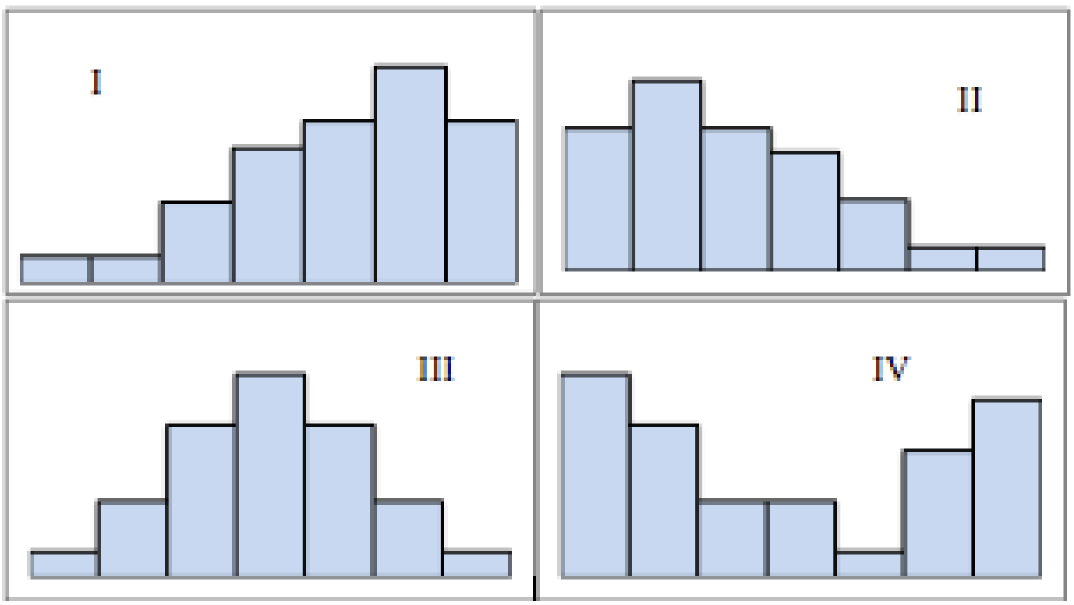
Name: ______________________ **Section: <u>MAT098/181C-</u>**

MAT098/181C FINAL EXAM (FORM D Key)

A scientific or graphing calculator is permitted. ***<u>Cellphones may not be used as calculators and must be off or on vibrate during the exam</u>****. Show all work on the test*

1. For each description, choose the histogram (I, II, III, IV) that matches the description. (8 pts)



a. The distribution of hours that students studied for an exam. Many students studied a lot. A similar number of students did not study very much.
IV

b. The distribution of heights of adult men.
III

c. The distribution of test scores on a very difficult final exam, in which most students have poor to average scores.
II

d. The distribution of ages at a skilled nursing facility, where most of the patients are elderly.
I

2. The table below shows the number of people in each group who survived or perished when the *Titanic* sank in April 1912. (Source: British Board of Trade report) One person is randomly selected. (10 pts)

	Group				
	1st class	**2nd class**	**3rd class**	**Crew**	**Total**
Survived	202	118	178	212	710
Perished	123	167	528	696	1514
Total	325	285	706	908	2224

a. What is the probability that the person survived?

$$P(\text{survived}) = \frac{710}{2224}$$

b. What is the probability that the person was in 3rd class, given that the person perished?

$$P(\text{3rd class}|\text{perished}) = \frac{528}{1514}$$

c. What is the probability that the person was in 1st class or survived?

$$P(\text{1st class or survived}) = \frac{325 + 710 - 202}{2224} = \frac{833}{2224}$$

3. During the 2016-2017 regular NBA season, the Boston Celtics won 53 games, the most among the teams in the Eastern Conference. The Golden State Warriors won 67 games, the most among the teams in the Western Conference. In the Eastern Conference, the mean games won was 39.6 with a standard deviation of 9.2 games. In the Western Conference, the mean games won was 42.4 with a standard deviation of 12.4 games. (10 pts)

    a. Convert each team's number of games won to a standard z score, rounded to 2 decimal places.

    $$\text{Celtics: } z = \frac{x - \mu}{\sigma} = \frac{53 - 39.6}{9.2} = 1.46$$

    $$\text{Warriors: } z = \frac{x - \mu}{\sigma} = \frac{67 - 42.4}{12.4} = 1.98$$

    b. Which team performed better compared to the other teams in its Conference? Explain your answer.

