PRIFYSGOL BANGOR
BANGOR UNIVERSITY

Arholiadau Diwedd Semester II 2010/2011
End of Semester II Examinations 2010/2011

Amser a ganiateir	:	2 awr
Time allowed	:	2 hours

PHP4006

ADVANCED STATISTICS

Please answer **ALL** questions from Section A of the paper, and **THREE** questions from Section B

Section A questions are each worth 1 mark.

Section B questions are each worth 25 marks.

Section B (Answer **THREE** questions)

1. A researcher wanted to know if gender affected how susceptible people were to distraction. The level of distraction was measured using a simple reaction time task (slower reaction time suggested higher distraction). She measured men and women of either 18-25 years, 26-30 years, and over 30 years in three conditions: one where they carried out a task with no distraction, one where they did the task but with the TV on and tuned to a sport channel, and in a third they completed the task while listening to sport on the radio.

    (a) State the name and design of the statistical test you intend to use?
    (3 marks)
    (b) Justify the use of your chosen test in (a). (6 marks)
    (c) State the null hypotheses of your test chosen in (a). (3 marks)
    (d) Explain what the significance level of the chosen test is. (3 marks)
    (e) State the assumptions that the chosen test requires to hold. (6 marks)
    (f) State what you would do if the assumptions do not hold. (4 marks)

2. A psychiatric problem is known to occur more frequently in older aged people. You are planning a study to determine the relationship between the occurrence of the problem and age. You are considering using a logistic regression analysis or a log-linear model to analyse your resulting data. Describe the types of data and the assumptions you would make that would be appropriate for:
    (a) Logistic Regression. (9 marks)
    (b) Log-Linear Model. (9 marks)
    (c) Outline a dataset and justify the use of ONE of the above techniques.
        DO NOT use a dataset used in the lectures or textbook (7 marks)

3. Child Psychologists are interested in the impact of television advertisements on the diets of children. The researchers obtain 3 samples of children (6-12 years): one group spend over 20 hours a week watching children's TV channels with advertisements; a second group a similar number of hours watching children's TV without advertisements; and a third control group are children who rarely watch TV at all. The quality of the children's diets are coded by nutritionists (blind to the experimental design) according to their sugar, salt, and fat content ranging from 0 (very good diet) to 5 (very poor diet). A short form of the SPSS output for their analysis is below.

(a) Explain the design of this experiment. (3 marks)

(b) Why did the researchers not rely on using multiple Mann-Whitney tests alone? (3 marks)

(c) What level of significance would be appropriate for their Mann-Whitney tests? (3 marks)

(d) Calculate the Pearson's r effect size for the statistically significant Mann-Whitney tests using the information below. (6 marks)

(e) Briefly summarise the main findings of the researcher's data. (6 marks)

(f) If the data was able to meet the assumptions of parametric testing, which statistical tests would have been appropriate here? What would be the statistical benefit of using these tests in such circumstances? (4 marks)

Descriptive Statistics

Quality of Diet

TV & Adverts	N	Valid	15
		Missing	0
	Median		4.0000
TV & No adverts	N	Valid	15
		Missing	0
	Median		1.0000
No TV	N	Valid	15
		Missing	0
	Median		1.0000

Test Statistics[a,b]

	Quality of Diet
Chi-square	12.340
df	2
Asymp. Sig.	.002

a. Kruskal Wallis Test

b. Grouping Variable: Advert Exposure

Mann-Whitney 1

Ranks

	Advert Exposure	N	Mean Rank	Sum of Ranks
Quality of Diet	TV & Adverts	15	20.63	309.50
	TV & No adverts	15	10.37	155.50
	Total	30		

Test Statistics^b

	Quality of Diet
Mann-Whitney U	35.500
Wilcoxon W	155.500
Z	-3.247
Asymp. Sig. (2-tailed)	.001
Exact Sig. [2*(1-tailed Sig.)]	.001^a

a. Not corrected for ties.

b. Grouping Variable: Advert Exposure

Mann-Whitney 2

Ranks

	Advert Exposure	N	Mean Rank	Sum of Ranks
Quality of Diet	TV & Adverts	15	19.90	298.50
	No TV	15	11.10	166.50
	Total	30		

