الزمن ثلاث ساعات	الفصل الدراسى الثانى	جامعة القاهرة
مادة تحليل بيانات STATA 613	تمهيدي ماجستير 23/06/2010	معهد الدراسات و البحوث الاحصائية

Part I

Answer all the following Questions

Question (1):

The sum of squares and products for a single-factor analysis of covariance follow. Draw appropriate conclusions for the test and Compute the common regression Coefficient. Where $F_{3,11,0.05} = 1.17$.

Source of Variation	Degree of Freedom	Sum of Squares and products		
		x	y	xy
Treatments	3	1500	1000	650
Error	12	6000	1200	550
Total	15	7500	2200	1200

Question (2):

a) Given the following regression model is :

$$y = 28.2 + 13.9x + 14d - 8.9dx$$

Where y is the spending, x is the income and d is a dummy presents the gender (1 if male and 0 for male). Write the regression equation for each gender. What is the spending for a female has income = 1000.

b) The price of selling a house depends on the area and number of rooms and the presence of the bathroom. The following data give the price, the area, number of rooms and the presence of the bathroom of 10 houses

price	10.000	20.000	30.000	40.000	50.000	60.000	70.000	80.000	90.000	100.000
area	800	1000	1200	1500	1800	2000	2200	2500	3000	3500
# of room	>2	>2	1	1	1	1	2	2	2	>2
bathroom	no	yes	no	no	yes	yes	no	no	yes	yes

Write the linear regression model between the price and the dependent variables without interaction. Write the matrix (X) and the vector (Y) numerically.

Question (3):

Age (x)	24	29	35	40	50	55	60	66	70	77
Ln[p/(1-p)]	-1.4	-1.1	-0.7	-0.8	-0.1	0.8	1	0.9	2.9	1.4

Use logistic regression to analyze these data where $\sum x^2$ = 28552 and $\sum xy = 351.9$, and find the percentage at age 80 years.

Part II

Using the attaché SPSS results answer the following

Question (1): Fill the Blank

The number of European Country is..... and the number of the Japanese Country is .....

The mean of the Previous Experience is .....

The maximum value and the minimum value of the education level are .....

The number of persons has 18 years of the education level is .....

The number of persons has less than 15 years of the education level is .....

Question (2): From the ANOVA table write the hypothesis and comment on the results

Question (3): Write the linear regression model between the current salary and both of the beginning salary and the dummy variable with and without interaction. Which one is the best? Why?

Between-Subjects Factors

		Value Label	N
Country of Origin	1	American	253
	2	European	73
	3	Japanese	79

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Previous Experience (months)	474	0	476	95.86	104.586
Valid N (listwise)	474				

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	21650.464	2270.173		9.537	.000
	Beginning Salary	1.399	.070	.645	19.976	.000
	J1	-13530.7	1448.667	-.336	-9.340	.000
	J2	-11803.7	1964.713	-.160	-6.008	.000

a. Dependent Variable: Current Salary

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	25665.114	2587.890		9.917	.000
	Beginning Salary	1.266	.081	.584	15.580	.000
	J1	-22874.1	3223.921	-.568	-7.095	.000
	J2	3443.252	16558.059	.047	.208	.835
	XJ1	.511	.156	.194	3.266	.001
	XJ2	-1.145	1.084	-.235	-1.056	.291

a. Dependent Variable: Current Salary

Educational Level (years)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	8	53	11.2	11.2	11.2
	12	190	40.1	40.1	51.3
	14	6	1.3	1.3	52.5
	15	116	24.5	24.5	77.0
	16	59	12.4	12.4	89.5
	17	11	2.3	2.3	91.8
	18	9	1.9	1.9	93.7
	19	27	5.7	5.7	99.4
	20	2	.4	.4	99.8
	21	1	.2	.2	100.0
	Total	474	100.0	100.0	

ANOVA

Horsepower

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	144039.1	2	72019.564	63.672	.000
Within Groups	447916.3	396	1131.102		
Total	591955.4	398			

مع اطيب التمنيات بالنجاح

السؤال الاول:

٣: تقوم مؤسسه صناعيه بدراسة امكانية توسيع الطاقه الانتاجيه للمصنع وذلك عن طريق شراء احدى آلتين مختلفتين بالحجم تبلغ الطاقه الانتاجيه لاحداهما ضعف الطاقه الانتاجيه للاخرى بينما تكلفة الاله الصغيرة هى ٦٠٠٠٠ دولار وتكلفه الاله الكبيرة ١٠٠٠٠٠ دولار.

