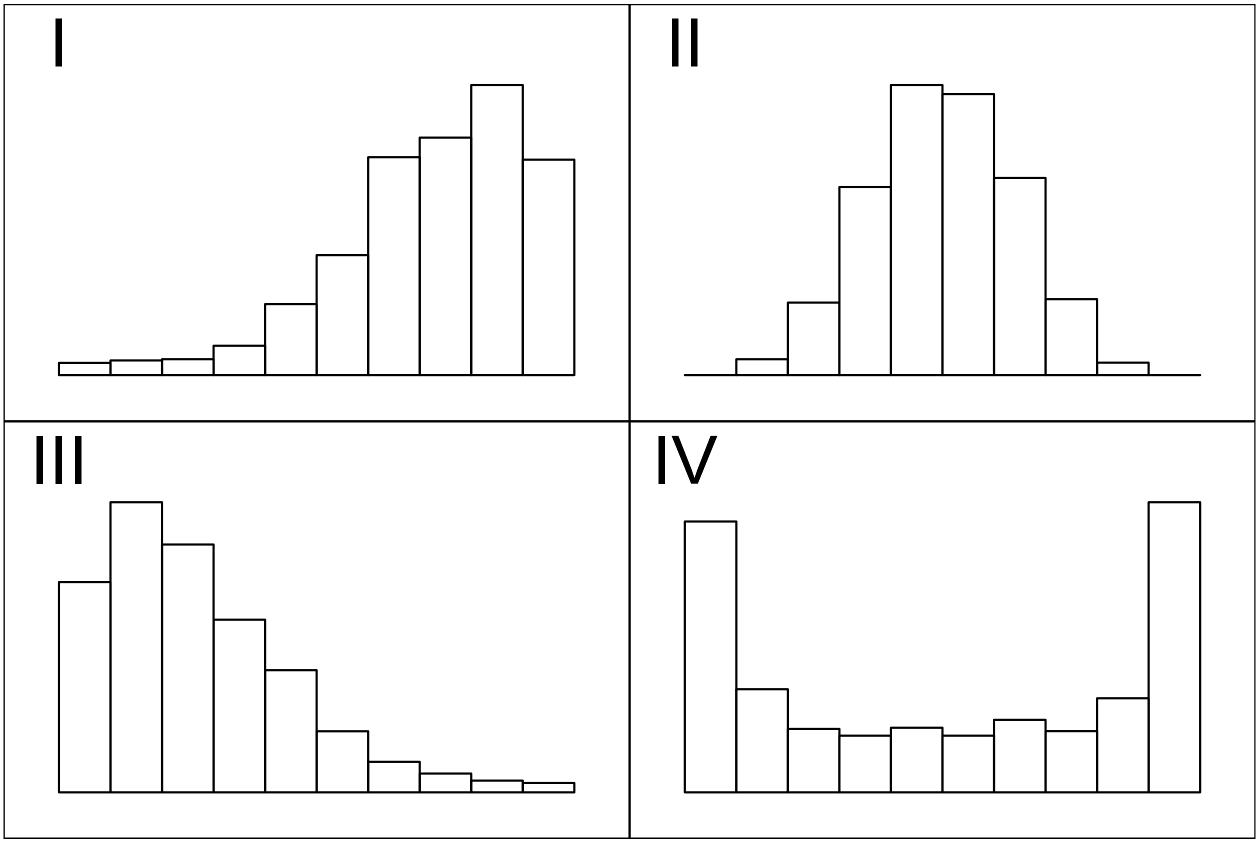
MAT-181 FINAL TAKE-HOME EXAM

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 heights of adult men

(c) The distribution of test scores on a very difficult exam, in which most students have poor to average scores, but a few did quite well.

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

2. (15 Points)

In a deck of strange cards, there are 515 cards. Each card has an image and a color. The amounts are shown in the table below.

	pink	violet	white	Total
kite	28	37	15	80
lamp	46	10	43	99
needle	45	41	34	120
pig	32	47	44	123
tree	18	26	49	93
Total	169	161	185	515

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

(b) What is the probability a random card is both a needle and violet?

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

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

(e) Is a kite or a lamp more likely to be white?

(f) What is the probability a random card is a needle given it is white?

(g) What is the probability a random card is either a lamp or violet (or both)?

Solution:

(a) $P(\text{pink}) = 0.328$

(b) $P(\text{needle and violet}) = 0.0796$

(c) $P(\text{violet given lamp}) = 0.101$

(d) $P(\text{tree}) = 0.181$

(e) $P(\text{white given kite}) = 0.188$ and $P(\text{white given lamp}) = 0.434$, so a lamp is more likely to be white than a kite is.

