Med Info 409 Fall, 2011

Final Exam Part 2

Instructions for Final Exam:

1. Please enter your last name and first name in the header above.
2. Save your document using the following naming convention: lastname_firstname_medinfo409_finalexam.doc (or .docx)
3. For SPSS output, please paste the appropriate tables in the Word document and make reference to each table and why you are including it. If a table is associated with one of the hypothesis testing steps, it must be clear which step and why you are including it there. Please also discuss and interpret as appropriate.
4. The Entire exam is worth 50 points.
5. **Please return the exam via Blackboard no later than 11:55 PM 12/4/11.**

Part 2.Each problem is worth 5 points. Paste SPSS tables in each problem where appropriate. (Please paste the table within the problem rather than append at the end of this document).

1. A researcher sampled 100 women from a medical record database and identified their smoking status and parity. Use the 5 steps of hypothesis testing to determine whether pregnancy and smoking status are independent.

Smoking Status	Parity	P
Before Pregnancy	No children	Children
Nonsmoker	5	55
Smoker	20	20

Step 1: State hypotheses.

H_0: Smoking status and parity are independent.

H_1: Smoking status and parity are not independent

$\alpha = 0.05$

Step 2: Both variables are categorical. Use χ^2 test of independence.

Step 3: Degrees of freedom = rows-1 * columns-1 = 1*1 = 1

Critical $\chi^2 = 3.84$ (Table B.5 with df = 1 and $\alpha = 0.05$)

Reject H_0 if $\chi^2 \geq 3.84$

Step 4: Calculate χ^2.

	Observed			
Prepregnancy smoking status		Parity		
	No Children	Children	Total	
Nonsmoker	5	55	60	
Smoker	20	20	40	
Total	25	75	100	
	Expected			
Prepregnancy smoking status		Parity		
	No Children	Children	Total	
Nonsmoker	15	45	60	
Smoker	10	30	40	
Total	25	75	100	
	Calculate χ^2			
	Non/No	Non/Yes	Smoker/No	Smoker/Yes
Observed	5	55	20	20
Expected	15	45	10	30
(O-E)	-10	10	10	-10
$(O-E)^2$	100	100	100	100
$(O-E)^2/E$	6.67	2.22	10	3.33
	$\chi^2 = 22.22$			

Step 5: Reject H_0 since 22.22 > 3.84. There is significant evidence, $\alpha = 0.05$, to show that smoking status and parity are not independent. The value 22.22 is associated with a p value smaller than 0.005, according to the χ^2 distribution table.

2. A medical student surveys adults to determine if gender and seat belt use are independent. Use the 5 steps of hypothesis testing to determine whether gender and seat belt use are independent.

Seat belt use	Gender M	F
Belt	60	70
No belt	40	30

Step 1: State hypotheses.

H_0: Gender and seat belt use are independent.

H_1: Gender and seat belt use are not independent.

$\alpha = 0.05$

Step 2: Both variables are categorical. Use χ^2 test of independence.

Step 3: Degrees of freedom = rows-1 * columns-1 = 1*1 = 1.

Critical χ^2 = 3.84 (Table B.5 with df = 1 and α = 0.05)

Reject H_0 if $\chi^2 \geq 3.84$.

Step 4: Calculate χ^2

	Observed			
Seat belt use		Gender		
	Male	Female	Total	
Belt	60	70	130	
No belt	40	30	70	
Total	100	100	200	
	Expected			
Seat belt use		Gender		
	Male	Female	Total	
Belt	65	65	130	
No belt	35	35	70	
Total	100	100	200	
	Calculate χ^2			
	M/ Belt	M/No belt	F/Belt	F/No belt
Observed	60	40	70	30
Expected	65	35	65	35
(O-E)	-5	5	5	-5
$(O-E)^2$	25	25	25	25
$(O-E)^2/E$	0.38	0.71	0.38	0.71
	χ^2 = 2.18			

Step 5: Do not reject H_0 since 2.18 < 3.84. We do not have significant evidence, α = 0.05, that there is an association between gender and seat belt use.

