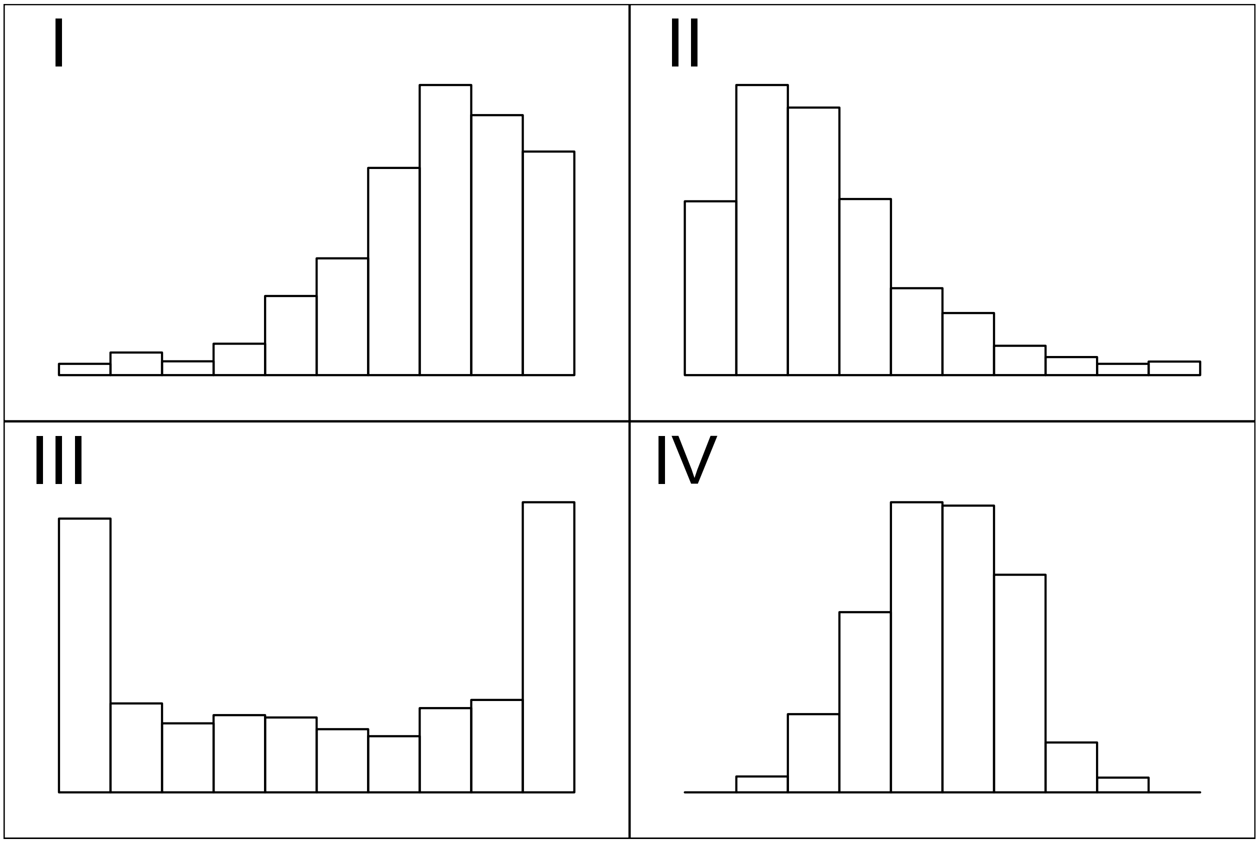
MAT-181 FINAL TAKE-HOME EXAM

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 lengths of newborn babies

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

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

(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) I
(c) II
(d) III

2. (15 Points)

   In a deck of strange cards, there are 709 cards. Each card has an image and a color. The amounts are shown in the table below.

	black	blue	green	red	teal	Total
gem	26	45	16	25	40	152
kite	39	15	12	30	36	132
quilt	27	18	35	23	31	134
shovel	14	46	34	37	49	180
tree	10	19	21	41	20	111
Total	116	143	118	156	176	709

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

   (b) What is the probability a random card is either a shovel or black (or both)?

