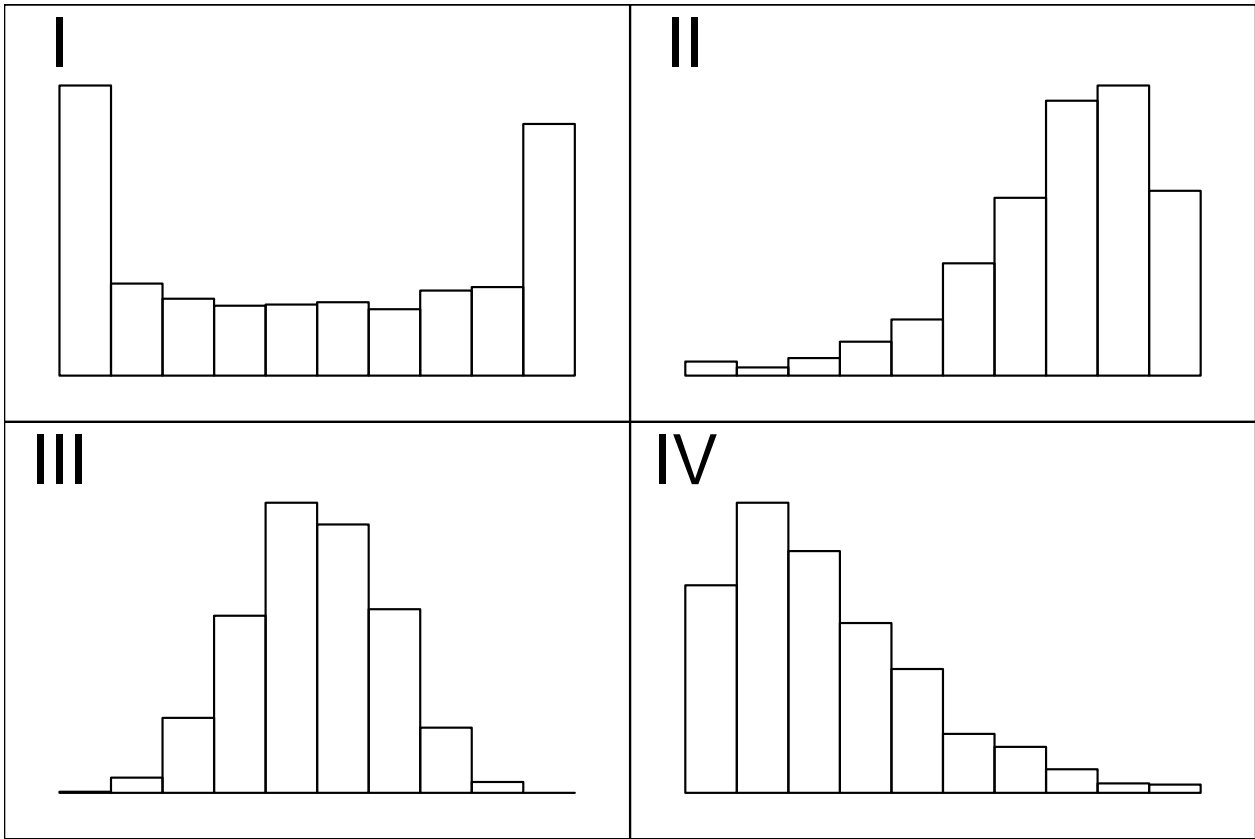# MAT-181 FINAL TAKE-HOME EXAM

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 ages at a skilled nursing facility, where most of the patients are elderly but a few are quite young.

(c) The distribution of weights of newborn babies

(d) 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.

Solution:

(a) I

(b) II

(c) III

(d) IV

2. (15 Points)

In a deck of strange cards, there are 674 cards. Each card has an image and a color. The amounts are shown in the table below.

	black	blue	gray	yellow	Total
gem	34	25	20	18	97
needle	46	38	40	47	171
rug	49	19	27	39	134
shovel	30	42	44	48	164
tree	33	12	13	50	108
Total	192	136	144	202	674

(a) What is the probability a random card is both a tree and gray?

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

(c) What is the probability a random card is black given it is a gem?