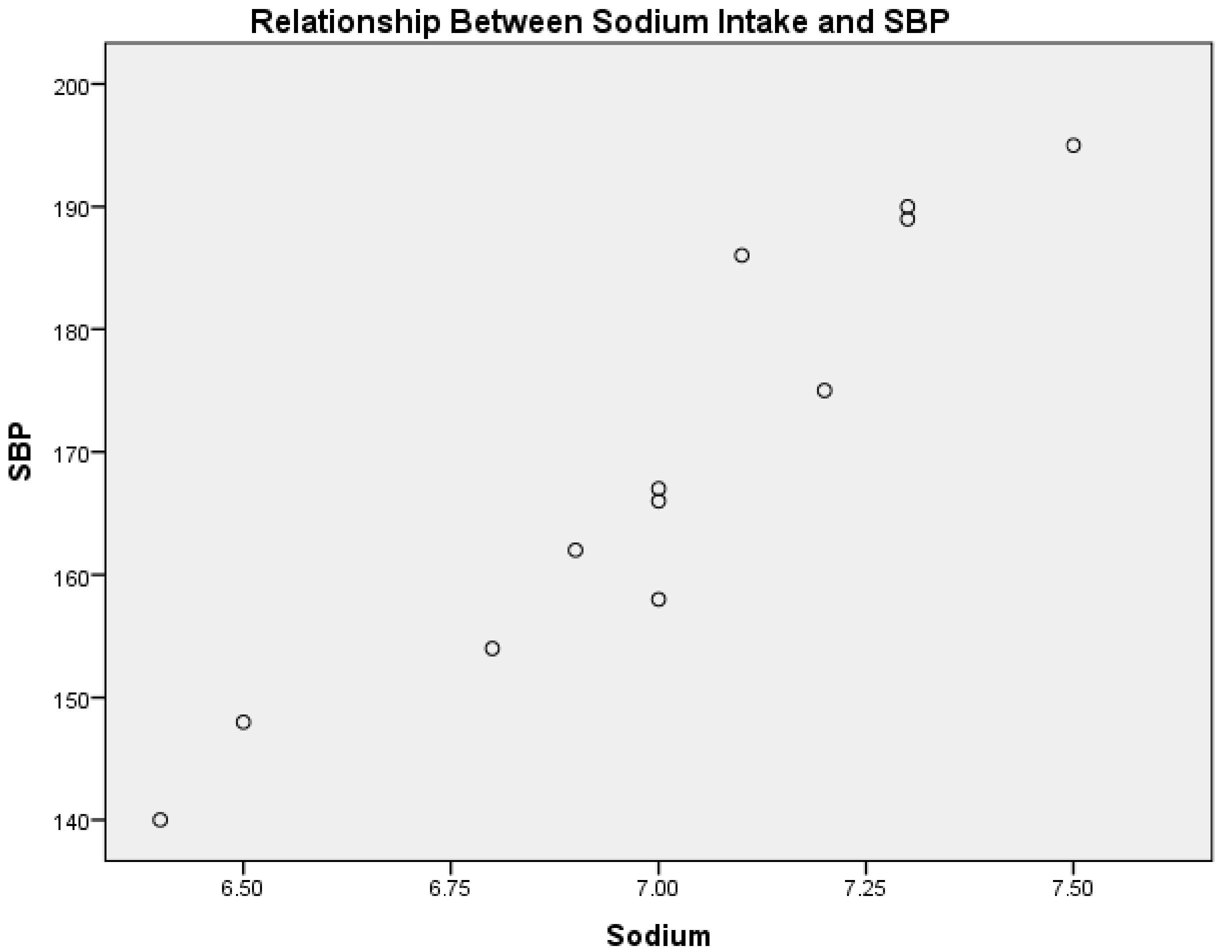
3. A researcher collected daily sodium intake and systolic blood pressure readings on 12 patients. Obtain the regression equation from SPSS and include the table you used for the data. Use the 5 step hypothesis testing approach to determine if the regression equation is significant at alpha = 0.05. What would be a likely blood pressure for a person with a sodium intake of 6.3: 7.6?

Sodium	SBP
6.8	154
7	167
6.9	162
7.2	175
7.3	190
7	158
7	166
7.5	195
7.3	189
7.1	186
6.5	148
6.4	140

Input Data Table:

Subject_ID	Sodium	SBP
1	6.8	154
2	7	167
3	6.9	162
4	7.2	175
5	7.3	190
6	7	158
7	7	166
8	7.5	195
9	7.3	189
10	7.1	186
11	6.5	148
12	6.4	140

Scatter Diagram:



Correlations:

Step 1: State Hypotheses.

H_0: ρ = 0

H_1: ρ ≠ 0, α = 0.05

Step 2: Two continuous variables; Pearson r correlation coefficient.

Step 3: Two sided test, use p value from SPSS at α = 0.05

Step 4: Test statistic from SPSS

Correlations

		Sodium	SBP
Sodium	Pearson Correlation	1	.942**
	Sig. (2-tailed)		.000
	N	12	12
SBP	Pearson Correlation	.942**	1
	Sig. (2-tailed)	.000	
	N	12	12

**. Correlation is significant at the 0.01 level (2-tailed).

$r = 0.942$ $p<0.01$

Step 5: Reject H_0 since $p < 0.01$. There is a strong positive correlation between sodium intake and systolic blood pressure.

