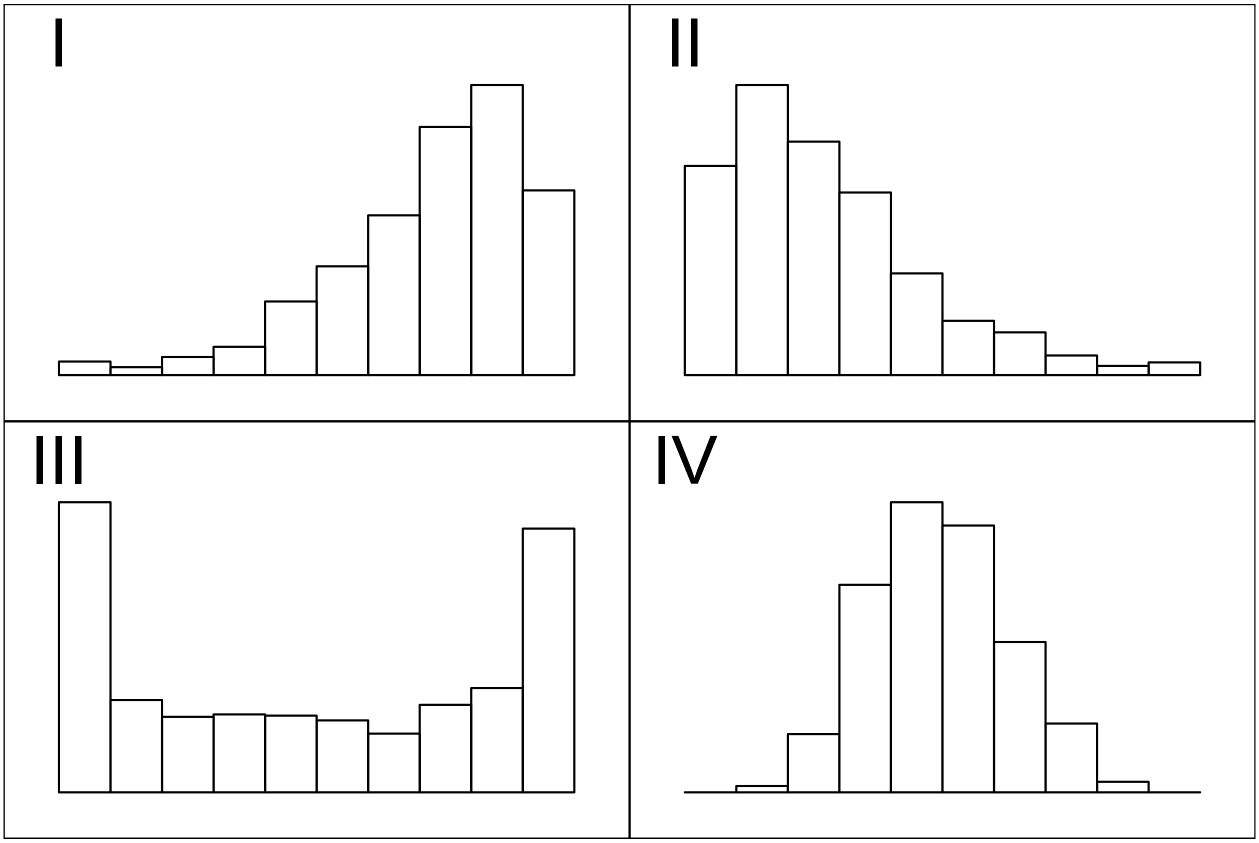
MAT-181 FINAL TAKE-HOME EXAM

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 quiz scores on an easy quiz. Most students did very well, but a few did poorly.

(b) The distribution of weights of newborn babies

(c) 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.

(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) I

(b) IV

(c) II

(d) III

2. (15 Points)

In a deck of strange cards, there are 420 cards. Each card has an image and a color. The amounts are shown in the table below.

	blue	gray	indigo	white	yellow	Total
horn	46	10	13	44	29	142
lamp	31	18	21	14	39	123
mop	11	47	15	45	37	155
Total	88	75	49	103	105	420

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

(b) What is the probability a random card is both a mop and yellow?

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

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

(e) What is the probability a random card is either a horn or gray (or both)?

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

(g) Is a lamp or a mop more likely to be white?

Solution:

(a) P(blue) = 0.21

(b) P(mop and yellow) = 0.0881

(c) P(mop given blue) = 0.125

(d) P(mop) = 0.369

(e) P(horn or gray) = 0.493

(f) P(blue given horn) = 0.324

(g) P(white given lamp) = 0.114 and P(white given mop) = 0.29, so a mop is more likely to be white than a lamp is.

