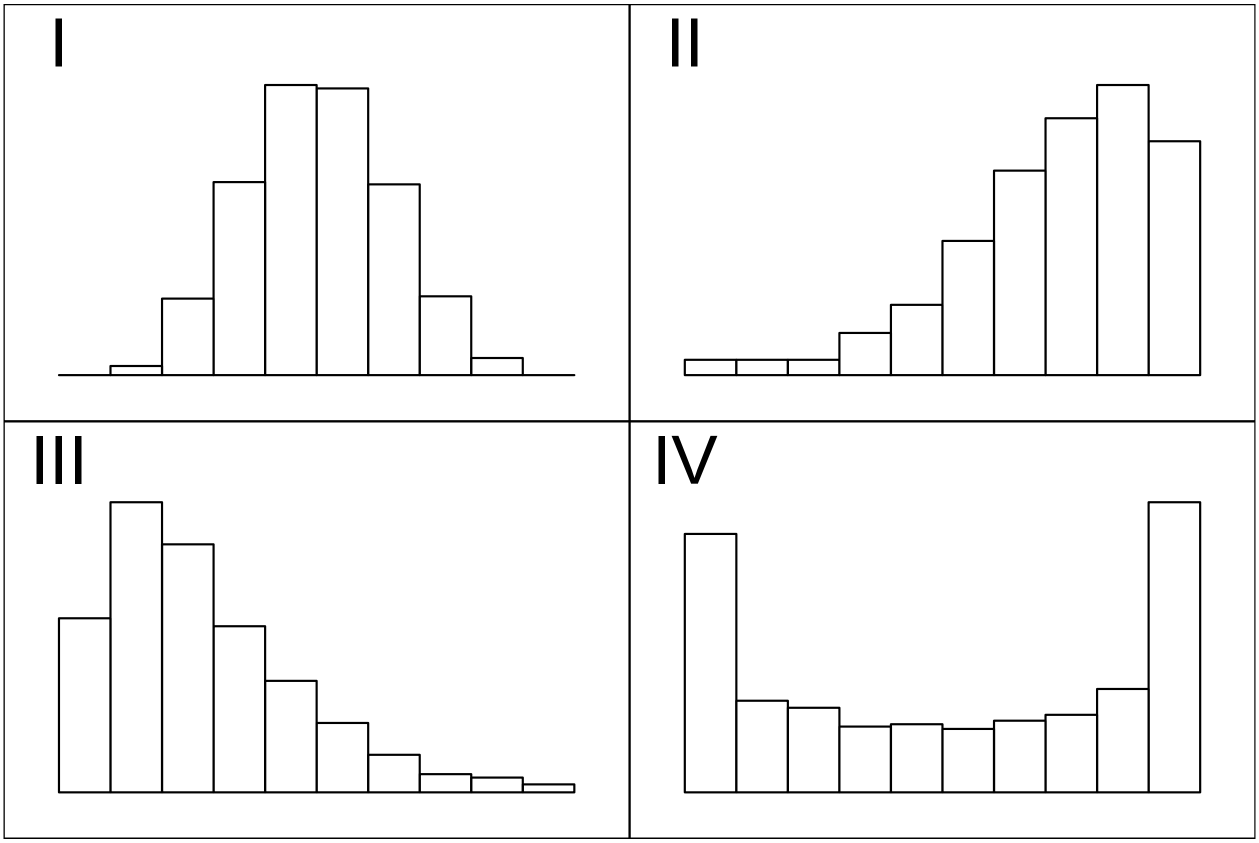
MAT-181 FINAL TAKE-HOME EXAM

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

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

(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 heights of adult women

Solution:

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

2. (15 Points)

In a deck of strange cards, there are 737 cards. Each card has an image and a color. The amounts are shown in the table below.

	orange	red	teal	violet	white	Total
bike	24	36	14	38	37	149
cat	35	15	49	31	30	160
gem	45	34	13	26	40	158
lamp	20	46	48	27	21	162
pig	10	33	18	22	25	108
Total	134	164	142	144	153	737

(a) Is a cat or a pig more likely to be violet?

(b) What is the probability a random card is either a gem or orange (or both)?

(c) What is the probability a random card is violet given it is a cat?

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

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

(f) What is the probability a random card is a bike given it is orange?

(g) What is the probability a random card is both a gem and white?

Solution:

(a) P(violet given cat) = 0.194 and P(violet given pig) = 0.204, so a pig is more likely to be violet than a cat is.

