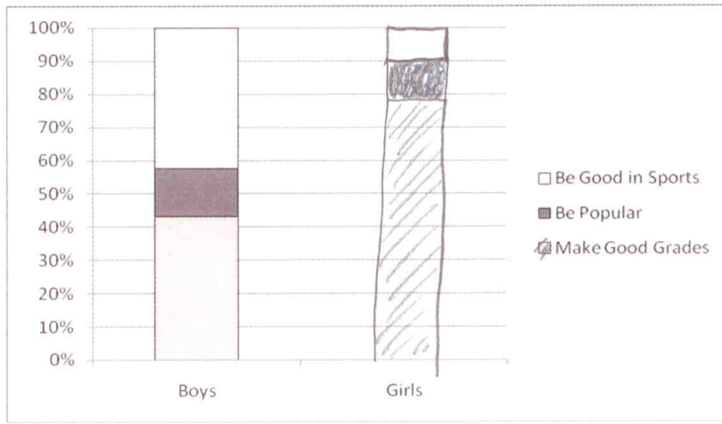
3/19/2009 MATH 271 Name: Solution
Dr. Lunsford Midterm Exam 100 Points Possible

Please show all of your work on this exam paper. To achieve the maximal amount of credit, please show all calculator input. Good luck!

Problem I. A study was performed to examine the personal goals of children in grades 4, 5, and 6. A random sample of students was selected from each of the grades 4, 5, and 6 from schools in Georgia. The students received a questionnaire regarding achieving personal goals. They were asked what they would most like to do at school: make good grades, be good at sports, or be popular. Results are presented by the gender of the child in the table below. The numbers in parentheses are the expected counts for a chi-square test of association based on these data. Please answer the following: (18 points total)

Gender			
Goal	Boys	Girls	Total
1-Make good grades	96 (144.19)	295 (246.81)	391
2-Be popular	32 (28.40)	45 (48.60)	77
3-Be good in sports	94 (49.42)	40 (84.58)	134
Total	222	380	602



(a) Above you are given a partially complete 100% stacked bar graph for these data. Complete the graph. (3 points)

(b) Explain (i.e. show the computation) how the expected count of 48.60 is obtained in the Girls/Be Popular cell. (3 points)

$$\frac{77(380)}{602} = 48.60$$

(c) Use your calculator to find the chi-square test statistic and p-value for these data. Clearly indicate your answers. (5 points)

$$\chi^2 = 89.97 \quad p = 2.9 \times 10^{-20}$$

(d) Based on your p-value in part (c), what is the conclusion of the test? Be sure to state the conclusion in context. (3 points)

Reject Null. There is an association between gender and what a student would most like to do at school for 4th, 5th, 6th graders in Georgia.

(e) If you have a significant result for these data can you reasonably conclude that gender determines what students would most like to do at school? Why or why not? (4 points)

No. Association does not imply causation. This is an observational study.

Problem II. You are designing a study to test the hypotheses

$$H_0 : \mu = 100 \text{ versus } H_a : \mu > 100$$

You plan to use a significance level of $\alpha = 0.05$. From past experience, you know that the population standard deviation is approximately $\sigma = 20$. Suppose there are $n = 15$ subjects for the study. Please answer the following questions. (20 points total)

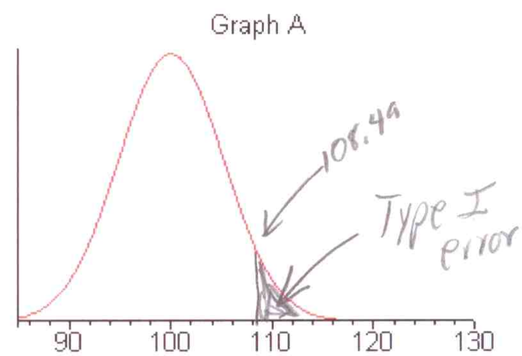
(a) What is the percent chance that you will commit a Type I error? (2 points) $\boxed{5\%}$