   (c) What is the probability a random card is teal given it is a tree?

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

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

   (f) Is a gem or a quilt more likely to be green?

   (g) What is the probability a random card is both a quilt and red?

Solution:

(a) P(red) = 0.22

(b) P(shovel or black) = 0.398

(c) P(teal given tree) = 0.18

(d) P(kite given blue) = 0.105

(e) P(tree) = 0.157

(f) P(green given gem) = 0.105 and P(green given quilt) = 0.261, so a quilt is more likely to be green than a gem is.

(g) P(quilt and red) = 0.0324

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	127	8
B	86	6
C	145	13
D	124	9

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

Type of fruit	Mass of specimen (g)
A	128
B	85.22
C	138.4
D	129.5

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	$z = \frac{128-127}{8}$	0.12
B	$z = \frac{85.22-86}{6}$	-0.13
C	$z = \frac{138.4-145}{13}$	-0.51
D	$z = \frac{129.5-124}{9}$	0.61

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

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 55.9 millimeters and a standard deviation of 9.3 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 57.9 and 62.6 millimeters?

Solution:

$$\mu = 55.9$$
$$\sigma = 9.3$$
$$x_1 = 57.9$$
$$x_2 = 62.6$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{57.9 - 55.9}{9.3} = 0.22$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{62.6 - 55.9}{9.3} = 0.72$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.7642 - 0.5871 = 0.1771$$

5. (10 points)

   A species of duck is known to have a mean weight of 296.6 grams and a standard deviation of 12 grams. A researcher plans to measure the weights of 36 of these ducks sampled randomly. What is the probability the **sample mean** will be between 299.1 and 301.1 grams?

Solution:

$$
\begin{aligned}
n &= 36 \\
\mu &= 296.6 \\
\sigma &= 12 \\
SE &= \frac{12}{\sqrt{36}} = 2 \\
x_1 &= 299.1 \\
x_2 &= 301.1
\end{aligned}
$$

$$
\begin{aligned}
z_1 &= \frac{x_1 - \mu}{SE} = \frac{299.1 - 296.6}{2} = 1.25 \\
z_2 &= \frac{x_2 - \mu}{SE} = \frac{301.1 - 296.6}{2} = 2.25 \\
P(x_1 < \overline{X} < x_2) &= P(z_1 < Z < z_2) = 0.9878 - 0.8944 = 0.0934
\end{aligned}
$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Catharus fuscescens*. She randomly samples 27 adults of *Catharus fuscescens*, resulting in a sample mean of 41.96 grams and a sample standard deviation of 5.09 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.

$$\begin{aligned} n &= 27 \\ \bar{x} &= 41.96 \\ s &= 5.09 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 27 - 1 \\ &= 26 \end{aligned}$$

Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 26$.

$$t^{\star} = 2.06$$

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

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} & \qquad UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 41.96 - 2.06 \times \frac{5.09}{\sqrt{27}} & &= 41.96 + 2.06 \times \frac{5.09}{\sqrt{27}} \\ &= 39.9 & &= 44 \end{aligned}$$

We are 95% confident that the population mean is between 39.9 and 44 grams.

$$CI = (39.9,\ 44)$$

7. (15 points)

   A student is taking a multiple choice test with 400 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 116 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.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.25(1-0.25)}{400}} = 0.0217$$