مع العلم ان الطاقه الانتاجيه للمصنع الحاليه قادرة على توفير متطلبات السوق الحاليه ، اما اذا ازداد الطلب بشكل معتدل عما عليه الان فان عمليه شراء الاله الصغيرة ستكون ملائمه، اما اذا ازداد الطلب الحالى بشكل كبير عندها سيكون المصنع بحاجه الى الة كبيرة او الى آلتين من النوع الصغير فاذا علمت ان:

١:الطلب على منتجات المؤسسه سيتم دوما تغطيه السوق والطاقه الاضافيه ستستخدم لتلبية الطلبيات العامه.

٢: التكلفه الاجماليه للمصنع يعيمد اولا على نوع الالات المشتراه وعلى منتجات المؤسسه فى المستقبل بالاضافه الى ان هناك تكلفه تعادل ٥٠٠ دولار تضاف لكل ١٠٠٠ دولار من سعر الاله الاصلى عندما يتم شراء اله لمقابله الطلب المتزايد او الطلبيات العاجله ،

واذا فرض ان المدير يعتقد بان هناك حظا اوفر لان يكون مستوى الطلب مرتفعا ويعادل ضعفى الحظ لان يكون مستوى الطلب معتدلا ، اما كون مستوى الطلب مستقرا فله نفس الحظ والفرصه المعطاة لمستوى الطلب المعتدل علما بان المدير يمكنه ان يكون قادر على اتخاذ قرارا من التوسع باحتمال ٠,٧ وغير قادر على اتخاذ قراره من التوسع باحتمال ٠,٣ .

والمطلوب

١: تشكيل مصفوفة العائد

٢: تحديد اى الافعال افضل مستخدما المعايير الخمسه (α=٠,٦)

٣: تحديد القرار الامثل على ضوء ما توفر من معلومات.

السؤال الثانى:

: يعتقد خبير التسويق فى شركه ما ان بسبه المستهلكين $B=\{b_1,b_2,b_3\}$ الذين يفضلون التصميم الجديد المنتج من قبل الشركه له التوزيح الاحتمالى التالى:

B	b_1	b_2	b_3
P(b)	0.4	0.3	0.3

فإذا استخدم المنتج الجديد فأن العائد المتوقع سينخفض بمقدار 7.5 دولار عند التصميم b_1 ويزداد بمفدار 5 دولار عند التصميم الثانى ويزداد بمقدار 10 عند التصميم الثالث ، اوجد القرار الامثل وقيمة اقصى ما يمكن دفعه للحصول على معلومات اضافيه كامله

ثم اوجد القرار الامثل عند غخذ عينه عشوائيه حجمها ٢ لدراسه حالة وضع الطلب على المنتج ذو التصميم الجديد والقيمه المتوقعه لمعلومات العينه وهل من مصلحة الشركه اجراء المعاينه ام لا حيث ان تكلفة المعاينه تساوى

١: ٢ دولار للوحده الواحده

٢: ٠,١ دولار للوحده الواحده.

السؤال الثالث:

أمام شركه للاستثمار بديلان للاختيار بينهما البديل الاول هو شراء فندق سياحى بمبلغ ٢٠ مليون دولار ويعتمد العائد منه على حركة السياحه

فأذا كانت هذه الحركه نشطه فأن العائد المتوقع ٦ مليون دولار فى السنه لمدة خمس سنوات بأحتمال ٠,٦ واذا كانت السياحه ضعيفه فان العلئد المتوقع ٢٤٠٠٠٠٠ دولار فى السنه لمدة خمس سنوات وبعد نهاية المدة يقدر قيمة الفندق ب ٥٦٠٠٠٠٠ دولار

والبديل الثانى هو شراء عمارة تكلفتها ١٨ مليون دولار ويبلغ ايرادها السنوى ٢ مليون دولار لمدة ثمانى سنوات ويقدر قيمة العمارة بعد ذلك ١٠ مليون دولار

والمطلوب تشكيل شجرة القرارات مع تحديد القرار الامثل.

