

INSTRUCTIONS: Read the questions carefully and completely. Answer each question and show work in the space provided. Partial credit will not be given if work is not shown. When asked to explain, describe, or comment, do so within the context of the problem. Be sure to include units when dealing with quantitative variables.

1. [18 pts] Short answer.

- a) [3] Statistics is about ... _____. (Fill in the blank with one word.)
- b) [3] A statistic is a summary of _____, while a parameter is a summary, or key number, of _____.
- c) [3] For a test of hypothesis, the P-value is 0.20. What should your decision be, and why?
- d) [3] For testing the hypothesis $H_0 : \mu_1 = \mu_2$ versus the alternative $H_0 : \mu_1 \neq \mu_2$, the value of the test statistic is 1.753 with $df = 15$. What is the P-value for the test?
- e) [3] Indicate whether each variable is categorical or quantitative. For quantitative variables, give the appropriate units.

Variable	Categorical	Quantitative (units)
Eye Color		
Speed of car		

- f) [3] The least squares regression equation relating the average outside temperature ($^{\circ}\text{C}$) and the amount of natural gas (ft^3) used in an uninsulated house is: $\text{predicted natural gas} = 6.854 - 0.393 \cdot \text{Temp}$. The linear relationship with average outside temperature explains 94.4% of the variation of in natural gas used. What is the value of the correlation coefficient relating average outside temperature and natural gas used?

2. [24 pts] Multiple Choice:

- a) ___ A 95% confidence interval for the mean distanced walked per day is from 2.5 miles to 4.3 miles. The sample mean distance walked per day is?
A: 2.5 miles B: 3.4 miles C: 3.0 miles D: cannot be determined
- b) ___ The correct interpretation of the interval in a) is?
A: I am 95% confident that the sample mean is in the interval.
B: 95% of the sample values are in the interval.
C: 95% of the population values are in the interval.
D: I am 95% confident that the population mean is in the interval.
- c) ___ The P-value is ...?
A: The probability that the null hypothesis is rejected.
B: The probability that the null hypothesis is true.
C: The probability of observing a value of the test statistic more extreme than the one observed when the null hypothesis is true.
D: The probability that the null hypothesis is not rejected.
- d) ___ Holding all other things constant, as the sample size decreases, the width of a confidence interval ...
A: Stays the same
B: Decreases.
C: Increases.
D: Cannot be determined.
- e) ___ Holding all other things constant, as the level of confidence decreases, the width of a confidence interval ...
A: Stays the same
B: Decreases.
C: Increases.
D: Cannot be determined.
- f) ___ You have calculated the correlation coefficient between two variables to be -0.95 . This would indicate?
A: A strong linear relationship.
B: A weak linear relationship.
C: No relationship.
D: No linear relationship.
- g) ___ The difference between an observed value of a response variable and the corresponding predicted value of the response is called?
A: An outlier.
B: An extrapolation.
C: An influential point.
D: A residual.
- h) ___ When you get two measurements on each individual in a single sample you have ...
A: Paired sample data.
B: Two independent sample data.
C: A single sample of categorical data.
D: Stratified random sample data.

