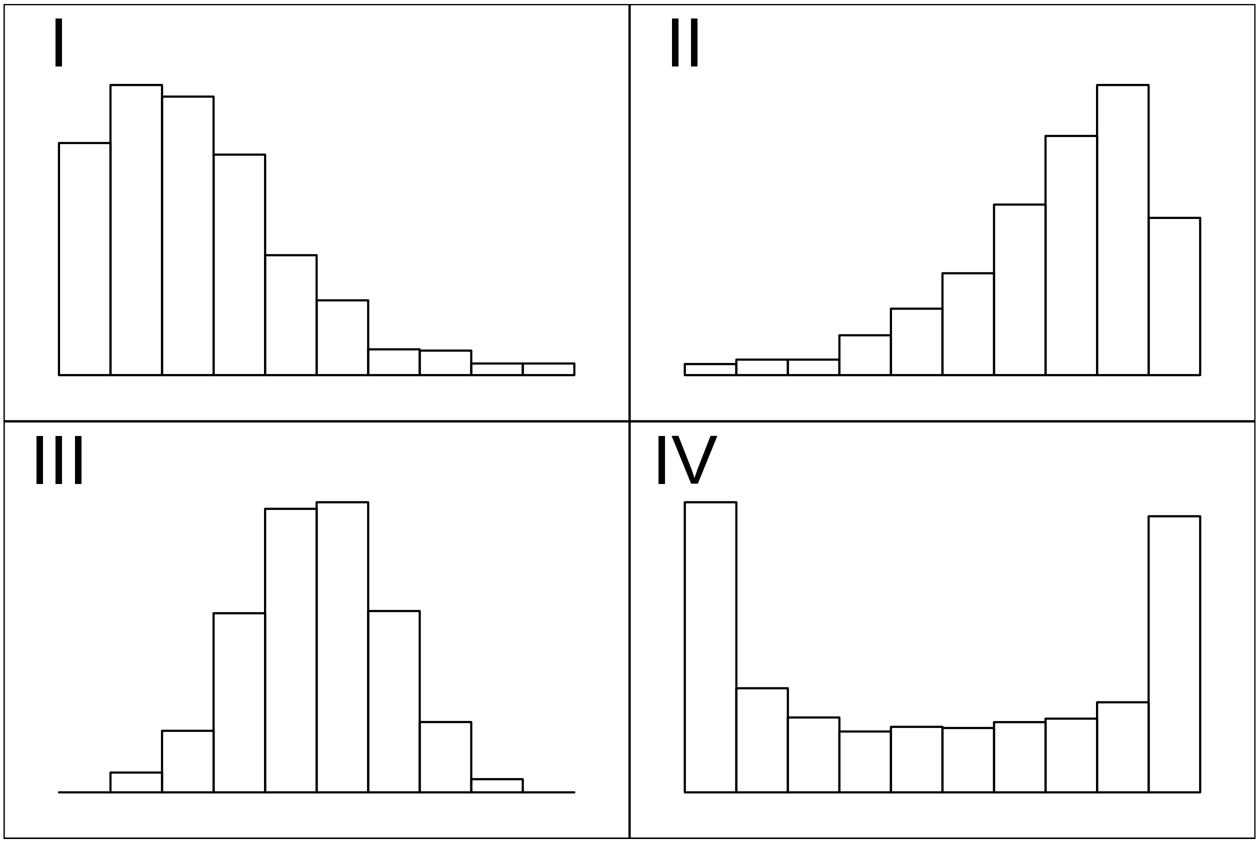
MAT-181 FINAL TAKE-HOME EXAM

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 NBA basketball players where only a few are high-paid superstars.

(b) 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.

(c) The distribution of heights of adult men

(d) The distribution of ages at a skilled nursing facility, where most of the patients are elderly but a few are quite young.

