar{x} = 6.17 - (-0.00462)(429) = 8.15$$

Our regression line:

$$y = 8.15 + (-0.00462)x$$

Make a plot.



10
8
6
4
2
0
y
0
200
400
600
800
1000
x

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.34$. If 81 trials occur, what is the probability of getting at least 23 but less than 36 successes?

   In other words, let $X \sim \text{Bin}(n = 81,\ p = 0.34)$ and find $P(23 \leq X < 36)$.

   Use a normal approximation along with the continuity correction.

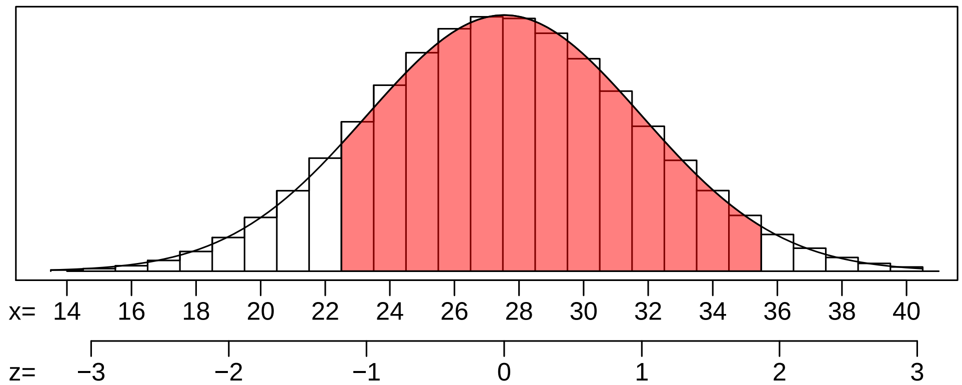
Solution: Find the mean.

$$\mu = np = (81)(0.34) = 27.54$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(81)(0.34)(1-0.34)} = 4.2634$$

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



Find the z scores.

$$z_1 = \frac{22.5 - 27.54}{4.2634} = -1.18$$

$$z_2 = \frac{35.5 - 27.54}{4.2634} = 1.87$$

Find the percentiles (from z-table).

$$\ell_1 = 0.119$$

$$\ell_2 = 0.9693$$

Calculate the probability.

$$P(23 \leq X < 36) = 0.9693 - 0.119 = 0.85$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 180$. You decide to run two-tail test on a sample of size $n = 9$ using a significance level $\alpha = 0.1$.

    You then collect the sample:

284.1	159.7	158.1	201.1	185.9
247.6	144	289.4	299.6	

    (a) Determine the p-value.

    (b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 180$$
$$H_A \quad \text{claims} \quad \mu \neq 180$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 218.833$$
$$s = 62.023$$

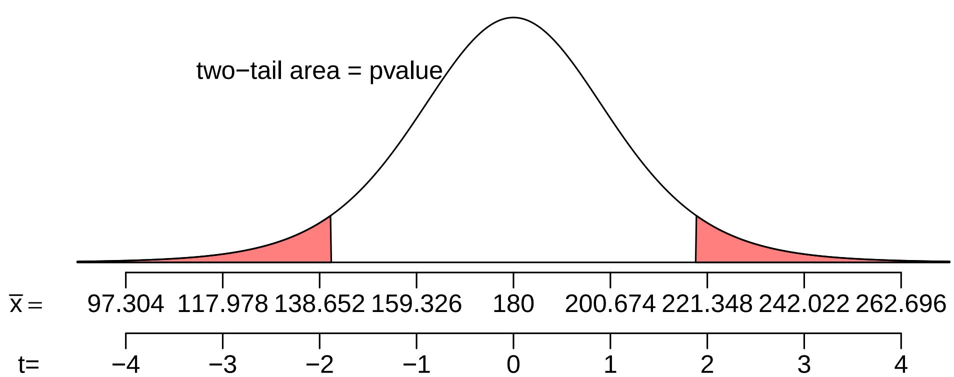
Determine the degrees of freedom.

$$df = 9 - 1 = 8$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{62.023}{\sqrt{9}} = 20.674$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{218.833 - 180}{20.674} = 1.88$$

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 $df = 8$.)

$$P(|T| > 2.31) = 0.05$$
$$P(|T| > 1.86) = 0.1$$

Basically, because t is between 2.31 and 1.86, 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.