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

(e) Is a needle or a shovel more likely to be gray?

(f) What is the probability a random card is either a tree or blue (or both)?

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

Solution:

(a) P(tree and gray) = 0.0193

(b) P(needle) = 0.254

(c) P(black given gem) = 0.351

(d) P(black) = 0.285

(e) P(gray given needle) = 0.234 and P(gray given shovel) = 0.268, so a shovel is more likely to be gray than a needle is.

(f) P(tree or blue) = 0.344

(g) P(gem given blue) = 0.184

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	88	15
B	91	13
C	130	5
D	131	12

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

Type of fruit	Mass of specimen (g)
A	89.35
B	71.24
C	122.4
D	127.2

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	$z = \frac{\|89.35-88\|}{15}$	0.09
B	$z = \frac{\|71.24-91\|}{13}$	1.52
C	$z = \frac{\|122.4-130\|}{5}$	1.53
D	$z = \frac{\|127.2-131\|}{12}$	0.32

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 104.1 millimeters and a standard deviation of 7.2 millimeters. If you pick a random leaf from that tree, what is the probability the length is between 87.9 and 106.7 millimeters?

Solution:

$$\mu = 104.1$$
$$\sigma = 7.2$$
$$x_1 = 87.9$$
$$x_2 = 106.7$$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{87.9 - 104.1}{7.2} = -2.25$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{106.7 - 104.1}{7.2} = 0.36$$

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2) = 0.6406 - 0.0122 = 0.6284$$

5. (10 points)

   A species of duck is known to have a mean weight of 210.6 grams and a standard deviation of 24 grams. A researcher plans to measure the weights of 144 of these ducks sampled randomly. What is the probability the **sample mean** will be between 211.1 and 213.1 grams?

Solution:

$$n = 144$$
$$\mu = 210.6$$
$$\sigma = 24$$
$$SE = \frac{24}{\sqrt{144}} = 2$$
$$x_1 = 211.1$$
$$x_2 = 213.1$$

$$z_1 = \frac{x_1 - \mu}{SE} = \frac{211.1 - 210.6}{2} = 0.25$$

$$z_2 = \frac{x_2 - \mu}{SE} = \frac{213.1 - 210.6}{2} = 1.25$$

$$P(x_1 < \overline{X} < x_2) = P(z_1 < Z < z_2) = 0.8944 - 0.5987 = 0.2957$$

6. (10 points)

An ornithologist wishes to characterize the average body mass of *Dendroica coronata*. She randomly samples 22 adults of *Dendroica coronata*, resulting in a sample mean of 11.96 grams and a sample standard deviation of 1.21 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.

$$n = 22$$
$$\bar{x} = 11.96$$
$$s = 1.21$$
$$\gamma = 0.95$$

Find the degrees of freedom.

$$df = n - 1$$
$$= 22 - 1$$
$$= 21$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.95$ and $df = 21$.

$$t^\star = 2.08$$

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

$$LB = \bar{x} - t^\star \frac{s}{\sqrt{n}} \qquad\qquad UB = \bar{x} + t^\star \frac{s}{\sqrt{n}}$$
$$= 11.96 - 2.08 \times \frac{1.21}{\sqrt{22}} \qquad\qquad = 11.96 + 2.08 \times \frac{1.21}{\sqrt{22}}$$
$$= 11.4 \qquad\qquad\qquad\qquad = 12.5$$

We are 95% confident that the population mean is between 11.4 and 12.5 grams.

$$CI = (11.4,\ 12.5)$$

7. (15 points)

A student is taking a multiple choice test with 500 questions. Each question has 4 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 142 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.

$$p_0 = \frac{1}{4} = 0.25$$

State the hypotheses.