(f) $P(\text{needle given white}) = 0.184$

(g) $P(\text{lamp or violet}) = 0.485$

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	72	5
B	149	13
C	77	8
D	133	9

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

Type of fruit	Mass of specimen (g)
A	76.75
B	166.7
C	76.76
D	147.4

   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	$z = \frac{\lvert 76.75-72\rvert}{5}$	0.95
B	$z = \frac{\lvert 166.7-149\rvert}{13}$	1.36
C	$z = \frac{\lvert 76.76-77\rvert}{8}$	0.03
D	$z = \frac{\lvert 147.4-133\rvert}{9}$	1.6

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 47.6 millimeters and a standard deviation of 8.8 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 43.3 and 55.9 millimeters?

Solution:

$$\mu = 47.6$$
$$\sigma = 8.8$$
$$x_1 = 43.3$$
$$x_2 = 55.9$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{43.3 - 47.6}{8.8} = -0.49$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{55.9 - 47.6}{8.8} = 0.94$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.8264 - 0.3121 = 0.5143$$

5. (10 points)

   A species of duck is known to have a mean weight of 215.3 grams and a standard deviation of 70 grams. A researcher plans to measure the weights of 196 of these ducks sampled randomly. What is the probability the **sample mean** will be between 204.8 and 208.3 grams?

Solution:

$$n = 196$$
$$\mu = 215.3$$
$$\sigma = 70$$
$$SE = \frac{70}{\sqrt{196}} = 5$$
$$x_1 = 204.8$$
$$x_2 = 208.3$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{204.8 - 215.3}{5} = -2.1$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{208.3 - 215.3}{5} = -1.4$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.0808 - 0.0179 = 0.0629$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Dendroica fusca*. She randomly samples 19 adults of *Dendroica fusca*, resulting in a sample mean of 13.1 grams and a sample standard deviation of 1.51 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.

$$\begin{aligned} n &= 19 \\ \bar{x} &= 13.1 \\ s &= 1.51 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 19 - 1 \\ &= 18 \end{aligned}$$

Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 18$.

$$t^{\star} = 2.1$$

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

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} & UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 13.1 - 2.1 \times \frac{1.51}{\sqrt{19}} & &= 13.1 + 2.1 \times \frac{1.51}{\sqrt{19}} \\ &= 12.4 & &= 13.8 \end{aligned}$$

We are 95% confident that the population mean is between 12.4 and 13.8 grams.

$$CI = (12.4,\ 13.8)$$

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 166 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.

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

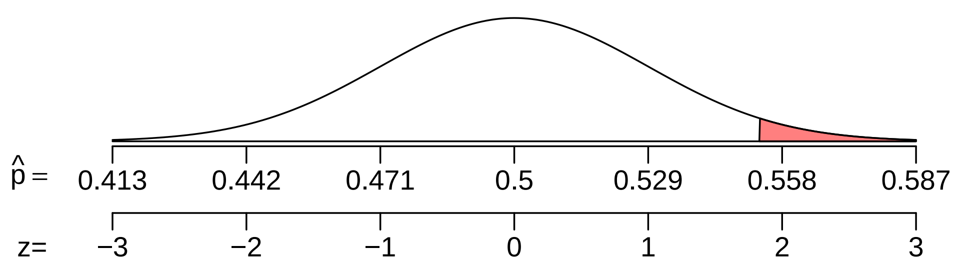
Determine the sample proportion.

$$\hat{p} = \frac{166}{300} = 0.553$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.553 - 0.5}{0.0289} = 1.83$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.553) \\ &= P(Z > 1.83) \\ &= 1 - P(Z < 1.83) \\ &= 0.0336 \end{aligned}$$

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 0.0336

(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
44	4.3	
85	4.8	
83	5.4	
28	3.1	
46	3.8	
62	5.1	
93	4.8	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

   (a) Complete the table.

