

Math 283 Practice Final

1. In a quality-control procedure, samples of parts from each of three plants were collected.

	Defective	Non-defective	Total
Plant 1	5		100
Plant 2	8		100
Plant 3	14		100
Total	27		300

- find the **expected** count for the number of defective parts from Plant 1.
- Set up null and alternative hypotheses to test whether all Plants have the same defective rate.
- Start the calculation of X^2 statistic from the upper-left cell (write only one term in the sum)
- The statistic X^2 was equal to 5.13. Make your conclusion based on this.

2. In a survey about tuition increase, 88 students out of 150 polled opposed the increase.

(a) Is there evidence that more than half of all students oppose the increase? Carry out all the steps of hypothesis test. Use significance level of 10%.

(b) How many students should be polled in order to achieve the margin of error no higher than $\pm 3\%$ (with confidence of 95%)?

3. Executives of an industrial plant want to determine which of two types of power – gas or electric – will produce more useful energy at the lower cost. "Energy cost" is calculated by taking the amount of money (in \$) invested, and dividing by the delivered amount of energy, in quadrillion BTU's. Random samples of 11 plants using electric utilities and 16 plants using gas utilities were taken. Electric averaged 85 \$/quad.BTU with the standard deviation of 45, and Gas averaged 106 \$/quad.BTU with the standard deviation of 56.

a) Find a 95% confidence interval (CI) for the difference between $\mu_1 =$ average Electric energy cost and $\mu_2 =$ average Gas energy cost.

b) Based on the interval, is there a significant difference between Electric and Gas energy cost?

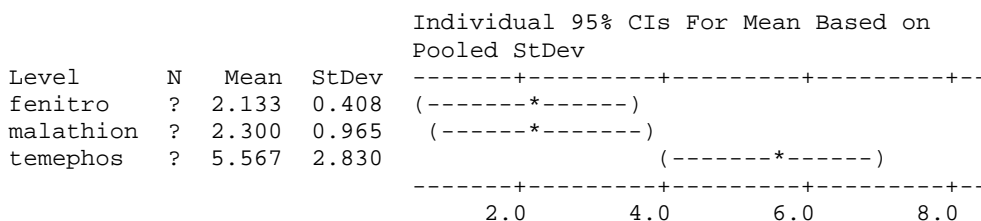
c) What assumptions are necessary for the above interval to work?

4. An article in the *Journal of the American Mosquito Control Association* [no kidding!] examined the resistance of mosquitoes against certain insecticides: temephos, malathion and fenitrothion. Each batch of mosquito larvae was exposed to one of three insecticides and the resistance ratio was computed.

One-way ANOVA: Resistance versus Insecticide

Analysis of Variance for Resistance

Source	DF	SS	MS	F	P
Insctde	?	44.97	?	?	?
Error	15	45.55	3.04		
Total	17	90.52			



- Fill in the ? marks in ANOVA table.
- Estimate the pooled standard deviation s_p
- How many mosquito batches were studied?
- Set up and test the null hypothesis.
- Compute a 95% confidence interval for the difference between Temephos and Malathion, based on pooled st. dev.
- Describe the Bonferroni procedure of simultaneously computing the CI's for differences between each pair of groups. At the confidence level of 94%, what is the t multiplier for the Bonferroni procedure?

5. Regression analysis was performed to see how taste of cheese depends on its chemical composition.

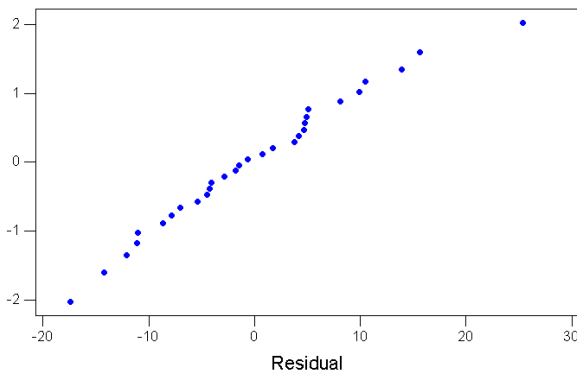
Regression Analysis: Taste versus Acetic, H2S, Lactic

The regression equation is
 Taste = - 28.9 + 0.33 Acetic + 3.91 H2S + 19.7 Lactic

Predictor	Coef	SE Coef	T	P
Constant	-28.88	19.74	-1.46	0.155
Acetic	0.328	4.460	0.07	0.942
H2S	3.912	1.248	3.13	?
Lactic	19.671	8.629	2.28	0.031

Source	DF	SS	MS	F	P
Regression	3	4994.5	1664.8	16.22	0.000
Residual Error	26	2668.4	102.6		
Total	29	7662.9			

- Estimate R^2 , the coefficient of determination.
- Which variables have an effect on taste of cheese? Explain.
- With what accuracy will one typically be able to predict the taste rating based on the above variables?
- Below is the normal probability plot of residuals. Are there serious concerns about the assumptions?



- Find a 90% confidence interval for the slope of variable Lactic.

6. A campus group is conducting a survey to determine the average percentage of fat in candy bars. Suppose that the actual average was 35% , with standard deviation $\pm 15\%$ hours. Assume Normal distribution.

