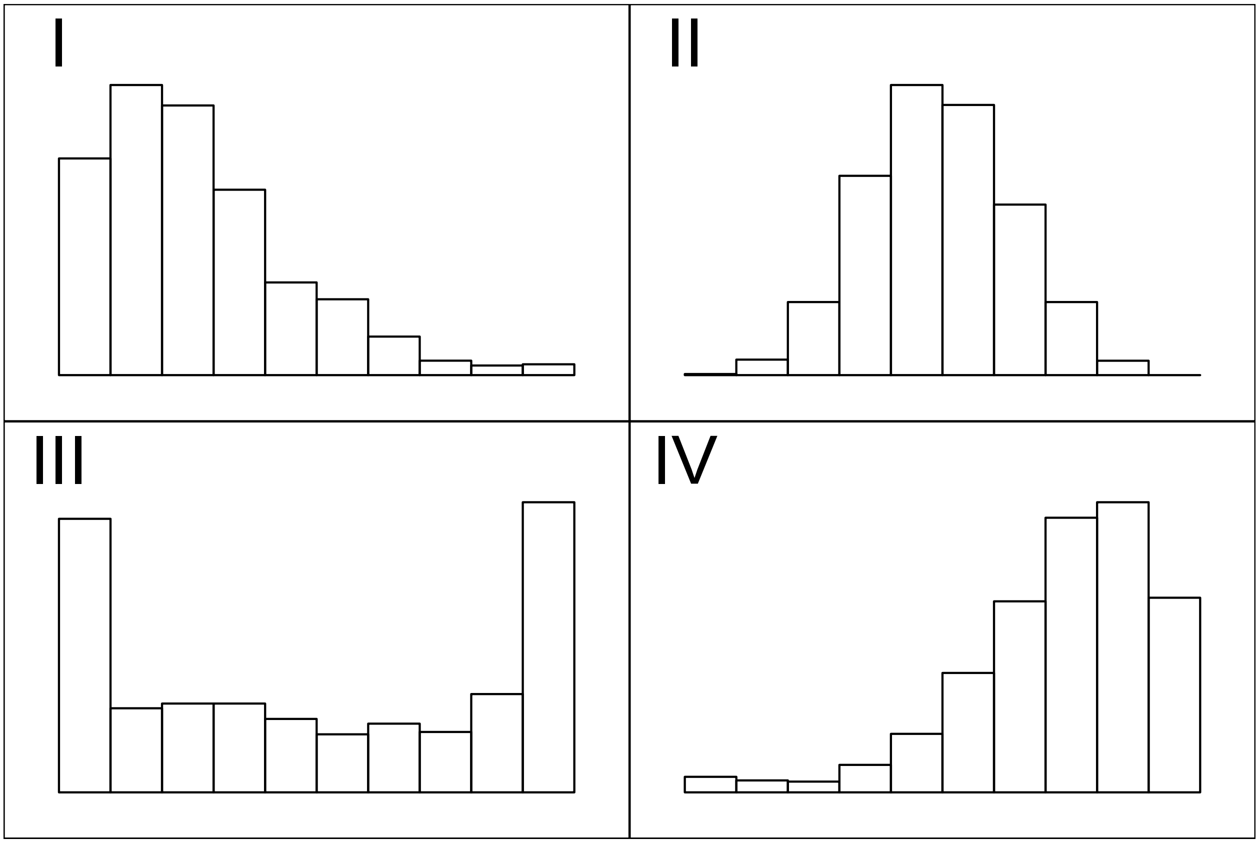
MAT-181 FINAL TAKE-HOME EXAM

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 test scores on a very difficult exam, in which most students have poor to average scores, but a few did quite well.

(b) The distribution of lengths of newborn babies

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

(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) II
(c) IV
(d) III

2. (15 Points)

In a deck of strange cards, there are 514 cards. Each card has an image and a color. The amounts are shown in the table below.

	blue	orange	pink	violet	yellow	Total
bike	10	49	25	29	11	124
gem	45	28	18	34	16	141
lamp	23	33	17	13	12	98
pig	14	46	26	24	41	151
Total	92	156	86	100	80	514

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

(b) What is the probability a random card is a lamp?