(b) P(gem or orange) = 0.335

(c) P(violet given cat) = 0.194

(d) P(white) = 0.208

(e) P(pig) = 0.147

(f) P(bike given orange) = 0.179

(g) P(gem and white) = 0.0543

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	4
B	73	5
C	101	14
D	93	12

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

Type of fruit	Mass of specimen (g)
A	115.2
B	68.25
C	120
D	110.4

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	$z = \frac{\lvert 115.2-113 \rvert}{4}$	0.54
B	$z = \frac{\lvert 68.25-73 \rvert}{5}$	0.95
C	$z = \frac{\lvert 120-101 \rvert}{14}$	1.36
D	$z = \frac{\lvert 110.4-93 \rvert}{12}$	1.45

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

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 59.4 millimeters and a standard deviation of 10 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 50.1 and 57.9 millimeters?

Solution:

$$\mu = 59.4$$
$$\sigma = 10$$
$$x_1 = 50.1$$
$$x_2 = 57.9$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{50.1 - 59.4}{10} = -0.93$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{57.9 - 59.4}{10} = -0.15$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.4404 - 0.1762 = 0.2642$$

5. (10 points)

   A species of duck is known to have a mean weight of 265.6 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 263.6 and 265.6 grams?

Solution:

$$n = 64$$
$$\mu = 265.6$$
$$\sigma = 20$$
$$SE = \frac{20}{\sqrt{64}} = 2.5$$
$$x_1 = 263.6$$
$$x_2 = 265.6$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{263.6 - 265.6}{2.5} = -0.8$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{265.6 - 265.6}{2.5} = 0$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.5 - 0.2119 = 0.2881$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Cistothorus palustris*. She randomly samples 31 adults of *Cistothorus palustris*, resulting in a sample mean of 9.87 grams and a sample standard deviation of 1.22 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.

$$n = 31$$
$$\bar{x} = 9.87$$
$$s = 1.22$$
$$\gamma = 0.95$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 31 - 1 \\ &= 30 \end{aligned}$$

Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 30$.

$$t^{\star} = 2.04$$

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

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} \\ &= 9.87 - 2.04 \times \frac{1.22}{\sqrt{31}} \\ &= 9.42 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 9.87 + 2.04 \times \frac{1.22}{\sqrt{31}} \\ &= 10.3 \end{aligned}$$

We are 95% confident that the population mean is between 9.42 and 10.3 grams.

$$CI = (9.42,\ 10.3)$$

7. (15 points)

A student is taking a multiple choice test with 600 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 169 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.

$$H_0 \quad \text{claims} \quad p = 0.25$$
$$H_A \quad \text{claims} \quad p > 0.25$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.25(1-0.25)}{600}} = 0.0177$$

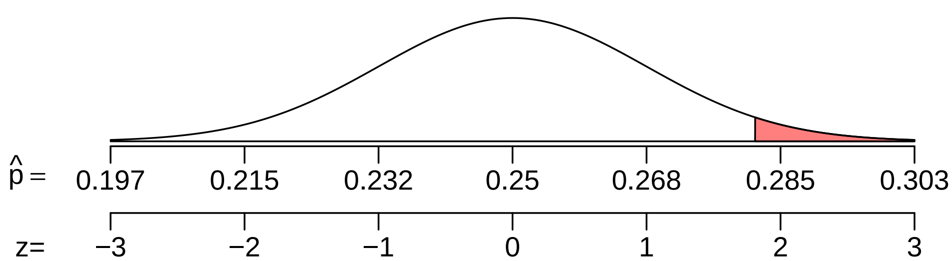
Determine the sample proportion.

$$\hat{p} = \frac{169}{600} = 0.282$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.282 - 0.25}{0.0177} = 1.81$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.282) \\ &= P(Z > 1.81) \\ &= 1 - P(Z < 1.81) \\ &= 0.0351 \end{aligned}$$

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

(c) The p-value is 0.0351

(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
260	77	
720	41	
300	67	
600	49	
610	28	
660	36	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

(a) Complete the table.

