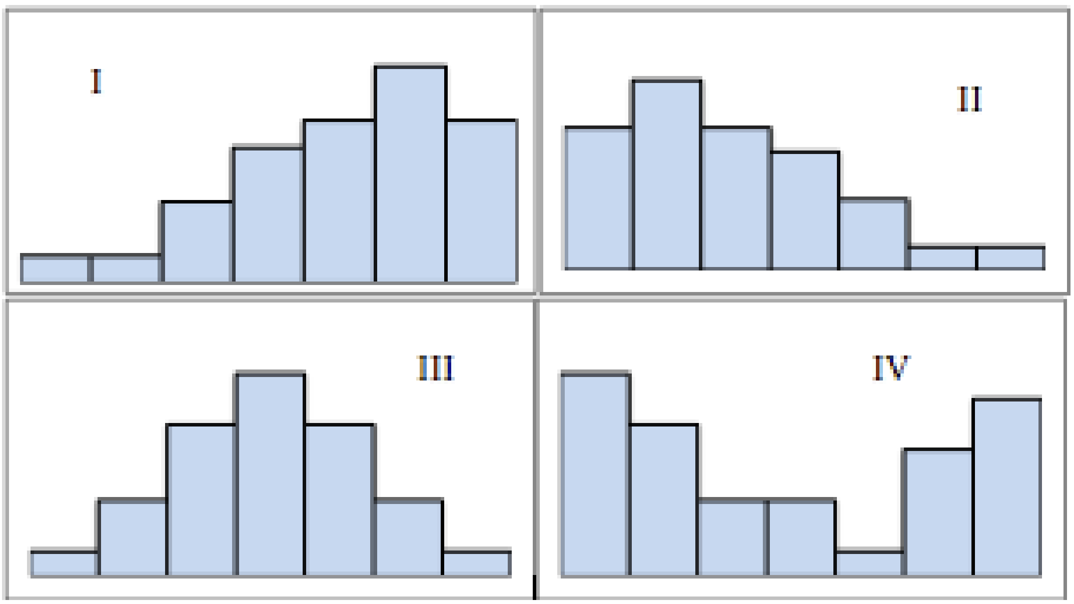
Name: ______________________ **Section: MAT098/181C-**

MAT098/181C FINAL EXAM (FORM C Key)

A scientific or graphing calculator is permitted. ***Cellphones may not be used as calculators and must be off or on vibrate during the exam****. 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 test scores on a very difficult exam, in which most students have poor to average scores.
   II

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

   c. The distribution of height of adult men.
   III

   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.
   IV

2. A bank conducts a survey to find out how people from two different age groups prefer to pay for most purchases. The results of the survey are shown in the table below. One person is randomly selected from the survey. (10 pts)

	Method of Payment				
Age group	**Cash**	**Debit Card**	**Credit Card**	**Mobile App**	**Total**
Under 40	34	62	21	7	124
40 and over	23	30	42	3	98
Total	57	92	63	10	222

a. What is the probability that the person prefers to pay with a credit card?

$$P(\text{credit card}) = \frac{63}{222}$$

b. What is the probability that person prefers to pay with cash, given that the person is at least 40 years old?

$$P(\text{cash}|\text{at least } 40) = \frac{23}{98}$$

c. What is the probability that the person is under 40 years old or prefers to pay with a debit card?

$$P(\text{under } 40 \text{ or debit card}) = \frac{124 + 92 - 62}{222} = \frac{154}{222}$$

3. Carl is working his first job and has saved \$6500 in his 401(k) retirement plan so far. Janet has worked for many years and has saved \$190,000 in her 401(k) plan. Assume that the mean amount saved by workers in Carl's age group is \$4000 with a standard deviation of \$2100, and the mean amount saved by workers in Janet's age group is \$203,000 with a standard deviation of \$50,000. (10 pts)

    a. Convert each savings amount to a standard z score, rounded to 2 decimal places.

$$\text{Carl: } z = \frac{x-\mu}{\sigma} = \frac{6500-4000}{2100} = 1.19$$

$$\text{Janet: } z = \frac{x-\mu}{\sigma} = \frac{190{,}000-203{,}000}{50{,}000} = -0.26$$

    b. Who has saved more money compared to other workers in the same age group? Explain your answer.

    Carl, because his z-score score is greater than Janet's z-score.