$$H_0 \quad \text{claims} \quad p = 0.25$$
$$H_A \quad \text{claims} \quad p > 0.25$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1 - p_0)}{n}} = \sqrt{\frac{0.25(1 - 0.25)}{500}} = 0.0194$$
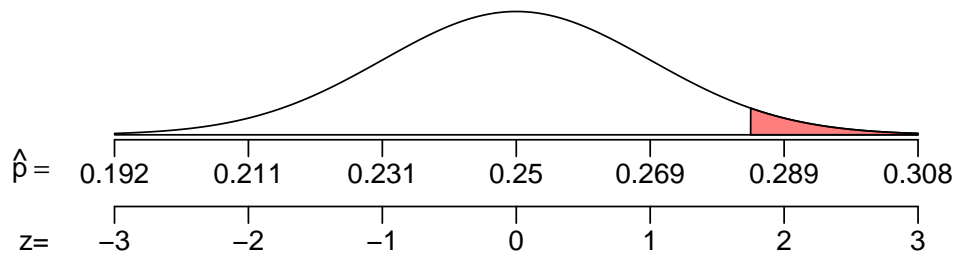
Determine the sample proportion.

$$\hat{p} = \frac{142}{500} = 0.284$$

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

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.284 - 0.25}{0.0194} = 1.75$$

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.

$$
\begin{aligned}
p\text{-value} &= P\,(\hat{p} > 0.284) \\
&= P(Z > 1.75) \\
&= 1 - P(Z < 1.75) \\
&= 0.0401
\end{aligned}
$$

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 $p = 0.25$ and H_A claims $p > 0.25$.

(c) The p-value is 0.0401

(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
98	13	
49	29	
42	54	
78	32	
89	44	
34	100	
$\sum x =$	$\sum y =$	$\sum xy =$
$\bar{x} =$	$\bar{y} =$	
$s_x =$	$s_y =$	

(a) Complete the table.

(b) Calculate the correlation coefficient (r) using the formula below.
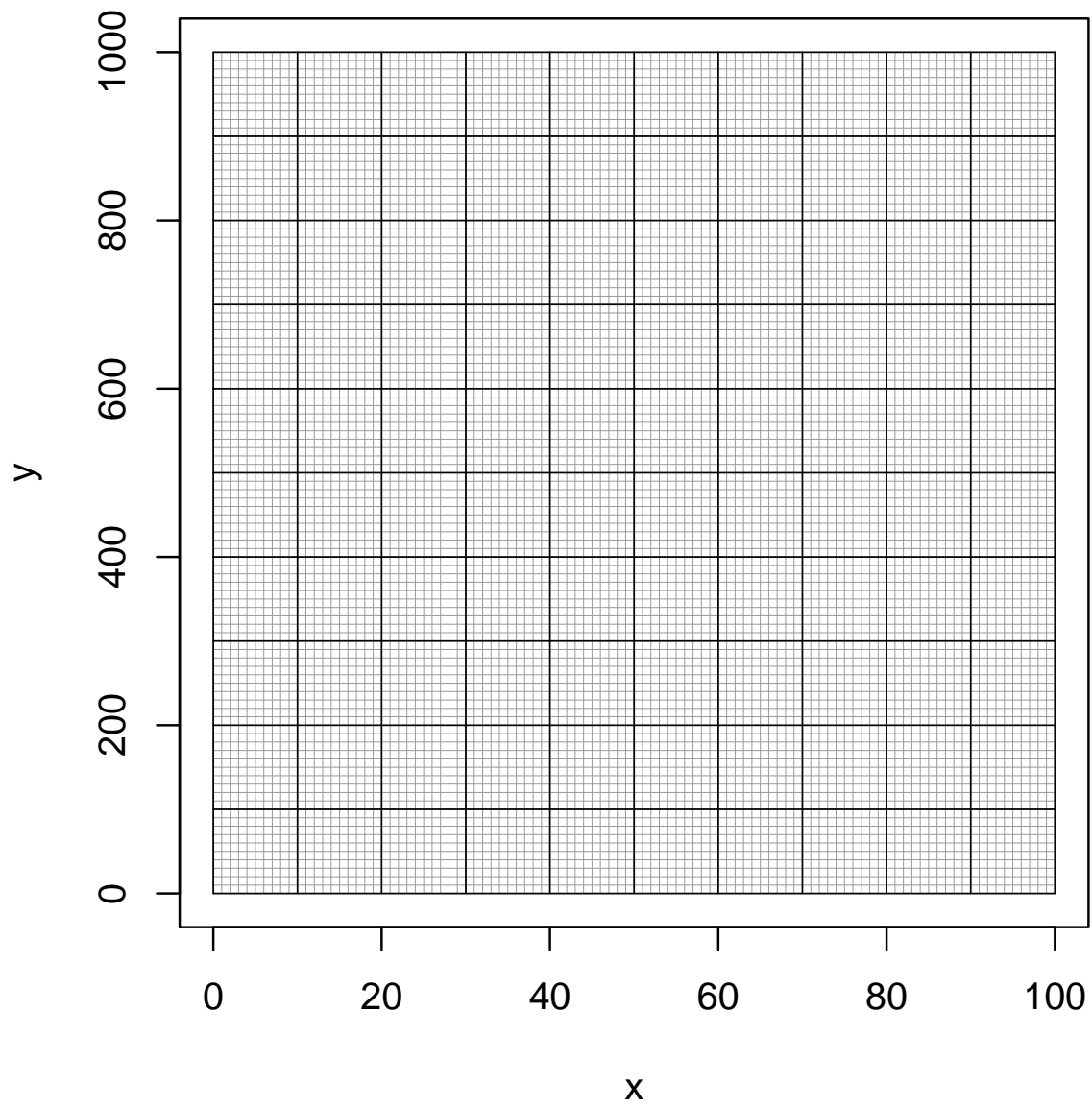
$$r = \frac{\sum xy - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

