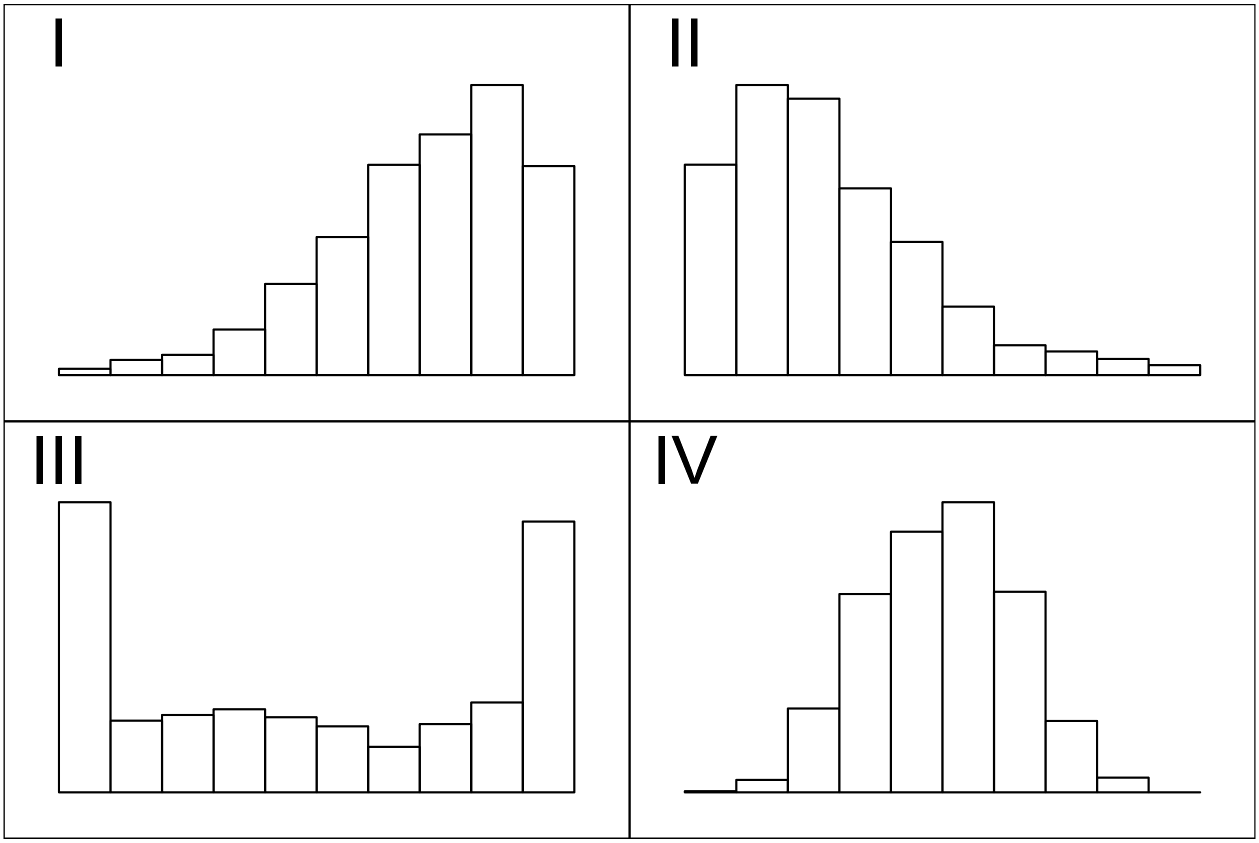
MAT-181 FINAL TAKE-HOME EXAM

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 women

(b) The distribution of annual income for school employees where a high percentage of employees are entry-level teachers and only a few are high-paid administrators.

(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 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.

Solution:

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

2. (15 Points)

In a deck of strange cards, there are 361 cards. Each card has an image and a color. The amounts are shown in the table below.

	gray	green	indigo	pink	Total
bike	14	13	24	47	98
jigsaw	29	48	28	12	117
shovel	22	43	32	49	146
Total	65	104	84	108	361

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

(b) Is a bike or a shovel more likely to be pink?

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

(d) What is the probability a random card is either a bike or gray (or both)?

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

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

(g) What is the probability a random card is both a bike and indigo?

Solution:

(a) P(pink given shovel) = 0.336

(b) P(pink given bike) = 0.48 and P(pink given shovel) = 0.336, so a bike is more likely to be pink than a shovel is.