Pre-Master Stat A 613-Statistics Time: 3 hrs	Final Exam Jan, 2016	Cairo University Institute of Statistical Studies & Research

Answer the following questions

Q1) According to the following results:

Group Statistics

	Gender	N	Mean	Std. Deviation
Current Salary	Male	258	$31,441.78	$19,499.214
	Female	216	$46,031.92	$17,558.021

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Current Salary	Equal variances assumed	119.6	.081	10.945	472	.000	$15,409.8	$1,407.9	$12,643.3	$18,176.4
	Equal variances not assumed			11.688	344.262	.000	$15,409.8	$1,318.4	$12,816.7	$18,002.9

- State the hypothesis of the test.
- What type of the above test ?
- Do women have more salary per hour than men?
- Do you have homogeneous population?
- The sample size equal to ----------

Q2) Consider the following table:

		Frequency	Percent	Valid Percent
Valid	Red	230	54.8	
	Green	75	17.9	
	Blue	85	20.2	
	Total	390	92.9	
Missing	System	30	7.1	
Total		420	100.0	

- Do you have missing data? How many?
- What is the percentage of Blue color?
- Calculate the valid percent
- Discuss the difference between percent and valid percent. Which one more accurate?

Q3) Consider the following results:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.897[a]	.805	.804	46.713

a. Predictors: (Constant), Horsepower

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-62.729-	6.779		-9.253-	.000
	Horsepower	2.459	.061	.897	40.500	.000

a. Dependent Variable: Engine Displacement (cu. inches)

- Classify the variables of the above results.
- State the hypothesis of the model.
- Write the equation of the model. Explain the coefficients
- How to evaluate the above model? Explain.

Q4)
- What is the range of salary of the custodial group?
- Construct 95% C.I. of the mean of clerical group?
- What is the mean salary of the manager group?
- State the hypothesis of the test.
- Under what condition, you construct the Post Hoc Test?
- Is there any difference between the three groups? Explain

	N	Mean	95% Confidence Interval for Mean		Minimum	Maximum
			Lower Bound	Upper Bound		
Clerical	363	$27,838.54	$27,057.40	$28,619.68	$15,750	$80,000
Custodial	27	$30,938.89	$30,102.37	$31,775.40	$24,300	$35,250
Manager	84	$63,977.80	$60,018.44	$67,937.16	$34,410	$135,000
Total	474	$34,419.57	$32,878.40	$35,960.73	$15,750	$135,000

ANOVA

Current Salary

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	89438483925.943	2	44719241962.971	434.481	.000
Within Groups	48478011510.397	471	102925714.459		
Total	137916495436.340	473			

Post Hoc Test

(I) Employment Category	(J) Employment Category	Mean Difference (I-J)	Std. Error	Sig.
Clerical	Custodial	-$3,100.349	$2,023.760	.277
	Manager	-$36,139.258*	$1,228.352	.000
Custodial	Clerical	$3,100.349	$2,023.760	.277
	Manager	-$33,038.909*	$2,244.409	.000
Manager	Clerical	$36,139.258*	$1,228.352	.000
	Custodial	$33,038.909*	$2,244.409	.000

Q5) Current Salary Stem-and-Leaf Plot has the form

```
 Frequency     Stem &  Leaf

    33.00        1 . 5666667778899999
   110.00        2  .000000111111111112222222222222233333333444444444444444
   115.00        2 . 55555555555666666666666677777777777777888888899999999999
    80.00        3 . 00000000000000001111111122233333333444444
    32.00        3 . 555556667778899
    20.00        4 . 00001233&
    12.00        4 . 55678&
    12.00        5 . 01244&
     7.00        5 . 556
    53.00 Extremes    (>=56750)

Stem width:     10000
Each leaf:       2 case(s)
```

- Is this data has normal shape? Discuss
- Do you have any outliers? How many?
- Which central measure will be better for this data.
- What is the interval mode of current salary?
- The stem width equal to .... while each leaf represents...

Q6) - What is the meaning of paired samples t-test?
- State the hypothesis of the test.
- The sample size equal to ----------
- What is the correlation between current and Beginning salary? Explain
- What is the mean of the beginning salary?
- Did current salary better than the beginning salary?

	Mean	N	Std. Deviation	Std. Error Mean
Current Salary	$34,419.57	474	$17,075.661	$784.311
Beginning Salary	$17,016.09	474	$7,870.638	$361.510

	N	Correlation	Sig.
Current Salary & Beginning Salary	474	.880	.000

	Paired Differences					Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		
				Lower	Upper	
Current Salary - Beginning Salary	$17,403.4	$10,814.6	$496.7	$16,427.4	$18,379.5	.000