(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
98	13	1274
49	29	1421
42	54	2268
78	32	2496
89	44	3916
34	100	3400
$\sum x = 390$	$\sum y = 272$	$\sum x_i y_i = 14775$
$\bar{x} = 65$	$\bar{y} = 45.33$	
$s_x = 26.76$	$s_y = 30.18$	

$$r = \frac{14775 - (6)(65)(45.33)}{(6-1)(26.76)(30.18)} = -0.719$$

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.

$$r_{\text{exact}} = -0.7193571$$

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:
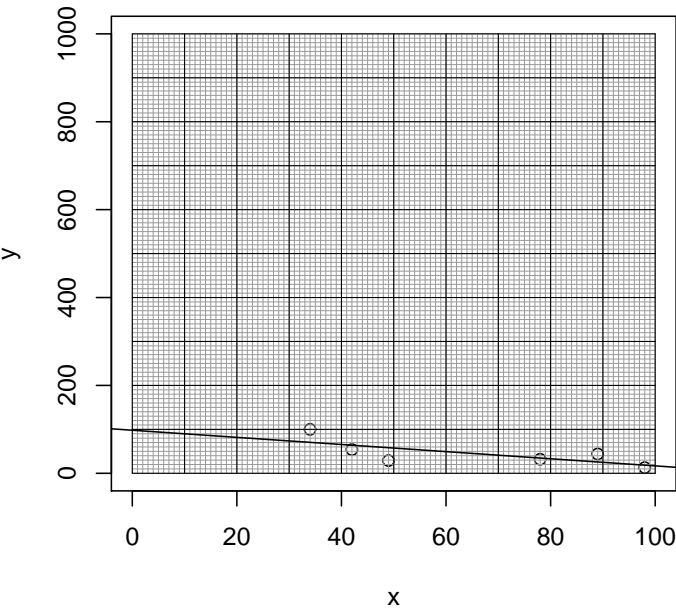
$$b = r\frac{s_y}{s_x} = -0.719 \cdot \frac{30.18}{26.76} = -0.811$$

$$a = \bar{y} - b\bar{x} = 45.3 - (-0.811)(65) = 98$$

Our regression line:

$$y = 98 + (-0.811)x$$

Make a plot.

9. (Extra credit: 5 points)

   Let each trial have a chance of success $p = 0.16$. If 86 trials occur, what is the probability of getting at least 15 but at most 21 successes?

   In other words, let $X \sim \text{Bin}(n = 86, p = 0.16)$ and find $P(15 \leq X \leq 21)$.