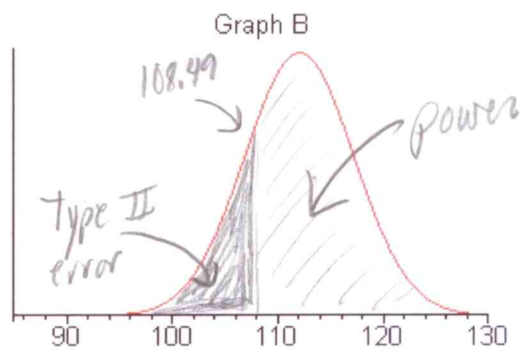
(b) What values of the sample mean will cause you to reject the null hypothesis? (4 points)

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{15}} = 5.16$

invNorm (.95, 100, 5.1639) = 108.49

$\boxed{\bar{x} \geq 108.49}$





(c) To your right you are given the graphs (on the same horizontal scale) of the sampling distributions of the sample mean if $\mu = 100$ (Graph A) and if $\mu = 112$ (Graph B). Shade the regions on the graphs that correspond to the probability of a Type I error, the probability of a Type II error, and the power of the test against the alternative $H_a : \mu = 112$. Be sure to clearly indicate which region is which. (6 points)

(d) Find the power of your test against the alternative $H_a : \mu = 112$? (4 points)

normalcdf (108.49, 1E99, 112, $\frac{20}{\sqrt{15}}$) = $\boxed{.7516}$

(e) Determine whether each of the following statements regarding increasing the value of the power is true or false. (1 point each, 4 total)

F The power will increase when we decrease the significance level of the test.

T The power will increase when we increase the sample size.

T The power will increase when we consider an alternative value that is farther from the null value (the value of the parameter of interest under the null hypothesis).

T The power would increase if we could reduce the assumed value of the standard deviation.

Problem III. The national distribution of scores on the Advanced Placement (AP) examination in Statistics (out of a possible 5 points) is provided in the following table:

Score	1 or 2	3	4	5
Percentage	40%	25%	23%	12%

You would like to compare the distribution of scores for students from Virginia to the national average. A simple random sample of 100 Virginia high school students who took the AP exam in Statistics is taken. The observed counts based on this sample of students are provided in the following table:

Score	1 or 2	3	4	5
Count	18	19	37	26

Which statistical test (that we have studied) would you use to make this comparison? Clearly state the null and alternative hypotheses for this test and identify the parameters in the context of this problem. (8 points) χ^2 goodness of fit test.

$H_0: p_{1 or 2} = 0.40,\ p_3 = 0.25,\ p_4 = 0.23,\ p_5 = 0.12$

H_a: At least one is different

Where p_i is the proportion of all Virginia students whose score on the AP exam in stats is in group i (i = 1or2, 3, 4, 5)

Problem IV. It is widely believed that a person's level of fitness prior to undergoing corrective knee surgery is an important factor in the speed of rehabilitation following surgery. A study was undertaken with 24 randomly selected male subjects who were to undergo such surgery. The physical fitness status of each subject before the surgery was categorized as being: below average (n = 8), average (n = 10), or above average (n = 6). Following surgery the number of days spent in physical therapy by each patient until fully rehabilitated was determined. Please answer the following questions: (16 points total)

(a) Please state all assumptions that need to be satisfied in order to conduct an ANOVA test with these data. Be sure to state these assumptions in the context of the problem. (6 points)

Population: All males who would undergo knee surgery

Variable: Number of days spent in physical therapy by patient until fully rehabilitated.

1. SRS from the population.
2. The variable is normally distributed for each fitness level population.
3. The standard deviation of the variable is the same for each fitness level population.

(b) Below you are given the partially completed ANOVA table for these data. Please complete the table. (10 points)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	P-Value
Status	2	672.0	336.0	16.96	0.000041
Error	21	416	19.81		
Total	23	1088.0			

Problem V. A study was conducted to compare five different training programs for improving endurance. Forty subjects were randomly divided into five groups of eight subjects in each group. A different training program was assigned to each group. After two months, the improvement in endurance was recorded for each subject. A one-way ANOVA was used to compare the five training programs, and the resulting *F* statistic was 3.69. Please answer the following questions: (18 points total)

(a) State the null and alternative hypotheses. Be sure to clearly indicate the parameters in the context of this problem. (6 points)

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

H_a: At least one is different

where μ_i is the mean improvement in endurance for training program i, $i = 1, \ldots, 5$.

