Pledge:

4/14/2005 MATH 271 Name: Solution
Dr. Lunsford Test 2 100 Points Possible

Problem I. A medical researcher wishes to see whether the pulse rates (in bpm) of smokers are higher than the pulse rates of nonsmokers, on average. Random samples of 100 smokers and 100 nonsmokers are selected and their pulse rates are measured. The summary of the data from these samples is given. Conduct the appropriate hypothesis test to determine if the researcher can conclude, at an $\alpha = 0.05$ level of significance, that smokers have higher pulse rates on average than nonsmokers? For your convenience (and mine) I have organized where to put your answers to this problem below. (18 points total)

Smokers	Nonsmokers
$\overline{X}_1 = 90$	$\overline{X}_2 = 88$
$s_1 = 5$	$s_2 = 6$
$n_1 = 100$	$n_2 = 100$

- Hypotheses (Clearly indicate the meaning of any symbols/variables you use): (6 points)

$H_0: \mu_1 = \mu_2$ $\mu_1 \equiv$ mean pulse rate of smokers
$H_1: \mu_1 > \mu_2$ $\mu_2 \equiv$ mean pulse rate of nonsmokers

- Test statistic and p-value of the test (Please indicate what statistical test you are using, and what function, including input, you use on your calculator to obtain the test statistic and p-value). (6 points)

Two Sample t-test w/ pooled variances. Input above stats to get $t = 2.5607$, $p = .005607$

- Conclusion (in terms of the hypotheses above and in terms of the context of this problem): (6 points)

Since the p-value $= .005 < \alpha = .05 \Rightarrow$ reject H_0 in favor of H_1. The data clearly support the claim that smokers have a higher pulse rate than nonsmokers, on average.

Problem II. To determine whether there is a relationship between the gender of an individual and the amount of alcohol consumed, the following data was collected from a random sample of 68 people: (16 points total)

(a) Supposing that the amount of alcohol a person consumes is independent of the individual's gender, complete the contingency table to your right by placing the expected number of observations (rounded to two decimal places) in the space below the observed number. (6 points)

	Alcohol Consumption			
Gender	Low	Moderate	High	Total
Male	10	9	8	27
	9.13	9.93	7.94	
Female	13	16	12	41
	13.87	15.07	12.06	
Total	23	25	20	68

[A] observed values

[B] output of χ^2 test - expected values.

(b) Conduct the appropriate hypothesis test to determine if the amount of alcohol a person consumes is independent of the individual's gender. Clearly indicate the hypotheses, what test you are using, the p-value of the test, and your conclusion(s) in the context of this problem. (10 points)

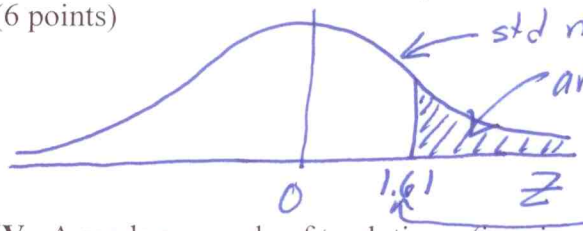
H_0: Amount of alcohol consumption independent of gender.
H_1: Amount of alcohol consumption not independent of gender.
χ^2 test for independence, $p = .868986 \Rightarrow$ fail to reject H_0.
$\Rightarrow$ Alcohol consump. indep. of gender.

Problem III. Let μ be the average amount of time it takes (in milliseconds) for a Klingon ship to reach warp speed. It is a well known fact that the average amount of time that a Starfleet Constitution-class ship takes to reach warp speed is 10 milliseconds. The Starfleet command has obtained warp-radar readings from 100 Klingon ships and wants to test the following the hypothesis $H_0 : \mu = 10$ against $H_1 : \mu > 10$. The data yields a z-score of 1.61 and a p-value of 0.1314. Assume this hypothesis test is taking place in the 23rd century (and thus Klingons are still Starfleet enemies) and that (amazingly!) the methods of hypothesis testing haven't changed all that much. Please answer the following: (13 points total)

Note: Other items to think about: Units, etc...

(a) If reaching warp speed faster gives one an advantage over their opponent, should the Starfleet commander be concerned? Justify your answer using the information above (hypotheses, p-value, etc.). Feel free to add any disclaimers or additional information you think might be helpful to know....(7 points) Performing a "straight up" hypothesis test @ the $\alpha = .10$ level, we would fail to reject H_0 in favor of the alternative since the p-value $> .10$. If I was Cpt. Kirk I would be concerned since this implies he will not have an advantage over Klingons in getting to warp speed faster (on average)!

(b) Draw a graphical representation of the p-value for this test. Clearly label everything relevant. (6 points)



Problem IV. A random sample of track times (in minutes) from three different kinds of compact disks is listed. We want to determine if there a significant difference in the average track time for the three different types. (12 points total)

Data in: L_1, L_2, L_3

JAZZ	CLASSICAL	POP
4.83	4.03	5.95
5.37	5.37	3.67
6.08	4.22	3.90
5.05	3.95	4.17
3.65	3.45	4.53
4.02	6.38	3.25
2.65		

(a) What statistical test will you use? What are the null and alternative hypotheses for this test? Clearly indicate what all parameters/variable names you use represent. (6 points) Will use One-Way ANOVA

$H_0 : \mu_J = \mu_C = \mu_P$

H_1: At least one of the means is not equal to the others.