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

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

(e) What is the probability a random card is either a bike or blue (or both)?

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

(g) Is a bike or a gem more likely to be blue?

Solution:

(a) P(pig given blue) = 0.152

(b) P(lamp) = 0.191

(c) P(orange) = 0.304

(d) P(blue given gem) = 0.319

(e) P(bike or blue) = 0.401

(f) P(gem and orange) = 0.0545

(g) P(blue given bike) = 0.0806 and P(blue given gem) = 0.319, so a gem is more likely to be blue than a bike 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	101	8
B	111	11
C	123	12
D	69	14

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

Type of fruit	Mass of specimen (g)
A	89.08
B	112.1
C	106.4
D	60.88

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	$z = \frac{89.08-101}{8}$	-1.49
B	$z = \frac{112.1-111}{11}$	0.1
C	$z = \frac{106.4-123}{12}$	-1.38
D	$z = \frac{60.88-69}{14}$	-0.58

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

4. (10 points)

   A tree's leaves were found to be normally distributed with a mean of 141.4 millimeters and a standard deviation of 2.7 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 144 and 147.2 millimeters?

Solution:

$$\mu = 141.4$$
$$\sigma = 2.7$$
$$x_1 = 144$$
$$x_2 = 147.2$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{144 - 141.4}{2.7} = 0.96$$
$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{147.2 - 141.4}{2.7} = 2.15$$
$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.9842 - 0.8315 = 0.1527$$

5. (10 points)

   A species of duck is known to have a mean weight of 111.5 grams and a standard deviation of 67.5 grams. A researcher plans to measure the weights of 81 of these ducks sampled randomly. What is the probability the **sample mean** will be between 96.5 and 128 grams?

Solution:

$$n = 81$$
$$\mu = 111.5$$
$$\sigma = 67.5$$
$$SE = \frac{67.5}{\sqrt{81}} = 7.5$$
$$x_1 = 96.5$$
$$x_2 = 128$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{96.5 - 111.5}{7.5} = -2$$
$$z_2 = \frac{x_2 - \mu}{SE} = \frac{128 - 111.5}{7.5} = 2.2$$
$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.9861 - 0.0228 = 0.9633$$

6. (10 points)

   An ornithologist wishes to characterize the average body mass of *Hylocichla mustelina*. She randomly samples 20 adults of *Hylocichla mustelina*, resulting in a sample mean of 57.29 grams and a sample standard deviation of 7.67 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.

$$\begin{aligned} n &= 20 \\ \bar{x} &= 57.29 \\ s &= 7.67 \\ \gamma &= 0.95 \end{aligned}$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 20 - 1 \\ &= 19 \end{aligned}$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$ and $df = 19$.

$$t^\star = 2.09$$

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

$$\begin{aligned} LB &= \bar{x} - t^\star \frac{s}{\sqrt{n}} & UB &= \bar{x} + t^\star \frac{s}{\sqrt{n}} \\ &= 57.29 - 2.09 \times \frac{7.67}{\sqrt{20}} & &= 57.29 + 2.09 \times \frac{7.67}{\sqrt{20}} \\ &= 53.7 & &= 60.9 \end{aligned}$$

We are 95% confident that the population mean is between 53.7 and 60.9 grams.

$$CI = (53.7,\ 60.9)$$

7. (15 points)

   A student is taking a multiple choice test with 200 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 110 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.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.5(1-0.5)}{200}} = 0.0354$$

