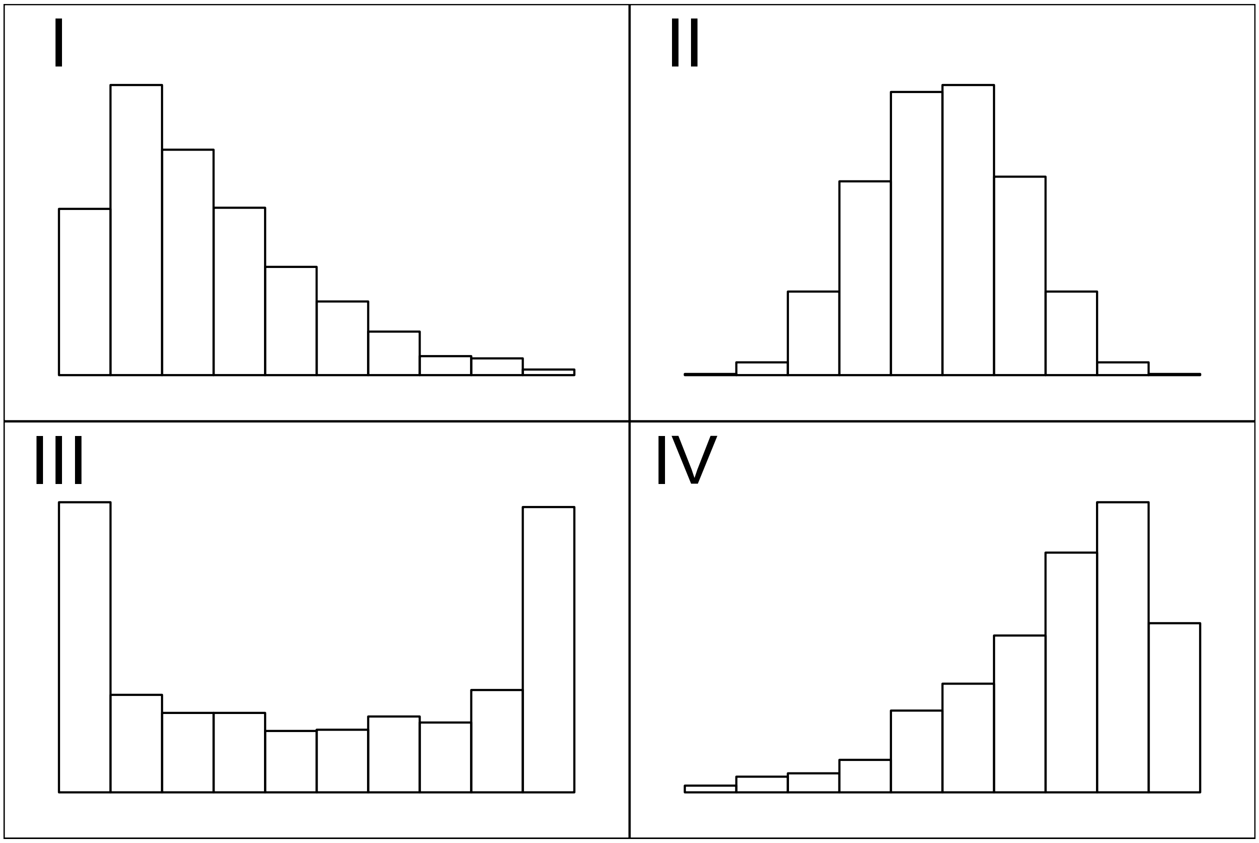
MAT-181 FINAL TAKE-HOME EXAM

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 hours that students studied for an exam when about half of students studied a lot and a similar number of students studied very little.

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

(d) The distribution of heights of adult men

Solution:

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

2. (15 Points)

In a deck of strange cards, there are 440 cards. Each card has an image and a color. The amounts are shown in the table below.

	indigo	red	violet	Total
dog	49	11	22	82
horn	21	38	42	101
needle	48	14	13	75
quilt	19	23	36	78
shovel	30	29	45	104
Total	167	115	158	440

(a) What is the probability a random card is red given it is a shovel?