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

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

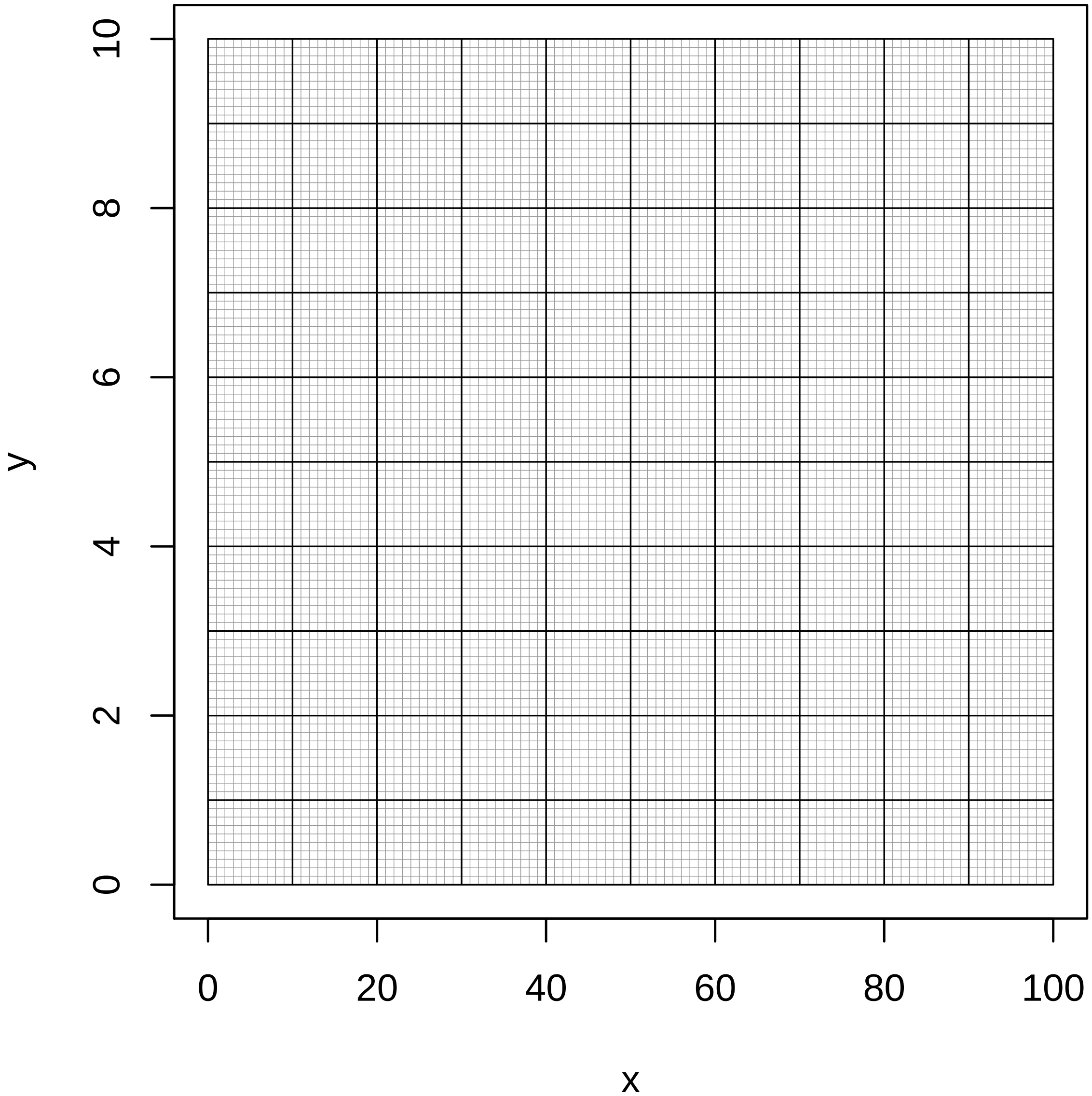
   (c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

   (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
44	4.3	189.2
85	4.8	408
83	5.4	448.2
28	3.1	86.8
46	3.8	174.8
62	5.1	316.2
93	4.8	446.4
$\sum x = 441$	$\sum y = 31.3$	$\sum x_i y_i = 2069.6$
$\bar{x} = 63$	$\bar{y} = 4.471$	
$s_x = 24.7$	$s_y = 0.7994$	

$$r = \frac{2069.6 - (7)(63)(4.471)}{(7-1)(24.7)(0.7994)} = 0.826$$

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.

$$r_{\text{exact}} = 0.82473$$

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:

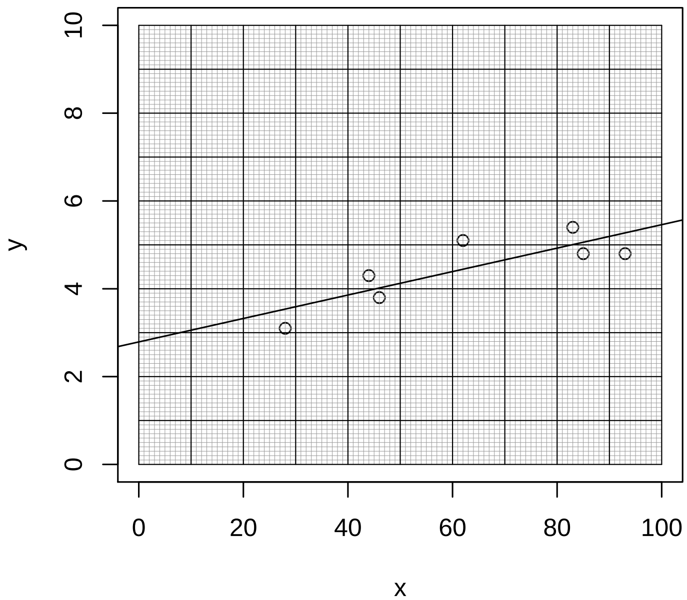
$$b = r\frac{s_y}{s_x} = 0.826 \cdot \frac{0.7994}{24.7} = 0.0267$$