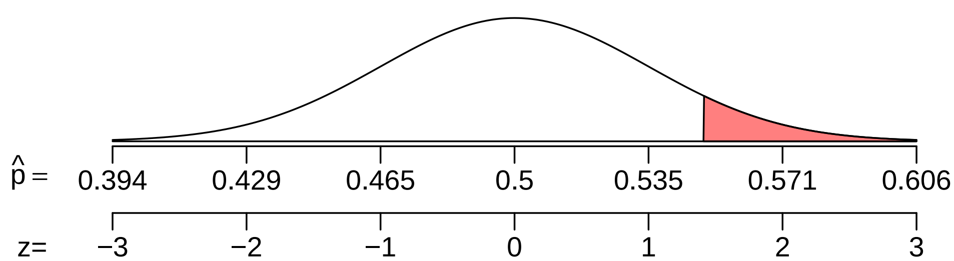
Determine the sample proportion.

$$\hat{p} = \frac{110}{200} = 0.55$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.55 - 0.5}{0.0354} = 1.41$$

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.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.55) \\ &= P(Z > 1.41) \\ &= 1 - P(Z < 1.41) \\ &= 0.0793 \end{aligned}$$

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

(c) The p-value is 0.0793

(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
58	760	
54	620	
93	470	
41	650	
90	480	
41	570	
34	740	
64	680	
25	800	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

(a) Complete the table.

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

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

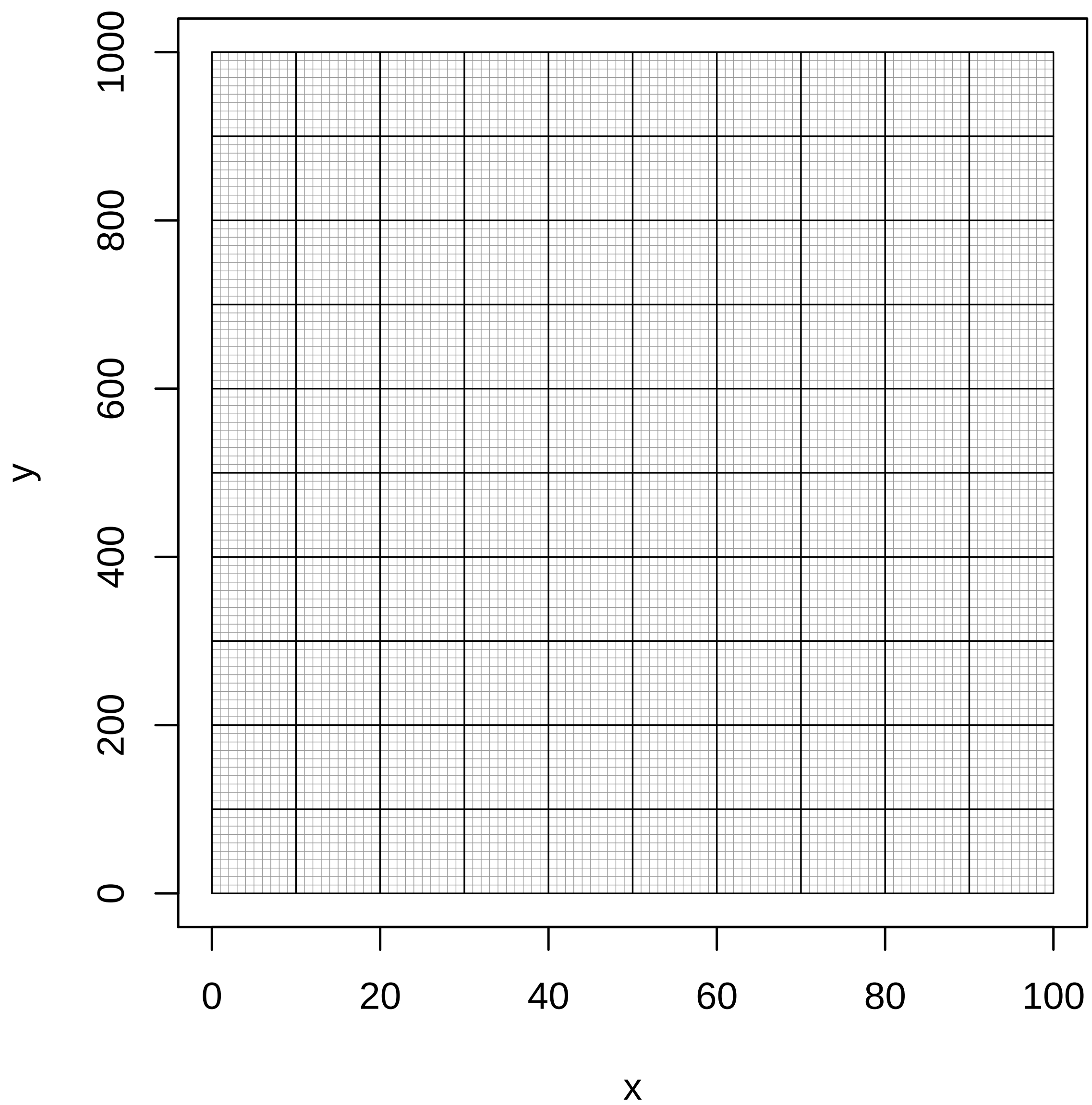
(c) The least-squares regression line will be represented as $y = a + bx$. Determine the parameters (b and a) using the formulas below.