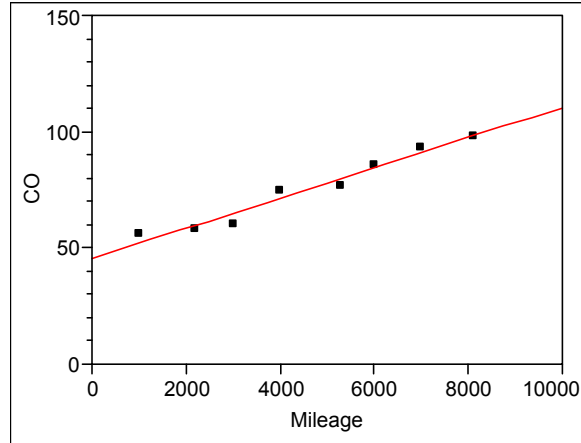
3. [12 pts] For each problem, determine whether the situation involves paired samples or two independent samples. Give the null and alternative hypotheses for the appropriate test. Be sure to define the parameter in the null hypothesis. **Do not attempt to perform the analysis.**
- a) [4] Airborne particles such as dust and some are an important part of air pollution. To evaluate particulate pollution, measurements were made every day for 40 days in the center of a small city and at a rural location about 10 miles west of the city. Because the winds usually blow from the west, we suspect that rural readings will be generally lower than the city readings.
- b) [4] In a randomized comparative experiment on the effect of dietary calcium on blood pressure, 54 healthy white males were divided at random into two groups of 27 each. One group received calcium; the other, a placebo. Blood pressures for all 54 men were taken before and after the study. The researchers believe the men in the calcium group will have a greater mean reduction in systolic blood pressure than the men taking the placebo.
- c) [4] In order to compare computer speeds, a set of benchmark programs are run and the CPU time required to complete each program is recorded. Six benchmark programs are selected. Each program is run on two computers and the CPU times are recorded. We wish to see if the two computers are the same or if there is a mean difference in times for the two computers.

4. [18 pts] Pollution from automobiles is a major concern, especially in high population density areas. The EPA conducted a study to see if there was a relationship between mileage on the car (miles) and carbon monoxide (CO) emissions (ppm). Below are the data. Also refer to the JMP output on the next page.

Mileage (miles)	1000	2200	3000	4000	5300	6000	7000	8100
CO (ppm)	56	58	60	75	77	86	93	98

- a) [2] What is the explanatory variable? What is the response variable?
- b) [3] Using the JMP output, give the least squares equation that relates CO to Mileage.
- c) [3] Give an interpretation of the intercept within the context of the problem.
- d) [4] Give an interpretation of the slope within the context of the problem.
- e) [3] Use the least squares regression equation to predict the CO and calculate the residual for the car with 6000 miles.
- f) [3] What does the plot of residuals versus mileage tell you about the fit of the linear model?

Least Squares Linear Regression Fit of CO on Mileage



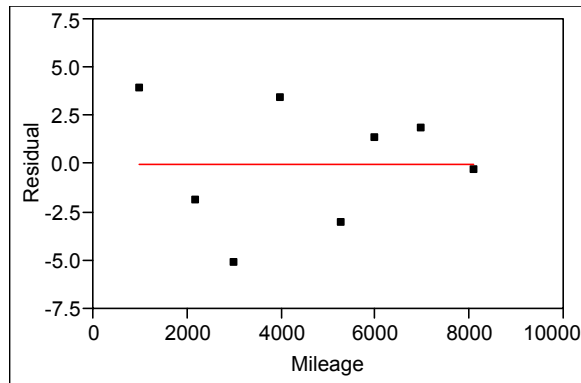
— Linear Fit

Summary of Fit

RSquare	0.961491
RSquare Adj	0.955073
Root Mean Square Error	3.447539
Mean of Response	75.375
Observations (or Sum Wgts)	8

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	45.620382	2.719459	16.78	<.0001
Mileage	0.0065037	0.000531	12.24	<.0001



5. [18] A business woman owns two fast food franchises in the same town. She wants to compare the average daily sales of the two. She chooses 5 days at random. On each day she records the daily sales (\$100) for each franchise. Below are the data.

Day	1	2	3	4	5	Mean	Std. Dev.
Franchise 1	1112	1212	1481	2017	2406	1645.6	551.4
Franchise 2	996	1238	1385	1867	2212	1539.6	492.4
Difference	116	-26	96	150	194	106.0	82.6