(b) What is the probability a random card is both a horn and red?

(c) What is the probability a random card is a quilt?

(d) What is the probability a random card is either a quilt or violet (or both)?

(e) Is a horn or a needle more likely to be red?

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

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

Solution:

(a) P(red given shovel) = 0.279

(b) P(horn and red) = 0.0864

(c) P(quilt) = 0.177

(d) P(quilt or violet) = 0.455

(e) P(red given horn) = 0.376 and P(red given needle) = 0.187, so a horn is more likely to be red than a needle is.

(f) P(needle given indigo) = 0.287

(g) P(red) = 0.261

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	98	7
B	146	6
C	149	10
D	102	8

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

Type of fruit	Mass of specimen (g)
A	91.14
B	150.7
C	149.2
D	101.2

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	$z = \frac{\lvert 91.14-98 \rvert}{7}$	0.98
B	$z = \frac{\lvert 150.7-146 \rvert}{6}$	0.79
C	$z = \frac{\lvert 149.2-149 \rvert}{10}$	0.02
D	$z = \frac{\lvert 101.2-102 \rvert}{8}$	0.1

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

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 33.4 millimeters and a standard deviation of 7.1 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 23.7 and 36 millimeters?

Solution:

$$\mu = 33.4$$
$$\sigma = 7.1$$
$$x_1 = 23.7$$
$$x_2 = 36$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{23.7 - 33.4}{7.1} = -1.37$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{36 - 33.4}{7.1} = 0.37$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.6443 - 0.0853 = 0.559$$

5. (10 points)

   A species of duck is known to have a mean weight of 268.2 grams and a standard deviation of 17.5 grams. A researcher plans to measure the weights of 49 of these ducks sampled randomly. What is the probability the **sample mean** will be between 268.2 and 273.7 grams?

Solution:

$$n = 49$$
$$\mu = 268.2$$
$$\sigma = 17.5$$
$$SE = \frac{17.5}{\sqrt{49}} = 2.5$$
$$x_1 = 268.2$$
$$x_2 = 273.7$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{268.2 - 268.2}{2.5} = 0$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{273.7 - 268.2}{2.5} = 2.2$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.9861 - 0.5 = 0.4861$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Catharus minimus*. She randomly samples 15 adults of *Catharus minimus*, resulting in a sample mean of 31.23 grams and a sample standard deviation of 4.94 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.

$$\begin{aligned} n &= 15 \\ \bar{x} &= 31.23 \\ s &= 4.94 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 15 - 1 \\ &= 14 \end{aligned}$$

Determine the critical t value, $t^{\star}$, such that $P(|T| < t^{\star}) = 0.95$ and $df = 14$.

$$t^{\star} = 2.14$$

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

$$\begin{aligned} LB &= \bar{x} - t^{\star}\frac{s}{\sqrt{n}} \\ &= 31.23 - 2.14 \times \frac{4.94}{\sqrt{15}} \\ &= 28.5 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^{\star}\frac{s}{\sqrt{n}} \\ &= 31.23 + 2.14 \times \frac{4.94}{\sqrt{15}} \\ &= 34 \end{aligned}$$

We are 95% confident that the population mean is between 28.5 and 34 grams.

$$CI = (28.5,\ 34)$$

7. (15 points)

   A student is taking a multiple choice test with 800 questions. Each question has 3 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 286 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.

$$p_0 = \frac{1}{3} = 0.333$$

State the hypotheses.

$$H_0 \quad \text{claims} \quad p = 0.333$$
$$H_A \quad \text{claims} \quad p > 0.333$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.333(1-0.333)}{800}} = 0.0167$$