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

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

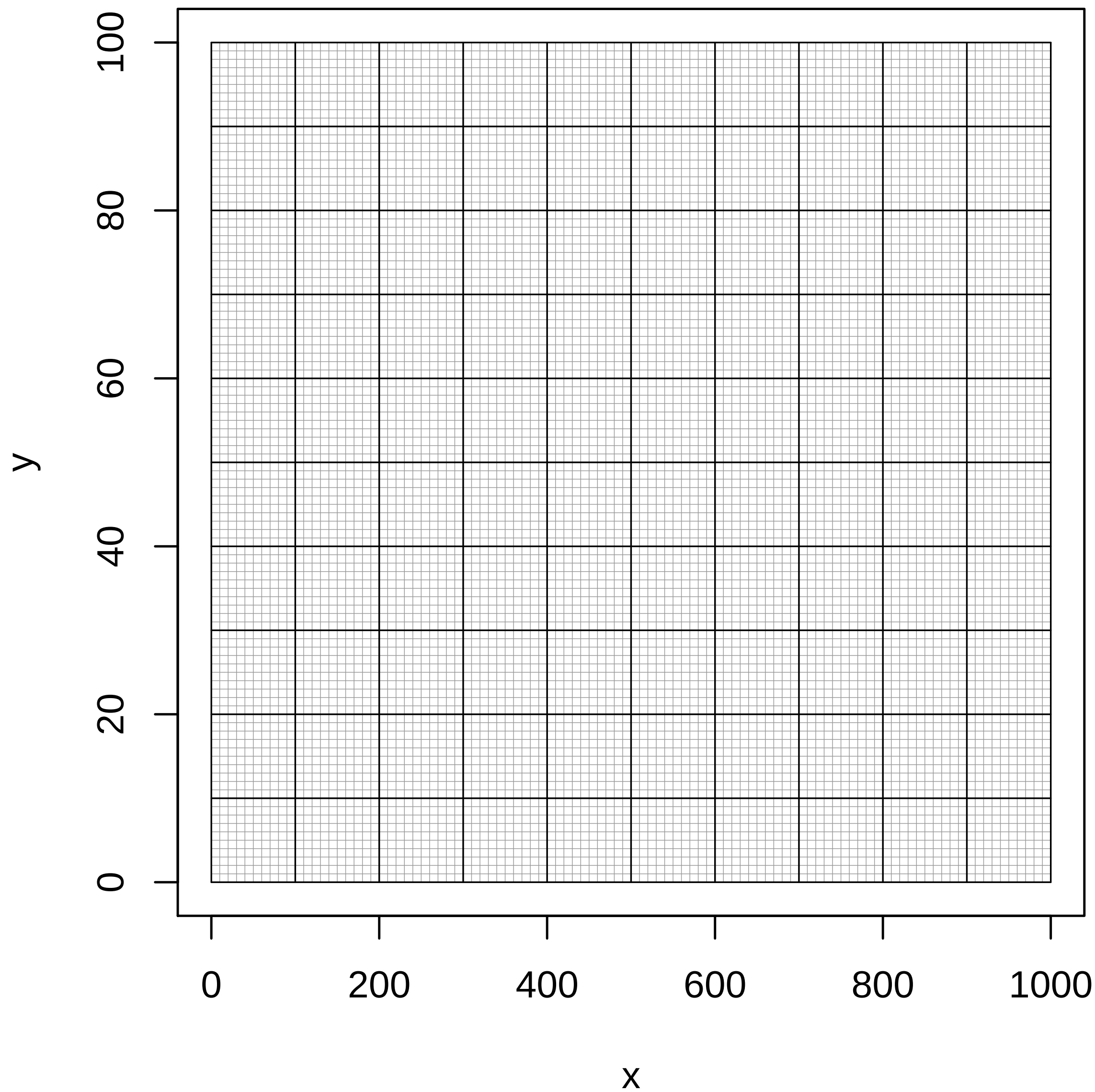
(c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

(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
260	77	20020
720	41	29520
300	67	20100
600	49	29400
610	28	17080
660	36	23760
$\sum x = 3150$	$\sum y = 298$	$\sum x_i y_i = 139880$
$\bar{x} = 525$	$\bar{y} = 49.67$	
$s_x = 194.9$	$s_y = 18.86$	

$$r = \frac{139880 - (6)(525)(49.67)}{(6-1)(194.9)(18.86)} = -0.902$$

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.

$$r_{\text{exact}} = -0.90131$$

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:

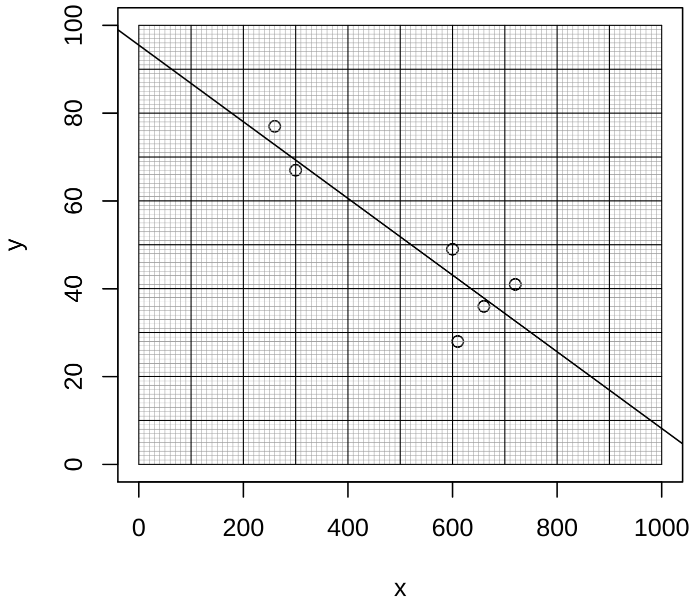
$$b = r\frac{s_y}{s_x} = -0.902 \cdot \frac{18.86}{194.9} = -0.0873$$

$$a = \bar{y} - b\bar{x} = 49.7 - (-0.0873)(525) = 95.5$$

Our regression line:

$$y = 95.5 + (-0.0873)x$$

Make a plot.



y
x
0
20
40
60
80
100
0
200
400
600
800
1000

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.58$. If 72 trials occur, what is the probability of getting at least 33 but less than 44 successes?