Solution:

(a) I

(b) IV

(c) III

(d) II

2. (15 Points)

In a deck of strange cards, there are 776 cards. Each card has an image and a color. The amounts are shown in the table below.

	green	orange	violet	white	yellow	Total
bike	41	16	35	48	49	189
dog	44	19	50	34	29	176
flower	37	38	47	15	20	157
gem	36	43	24	14	42	159
wheel	12	21	25	27	10	95
Total	170	137	181	138	150	776

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

(b) What is the probability a random card is a flower given it is yellow?

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

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

(e) What is the probability a random card is either a gem or violet (or both)?

(f) What is the probability a random card is both a wheel and orange?

(g) Is a dog or a gem more likely to be white?

Solution:

(a) P(violet) = 0.233

(b) P(flower given yellow) = 0.133

(c) P(green given bike) = 0.217

(d) P(bike) = 0.244

(e) P(gem or violet) = 0.407

(f) P(wheel and orange) = 0.0271

(g) P(white given dog) = 0.193 and P(white given gem) = 0.0881, so a dog is more likely to be white than a gem 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	91	9
B	99	6
C	62	13
D	97	15

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

Type of fruit	Mass of specimen (g)
A	90.82
B	103.1
C	72.79
D	97.45

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	$z = \frac{\lvert 90.82-91 \rvert}{9}$	0.02
B	$z = \frac{\lvert 103.1-99 \rvert}{6}$	0.69
C	$z = \frac{\lvert 72.79-62 \rvert}{13}$	0.83
D	$z = \frac{\lvert 97.45-97 \rvert}{15}$	0.03

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

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 138 millimeters and a standard deviation of 9.2 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 116.2 and 121.9 millimeters?

Solution:

$$\mu = 138$$
$$\sigma = 9.2$$
$$x_1 = 116.2$$
$$x_2 = 121.9$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{116.2 - 138}{9.2} = -2.37$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{121.9 - 138}{9.2} = -1.75$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.0401 - 0.0089 = 0.0312$$

5. (10 points)

   A species of duck is known to have a mean weight of 178.1 grams and a standard deviation of 33 grams. A researcher plans to measure the weights of 121 of these ducks sampled randomly. What is the probability the **sample mean** will be between 170.6 and 179.6 grams?

Solution:

$$n = 121$$
$$\mu = 178.1$$
$$\sigma = 33$$
$$SE = \frac{33}{\sqrt{121}} = 3$$
$$x_1 = 170.6$$
$$x_2 = 179.6$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{170.6 - 178.1}{3} = -2.5$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{179.6 - 178.1}{3} = 0.5$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.6915 - 0.0062 = 0.6853$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Ammodramus maritimus*. She randomly samples 30 adults of *Ammodramus maritimus*, resulting in a sample mean of 21.68 grams and a sample standard deviation of 1.57 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.

$$\begin{aligned} n &= 30 \\ \bar{x} &= 21.68 \\ s &= 1.57 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 30 - 1 \\ &= 29 \end{aligned}$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$ and $df = 29$.

$$t^\star = 2.05$$

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

$$\begin{aligned} LB &= \bar{x} - t^\star \frac{s}{\sqrt{n}} & UB &= \bar{x} + t^\star \frac{s}{\sqrt{n}} \\ &= 21.68 - 2.05 \times \frac{1.57}{\sqrt{30}} & &= 21.68 + 2.05 \times \frac{1.57}{\sqrt{30}} \\ &= 21.1 & &= 22.3 \end{aligned}$$

We are 95% confident that the population mean is between 21.1 and 22.3 grams.

$$CI = (21.1,\ 22.3)$$

7. (15 points)

A student is taking a multiple choice test with 1000 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 529 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.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{1000}} = 0.0158$$