    Warriors, because their z-score is greater than the Celtics' z-score.

4. An investor monitors her stock investment's performance every year. Suppose the investment has a mean yearly change in value of 7.6% with a standard deviation of 4.3%. The distribution is a normal distribution. One year is selected at random. Find the probability that the investment's change in value was between 8% and 10%. (10 pts)

$$P(8 < x < 10) = P\left(\frac{8 - 7.6}{4.3} < z < \frac{10 - 7.6}{4.3}\right)$$

$$= P(0.09 < z < 0.56)$$

$$= 0.7123 - 0.5359$$

$$= 0.1764$$

5. For the first 12 seasons of the reality TV competition show *The Voice*, the mean age of the contestants was 24.4 years with a standard deviation of 7.0 years. Suppose that 50 contestants are selected at random. What is the probability that the ***mean age*** $\overline{\boldsymbol{x}}$ of the contestants is less than 22 years? (10 pts)

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma/\sqrt{n}} = \frac{22 - 24.4}{7.0/\sqrt{50}} = -2.42$$

$$P(\bar{x} < 22) = P(z < -2.42) = 0.0078$$

6. In a simple random sample of 4000 households, 63% reported that they have at least one person in the home who plays video games regularly (at least 3 hours per week). Construct a 90% confidence interval estimate for the proportion of households with at least one person who pays video games regularly. Round the lower limit and upper limit to 3 decimal places. (12 pts)

$$\hat{p} \pm Z_c \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.63 \pm 1.645 \cdot \sqrt{\frac{0.63(0.37)}{4000}} = 0.63 \pm 0.013$$

$$= (0.617, 0.643)$$

7. In a 2016 Los Angeles Times poll, 974 out of 1909 registered California voters said that they opposed a ballot measure to repeal the state's death penalty. Test the claim that the majority (more than 50%) of California voters opposed the measure to repeal the death penalty. Assume a significance level of 0.01. (20 pts)

   a. Express the original claim in symbolic form.

      $p > 0.50$

   b. Identify the null and alternative hypotheses in symbolic form.

      $H_0: p = 0.50$
      $H_1: p > 0.50$

   c. Is the test a left-tailed, right-tailed, OR two-tailed test?

      right tailed

   d. What is the value of the test statistic, rounded to 2 decimal places?

      $$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.51 - 0.50}{\sqrt{\frac{0.50(1-0.50)}{1909}}} = 0.87$$

   e. What is the P-value?

      $\text{P-value} = 1 - 0.8078 = 0.1922$

   f. State the conclusion about the null hypothesis: reject H_0 OR fail to reject H_0? Explain your answer.

      Fail to reject H_0, because the P-value is greater than the significance level.

   g. State a final conclusion (in non-technical language) that addresses the original claim.

      There is not sufficient evidence to support the claim that the majority of California voters opposed the measure to repeal the death penalty.

The table below shows the mean commuting times (in minutes) of workers and the home ownership rate (as a percent) in Suffolk County, MA from 2010 to 2015. (20 pts)

Year	2010	2011	2012	2013	2014	2015
Mean commuting time (x)	28.5	28.7	28.9	29.0	29.5	30.2
Home ownership rate (y)	39.5	38.8	38.5	37.8	37.8	37.5

a. Use your calculator to find the correlation coefficient r, rounded to 3 decimal places. What does r tell you about the ***direction*** and ***strength*** of the relationship between x and y? Assume a significance level of 0.05.

$$r = -0.846$$

Negative, Strong.

b. Use your calculator to find the following, rounded to 2 decimal places.