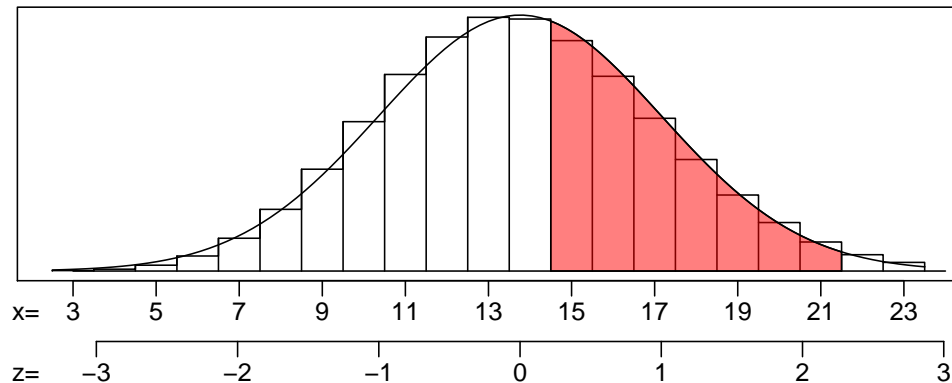   Use a normal approximation along with the continuity correction.

Solution: Find the mean.

$$\mu = np = (86)(0.16) = 13.76$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(86)(0.16)(1-0.16)} = 3.3998$$

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

Find the z scores.

$$z_1 = \frac{14.5 - 13.76}{3.3998} = 0.22$$

$$z_2 = \frac{21.5 - 13.76}{3.3998} = 2.28$$

Find the percentiles (from z-table).

$$\ell_1 = 0.5871$$

$$\ell_2 = 0.9887$$

Calculate the probability.

$$P(15 \leq X \leq 21) = 0.9887 - 0.5871 = 0.4016$$

10. (Extra credit: 5 points)

    A null hypothesis claims a population has a mean $\mu = 240$. You decide to run two-tail test on a sample of size $n = 10$ using a significance level $\alpha = 0.01$.

    You then collect the sample:

239.5	279.5	346.7	324.2	322.3
297.7	209	258.8	262.8	321.3

    (a) Determine the p-value.
    (b) Do you reject the null hypothesis?

Solution: State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 240$$

$$H_A \quad \text{claims} \quad \mu \neq 240$$

Find the mean and standard deviation of the sample.
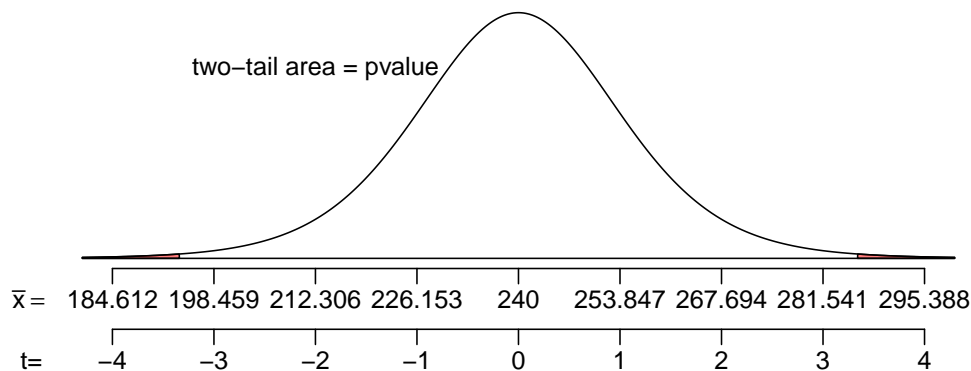
$$\bar{x} = 286.18$$

$$s = 43.787$$

Determine the degrees of freedom.

$$df = 10 - 1 = 9$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{43.787}{\sqrt{10}} = 13.847$$

Make a sketch of the null's sampling distribution.

two–tail area = pvalue

$\bar{x} =$	184.612	198.459	212.306	226.153	240	253.847	267.694	281.541	295.388
t=	−4	−3	−2	−1	0	1	2	3	4

Find the *t* score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{286.18 - 240}{13.847} = 3.34$$

Find the *p*-value.

$$p\text{-value} = P(|T| > 3.34)$$

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.)

$$P(|T| > 3.69) = 0.005$$

$$P(|T| > 3.25) = 0.01$$

Basically, because *t* is between 3.69 and 3.25, 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.