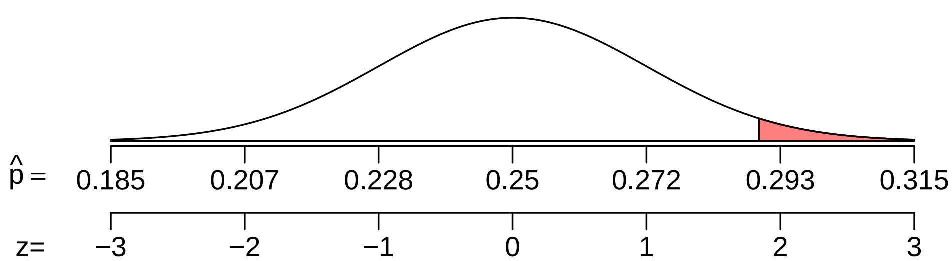
Determine the sample proportion.

$$\hat{p} = \frac{116}{400} = 0.29$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.29 - 0.25}{0.0217} = 1.84$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.29) \\ &= P(Z > 1.84) \\ &= 1 - P(Z < 1.84) \\ &= 0.0329 \end{aligned}$$

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 0.0329
(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
23	4.1	
94	7.5	
32	5.7	
25	4.7	
59	5.4	
10	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.

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

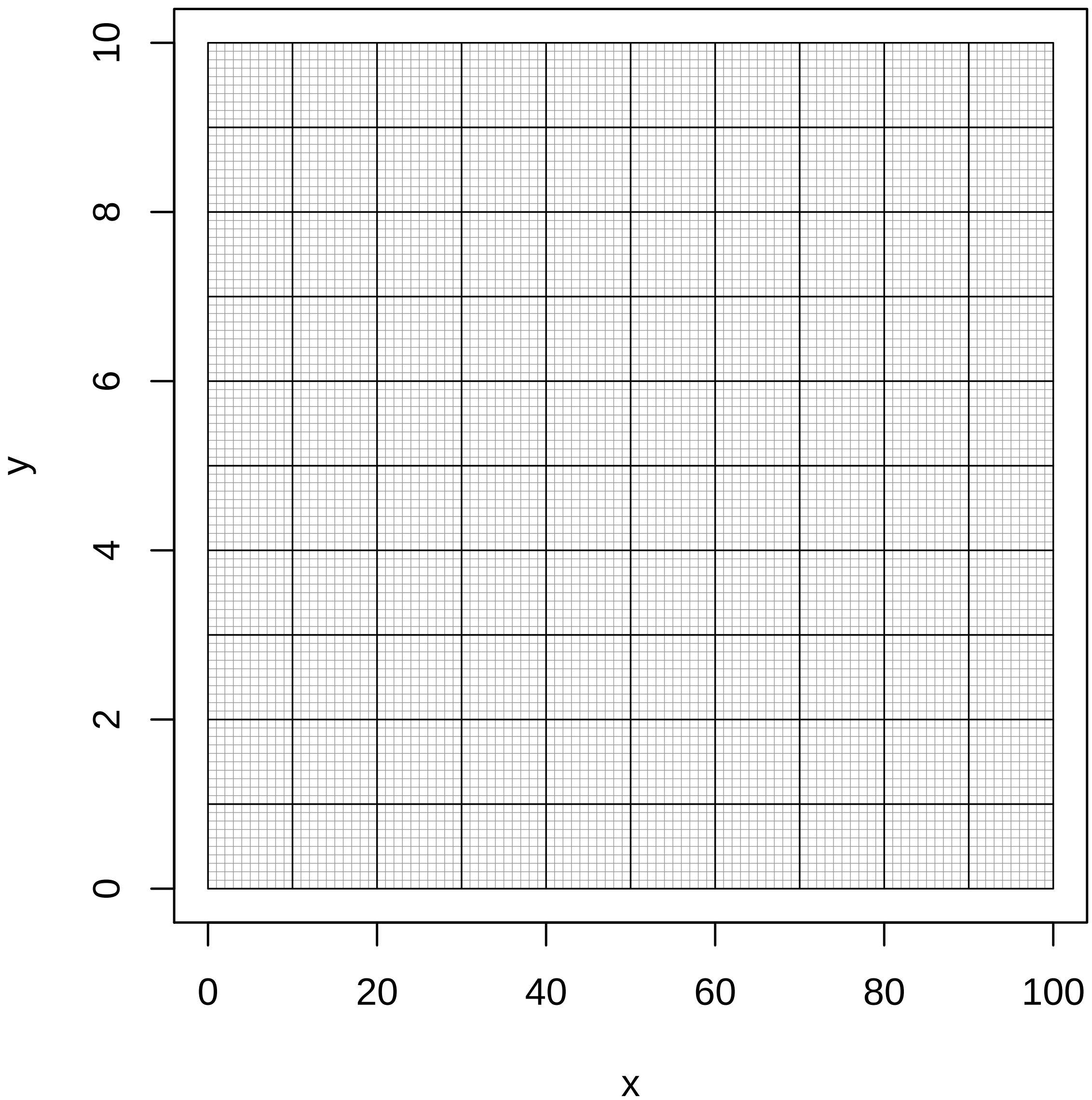
(c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