(c) P(jigsaw given green) = 0.462

(d) P(bike or gray) = 0.413

(e) P(indigo) = 0.233

(f) P(jigsaw) = 0.324

(g) P(bike and indigo) = 0.0665

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	113	10
B	96	4
C	125	11
D	104	14

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

Type of fruit	Mass of specimen (g)
A	105.5
B	92.2
C	130.1
D	120.4

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	$z = \frac{105.5-113}{10}$	-0.75
B	$z = \frac{92.2-96}{4}$	-0.95
C	$z = \frac{130.1-125}{11}$	0.46
D	$z = \frac{120.4-104}{14}$	1.17

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

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 163.5 millimeters and a standard deviation of 3.7 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 160.5 and 163.7 millimeters?

Solution:

$$\mu = 163.5$$
$$\sigma = 3.7$$
$$x_1 = 160.5$$
$$x_2 = 163.7$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{160.5 - 163.5}{3.7} = -0.81$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{163.7 - 163.5}{3.7} = 0.05$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.5199 - 0.209 = 0.3109$$

5. (10 points)

   A species of duck is known to have a mean weight of 242.4 grams and a standard deviation of 35 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 233.4 and 252.9 grams?

Solution:

$$n = 49$$
$$\mu = 242.4$$
$$\sigma = 35$$
$$SE = \frac{35}{\sqrt{49}} = 5$$
$$x_1 = 233.4$$
$$x_2 = 252.9$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{233.4 - 242.4}{5} = -1.8$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{252.9 - 242.4}{5} = 2.1$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.9821 - 0.0359 = 0.9462$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Catharus fuscescens*. She randomly samples 13 adults of *Catharus fuscescens*, resulting in a sample mean of 42.23 grams and a sample standard deviation of 4.33 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.

$$\begin{aligned} n &= 13 \\ \bar{x} &= 42.23 \\ s &= 4.33 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 13 - 1 \\ &= 12 \end{aligned}$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$ and $df = 12$.

$$t^\star = 2.18$$

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

$$\begin{aligned} LB &= \bar{x} - t^\star \frac{s}{\sqrt{n}} \\ &= 42.23 - 2.18 \times \frac{4.33}{\sqrt{13}} \\ &= 39.6 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^\star \frac{s}{\sqrt{n}} \\ &= 42.23 + 2.18 \times \frac{4.33}{\sqrt{13}} \\ &= 44.8 \end{aligned}$$

We are 95% confident that the population mean is between 39.6 and 44.8 grams.

$$CI = (39.6,\ 44.8)$$

7. (15 points)

   A student is taking a multiple choice test with 300 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 165 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.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{300}} = 0.0289$$

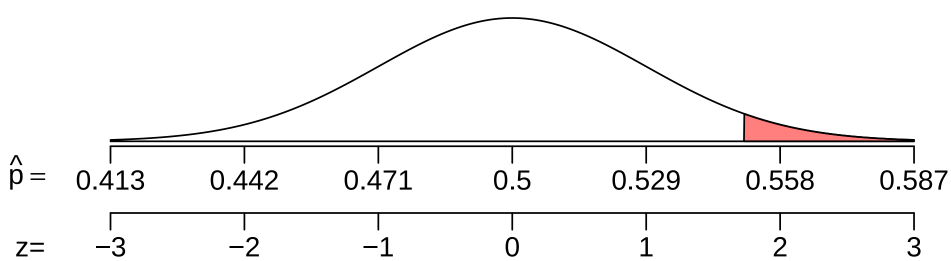
Determine the sample proportion.

$$\hat{p} = \frac{165}{300} = 0.55$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.55 - 0.5}{0.0289} = 1.73$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.55) \\ &= P(Z > 1.73) \\ &= 1 - P(Z < 1.73) \\ &= 0.0418 \end{aligned}$$

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 0.0418

(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
87	3.5	
20	6.8	
28	7.6	
46	5.9	
91	3.4	
32	6.7	
14	7.2	
98	3.3	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

   (a) Complete the table.

