CREDIT: The questions on this document were written by Erik Packard, PhD, Associate Professor of Mathematics at Colorado Mesa University.

1. Here are sample data about their hit from the first 34 games for CMU women's softball team this season (they are 34-0) for two players, Brooke Hodgson and Kaila Jacobi.

	Singles	Doubles	Triples	Homeruns
Brooke Hodgson	35	18	2	12
Kaila Jacobi	14	15	1	5

   A) (7 pts) Give the distribution in percentages of the different type of hits for Brooke Hodgson.

   B) (6 pts) Which player had a higher percent of their hits as doubles?

2. (14 pts) A candy comes in 4 colors, RED, GREEN, BLUE, YELLOW. The manufacturer claims that 30% are RED, 30% are GREEN, 20% are BLUE, and 20% are YELLOW. Can we conclude at the 5% level of significance the distribution is different than stated? DATA: A random sample of candies had 58 RED, 50 GREEN, 25 BLUE, and 17 YELLOW. Make sure to give the critical value(s) from the table and the test statistic from the data as well as the Yes or No answer.

3. (5 pts) Suppose somebody else collects data for #2. What is the chance they will conclude the distribution differs from what the manufacturer states when in fact the manufacturer is correct?

4. (5 pts) Suppose somebody else collects data for #2. What is the chance they will not conclude the distribution differs from what the manufacturer states when in fact the manufacturer is wrong?

5. (14 pts) Can we conclude at the 1% significance level there is any relationship between gender and favorite flavor of ice cream (chosen from vanilla, chocolate, and strawberry)? The table below gives results from a random sample. Make sure to give the critical value(s) from the table and the test statistic from the data as well as the Yes or No answer.

	Vanilla	**Chocolate**	**Strawberry**
Men	35	65	20
Women	70	30	80

6. (10 pts) Suppose someone else collects data for #5 and they get $(\chi^2)_{\text{Data}} = 5.991$. Give their *p*-value and its meaning in everyday terms.

7. (14 pts) Speeds of cars are measured along I-70 near Loma, Colorado. Assume the data is random from normal populations with equal variances. Can we conclude at the 5% significance level that there is any difference in the average speed of cars with license plates from Colorado, Utah, and California? Make sure to give the critical value(s) from the table and the test statistic from the data as well as the Yes or No answer.

Colorado	Utah	California
$n = 92$	$n = 21$	$n = 10$
$\bar{x} = 76$	$\bar{x} = 79$	$\bar{x} = 82$
$s = 3.4$	$s = 4.5$	$s = 6.1$

8. The number of wins college football teams (FW) and college basketball teams (BW) had last season is given for 8 teams picked at random.

	Wisconsin	**Texas**	**Missouri**	**Colorado State**
$x =$ FW	13	7	7	7
$y =$ BW	15	19	20	11

	Arkansas State	**Southern Miss.**	**Florida Int.**	**Idaho**
$x =$ FW	7	8	8	4
$y =$ BW	11	16	14	22

$$\sum x = 61 \qquad \sum y = 128 \qquad \sum x^2 = 509 \qquad \sum y^2 = 2164 \qquad \sum xy = 950$$

$$\sum x^2 - \frac{(\sum x)^2}{n} = 43.875 \qquad \sum y^2 - \frac{(\sum y)^2}{n} = 116 \qquad \sum xy - \frac{\sum x \sum y}{n} = -26$$

$$\sqrt{\frac{\sum y^2 - b\sum y - m\sum xy}{n-2}} = 4.095 \qquad \sqrt{1 + \frac{1}{8} + \frac{(1-7.625)^2}{\sum x^2 - \frac{(\sum x)^2}{n}}} = 1.458$$

$$\sqrt{\frac{1}{8} + \frac{(1-7.625)^2}{\sum x^2 - \frac{(\sum x)^2}{n}}} = 1.061$$

Please see the next page…

A) (4 pts) Give a scatterplot.

B) (5 pts) Find *r*, the linear correlation coefficient.

C) (6 pts) Find the line of best fit and graph it on the scatterplot, by plotting two points.

D) (4 pts) Use the line of best fit to estimate the BW (basketball wins) for a university that had 1 FW (football wins).

E) (3 pts) What do you think of the answer in part D), explain.

F) (4 pts) Interpret the slope in everyday terms.

G) (4 pts) What percent of the differences in BW can be explained by the regression line on FW?

H) (6 pts) Give a 95% CI for the BW for a university that had 1 FW.

I) (6 pts) Is there good evidence at the 10% significance level that ρ, the population linear correlation coefficient is not 0? Make sure to give the critical value(s) from the table and the test statistic from the data as well as the Yes or No answer.

The rest are worth 1 point each.

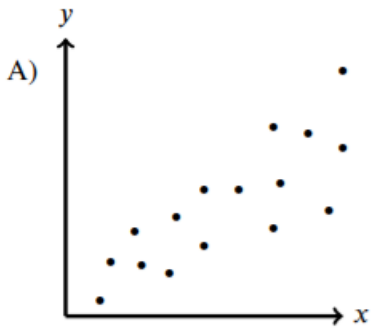
9. What does the least squares line minimize?

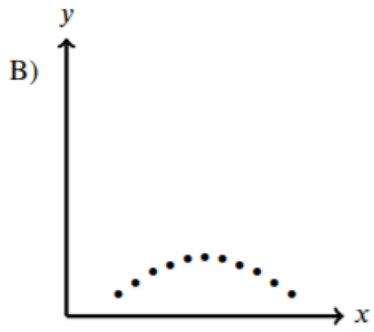
10. If there is a negative relationship, then *r* will be negative in part because bigger than average x's will correspond to smaller than average y's making $\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{s_y}\right)$ the product of a positive and a _____ which is negative.

11. Name three other relationships besides linear. Such as exponential, logistic, and _____.

12. Is *r* sensitive to outliers?

13. Which scatter plot shows a stronger relationship?





14. Give an example in which there is a strong association between x and y, but there is no cause and effect.

15. Do you think that people with an agenda will still try to show x affects y even if the setting is too complex with many variables interacting?

16. There is a strong correlation between education and wealth. Give a possible lurking variable that could explain this without having education have a cause and effect on wealth.

17. There is a strong correlation between education and wealth. Give a possible lurking variable that could explain this without having education have a cause and effect on wealth.

18. If a person is motivated, they are likely to become wealthy and become educated. Do you think that motivation explains all the association between and wealth, so in fact there is no cause and effect?

19. If a person is motivated, they are likely to become wealthy and educated. Do you think that motivation explains part of the association between education and wealth, so in fact the cause and effect still exist, but it's not as strong as many might think?

20. Give an example of Simpson's Paradox.

21. The 4 assumptions for ANOVA are normal populations, _____, independent samples and simple random samples (SRS's).

22. The 4 assumptions for ANOVA are normal populations, equal variances, independent samples and _____.

23. If you do a good job of collecting data from different sources, the data will vary for only two reasons, those are source and _____.

24. Variance due to factor is a weighted variance of the sample _____.

25. Variance due to error is a weighted mean of the sample _____.