(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
23	4.1	94.3
94	7.5	705
32	5.7	182.4
25	4.7	117.5
59	5.4	318.6
10	3.3	33
$\sum x = 243$	$\sum y = 30.7$	$\sum x_i y_i = 1450.8$
$\bar{x} = 40.5$	$\bar{y} = 5.117$	
$s_x = 30.83$	$s_y = 1.457$	

$$r = \frac{1450.8 - (6)(40.5)(5.117)}{(6-1)(30.83)(1.457)} = 0.923$$

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.

$$r_{\text{exact}} = 0.923811$$

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:

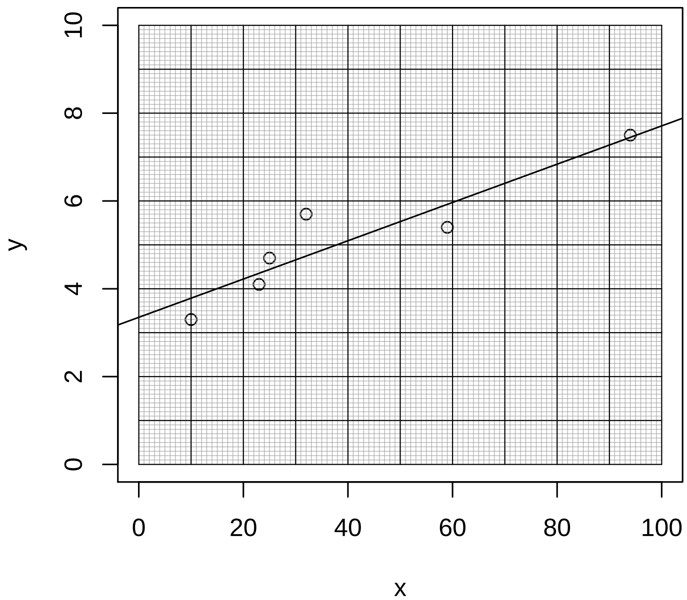
$$b = r\frac{s_y}{s_x} = 0.923 \cdot \frac{1.457}{30.83} = 0.0436$$

$$a = \bar{y} - b\bar{x} = 5.12 - (0.0436)(40.5) = 3.35$$

Our regression line:

$$y = 3.35 + (0.0436)x$$

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 $p = 0.51$. If 111 trials occur, what is the probability of getting more than 48 but at most 55 successes?

   In other words, let $X \sim \text{Bin}(n = 111,\ p = 0.51)$ and find $P(48 < X \leq 55)$.

   Use a normal approximation along with the continuity correction.