$\bar{x} =$ 29.13 $\bar{y} =$ 38.32 $s_x =$ 0.62 $s_y =$ 0.76

c. Find the equation for the regression line, $y = a + bx$ using the following, rounded to 2 decimal places:

$$b = r\frac{s_y}{s_x} \quad \text{and} \quad a = \bar{y} - b\bar{x}$$

$$b = -0.846 \cdot \frac{0.76}{0.62} = -1.04$$

$$a = 38.32 - (-1.04)(29.13) = 68.62$$

$$y = 68.62 - 1.04x$$

d. Predict the home ownership rate if the mean commuting time is 40 minutes. What is a potential problem in predicting the home ownership rate for a much longer commuting time?

$$y = 68.62 - 1.04(40) = 27.02\%$$

Possible answers: The equation has a negative slope, so the home ownership rate would eventually be 0%, which is unlikely to happen; The relationship might not be linear, because the home ownership rate could stabilize or even increase over much longer commuting times.

<u>*EXTRA CREDIT</u>: Assume that 85% of adults met with a health professional last year. A hospital survey showed that 1742 out of 2000 adults met with a health professional last year. (a) Calculate *np* and *nq*, and then explain why it is reasonable to approximate this distribution by a normal distribution. (b) What is the probability that, among 2000 adults, more than 1742 of them met with a health professional last year? (3 pts)

$$np = 2000 \bullet 0.85 = 1700$$
$$nq = 2000 \bullet 0.15 = 300$$

Because np ≥ 5 and nq ≥ 5, we can approximate this distribution by a normal distribution.

$$\mu = np = 2000 \bullet 0.85 = 1700$$

$$\sigma = \sqrt{2000 \bullet 0.85 \bullet 0.15} = 16.0$$

$$P(x_B > 1742) = P(x_N > 1742.5)$$

$$= P\left(z > \frac{1742.5 - 1700}{16.0}\right)$$

$$= P(z > 2.66)$$

$$= 1 - 0.9961$$

$$= 0.0039$$

<u>*EXTRA CREDIT</u>: A marketing analyst for a clothing company wants to estimate the percentage of teens (ages 13-17) who use Snapchat. Assume that 75% of teens use Snapchat. How many teens must the analyst survey in order to be 95% confident that the estimate is within 4 percentage points of the true population proportion? (3 pts)

$$n = \hat{p} \cdot \hat{q} \cdot \left(\frac{Z_c}{E}\right)^2$$

$$n = 0.75 \cdot 0.25 \cdot \left(\frac{1.96}{0.04}\right)^2$$

$$\approx 450.2$$

$$\approx 451 \text{ (always round up)}$$

z-score

$$z = \frac{x - \mu}{\sigma}$$

Central Limit Theorem

Mean of the sample mean is $\mu_{\bar{x}} = \mu$

Standard deviation of the sample mean is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

z-score for sample mean

$$z = \frac{x - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

Binomial Distribution

Mean: $\mu = np$

Standard Deviation: $\sigma = \sqrt{np(1-p)}$

Sampling Distribution of Sample Proportion

Mean: p

Standard Deviation: $\sigma = \sqrt{\frac{p(1-p)}{n}}$

Confidence Interval for Population Parameters

Concept	Population Proportion p	Population Mean μ	
confidence interval formula	$\hat{p} \pm Z_c \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$	σ known $\bar{x} \pm Z_c \cdot \dfrac{\sigma}{\sqrt{n}}$	σ unknown $\text{df} = n - 1$ $\bar{x} \pm T_c \cdot \dfrac{s}{\sqrt{n}}$
sample size formula	$\hat{p} = \frac{x}{n}$ known $n = \hat{p} \cdot \hat{q} \cdot \left(\dfrac{Z_c}{E}\right)^2$ $\hat{p}$ unknown $n = \dfrac{1}{4} \cdot \left(\dfrac{Z_c}{E}\right)^2$	$n = \left(\dfrac{Z_c \cdot \sigma}{E}\right)^2$	

- 90% confidence interval: $Z_c \approx 1.645$
- 95% confidence interval: $Z_c \approx 1.960$
- 99% confidence interval: $Z_c \approx 2.576$

Hypothesis Testing

Concept	Population Proportion p	Population Mean μ	
test statistics	$z = \dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$	σ known $z = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$	σ unknown $\text{df} = n - 1$ $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

- If the P-value $< \alpha$, we reject the null hypothesis.
- If the P-value $\geq \alpha$, we fail to reject the null hypothesis.