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

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

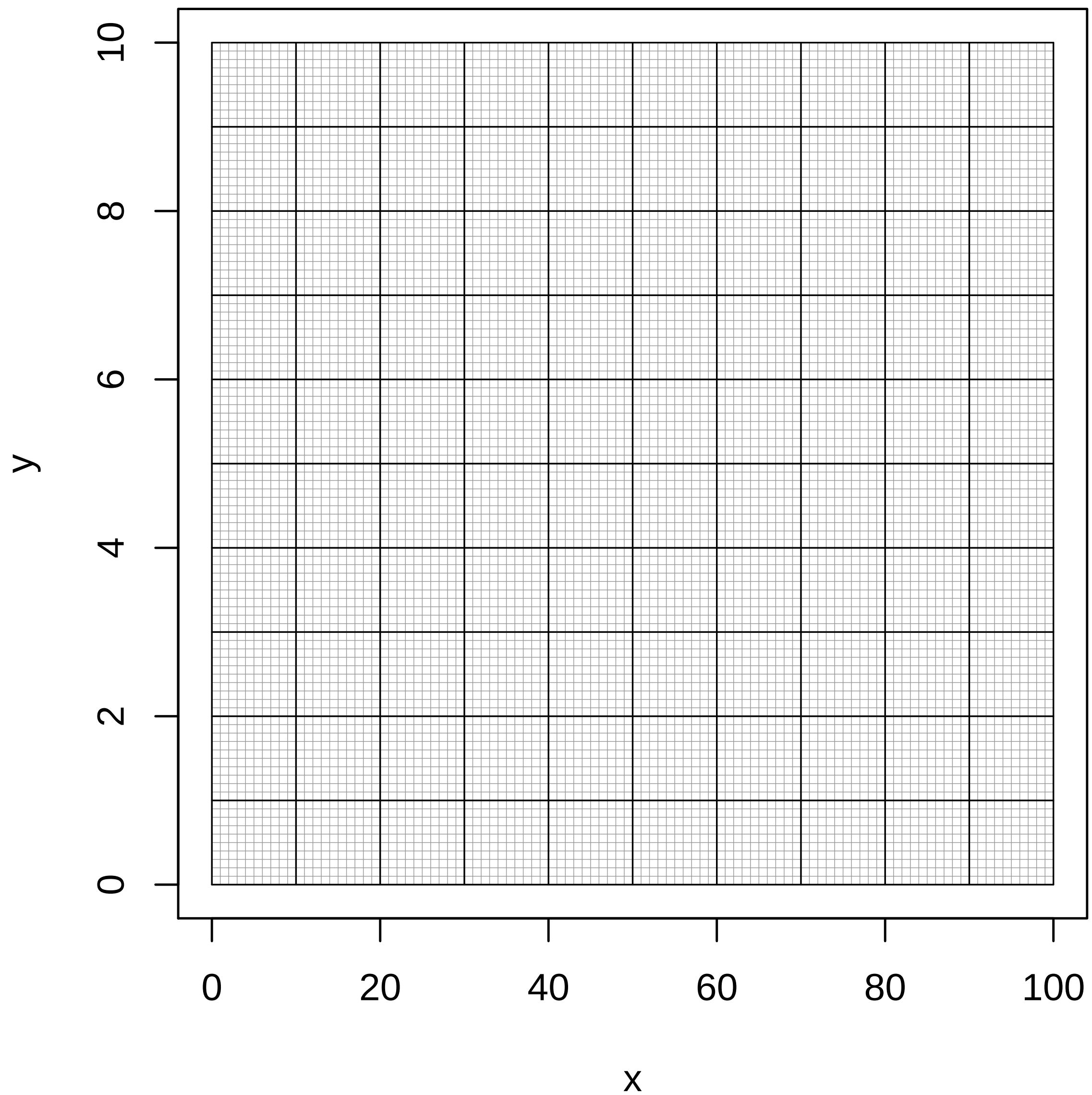
   (c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

   (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
87	3.5	304.5
20	6.8	136
28	7.6	212.8
46	5.9	271.4
91	3.4	309.4
32	6.7	214.4
14	7.2	100.8
98	3.3	323.4
$\sum x = 416$	$\sum y = 44.4$	$\sum x_i y_i = 1872.7$
$\bar{x} = 52$	$\bar{y} = 5.55$	
$s_x = 34.52$	$s_y = 1.845$	

$$r = \frac{1872.7 - (8)(52)(5.55)}{(8-1)(34.52)(1.845)} = -0.978$$

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.

$$r_{\text{exact}} = -0.9783182$$

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:

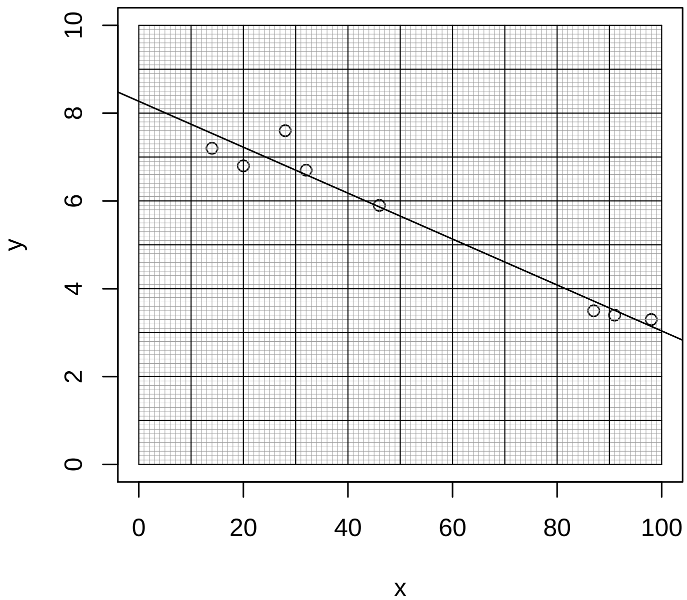
$$b = r\frac{s_y}{s_x} = -0.978 \cdot \frac{1.845}{34.52} = -0.0523$$

$$a = \bar{y} - b\bar{x} = 5.55 - (-0.0523)(52) = 8.27$$

Our regression line:

$$y = 8.27 + (-0.0523)x$$

Make a plot.