Linear Regression:

Step 1: State Hypotheses.

H_0: $\rho = 0$

H_1: $\rho \neq 0$, $\alpha = 0.05$

Step 2: Two continuous variables; Pearson linear regression analysis

Step 3: Two sided test, use p value from SPSS at $\alpha = 0.05$

Step 4: Test statistic from SPSS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.942[a]	.888	.877	6.283

a. Predictors: (Constant), Sodium

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3136.877	1	3136.877	79.457	.000[a]
	Residual	394.789	10	39.479		
	Total	3531.667	11			

a. Predictors: (Constant), Sodium

b. Dependent Variable: SBP

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-198.026	41.233		-4.803	.001
	Sodium	52.456	5.885	.942	8.914	.000

a. Dependent Variable: SBP

Step 5: Reject H_0 since $p < 0.001$. There is a strong linear association between sodium intake and systolic blood pressure. R = 0.942. $R^2 = 0.888$, so 89% of the variation in systolic blood pressure is explained by sodium intake.

Regression Equation:
SBP = -198.026 + 52.456*sodium
For a sodium intake of 6.3, the predicted SBP is 132.447 rounds to 132.
For a sodium intake of 7.6, the predicted SBP is 200.640 rounds to 201.

24. Two health inspectors rate 11 hospitals on cleanliness. Determine whether their rankings are comparable, using the appropriate nonparametric test and the 5 steps of hypothesis testing.

Inspector 1	Inspector 2
2	1
3	3
2	3
3	2
1	2
4	5
5	4
3	2
1	1
3	4
4	3

Step 1: State hypotheses.

H_0: There is no difference in ratings between the two inspectors.

H_1: There is a difference in ratings between the two inspectors.

$\alpha = 0.05$

Step 2: There are two independent groups, each with small sample size. Use Wilcoxon Rank Sum test (Mann-Whitney U test).
Step 3: Reject H_0 if $p < 0.05$ using SPSS output.
Step 4:

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Rating is the same across categories of Inspector_Number.	Independent-Samples Mann-Whitney U Test	.847[1]	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

[1] Exact significance is displayed for this test.

Descriptives

	Inspector_Number			Statistic	Std. Error
Rating	1	Mean		2.82	.377
		95% Confidence Interval for Mean	Lower Bound	1.98	
			Upper Bound	3.66	
		5% Trimmed Mean		2.80	
		Median		3.00	
		Variance		1.564	
		Std. Deviation		1.250	
		Minimum		1	
		Maximum		5	
		Range		4	
		Interquartile Range		2	
		Skewness		.037	.661
		Kurtosis		-.468	1.279

2	Mean		2.73	.384
	95% Confidence Interval for Mean	Lower Bound	1.87	
		Upper Bound	3.58	
	5% Trimmed Mean		2.70	
	Median		3.00	
	Variance		1.618	
	Std. Deviation		1.272	
	Minimum		1	
	Maximum		5	
	Range		4	
	Interquartile Range		2	
	Skewness		.265	.661
	Kurtosis		-.625	1.279

Step 5: Do not reject the null hypothesis since $p = 0.847$. We do not have significant evidence, $\alpha = 0.05$, that there is a difference between the ratings of inspector 1 and the ratings of inspector 2. Therefore the ratings of the two inspectors are comparable.

25. A medical informatics manager reviews sick day use in her department. She identifies that employees in her small department used the following amount of sick time in the past year:

Men: 5,10,2,0,6,4,5,15
Women: 8, 9,3,5,0,4,15

The statistician indicated she should use a nonparametric test to determine if there is a difference in sick time use between the men and women in her department. Use the 5-step approach to hypothesis testing and alpha=0.05 to determine if there is a difference.

Step 1: State hypotheses.

H_0: There is no difference in sick time use between men and women

H_1: There is a difference in sick time use between men and women

$\alpha = 0.05$

Step 2: There are two independent groups, each with small sample size. Use Wilcoxon Rank Sum test (Mann-Whitney U test).

Step 3: Reject H_0 if $p < 0.05$ using SPSS output.

Step 4:

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Sick_Days is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	1.000[1]	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

[1] Exact significance is displayed for this test.