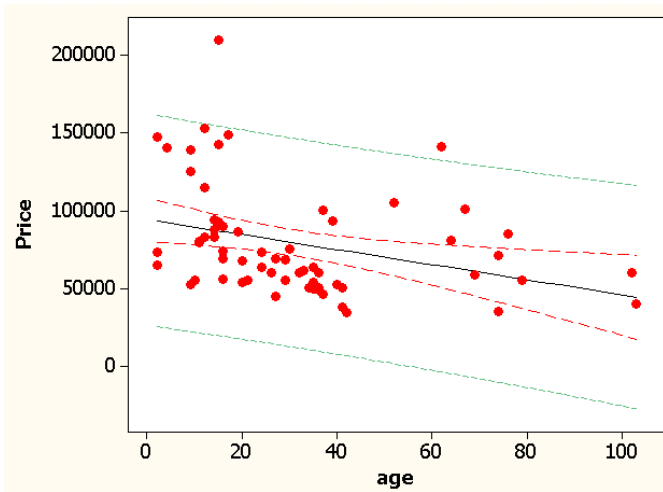
- What is the probability that any given candy bar will have more than 40% fat?
- What is the probability that the **average** fat content in a sample of 36 bars is above 40%?
- Give an interval symmetric about 35% that would contain about 68% of all the sample values.
- Compute the 40th percentile of fat content.

7. For a brand of diet cola, a blind experiment was conducted. Six respondents were given a regular and a diet cola in two unmarked containers, and were asked to rate the taste. The results are below.

Taster no.	1	2	3	4	5	6
Diet scores	7	5	6	1	2	6
Regular scores	5	9	8	5	6	5

- Test whether the diet and regular cola are rated differently by consumers.
- Find the 1st quartile of the Diet scores.

8. Regression



Questions are based on the following graph presenting the data on Price vs. Age of several houses.

- i. Estimate the price of an 80-year old house.
 - (a) 60,000
 - (b) 100,000
 - (c) 50,000
 - (d) 0
- ii. The estimated regression equation is
 - (a) $Y = 94,000 + 484 * X$
 - (b) $Y = 60,000 + 484 * X$
 - (c) $Y = 94,000 - 484 * X$
 - (d) $Y = 60,000 - 484 * X$
- iii. correlation coefficient between X and Y is
 - (a) 0.84
 - (b) 0.33
 - (c) -0.33
 - (d) -0.84
- iv. Visually, estimate the value of s (residual st.dev.)
- v. Why is the Confidence band wider on the right side of the plot? Explain.

9. Multiple choice - Circle the best answer.

- i. In a study of heart surgery, one issue was the effect of drugs called beta-blockers on the pulse rate of patients. The subjects were divided in three groups (beta-blocker A, beta-blocker B and placebo) and the pulse rates (beats per minute) were measured. What kind of statistical analysis would we perform to compare the effect of beta-blockers?
 - (a) two-sample t-test
 - (b) z-test for proportion
 - (c) ANOVA
 - (d) matched pairs t-test
- ii. To compare the gas prices on weekdays and weekends, prices for Tuesday and Friday were recorded for 10 randomly chosen weeks. What kind of statistical analysis would we perform to conclude if the Tuesday and Friday prices were different?
 - (a) two-sample t-test
 - (b) z-test for proportion
 - (c) ANOVA
 - (d) matched pairs t-test
- iii. Which of these measures is resistant to outliers?
 - (a) Mean
 - (b) Variance
 - (c) standard deviation
 - (d) median
 - (e) none of the above
- iv. For a sample of 36 bags of M&M's, a 95% confidence interval for μ = the average number of M&M's per bag was computed and equaled to [23.6; 24.8].
Circle all that apply:
 - (a) The sample mean was 24.0
 - (b) we are 95% sure that μ will belong to the above interval
 - (c) for 95% bags the number of candies in a bag will belong to the above interval
 - (d) the population mean μ cannot be above 25
 - (e) the C.I. was computed incorrectly because both numbers are fractions

10. True/False (no explanations needed, but they won't hurt).

- a) If a population of data values is symmetric, then any sample of data values from the population must be symmetric.
- b) As n increases the width of a confidence interval decreases.
- c) I construct a 95% confidence interval and get lower limit 10.25. If instead I had constructed a 99% confidence interval based on the same data then the lower limit would be higher than 10.25.
- d) When testing the hypothesis $H_0 : \mu = 5$, if we accepted H_0 , this would mean that μ is actually equal to 5.
- e) The X^2 test statistic is always nonnegative.
- f) The standard deviation and IQR are measures of spread.
- g) We should not use t-test when the sample is small and strongly skewed
- h) When finding a 95% C.I. based on a sample of size 8, my t^* equals 2.306
- i) If we rejected the null hypothesis at the level $\alpha = 0.10$ then we should necessarily reject it at the level $\alpha = 0.05$ also.
- j) We want to make inference on the average salary of male residents of U.S. An appropriate method of data collection would be to randomly sample 100 single and 100 married men.
- k) If we reject the null hypothesis in ANOVA then we can conclude that **all** of the group means are unequal.
- l) Correlation of 0.4 between two variables is always statistically significant.
- m) If two variables X and Y are significantly correlated then changes in X must lead to changes in Y.
- n) When finding an average sentence length, an appropriate method of data collection is to point a pen randomly on the page and pick the sentence that it hits.
- o) If $s_x = 0.5$, $s_y = 2$ and correlation $r = 0.8$, then the slope of regression line is 3.2