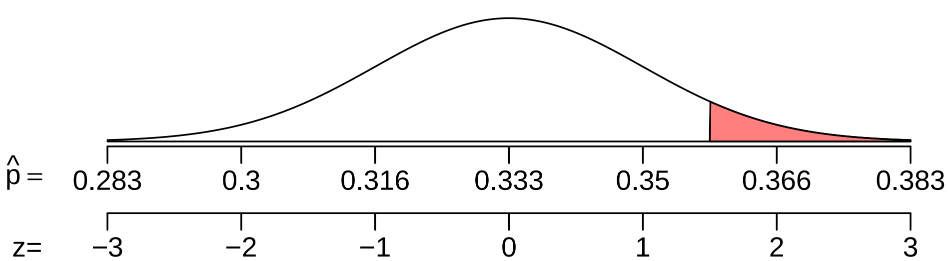
Determine the sample proportion.

$$\hat{p} = \frac{286}{800} = 0.358$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.358 - 0.333}{0.0167} = 1.5$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.358) \\ &= P(Z > 1.5) \\ &= 1 - P(Z < 1.5) \\ &= 0.0668 \end{aligned}$$

Compare p-value to α (which is 0.05).

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

Make the conclusion: we retain the null hypothesis.

We think the student might just be guessing.

(a) Right tail (one-tail) proportion test

(b) Hypotheses: H_0 claims $p = 0.333$ and H_A claims $p > 0.333$.

(c) The p-value is 0.0668

(d) We retain the null hypothesis.

(e) We think the student might just be guessing.

8. (20 points) [Note: this question uses 2 pages.]

You have collected the following data:

x	y	xy
9	530	
1.3	540	
5.4	300	
4.6	550	
7.4	440	
1.1	950	
5.1	500	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

(a) Complete the table.

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

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

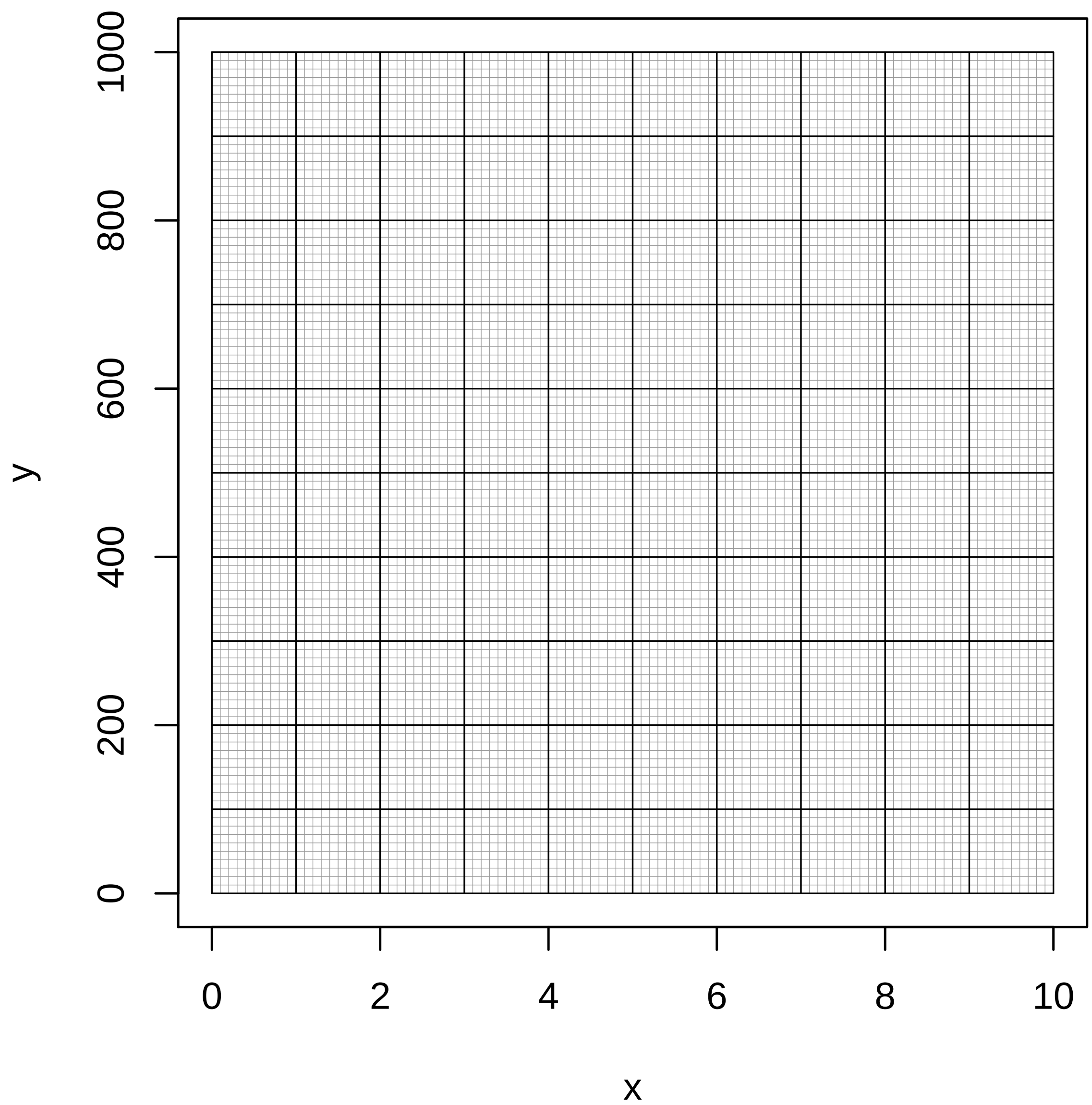
(c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