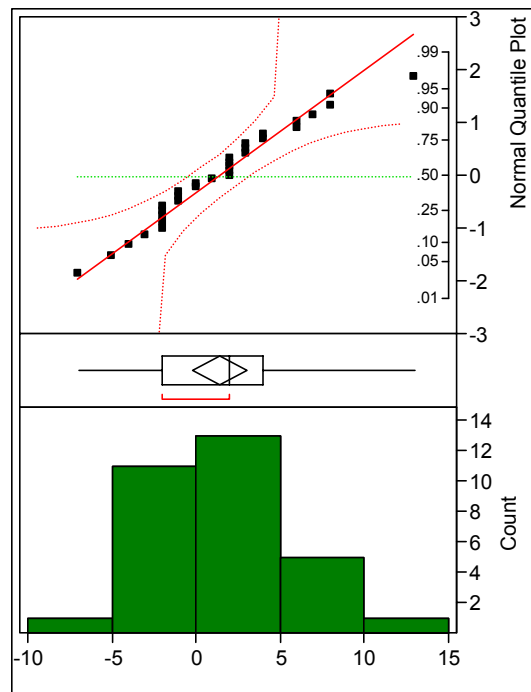
- a) [3] Explain why this is paired sample data.
- b) [8] Construct a 95% confidence interval on the mean difference in daily franchise sales. Note: the nearly normal condition is met, but you should check the other conditions.
- c) [4] Based on your confidence interval in b), do the two franchises differ in terms of population mean daily sales? Explain briefly.
- d) [3] Explain what 95% confidence is. In your explanation you may not use the words confidence or confident.

6. [22 pts] Does regular exercise lead to lower resting pulse rates? A random sample of 20 college students was selected. Each student was asked whether or not they exercised regularly (at least 30 minutes of aerobic exercise 3 times a week). The resting pulse rate for each student was also taken. Below are the summarized data.

	Number	Mean	Std. Dev.
No regular exercise	8	72.0	9.13
Regular exercise	12	65.0	7.01

- a) [4] Is this an observational study or an experiment? Explain briefly.
- b) [4] Why is this two independent samples data?
- c) [14] Determine whether the mean resting pulse rate for all college students who exercise regularly is lower than that of all college students who do not exercise regularly. Be sure to include the steps for a test of hypothesis. Note: you do not need to check conditions for this problem. Use $df = 16$.

Distribution of Difference = Helium Ball Distance – Air Ball Distance



Test Mean=value

		t Test	
Hypothesized Value	0	Test Statistic	1.8331
Actual Estimate	1.41935	Prob > t	0.0767
df	30	Prob > t	0.0384
Std Dev	4.31103	Prob < t	0.9616

Final Exams will not be graded until the end of finals week. Course grades will be posted outside my office door on Friday afternoon, December 16. I keep the final exam papers for one semester. If you wish to pick up your final exam you can do so during final exams Spring 2006.

Formulas

$$\bar{y} = \frac{\sum y}{n} \quad s_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n-1}} \quad \bar{x} = \frac{\sum x}{n} \quad s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}$$

$$r = \frac{\sum z_x z_y}{n-1} \quad z_x = \frac{x - \bar{x}}{s_x} \quad z_y = \frac{y - \bar{y}}{s_y}$$

$$b_1 = r \frac{s_y}{s_x} \quad b_0 = \bar{y} - b_1 \bar{x} \quad \hat{y} = b_0 + b_1 x \quad \text{residual} = y - \hat{y}$$

Sampling Distribution of \bar{y} :

Mean: μ

Standard Error: $SE(\bar{y}) = \frac{s}{\sqrt{n}}$

Confidence interval for μ :

$$\bar{y} - t^* \frac{s}{\sqrt{n}} \quad \text{to} \quad \bar{y} + t^* \frac{s}{\sqrt{n}}$$

Test Statistic:

$$t = \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}} \quad df = n - 1$$

Confidence interval for $\mu_1 - \mu_2$:

$$(\bar{y}_1 - \bar{y}_2) - t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{to} \quad (\bar{y}_1 - \bar{y}_2) + t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad t = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Test Statistic:

Confidence interval for μ_d :

$$\bar{d} - t^* \frac{s_d}{\sqrt{n_d}} \quad \text{to} \quad \bar{d} + t^* \frac{s_d}{\sqrt{n_d}}$$

Test Statistic:

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n_d}}} \quad df = n_d - 1$$