4. In a recent study of commuters in Berlin, Germany, the mean commuting time was 34.5 minutes with a standard deviation of 24.2 minutes. The distribution of commuting times is a normal distribution. Find the probability that a randomly selected worker had a commuting time between 20 minutes and 30 minutes. (10 pts)

$$P(20 < x < 30) = P\left(\frac{20-34.5}{24.2} < z < \frac{30-34.5}{24.2}\right)$$

$$= P(-0.60 < z < -0.19)$$

$$= 0.4247 - 0.2743$$

$$= 0.1504$$

5. As of July 3, 2017, the members of the United States Senate had a mean age of 63.2 years with a standard deviation of 10.6 years. What is the probability that the ***mean age $\overline{x}$*** for a simple random sample of 35 Senators is less than 60 years? (10 pts)

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma/\sqrt{n}} = \frac{60 - 63.2}{10.6/\sqrt{35}} = -1.79$$

$$P(\bar{x} < 60) = (P(z < -1.79) = 0.0367$$

6. In a Suffolk University poll conducted in October 2016, 500 people in Massachusetts were asked whether they favored expanding charter schools in the state, and 45.4% said yes. Construct a 95% confidence interval estimate for the proportion of people who favored the expansion of charters schools in the state. Round the lower limit and upper limit of the interval to 3 decimal places. (12 pts)

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

$$0.454 \pm 1.96 \cdot \sqrt{\frac{0.454(0.546)}{500}} = 0.454 \pm 0.044$$

$$= (0.410, 0.498)$$

7. It is widely believed that the mean length of a song is 3.5 minutes. Assume that the population standard deviation is known to be 0.75 minute. A simple random sample of 50 songs shows that its mean length is 3.8 minutes. Test the claim that the mean length of a song is greater than 3.5 minutes. Assume a significance level of 0.10. (20 pts)

    a. Express the original claim in symbolic form.

    $\mu > 3.5$

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

    $H_0: \mu = 3.5$
    $H_1: \mu > 3.5$

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

    right tailed

    d. Do you use the standard normal distribution OR the Student t distribution? Explain your answer.

    Standard normal distribution, because n > 30 and σ is known.

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

    $$z = \frac{x - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3.8 - 3.5}{\frac{0.75}{\sqrt{50}}} \approx 2.83$$

    f. What is the P-value?

    P-value $= 1 - 0.9977 = 0.0023$

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

    Reject H_0, because the P-value is less than the significance level.

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

    There is sufficient evidence to support the claim that the mean length of a song is greater than 3.5 minutes.

8. The table below shows an airline's lowest price *y* for a round-trip ticket from Boston, MA to Los Angeles, CA when purchased *x* days in advance of the flight date. (20 pts)

Days ticket purchased in advance (x)	2	7	21	28	56	77
Dollar price of ticket (*y*)	536	442	360	331	302	294

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.

$$r = -0.840$$

Negative, Strong.

b. Use your calculator to find the following, rounded to 1 decimal place.

$\bar{x} =$ 31.8 $\quad\quad$ $\bar{y} =$ 377.5 $\quad\quad$ $s_x =$ 29.2 $\quad\quad$ $s_y =$ 94.3

c. Find the equation for the regression line, $y = a + bx$ using the formulas below. Round values to 2 decimal places.

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

$$b = -0.840 \cdot \frac{94.3}{29.2} = -2.71$$

$$a = 377.5 - (-2.71)(31.8) = 463.68$$

$$y = 463.68 - 2.71x$$

d. Predict the price of a ticket if it is purchased 100 days in advance. What is a potential problem in predicting the price of ticket purchased far in advance?

$$y = 463.68 - 2.71(100) = \$192.68$$

Possible answers: The equation has a negative slope, so the ticket price would eventually be $0, which is highly unlikely; The price of a ticket can depend on other factors, e.g. whether the travel dates fall around holidays.

<u>*EXTRA CREDIT</u>: Assume that 53% of customers tip their taxi drivers. A taxi company found that among its last 90 customers, 50 of them tipped their drivers. (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 90 customers, at least 50 of them tip their taxi drivers? (3 pts)

$$np = 90 \bullet 0.53 = 47.7$$
$$nq = 90 \bullet 0.47 = 42.3$$

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

$$\mu = 90 \bullet 0.53 = 47.7$$

$$\sigma = \sqrt{90 \bullet 0.53 \bullet 0.47} = 4.7$$

$$P(x_B \geq 50) = P(x_N > 49.5)$$

$$= P\left(z > \frac{49.5 - 47.7}{4.7}\right)$$

$$= P(z > 0.38)$$

$$= 1 - 0.6480$$

$$= 0.3520$$

<u>*EXTRA CREDIT</u>: A manager for a start-up company wants to estimate the percentage of young adults (ages 18-29) who use Instagram. An article stated that 55% of young adults use Instagram. How many young adults must the manager survey in order to be 99% confident that the estimate is within 3 percentage points of the true population proportion? (3 pts)

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

$$n = 0.55 \cdot 0.45 \cdot \left(\frac{2.576}{0.03}\right)^2$$

$$\approx 1824.8$$

$$\approx 1825 \text{ (always round up)}$$

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 $\mathrm{df} = n - 1$ $\bar{x} \pm T_c \cdot \dfrac{s}{\sqrt{n}}$
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 $\mathrm{df} = n - 1$ $t = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

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