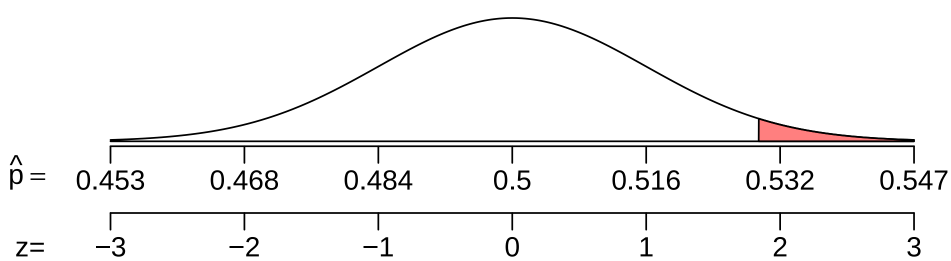
Determine the sample proportion.

$$\hat{p} = \frac{529}{1000} = 0.529$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.529 - 0.5}{0.0158} = 1.84$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.529) \\ &= P(Z > 1.84) \\ &= 1 - P(Z < 1.84) \\ &= 0.0329 \end{aligned}$$

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 $p = 0.5$ and H_A claims $p > 0.5$.

(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
4.4	2.9	
5.3	2.5	
7.2	1.9	
5.7	2.1	
6.8	4.7	
2.8	5.6	
2.8	3	
3.3	6.8	
4.9	2.7	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

(a) Complete the table.

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

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

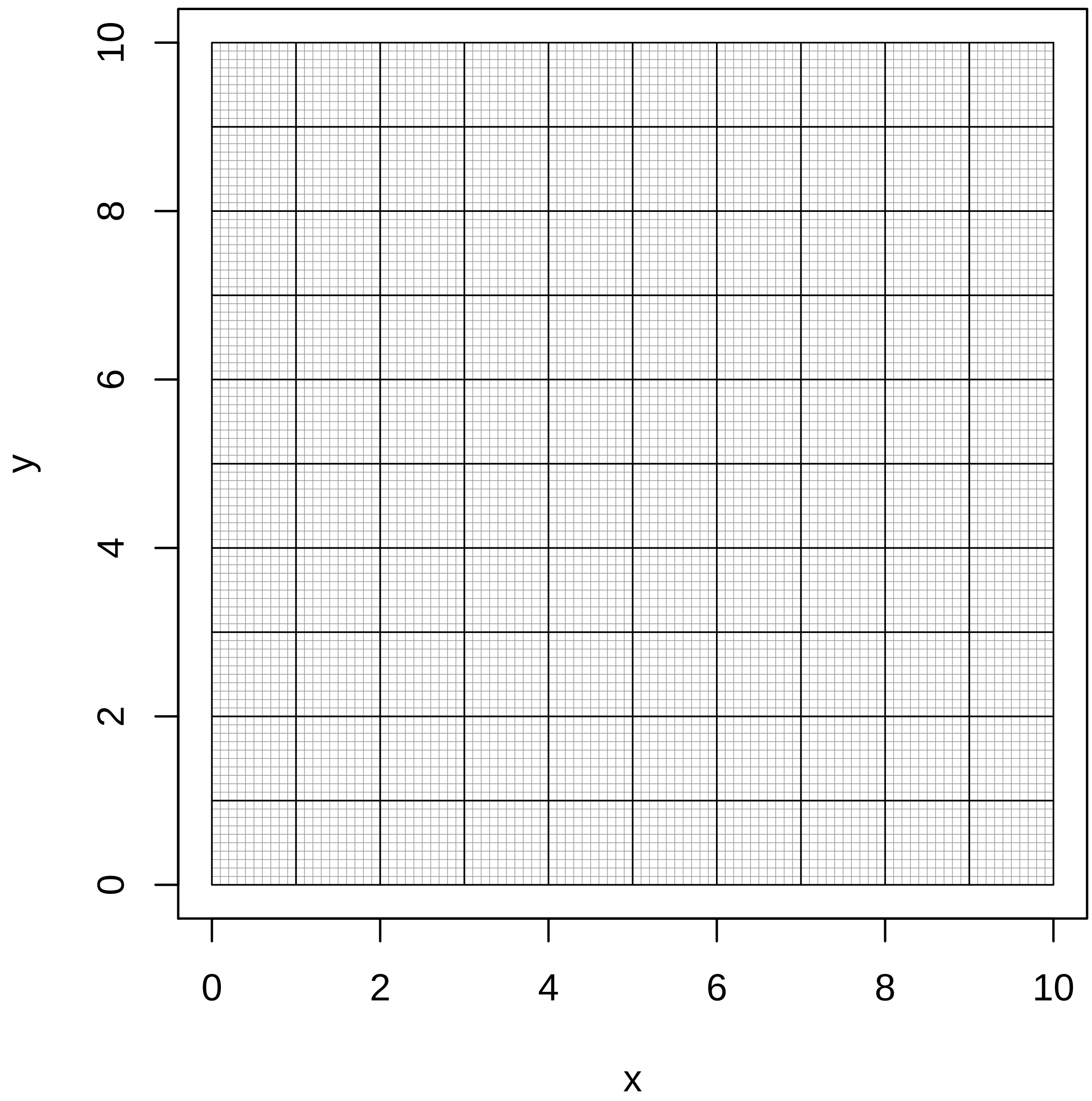
(c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