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

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

(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
58	760	44080
54	620	33480
93	470	43710
41	650	26650
90	480	43200
41	570	23370
34	740	25160
64	680	43520
25	800	20000
$\sum x = 500$	$\sum y = 5770$	$\sum x_i y_i = 303170$
$\bar{x} = 55.56$	$\bar{y} = 641.1$	
$s_x = 23.69$	$s_y = 118.1$	

$$r = \frac{303170 - (9)(55.56)(641.1)}{(9-1)(23.69)(118.1)} = -0.778$$

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.

$$r_{\text{exact}} = -0.7770319$$

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:

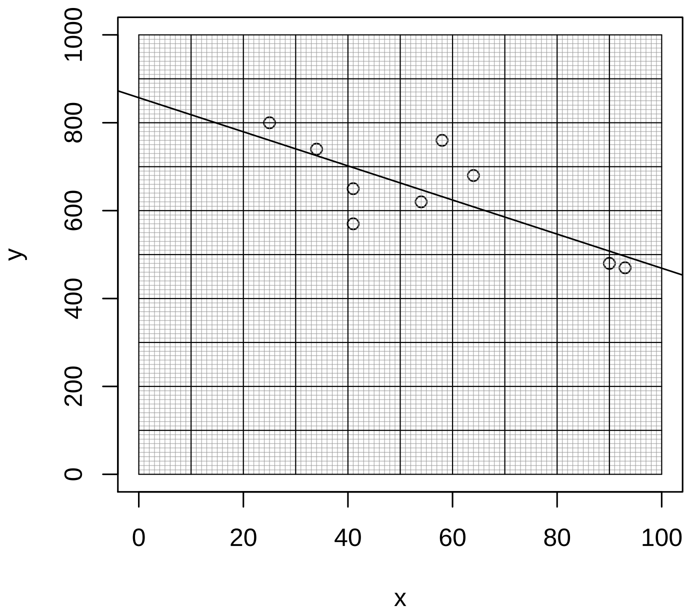
$$b = r\frac{s_y}{s_x} = -0.778 \cdot \frac{118.1}{23.69} = -3.88$$

$$a = \bar{y} - b\bar{x} = 641 - (-3.88)(55.6) = 857$$

Our regression line:

$$y = 857 + (-3.88)x$$

Make a plot.



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

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.46$. If 174 trials occur, what is the probability of getting more than 64 but less than 87 successes?

   In other words, let $X \sim \text{Bin}(n = 174,\ p = 0.46)$ and find $P(64 < X < 87)$.

   Use a normal approximation along with the continuity correction.