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	71	7
B	73	4
C	60	14
D	110	6

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

Type of fruit	Mass of specimen (g)
A	77.51
B	70.92
C	64.76
D	106.3

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	$z = \frac{77.51-71}{7}$	0.93
B	$z = \frac{70.92-73}{4}$	-0.52
C	$z = \frac{64.76-60}{14}$	0.34
D	$z = \frac{106.3-110}{6}$	-0.61

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

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 30.5 millimeters and a standard deviation of 2.6 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 24.2 and 29.6 millimeters?

Solution:

$$\mu = 30.5$$
$$\sigma = 2.6$$
$$x_1 = 24.2$$
$$x_2 = 29.6$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{24.2 - 30.5}{2.6} = -2.42$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{29.6 - 30.5}{2.6} = -0.35$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.3632 - 0.0078 = 0.3554$$

5. (10 points)

   A species of duck is known to have a mean weight of 217.1 grams and a standard deviation of 25 grams. A researcher plans to measure the weights of 100 of these ducks sampled randomly. What is the probability the **sample mean** will be between 212.6 and 216.6 grams?

Solution:

$$n = 100$$
$$\mu = 217.1$$
$$\sigma = 25$$
$$SE = \frac{25}{\sqrt{100}} = 2.5$$
$$x_1 = 212.6$$
$$x_2 = 216.6$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{212.6 - 217.1}{2.5} = -1.8$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{216.6 - 217.1}{2.5} = -0.2$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.4207 - 0.0359 = 0.3848$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Catharus fuscescens*. She randomly samples 16 adults of *Catharus fuscescens*, resulting in a sample mean of 42.97 grams and a sample standard deviation of 3.95 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.

$$\begin{aligned} n &= 16 \\ \bar{x} &= 42.97 \\ s &= 3.95 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 16 - 1 \\ &= 15 \end{aligned}$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$ and $df = 15$.

$$t^\star = 2.13$$

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

$$\begin{aligned} LB &= \bar{x} - t^\star \frac{s}{\sqrt{n}} & UB &= \bar{x} + t^\star \frac{s}{\sqrt{n}} \\ &= 42.97 - 2.13 \times \frac{3.95}{\sqrt{16}} & &= 42.97 + 2.13 \times \frac{3.95}{\sqrt{16}} \\ &= 40.9 & &= 45.1 \end{aligned}$$

We are 95% confident that the population mean is between 40.9 and 45.1 grams.

$$CI = (40.9,\ 45.1)$$

7. (15 points)

A student is taking a multiple choice test with 500 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 115 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.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.2(1-0.2)}{500}} = 0.0179$$

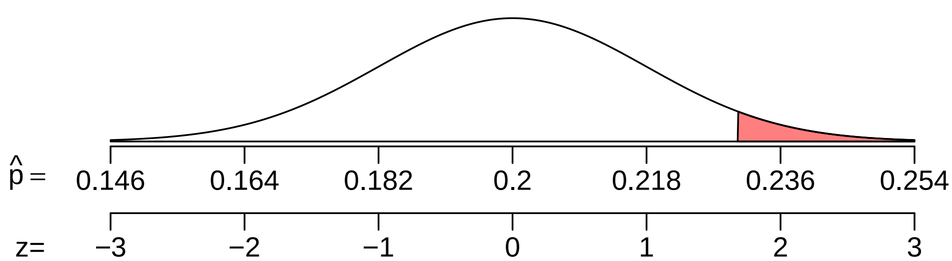
Determine the sample proportion.

$$\hat{p} = \frac{115}{500} = 0.23$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.23 - 0.2}{0.0179} = 1.68$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.23) \\ &= P(Z > 1.68) \\ &= 1 - P(Z < 1.68) \\ &= 0.0465 \end{aligned}$$

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 $p = 0.2$ and H_A claims $p > 0.2$.
(c) The p-value is 0.0465
(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
860	74	
120	28	
820	68	
340	22	
470	48	
960	67	
260	30	
120	25	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

(a) Complete the table.