Descriptives

	Gender			Statistic	Std. Error
Sick_Days	Men	Mean		5.88	1.663
		95% Confidence Interval for Mean	Lower Bound	1.94	
			Upper Bound	9.81	
		5% Trimmed Mean		5.69	
		Median		5.00	
		Variance		22.125	
		Std. Deviation		4.704	
		Minimum		0	
		Maximum		15	
		Range		15	
		Interquartile Range		7	
		Skewness		1.027	.752
		Kurtosis		1.161	1.481

	Women	Mean		6.29	1.848
		95% Confidence Interval for Mean	Lower Bound	1.76	
			Upper Bound	10.81	
		5% Trimmed Mean		6.15	
		Median		5.00	
		Variance		23.905	
		Std. Deviation		4.889	
		Minimum		0	
		Maximum		15	
		Range		15	
		Interquartile Range		6	
		Skewness		.776	.794
		Kurtosis		.755	1.587

Step 5: Do not reject H_0 since p = 1.00. We do not have significant evidence, $\alpha = 0.05$, that there is a difference in sick time use between men and women. The mean and median of group 0 (men) are not significantly different from the mean and median of group 1 (women).

26. Create your own test question that can be answered using ANOVA and the 5-step approach. Write the question, your response and the data here.

The United Federation of Planets Starfleet Procurement Service needs to award a new contract for purchase of the dilithium crystals that power their starships. To evaluate potential suppliers, the service sends survey teams to five planets: Vulcan, Klingon, Romulus, Rigel, and Arcturus. Each survey team retrieves six random samples of dilithium from its assigned planet and measures the total energy density (kilojoules/microgram) of each sample. The data obtained are as follows:

Source		Vulcan	Klingon	Romulan	Rigelian	Arcturan
		24	14	11	7	19
		15	7	9	7	24
		21	12	7	4	19
		27	17	13	7	15
		33	14	12	12	10
		23	16	18	18	20

Use the five step approach to hypothesis testing to determine whether there is evidence that certain planets can supply higher energy dilithium.

Step 1: State hypotheses.

H_0: The mean energy density of the dilithium is the same for all five planets.

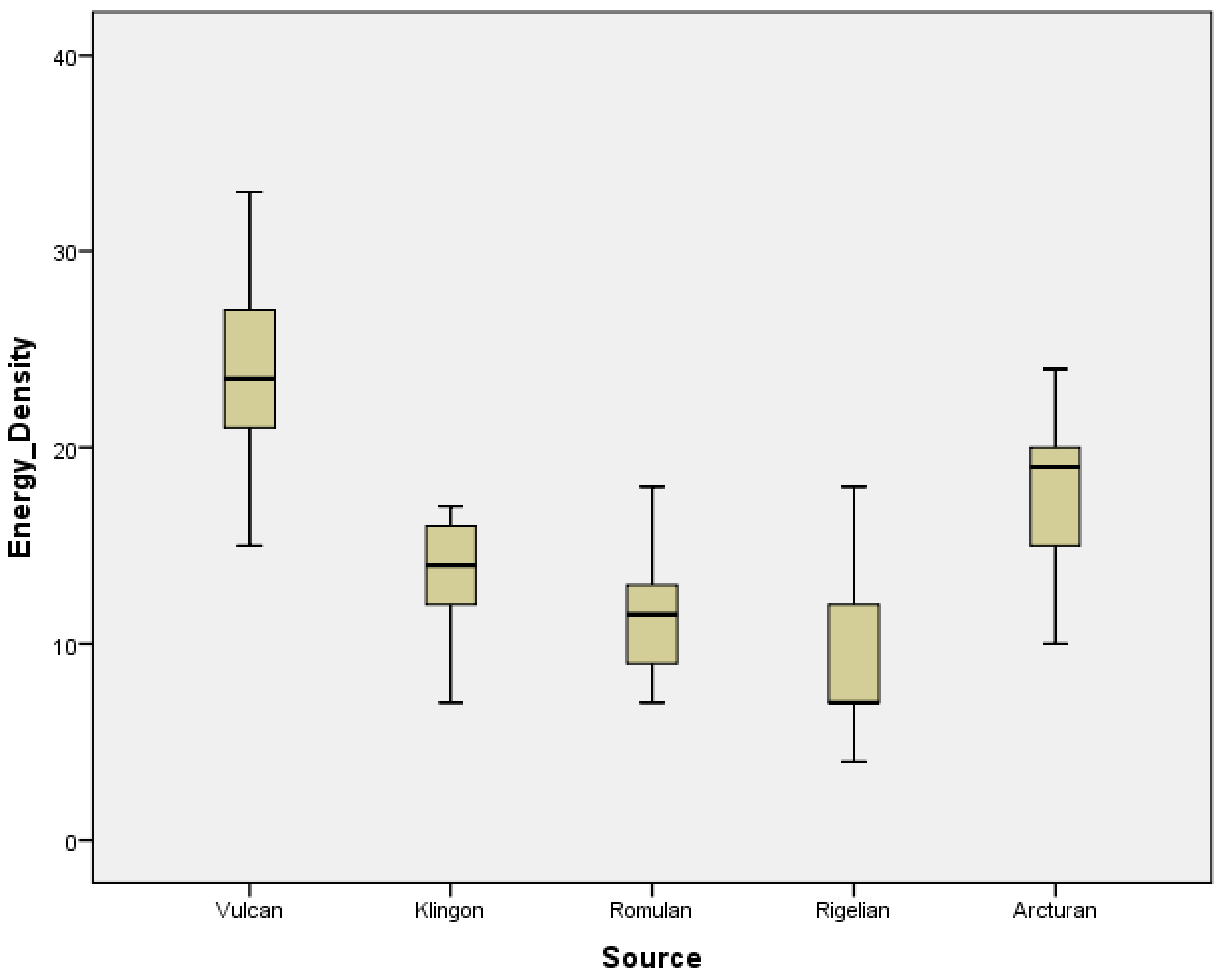
H_1: The mean energy density of the dilithium is not the same for all planets