(b) Use your calculator to find the p-value for the test. Be sure to show all calculator input. (4 points)

Fcdf(3.69, 1E99, 4, 35) = 0.0131

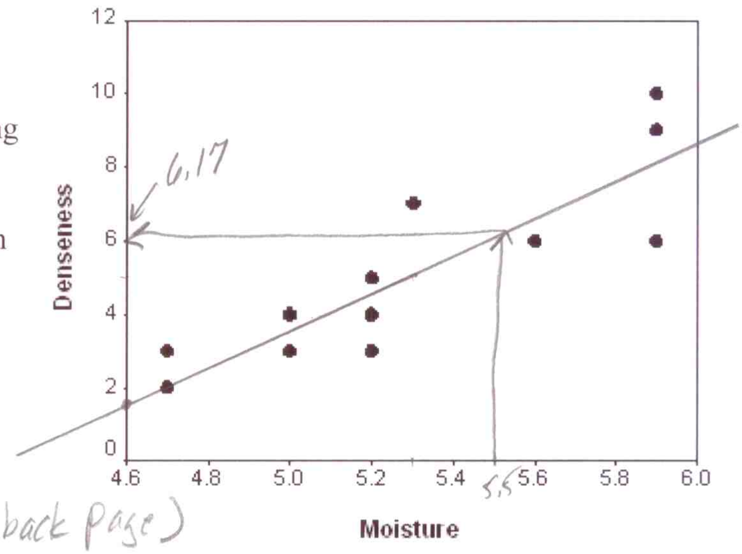
(c) At a significance level of 0.05, what is the appropriate conclusion of the ANOVA test? Be sure to state your conclusion in the context of this problem. (4 points)

Reject null. At least one of the training programs has a mean improvement in endurance that is different from the others.

(d) Regardless of your answer in part (c), if you obtained a significant result via the ANOVA, can you reasonably claim that the training programs caused differences in the mean amount of improvement? Why or why not? (4 points)

Yes. This was an experimental study, i.e. subjects were randomly assigned to the 5 training programs, thus eliminating confounding variables.

Problem VI. The moisture in the wet mix of cement is considered to have an effect on the denseness of the finished product. The moisture of the mix was controlled at various levels, and the denseness of the finished product was measured. The data were entered into SPSS. Below you are given a scatterplot of the data (including a graph of the least squares regression line) and on the next page you are given SPSS output from a regression analysis. Please answer the following questions. (20 points total)



(a) What is the equation of the least squares regression line? If you use variable names such as *x* and *y* please indicate which is moisture and which is denseness. (4 points)

Denseness = 5 * moisture − 21.33

from coeff table (see back page)

(b) Use the regression line to predict the denseness for a moisture level of 5.5. Show this prediction on the graph above. Does this mean every batch of concrete with a moisture level of 5.5 will have this predicted denseness? Explain. (5 points)

Denseness = 5(5.5) − 21.33 = 6.17

No. The regression line predicts the expected (or average) denseness for a moisture level of 5.5. Individual levels will vary about the line.

(c) What is the value of the correlation between moisture and denseness? (2 points)

$r = 0.868$ ← from Model Summary Table

(d) Do you think this linear model is a good predictive model of denseness given moisture? Why or why not? Hint: You should appeal to the coefficient of variation to answer this question. (4 points)

Yes. $R^2 = 0.75$ thus 75% of the variation in denseness is explained by the linear regression on moisture.

(e) Find the residual for the data point (5.6,6). (3 points)

e_1 = observed − expected = $6 - (5 * (5.6) - 21.33) = \boxed{-0.67}$

(f) What is an estimate of the variance of the residuals? (2 points)

$(1.310)^2 = 1.7161$

Descriptive Statistics

	Mean	Std. Deviation	N
DENSENESS	5.17	2.52	12
MOISTURE	5.30	.44	12

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.868	.754	.729	1.310

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	52.500	1	52.500	30.583	.00025
	Residual	17.167	10	1.717		
	Total	69.667	11			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-21.33	4.81		-4.44	.00126
	MOISTURE	5.00	.90	.868	5.53	.00025