(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
9	530	4770
1.3	540	702
5.4	300	1620
4.6	550	2530
7.4	440	3256
1.1	950	1045
5.1	500	2550
$\sum x = 33.9$	$\sum y = 3810$	$\sum x_i y_i = 16473$
$\bar{x} = 4.843$	$\bar{y} = 544.3$	
$s_x = 2.91$	$s_y = 198.9$	

$$r = \frac{16473 - (7)(4.843)(544.3)}{(7-1)(2.91)(198.9)} = -0.57$$

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.

$$r_{\text{exact}} = -0.5695996$$

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:

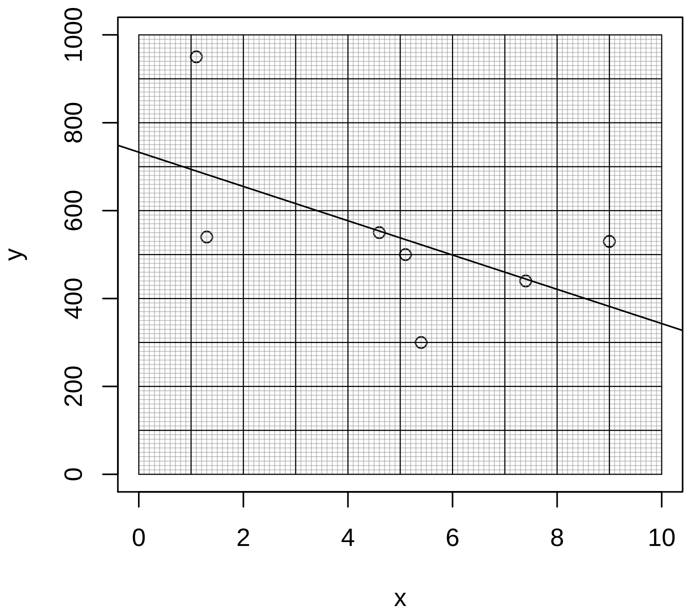
$$b = r\frac{s_y}{s_x} = -0.57 \cdot \frac{198.9}{2.91} = -39$$

$$a = \bar{y} - b\bar{x} = 544 - (-39)(4.84) = 733$$

Our regression line:

$$y = 733 + (-39)x$$

Make a plot.


y
x
0
200
400
600
800
1000
0
2
4
6
8
10

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.77$. If 165 trials occur, what is the probability of getting at least 121 but at most 130 successes?

