Final

You are required to show your work in all problems except problem 4.

Problem 1: A box contains one red marble and nine green ones. Five draws are made at random with replacement. The chance that exactly two draws will be red is:

\[ 10 \times \left( \frac{1}{10} \right)^2 \left( \frac{9}{10} \right)^3 \]

Is the addition rule used in deriving this formula? Answer yes or no, and explain carefully.

Yes, the addition rule is used in this formula, since the chance of each combination in the five draws are mutually exclusive events. To calculate the final probability, we add the probabilities using the addition rule. This is accounted for in the combinatorial number \( \frac{5!}{2! \cdot 3!} = 10 \)

Problem 2: A simple random sample of size 400 was taken from the population of all manufacturing establishments in California. 15 establishments in the sample had 100 employees or more.

1. Estimate the percentage of manufacturing establishments with 100 employees or more.
   \[ \frac{15}{400} \times 100 \%. = 3.75 \% \] (Sample percent) (1 pt)

2. Put a give-or-take number to this estimate.
   \[ SE \ of \ sample \ percentage = \sqrt{\frac{0.0376 \times 0.9625}{400}} \]
   \[ = \frac{0.19}{20} \times 100 \% = 0.95 \% \] (1.5 pt)
(3) Give a 95% confidence interval to the population percentage. 

\[ \left( 3.75 - 2 \times 0.95, \ 3.75 + 2 \times 0.95 \right) \]

\[ = (1.85, 5.65) \]

Problem 3: One hundred draws are made at random with replacement from a box of tickets. Each ticket has a number written on it. The average of the draws is 22.7, and the SD is 10. Someone claims that the average of the box is 20.

(1) Formulate the null and the alternative hypotheses needed to test this claim. (1 pt)

\[ H_0: \mu = 20 \quad (\text{claim}) \]

\[ H_1: \mu > 20 \]

(2) Calculate the appropriate test statistics and the corresponding P-value. (2 pts)

\[ Z = \frac{22.7 - 20}{10/\sqrt{100}} = \frac{2.7}{1} = 2.7 \]

(1 pt)

(3) What is your conclusion?

\[ \text{Reject } H_0 \]

Data provides enough evidence to reject the claim that the average of the draws is 20.

\[ P\text{-Value} = \left(1 - 0.943\right)/2 = 0.00345 \]

\[ < 0.05 \]
Problem 4: True or false:

1. If the correlation coefficient is 0.90, then 90% of the points are highly correlated.
   False. This means the data points are highly correlated. (1 pt)

2. For a very large data set, a t-test will produce almost the same results as a z-test.
   True. (1 pt)

3. The P-value of a test equals its observed significance level.
   True. (1 pt)

4. If the correlation coefficient is -0.80, below-average values of the dependent variable are associated with below-average values of the independent variable.
   False: Below-average values of the dependent variable are associated with above-average values. (1 pt)

5. In the presence of non-linear association the correlation coefficient is close to 1. of indep. var. False. In the presence of non-linear association the correlation coefficient varies and might be close to zero. (1 pt)
Problem 5: The following results show the relationship between education (years of schooling completed) and systolic blood pressure for men age 25-34:
average education \(\sim 13\) years, \(SD \sim 3\) years
average blood pressure \(\sim 119\) mm, \(SD \sim 11\) mm, \(r = -0.1\)

1. Using the normal approximation, about what percentage of men have blood pressure greater than 110?
   \[
   Z = \frac{110 - 119}{11} = -0.82 \\
   \text{Percentage is } \approx 60 + (100 - 60)/2 = 80\% 
   \]

2. What is the slope of the regression line which predicts blood pressure from education?
   \[
   \frac{\sqrt{\text{SD}_y}}{\text{SD}_x} = \text{slope} = -0.4 \times \frac{14}{13} \\
   = -0.085 
   \]

3. What is the R.M.S. error of the regression line?
   \[
   \text{RMS} = \sqrt{1 - (\rho)^2} \times 14 = \sqrt{0.99} \times 14 \\
   = 10.94 
   \]

4. What is the predicted blood pressure if the numbers of years of education is 15?
   \[
   (15 - 13)/3 = 2/3 \times (-0.1) \times 14 = -0.73 \\
   \text{Predicted blood pressure} = 119 - 0.73 = 118.27 
   \]

5. Of men with 13 years of education, about what percentage were in the top 10% of blood pressure?
   \[
   Z = \frac{(BP - 119)}{11} = 1.3 \\
   \text{Top 10%} \\
   \text{is equivalent to a BP above 133.3 mm} \\
   \]

- 13 years of education corresponds to 119 mm BP.

   \[
   (133.3 - 119)/10.94 = 1.307 \text{ S.D.} \\
   \text{Percentage is } (100 - 80.64)/2 = 9.68\% 
   \]
Problem 6: Calculate the coefficient of correlation for the data in the following tables

(1) \[ x \] | 0 0 1 1 1 1 3  \\ \[ y \] | 1 3 7 9 13 15  

(2) \[ x \] | 0 0 5 5 5 15  \\ \[ y \] | 1 3 7 9 13 15  

(3) \[ x \] | 0 0 5 5 5 15  \\ \[ y \] | -1 1 5 7 11 13  

(1) Average of \( x = 1 \)  
\( SD_x = 1 \)  
\( \mu_x = 8 \)  
\( SD_y = 5 \)  

\[
\begin{array}{c|cccccc}
\text{S} & -1 & -1 & 0 & 0 & 0 & 2 \\
\text{S} & -1.4 & -1 & -0.2 & 0.2 & 1.4 & 1.4 \\
\end{array}
\]

\[
\begin{align*}
\text{r} &= \frac{(-1)(-1.4) + (1)(-1) + 2 \times 1.4}{6} \\
&= 0.87
\end{align*}
\]

(2) \( x \) equal than before multiplied by 5  
\( r = 0.87 \)

(3) \( y \) equal than before minus 2  
\( r = 0.87 \)
Problem 7: A coin is tossed 100 times. Using the normal approximation:

(1) Estimate the chance of getting 60 heads. (2 pt)

(2) Estimate the chance of getting between 40 and 60 heads.

Box Model: 0 1

Average of Box = \( \frac{1}{2} \), SD of Box = \( \sqrt{\frac{1}{2} \times \frac{1}{2}} \)

Expected Value = \( 100 \times \frac{1}{2} = \frac{50}{1} \)

Standard Error = \( \sqrt{100 \times \frac{1}{2}} = \frac{5}{1} \)

\( Z \) : 60 is in the middle point of a rectangle with base 59.5 - 60.5. Convert base to SVs: \( \frac{59.5 - 50}{5} = 1.9 \), \( \frac{60.5 - 50}{5} = 2.1 \)

The area between 1.9 and 2.1 is \( 98.215 - 97.31 = 0.905\% \)

2) The chance is equal to the probability of getting between 40 & 60 heads. This is approximated by the area under the normal curve (39.5, 60.5).

In standard units this corresponds to the interval (-2.1, 2.1) which is 96.42%.

Problem 8 (Bonus): Fill in the blanks.

(a) The chance error is in the observed value. (Options: observed, expected).
(b) The confidence interval is for the population percentage. (Options: sample, population)

(2 pts)