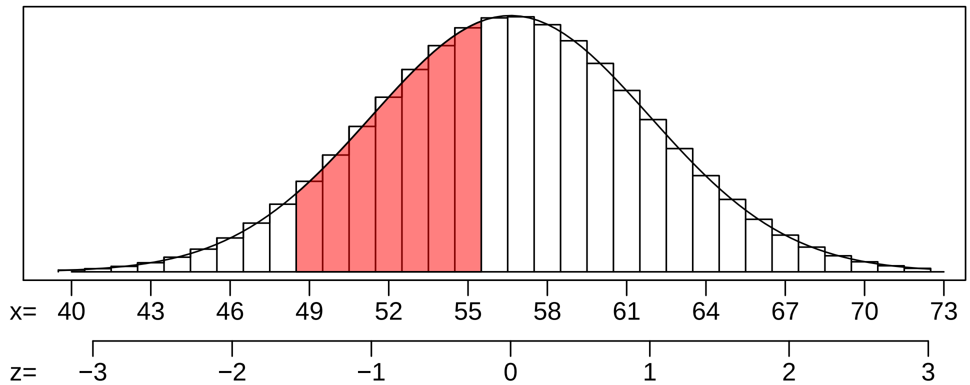
Solution: Find the mean.

$$\mu = np = (111)(0.51) = 56.61$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(111)(0.51)(1-0.51)} = 5.2668$$

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



Find the z scores.

$$z_1 = \frac{48.5 - 56.61}{5.2668} = -1.54$$

$$z_2 = \frac{55.5 - 56.61}{5.2668} = -0.21$$

Find the percentiles (from z-table).

$$\ell_1 = 0.0618$$

$$\ell_2 = 0.4168$$

Calculate the probability.

$$P(48 < X \leq 55) = 0.4168 - 0.0618 = 0.355$$

10. (Extra credit: 5 points)

A null hypothesis claims a population has a mean $\mu = 80$. You decide to run two-tail test on a sample of size $n = 10$ using a significance level $\alpha = 0.05$.

You then collect the sample:

89.3	71.8	89.8	92.1	93.7
75.8	72.4	100.1	120.5	96.4

(a) Determine the p-value.

(b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \text{ claims } \mu = 80$$

$$H_A \text{ claims } \mu \neq 80$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 90.19$$

$$s = 14.66$$

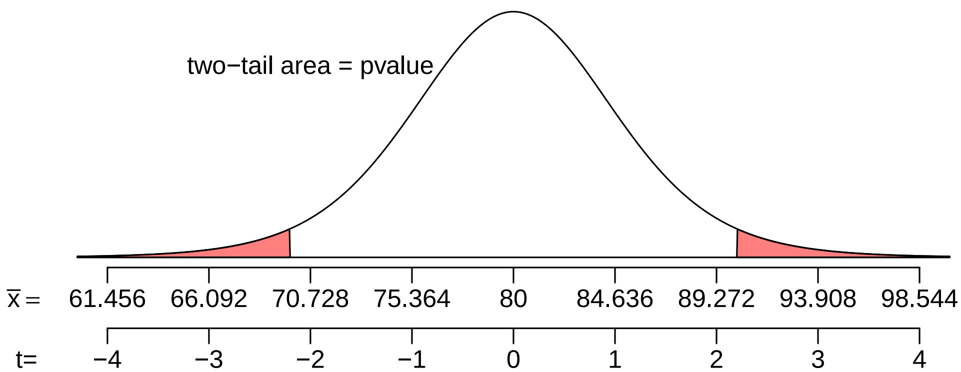
Determine the degrees of freedom.

$$df = 10 - 1 = 9$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{14.66}{\sqrt{10}} = 4.636$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{90.19 - 80}{4.636} = 2.2$$

Find the p-value.

$$p\text{-value} = P(|T| > 2.2)$$

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 $df = 9$.)

$$P(|T| > 2.26) = 0.05$$

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

Basically, because t is between 2.26 and 1.83, 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 ($\alpha = 0.05$).

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

No, we do not reject the null hypothesis.

(a) $0.05 < p\text{-value} < 0.1$

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