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

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.65$. If 70 trials occur, what is the probability of getting more than 40 but less than 46 successes?

   In other words, let $X \sim \text{Bin}(n = 70,\ p = 0.65)$ and find $P(40 < X < 46)$.

   Use a normal approximation along with the continuity correction.

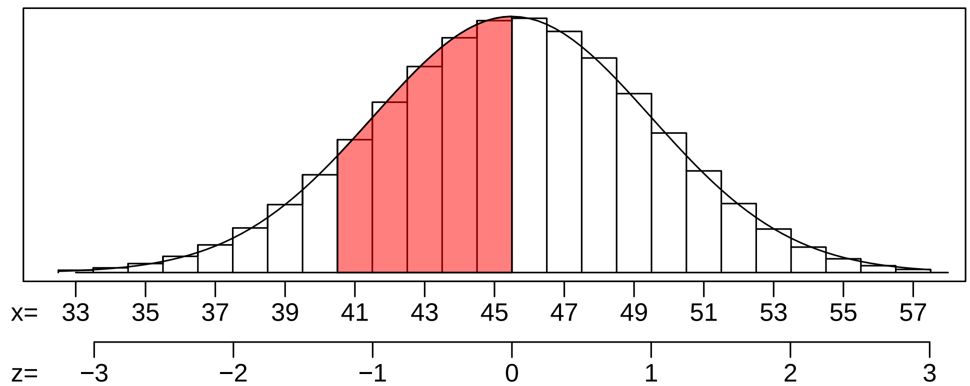
Solution: Find the mean.

$$\mu = np = (70)(0.65) = 45.5$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(70)(0.65)(1-0.65)} = 3.9906$$

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



Find the z scores.

$$z_1 = \frac{40.5 - 45.5}{3.9906} = -1.25$$

$$z_2 = \frac{45.5 - 45.5}{3.9906} = 0$$

Find the percentiles (from z-table).

$$\ell_1 = 0.1056$$

$$\ell_2 = 0.5$$

Calculate the probability.

$$P(40 < X < 46) = 0.5 - 0.1056 = 0.394$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 130$. You decide to run two-tail test on a sample of size $n = 8$ using a significance level $\alpha = 0.1$.

    You then collect the sample:

135.3	182	215.1	118.5	146.6
166.9	211.2	96.4		

    (a) Determine the p-value.
    (b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 130$$
$$H_A \quad \text{claims} \quad \mu \neq 130$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 159$$
$$s = 42.658$$

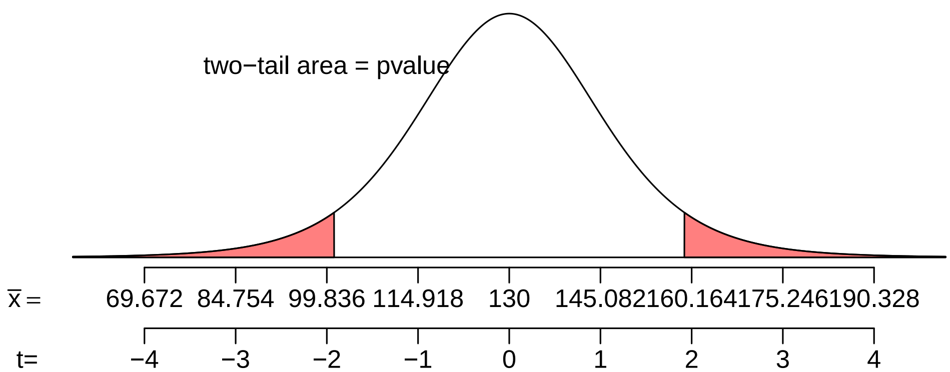
Determine the degrees of freedom.

$$df = 8 - 1 = 7$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{42.658}{\sqrt{8}} = 15.082$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{159 - 130}{15.082} = 1.92$$

Find the p-value.

$$p\text{-value} = P(|T| > 1.92)$$

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 $df = 7$.)

$$P(|T| > 2.36) = 0.05$$
$$P(|T| > 1.89) = 0.1$$

Basically, because t is between 2.36 and 1.89, 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.