$$a = \bar{y} - b\bar{x} = 4.47 - (0.0267)(63) = 2.79$$

Our regression line:

$$y = 2.79 + (0.0267)x$$

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 $p = 0.76$. If 180 trials occur, what is the probability of getting more than 131 but at most 141 successes?

   In other words, let $X \sim \text{Bin}(n = 180,\ p = 0.76)$ and find $P(131 < X \leq 141)$.

   Use a normal approximation along with the continuity correction.

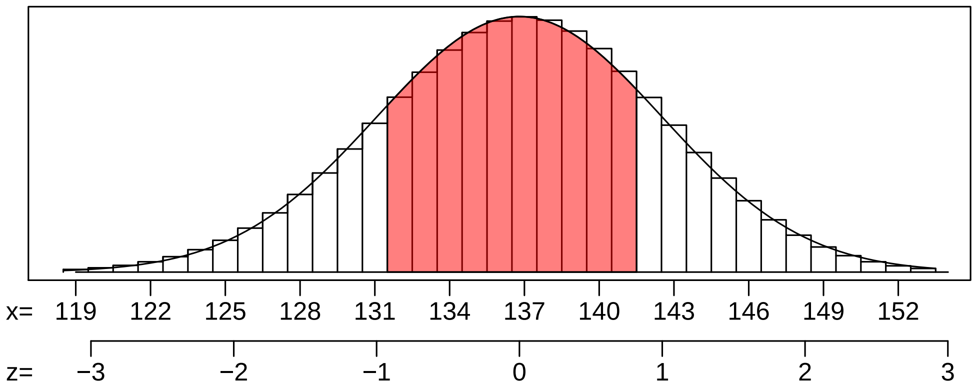
Solution: Find the mean.

$$\mu = np = (180)(0.76) = 136.8$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(180)(0.76)(1-0.76)} = 5.7299$$

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



Find the z scores.

$$z_1 = \frac{131.5 - 136.8}{5.7299} = -0.92$$

$$z_2 = \frac{141.5 - 136.8}{5.7299} = 0.82$$

Find the percentiles (from z-table).

$$\ell_1 = 0.1788$$

$$\ell_2 = 0.7939$$

Calculate the probability.

$$P(131 < X \leq 141) = 0.7939 - 0.1788 = 0.615$$

10. (Extra credit: 5 points)

A null hypothesis claims a population has a mean $\mu = 120$. 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:

187.6	135.5	107.8	116.2	125.1
146.3	164	152.3	109	131.9

(a) Determine the p-value.

(b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 120$$

$$H_A \quad \text{claims} \quad \mu \neq 120$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 137.57$$

$$s = 25.486$$

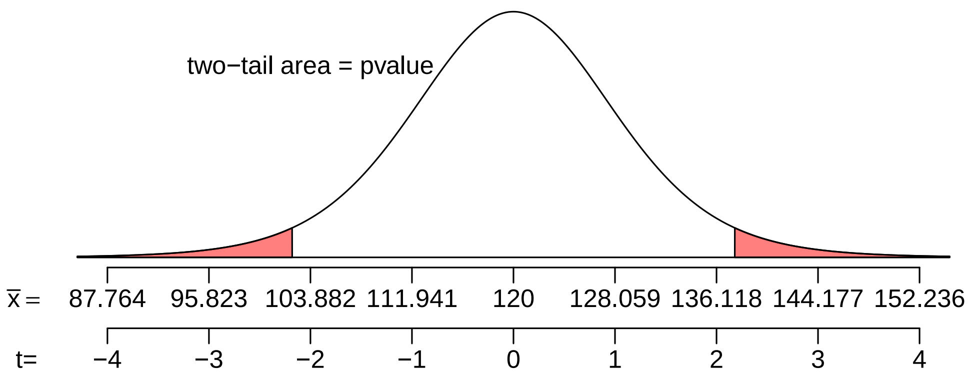
Determine the degrees of freedom.

$$df = 10 - 1 = 9$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{25.486}{\sqrt{10}} = 8.059$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{137.57 - 120}{8.059} = 2.18$$

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 $df = 9$.)

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

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

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.