(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
4.4	2.9	12.76
5.3	2.5	13.25
7.2	1.9	13.68
5.7	2.1	11.97
6.8	4.7	31.96
2.8	5.6	15.68
2.8	3	8.4
3.3	6.8	22.44
4.9	2.7	13.23
$\sum x = 43.2$	$\sum y = 32.2$	$\sum x_i y_i = 143.37$
$\bar{x} = 4.8$	$\bar{y} = 3.578$	
$s_x = 1.629$	$s_y = 1.712$	

$$r = \frac{143.37 - (9)(4.8)(3.578)}{(9-1)(1.629)(1.712)} = -0.502$$

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.

$$r_{\text{exact}} = -0.5013372$$

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:

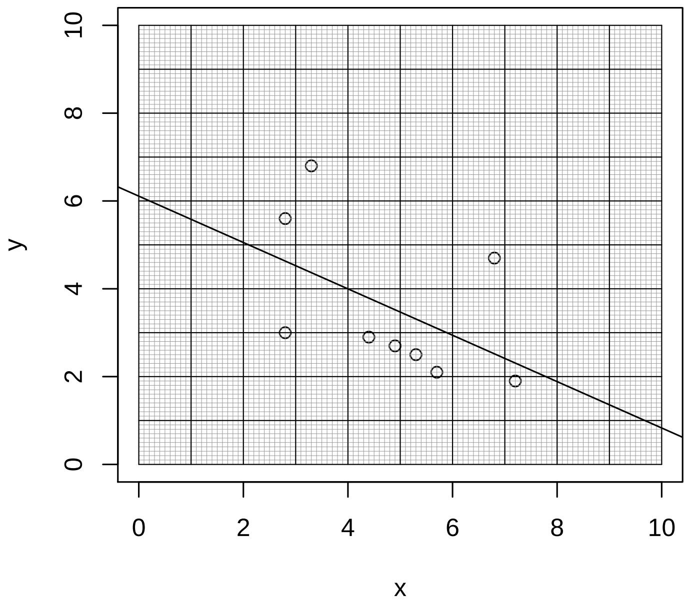
$$b = r\frac{s_y}{s_x} = -0.502 \cdot \frac{1.712}{1.629} = -0.528$$

$$a = \bar{y} - b\bar{x} = 3.58 - (-0.528)(4.8) = 6.11$$

Our regression line:

$$y = 6.11 + (-0.528)x$$

Make a plot.



10
8
6
4
2
0
y
0
2
4
6
8
10
x

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.44$. If 158 trials occur, what is the probability of getting more than 58 but at most 70 successes?

   In other words, let $X \sim \text{Bin}(n = 158,\ p = 0.44)$ and find $P(58 < X \leq 70)$.