$\alpha = 0.05$

Step2: Select the test statistic. Use ANOVA because there are five independent groups and the goal is to compare the energy density values among the five groups.

Step 3: Generate the decision rule. $Df_1 = k-1 = 5-1 = 4$. $Df_2 = N-k = 30-5 = 25$. The critical value of F with 4 and 25 degrees of freedom at a 5% significance level is 2.76. Reject H_0 if $F > 2.76$. Do not reject H_0 if $F < 2.76$.

Step 4: Compute the value of the test statistics.



ANOVA

Energy_Density

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	803.000	4	200.750	9.008	.000
Within Groups	557.167	25	22.287		
Total	1360.167	29			

Multiple Comparisons

Energy_Density

Tukey HSD

(I) Source	(J) Source	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Vulcan	Klingon	10.500*	2.726	.006	2.50	18.50
	Romulan	12.167*	2.726	.001	4.16	20.17
	Rigelian	14.667*	2.726	.000	6.66	22.67
	Arcturan	6.000	2.726	.212	-2.00	14.00
Klingon	Vulcan	-10.500*	2.726	.006	-18.50	-2.50
	Romulan	1.667	2.726	.972	-6.34	9.67
	Rigelian	4.167	2.726	.554	-3.84	12.17
	Arcturan	-4.500	2.726	.481	-12.50	3.50
Romulan	Vulcan	-12.167*	2.726	.001	-20.17	-4.16
	Klingon	-1.667	2.726	.972	-9.67	6.34
	Rigelian	2.500	2.726	.888	-5.50	10.50
	Arcturan	-6.167	2.726	.190	-14.17	1.84
Rigelian	Vulcan	-14.667*	2.726	.000	-22.67	-6.66
	Klingon	-4.167	2.726	.554	-12.17	3.84
	Romulan	-2.500	2.726	.888	-10.50	5.50
	Arcturan	-8.667*	2.726	.029	-16.67	-.66
Arcturan	Vulcan	-6.000	2.726	.212	-14.00	2.00
	Klingon	4.500	2.726	.481	-3.50	12.50
	Romulan	6.167	2.726	.190	-1.84	14.17
	Rigelian	8.667*	2.726	.029	.66	16.67

*. The mean difference is significant at the 0.05 level.

Descriptives

Energy_Density

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Vulcan	6	23.83	6.014	2.455	17.52	30.14	15	33
Klingon	6	13.33	3.559	1.453	9.60	17.07	7	17
Romulan	6	11.67	3.777	1.542	7.70	15.63	7	18
Rigelian	6	9.17	5.037	2.056	3.88	14.45	4	18
Arcturan	6	17.83	4.792	1.956	12.80	22.86	10	24
Total	30	15.17	6.849	1.250	12.61	17.72	4	33

Step 5: Reject H_0 since $F = 9.008$ (is > 2.76) and $p < 0.001$. We have significant evidence, $\alpha = 0.05$, that at least one pair of means are different. The follow up Tukey HSD analysis shows that:

- The mean energy density of Vulcan dilithium is significantly higher than that of Klingon, Romulan, or Rigelian, but not significantly higher than Arcturan.
- The mean energy density of Arcturan dilithium is significantly higher than that of Rigelian.
- Overall, the Vulcan dilithium looks like the highest quality product in the group, but you knew that all along, didn’t you? Live long and prosper. ☺