   In other words, let $X \sim \text{Bin}(n = 165,\ p = 0.77)$ and find $P(121 \leq X \leq 130)$.

   Use a normal approximation along with the continuity correction.

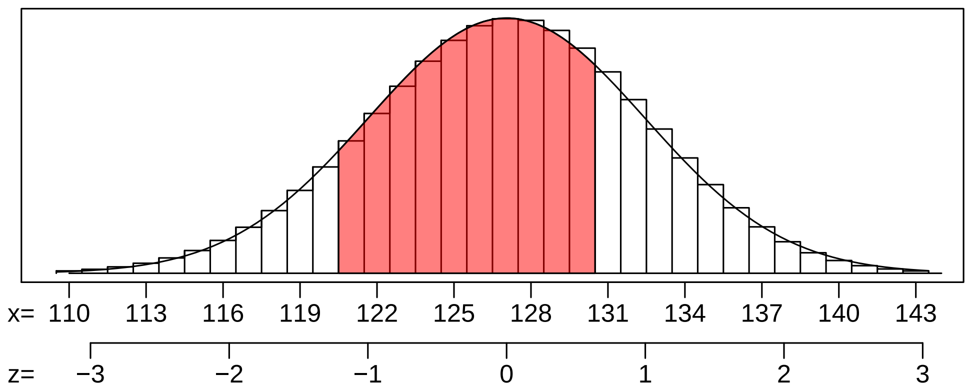
Solution: Find the mean.

$$\mu = np = (165)(0.77) = 127.05$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(165)(0.77)(1-0.77)} = 5.4057$$

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



Find the z scores.

$$z_1 = \frac{120.5 - 127.05}{5.4057} = -1.21$$

$$z_2 = \frac{130.5 - 127.05}{5.4057} = 0.64$$

Find the percentiles (from z-table).

$$\ell_1 = 0.1131$$

$$\ell_2 = 0.7389$$

Calculate the probability.

$$P(121 \leq X \leq 130) = 0.7389 - 0.1131 = 0.6258$$

10. (Extra credit: 5 points)

A null hypothesis claims a population has a mean $\mu = 150$. 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:

149.5	153.8	149.1	155.3	162.2
149.6	150.9	153.6	148.6	152.2
153				

(a) Determine the p-value.

(b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 150$$

$$H_A \quad \text{claims} \quad \mu \neq 150$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 152.527$$

$$s = 3.897$$

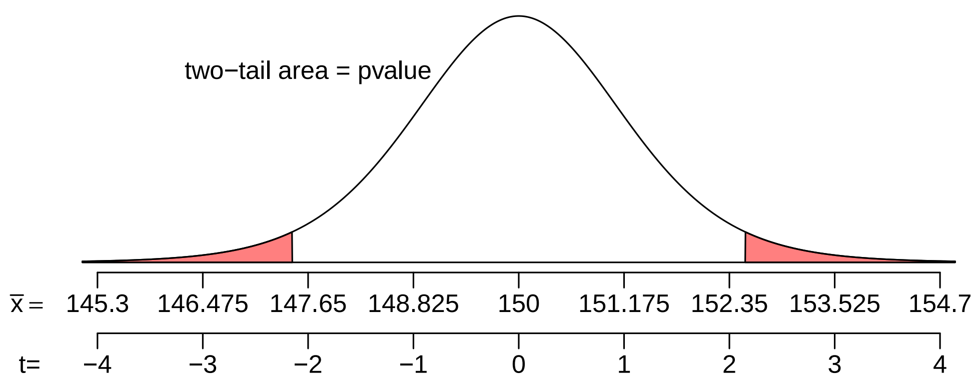
Determine the degrees of freedom.

$$df = 11 - 1 = 10$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{3.897}{\sqrt{11}} = 1.175$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{152.527 - 150}{1.175} = 2.15$$

Find the p-value.

$$p\text{-value} = P(|T| > 2.15)$$

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.