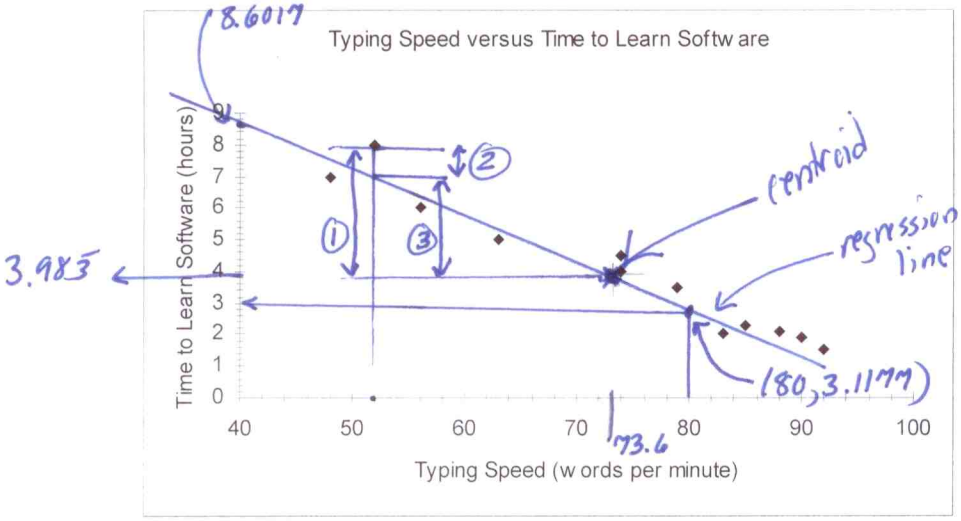
where μ_J, μ_C, and μ_P are the mean track times for Jazz, Classical, & Pop CDs respectively.

(b) What is the p-value of this test? Using this p-value, state the results of the test in terms of the hypotheses above and in the context of the problem. Clearly indicate what calculator function you use including input. (6 points)

ANOVA (L_1, L_2, L_3) gives $p = .853$ $\therefore$ we fail to reject the null hypothesis. Thus the data indicates that there is no significant difference in the mean track times for jazz, classical, and pop CDs.

Problem V. The designers of a new word processing program want to know whether the time (in hours) that it takes a new user of their software to learn the program is related to the typing speed of the user (in words per minute). The data are shown below along with a scatterplot. Please answer the following questions: (41 points total)

SPEED(X)	TIME(Y)
48	7
74	4
52	8
79	3.5
83	2
56	6
85	2.3
63	5
88	2.1
74	4.5
90	1.9
92	1.5



(a) Find the centroid of the data and plot it on the scatterplot above. Clearly indicate your answer below. (5 points)

$(\bar{x} = 73.\bar{6}, \bar{y} = 3.98\bar{3})$

(b) Compute the value of the correlation coefficient and the coefficient of determination of the data. Clearly indicate your answers. (6 points)

$r = -.9742$ → $r^2 = .9498$

(c) Conduct the appropriate hypothesis test to determine if there is significant negative correlation between the two variables. Clearly indicate your hypotheses, the value of the test statistic, the p-value of the test, and your conclusion. (10 points)

$H_0: \rho = 0$
$H_1: \rho < 0$
$t = -13.6389$ → p-value $= 4.345 \times 10^{-8}$ ∴ Reject Null in favor of alternative. There is significant negative correlation between the two variables.

(d) Find and accurately plot the equation of the least squares regression line. Clearly indicate your answer. Note that the vertical axis is at $x = 40$. (6 points)

$y = a + bx$ where $a = 14.085652\overline{17}$ and $b = -0.13\overline{71}$
@ $x = 40$ we have $y = 8.6017$ → see graph

(e) Use the data point $(x, y) = (52, 8)$ to compute and graphically show the following values (Note that y' is obtained via the regression line). Identify each difference below as to whether it contributes to the total variation, the explained variation (due to regression), and/or the residual (or unexplained) variation. (6 points)

$y - \bar{y} = 4.01\overline{6}$ total ①

$y - y' = 1.044$ unexplained (residual) ②

$y' - \bar{y} = 2.973$ explained (regression) ③

Problem V., continued

(f) For users who have a typing speed of 80 words per minute, how long will it take them, on average, to learn this software? Show this on your graph above. (4 points)

→ see graph

$y = 14.0857 - .1371(80) = 3.1177$ hours

(g) Below is the ANOVA portion of the Excel output from the regression analysis for this data. Show how you can compute the proportion of the total variation that is explained via the regression equation using this output. (4 points)

ANOVA

	df	SS	MS	F	Significance F
Regression	1	49.02139	49.02139	186.0197	8.69E-08
Residual	10	2.635279	0.263528		
Total	11	51.65667			

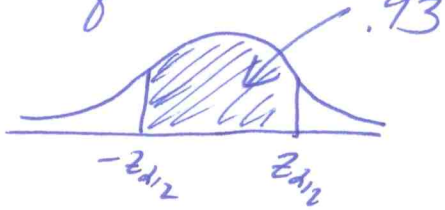
$\frac{49.02139}{51.65667} = .94898471 = r^2$ ← ★!

Note!

Bonus The Alliance for Hunger wants to estimate the proportion of Americans that believe the United States does not spend enough money on fighting hunger at home. They would like to have a maximum error of ±2.3% with 93% confidence. Assuming they are able to get a good random sample, what is the smallest sample size they could use to achieve these results? You must show all work (what formula you use and what values you plug in) to get credit for this problem. (5 points)

Use $n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$

We don't have an estimate for $\hat{p}$ & $\hat{q}$ so use $\frac{1}{2}$ for both.



$z_{\alpha/2} = \text{invNorm}(.965) = 1.81191$

(or use $-\text{invNorm}(.035)$)

$\therefore n = \frac{1}{4}\left(\frac{1.81191}{.023}\right)^2 = 1551.52$

$\Rightarrow$ **$n = 1552$**