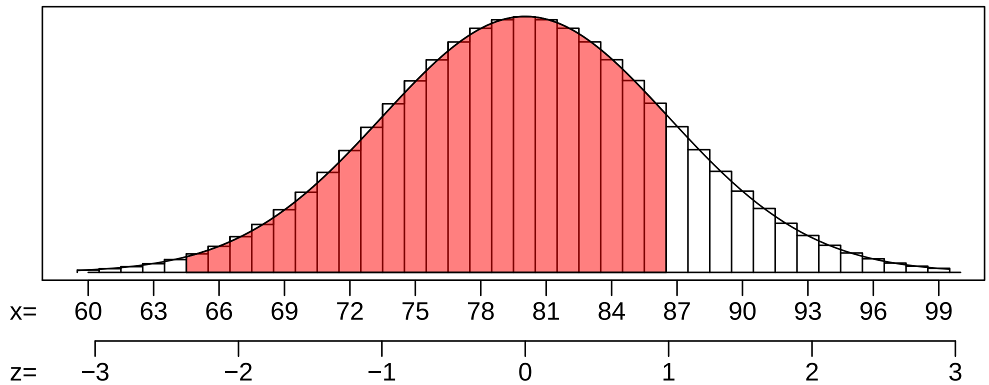
Solution: Find the mean.

$$\mu = np = (174)(0.46) = 80.04$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(174)(0.46)(1-0.46)} = 6.5743$$

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



Find the z scores.

$$z_1 = \frac{64.5 - 80.04}{6.5743} = -2.36$$

$$z_2 = \frac{86.5 - 80.04}{6.5743} = 0.98$$

Find the percentiles (from z-table).

$$\ell_1 = 0.0091$$

$$\ell_2 = 0.8365$$

Calculate the probability.

$$P(64 < X < 87) = 0.8365 - 0.0091 = 0.827$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 60$. You decide to run two-tail test on a sample of size $n = 9$ using a significance level $\alpha = 0.01$.

    You then collect the sample:

57.1	74.5	65	61.3	69.1
76.7	62.8	70.2	78.7	

    (a) Determine the p-value.

    (b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \text{ claims } \mu = 60$$

$$H_A \text{ claims } \mu \neq 60$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 68.378$$

$$s = 7.39$$

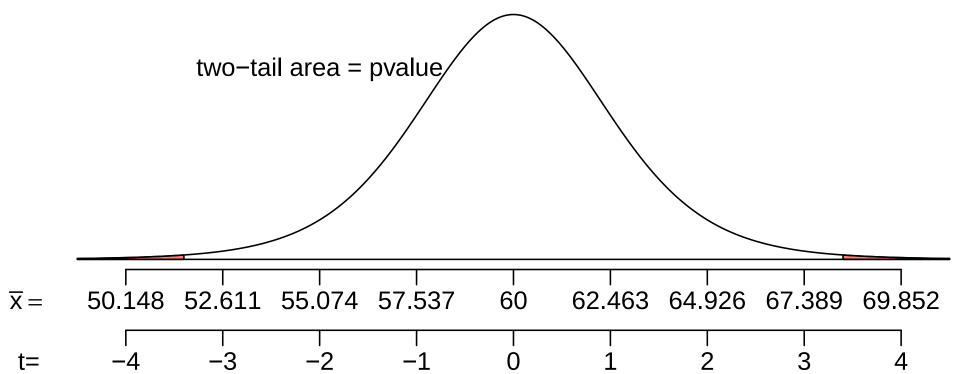
Determine the degrees of freedom.

$$df = 9 - 1 = 8$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{7.39}{\sqrt{9}} = 2.463$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{68.378 - 60}{2.463} = 3.4$$

Find the p-value.

$$p\text{-value} = P(|T| > 3.4)$$

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 $df = 8$.)

$$P(|T| > 3.83) = 0.005$$

$$P(|T| > 3.36) = 0.01$$

Basically, because t is between 3.83 and 3.36, we know the p-value is between 0.005 and 0.01.

$$0.005 < p\text{-value} < 0.01$$

Compare the p-value and the significance level ($\alpha = 0.01$).

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

Yes, we reject the null hypothesis.

(a) $0.005 < p\text{-value} < 0.01$

(b) Yes, we reject the null hypothesis.