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

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

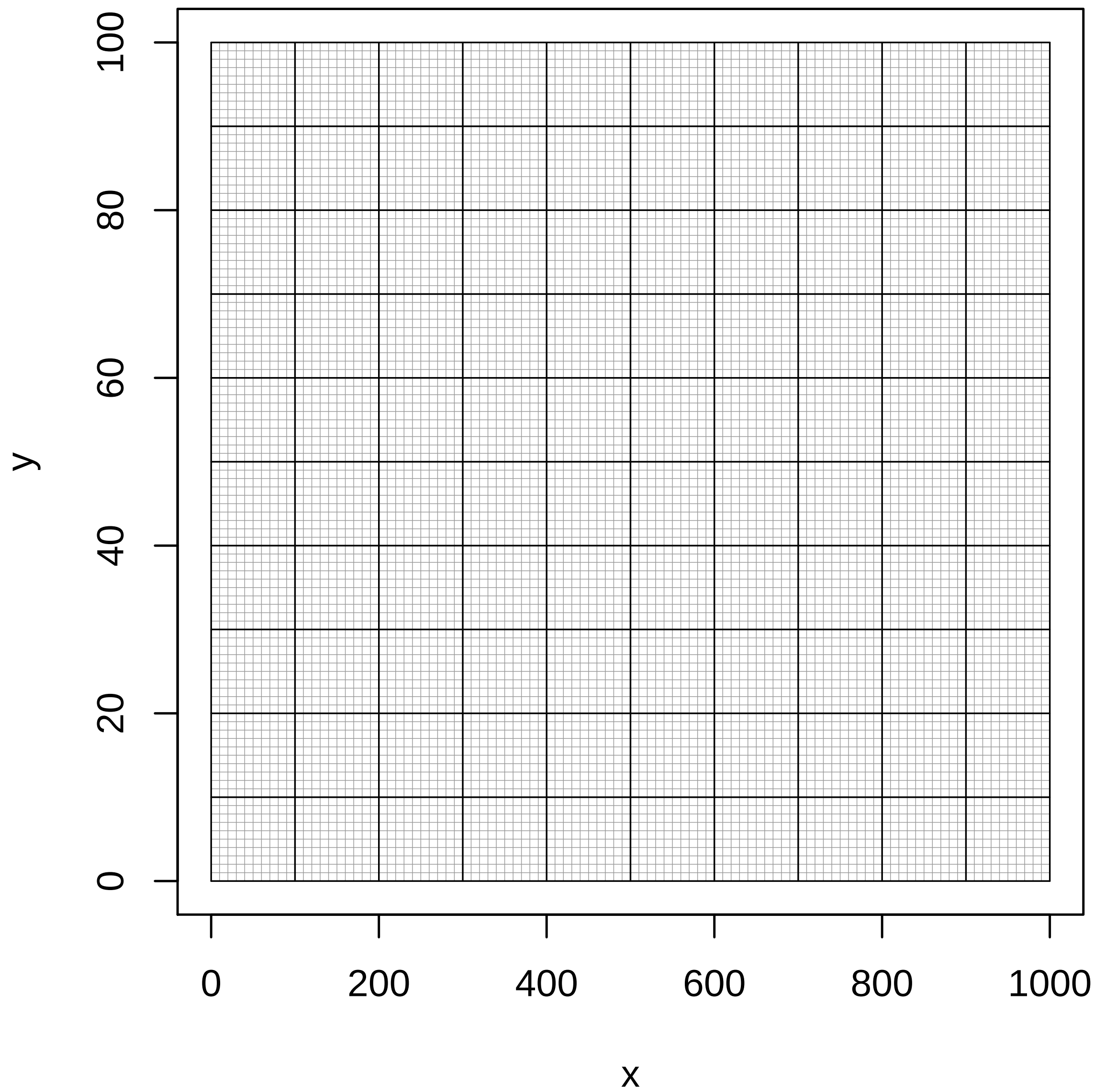
(c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

(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
860	74	63640
120	28	3360
820	68	55760
340	22	7480
470	48	22560
960	67	64320
260	30	7800
120	25	3000
$\sum x = 3950$	$\sum y = 362$	$\sum x_i y_i = 227920$
$\bar{x} = 493.8$	$\bar{y} = 45.25$	
$s_x = 341.5$	$s_y = 21.73$	

$$r = \frac{227920 - (8)(493.8)(45.25)}{(8-1)(341.5)(21.73)} = 0.946$$

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.

$$r_{\text{exact}} = 0.9468846$$

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:

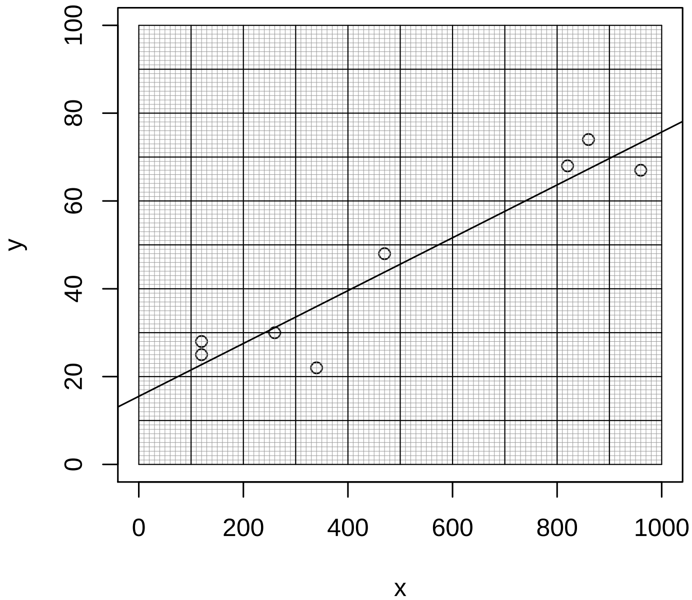
$$b = r\frac{s_y}{s_x} = 0.946 \cdot \frac{21.73}{341.5} = 0.0602$$

$$a = \bar{y} - b\bar{x} = 45.2 - (0.0602)(494) = 15.5$$

Our regression line:

$$y = 15.5 + (0.0602)x$$

Make a plot.