   Use a normal approximation along with the continuity correction.

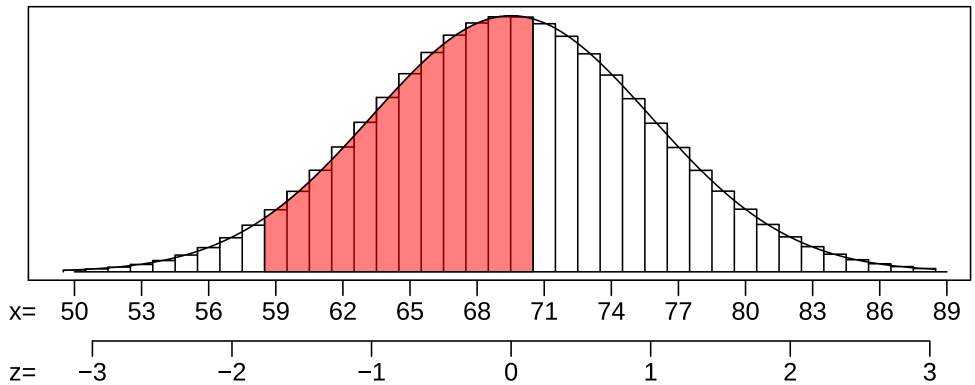
Solution: Find the mean.

$$\mu = np = (158)(0.44) = 69.52$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(158)(0.44)(1-0.44)} = 6.2395$$

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



Find the z scores.

$$z_1 = \frac{58.5 - 69.52}{6.2395} = -1.77$$

$$z_2 = \frac{70.5 - 69.52}{6.2395} = 0.16$$

Find the percentiles (from z-table).

$$\ell_1 = 0.0384$$

$$\ell_2 = 0.5636$$

Calculate the probability.

$$P(58 < X \leq 70) = 0.5636 - 0.0384 = 0.526$$

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 $n = 11$ using a significance level $\alpha = 0.05$.

You then collect the sample:

191.9	145	133.6	126.4	106.6
147.6	176	109	188.3	200.1
145.2				

(a) Determine the p-value.

(b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \text{ claims } \mu = 130$$

$$H_A \text{ claims } \mu \neq 130$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 151.791$$

$$s = 32.89$$

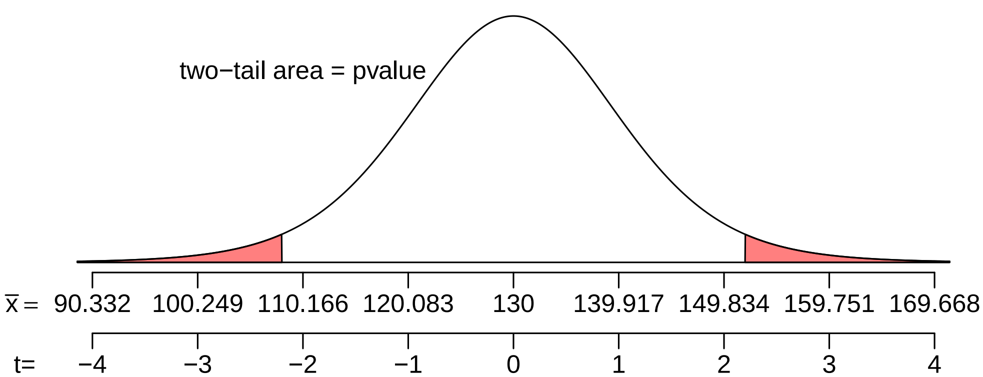
Determine the degrees of freedom.

$$df = 11 - 1 = 10$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{32.89}{\sqrt{11}} = 9.917$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{151.791 - 130}{9.917} = 2.2$$

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 $df = 10$.)

$$P(|T| > 2.23) = 0.05$$

$$P(|T| > 1.81) = 0.1$$

Basically, because t is between 2.23 and 1.81, 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.