

Instructions: *Include all relevant work to get full credit. Encircle your final answers.*

1. A market analyst wished to see whether consumers have any preference among five flavors of a new fruit soda. A sample of 120 people provide these data:

Cherry	Strawberry	Orange	Lime	Grape
32	36	20	18	14

- a. Using appropriate mathematical symbols, formulate the most appropriate null and alternative hypotheses to test if consumers have equal preference among the five flavors. [Hint: *The sum of the proportions should equal 1.*] [2]
- b. Using $\alpha = 0.01$, define your rejection rule. [2]
- c. Compute the observed value of the appropriate test statistic. [6]
- d. Do you reject the null hypothesis? Write a practical conclusion. [3]

2. A sociologist wishes to see whether the number of years of college a person has completed is related to her or his place of residence. A sample of 400 people is selected and classified as shown in the table below:

Location	No college	4-year degree	Advanced degree	Total
Urban	60	52	40	152
Suburban	32	??	??	140
Rural	24	??	??	108
Total	116	164	120	400

- a. In the context of the problem, formulate the appropriate null and alternative hypotheses. [2]
- b. Using $\alpha = 0.01$, define your rejection rule. [2]

- c. Compute the missing expected counts. Use 2 decimal places.

[4]

Location	No college	4-year degree	Advanced degree	Total
Urban	44.08	62.32	45.60	152
Suburban	40.60	_____	_____	140
Rural	31.32	_____	_____	108
Total	116	164	120	400

- d. The table below shows the contribution of each cell to the value of the test statistic. Fill up the missing contributions then compute the observed value of the test statistic.

[6]

Location	No college	4-year degree	Advanced degree
Urban	5.74	_____	_____
Suburban	_____	1.54	7.72
Rural	_____	8.78	4.74

- e. Do you reject the null hypothesis? Write a practical conclusion.

[3]

3. **Income and road rage.** Is a driver's propensity to engage in road rage related to his or her income? Researchers at Mississippi State University attempted to answer this question by conducting a survey of a representative sample of U.S. adult drivers. Based on how often each driver engaged in certain road rage behaviors (e.g., making obscene gestures, tailgating, and thinking about physically hurting another driver), a road rage score was assigned. (Higher scores indicate a greater pattern of road rage behavior.) The drivers were grouped according to their annual income: under \$30,000, between \$30,000 and \$60,000, and over \$60,000. The data are summarized in the table below.

Income Group	n_i	Average (\bar{x}_i)	Sample S.D. (s_i)
Under \$30,000	48	4.60	1.25
\$30,000 to \$60,000	45	5.08	1.32
Over \$60,000	30	5.32	1.42

- a. Identify the following:

i. Experimental units: _____

[1]

ii. Response Variable: _____

[1]

iii. Factor: _____

[1]

iv. Treatments: _____

[1]

- b. Using mathematical symbols, formulate the appropriate null and alternative hypotheses for this problem. Clearly define one of the parameters you used in the null hypothesis.

[3]

c. Construct the ANOVA Table.

ANOVA Table

[10]

Source	DF	Sum of Squares	Mean Square	F
Treatment				
Error				xxxxx
Total			xxxxx	xxxxx

d. Using $\alpha = 0.01$, define your rejection rule.

[2]

e. Do you reject the null hypothesis? Write a practical conclusion.

[3]

f. What are the assumptions for the ANOVA model?

[2]

i. The samples came from populations with equal standard deviations.

ii. _____

iii. _____

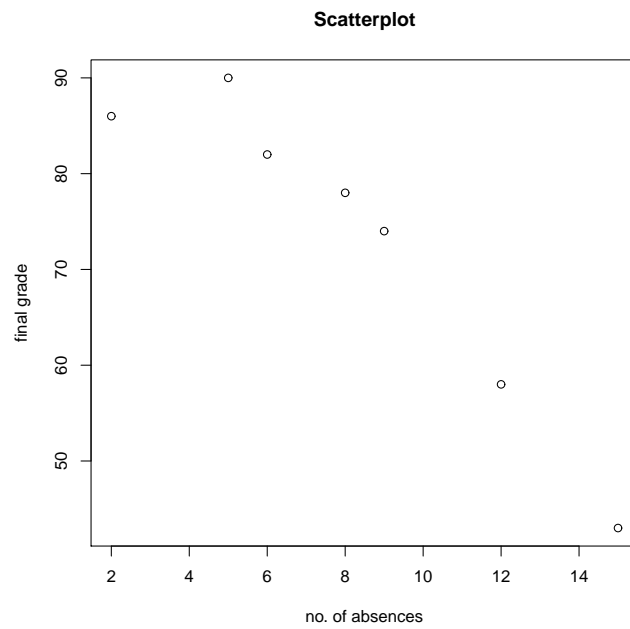
g. Based on our data, was it reasonable to assume equal standard deviations? Explain.

[2]

4. Data from a random sample of 7 students are used to examine the relation between the number of absences(x) and final grade (y) in an elementary statistics course.

The summary statistics for the available data are given below.

$$\begin{aligned}\sum x &= 57 & \sum x^2 &= 579 \\ \sum y &= 511 & \sum y^2 &= 38,993 \\ n &= 7 & \sum xy &= 3,745\end{aligned}$$



a. Compute the value of SS_{xx} , SS_{yy} , and SS_{xy} . [6]

b. Determine the correlation coefficient r and the coefficient of determination r^2 . Explain the meaning of r^2 in the context of this problem. [5]

c. Determine the regression line. [5]

d. On average, what happens to a student's final grade in an elementary statistics course if he/she misses an additional class? [2]

e. Calculate the value of $s = \hat{\sigma}$, the estimate of σ_ϵ . [3]

f. If the person with 2 absences had a final grade of 86 (see plot), determine the residual for this person. Use the $\hat{\sigma}$ that you obtained in the previous problem to check if this point is an outlier. [5]

g. Construct and interpret a 90% confidence interval for b . [7]

h. Using the confidence interval for b that you obtained in the previous problem, test $H_0 : b = 0$ vs. $H_1 : b \neq 0$. Explain how you arrived at your conclusion. What is the significance level of your test? [3]

5. In the simple linear model, $Y_i = a + bx_i + \epsilon_i$, where ϵ_i are independent and identically distributed with mean $E(\epsilon_i) = 0$ and variance $V(\epsilon_i) = \sigma^2$, show that $\hat{b} = \frac{SS_{xy}}{SS_{xx}}$ is an unbiased estimator of b . [Note: In this setup, x_i is considered constant and so $E(Y_i) = a + bx_i$.]

[10]