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

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.81$. If 238 trials occur, what is the probability of getting at least 188 but at most 199 successes?

   In other words, let $X \sim \text{Bin}(n = 238,\ p = 0.81)$ and find $P(188 \leq X \leq 199)$.

   Use a normal approximation along with the continuity correction.

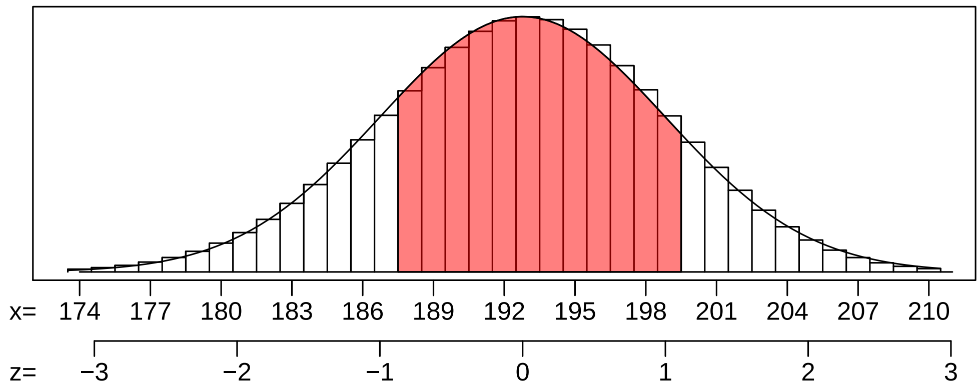
Solution: Find the mean.

$$\mu = np = (238)(0.81) = 192.78$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(238)(0.81)(1-0.81)} = 6.0521$$

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



Find the z scores.

$$z_1 = \frac{187.5 - 192.78}{6.0521} = -0.87$$

$$z_2 = \frac{199.5 - 192.78}{6.0521} = 1.11$$

Find the percentiles (from z-table).

$$\ell_1 = 0.1922$$

$$\ell_2 = 0.8665$$

Calculate the probability.

$$P(188 \leq X \leq 199) = 0.8665 - 0.1922 = 0.6743$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 90$. You decide to run two-tail test on a sample of size $n = 12$ using a significance level $\alpha = 0.05$.

    You then collect the sample:

99.1	98.1	90.5	99.8	88.7
85	108.8	115.6	97.8	99.9
79.6	92.1			

    (a) Determine the p-value.

    (b) Do you reject the null hypothesis?

Solution: State the hypotheses.

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

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

Find the mean and standard deviation of the sample.

$$\bar{x} = 96.25$$

$$s = 9.922$$

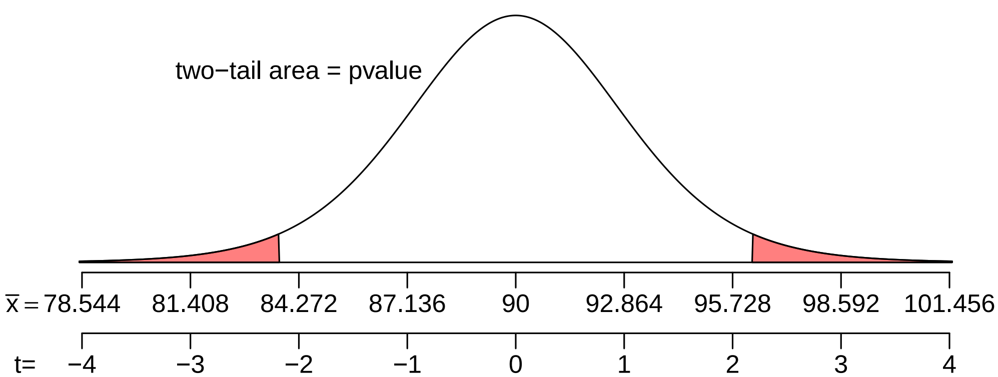
Determine the degrees of freedom.

$$df = 12 - 1 = 11$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{9.922}{\sqrt{12}} = 2.864$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{96.25 - 90}{2.864} = 2.18$$

Find the p-value.

$$p\text{-value} = P(|T| > 2.18)$$

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 $df = 11$.)

$$P(|T| > 2.2) = 0.05$$

$$P(|T| > 1.8) = 0.1$$

Basically, because t is between 2.2 and 1.8, 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.