   In other words, let $X \sim \text{Bin}(n = 72,\ p = 0.58)$ and find $P(33 \leq X < 44)$.

   Use a normal approximation along with the continuity correction.

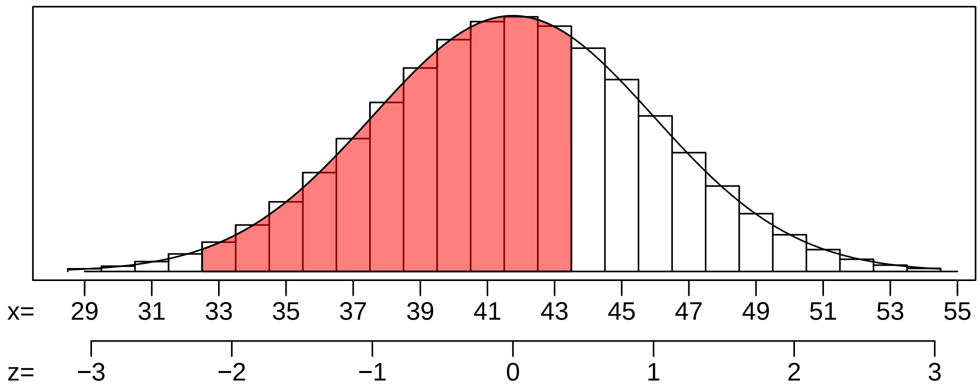
Solution: Find the mean.

$$\mu = np = (72)(0.58) = 41.76$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(72)(0.58)(1-0.58)} = 4.188$$

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



Find the z scores.

$$z_1 = \frac{32.5 - 41.76}{4.188} = -2.21$$

$$z_2 = \frac{43.5 - 41.76}{4.188} = 0.42$$

Find the percentiles (from z-table).

$$\ell_1 = 0.0136$$

$$\ell_2 = 0.6628$$

Calculate the probability.

$$P(33 \leq X < 44) = 0.6628 - 0.0136 = 0.649$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 170$. You decide to run two-tail test on a sample of size $n = 8$ using a significance level $\alpha = 0.05$.

    You then collect the sample:

201.3	156.7	187.4	169.1	211.5
215.3	214.1	166.1		

    (a) Determine the *p*-value.

    (b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 170$$

$$H_A \quad \text{claims} \quad \mu \neq 170$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 190.188$$

$$s = 23.707$$

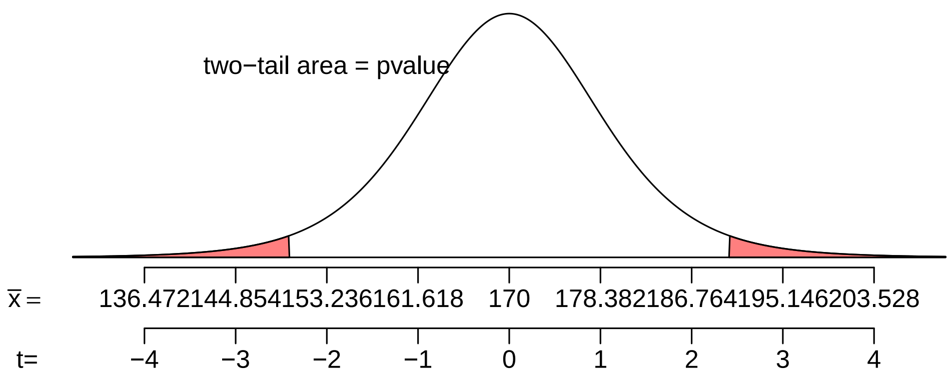
Determine the degrees of freedom.

$$df = 8 - 1 = 7$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{23.707}{\sqrt{8}} = 8.382$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{190.188 - 170}{8.382} = 2.41$$

Find the p-value.

$$p\text{-value} = P(|T| > 2.41)$$

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 $df = 7$.)

$$P(|T| > 2.52) = 0.04$$

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

Basically, because t is between 2.52 and 2.36, we know the p-value is between 0.04 and 0.05.

$$0.04 < p\text{-value} < 0.05$$

Compare the p-value and the significance level ($\alpha = 0.05$).

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

Yes, we reject the null hypothesis.

(a) $0.04 < p\text{-value} < 0.05$

(b) Yes, we reject the null hypothesis.