Test Statistics^b

	Quality of Diet
Mann-Whitney U	46.500
Wilcoxon W	166.500
Z	-2.781
Asymp. Sig. (2-tailed)	.005
Exact Sig. [2*(1-tailed Sig.)]	.005^a

a. Not corrected for ties.

b. Grouping Variable: Advert Exposure

Mann-Whitney 3

Ranks

	Advert Exposure	N	Mean Rank	Sum of Ranks
Quality of Diet	TV & No adverts	15	15.17	227.50
	No TV	15	15.83	237.50
	Total	30		

Test Statistics[b]

	Quality of Diet
Mann-Whitney U	107.500
Wilcoxon W	227.500
Z	-.214
Asymp. Sig. (2-tailed)	.831
Exact Sig. [2*(1-tailed Sig.)]	.838[a]

a. Not corrected for ties.

b. Grouping Variable: Advert Exposure

4. The data below are on glucose control in diabetic patients. Good control is measured by a low value of Glucose (the outcome or Y value).

outcome	G	Glucose in the blood
predictors	K	Knowledge of the illness
	F	Measure of attribution called fatalistic externalism
	D	Duration of the illness in months
	S	Length of schooling 0 – less than 13 years, 1 – more than 13 years

The variables above were measured on diabetic patients. A theory proposes that glucose is directly affected by knowledge and duration, knowledge is directly affected by fatalism, and fatalism is directly affected by duration. The theory does not propose any effect of schooling.

The output below is taken from the use of forward regressions to predict each of glucose, knowledge, and fatalism from the variables suggested by the above theory. Schooling was included as a possible predictor in each regression.

(a) Draw the path diagram that the above model proposes.   (5 marks)

(b) Comment on the correlation coefficients between schooling and the other 4
   variables. (3 marks)

(c) Using the model in Regression 1 what equation would you use to predict glucose,
   and predict the Glucose level of a person with Knowledge = 39, Duration = 150,
   Schooling = 0? (7 marks)

(d) Use the regression outputs to enter the Path Coefficients for the original
   theoretical model (write them in the path diagram you drew earlier)
   (4 marks)

(e) Use the data below to specify a second model which better accounts for all the
   data, and enter the appropriate path coefficients. (6 marks)

Correlations

Correlations

		glucose	knowledge	fatalism	duration	schooling
glucose	Pearson Correlation	1	-.341**	.150	-.122	-.319**
	Sig. (2-tailed)	.	.004	.223	.320	.008
	N	68	68	68	68	68
knowledge	Pearson Correlation	-.341**	1	-.491**	-.110	.333**
	Sig. (2-tailed)	.004	.	.000	.371	.005
	N	68	68	68	68	68
fatalism	Pearson Correlation	.150	-.491**	1	.275*	-.257*
	Sig. (2-tailed)	.223	.000	.	.023	.035
	N	68	68	68	68	68
duration	Pearson Correlation	-.122	-.110	.275*	1	-.250*
	Sig. (2-tailed)	.320	.371	.023	.	.040
	N	68	68	68	68	68
schooling	Pearson Correlation	-.319**	.333**	-.257*	-.250*	1
	Sig. (2-tailed)	.008	.005	.035	.040	.
	N	68	68	68	68	68

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

*. Correlation is significant at the 0.05 level (2-tailed).

Regression 1

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	knowledge	.	Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: glucose

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.341^a	.116	.103	18.810

a. Predictors: (Constant), knowledge

b. Dependent Variable: glucose

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3067.774	1	3067.774	8.670	.004^a
	Residual	23352.990	66	353.833		
	Total	26420.765	67			

a. Predictors: (Constant), knowledge

b. Dependent Variable: glucose

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	126.438	11.436		11.056	.000			
	knowledge	-.932	.317	-.341	-2.945	.004	-.341	-.341	-.341

a. Dependent Variable: glucose

Excluded Variables[b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	fatalism	-.023[a]	-.173	.863	-.021	.759
	duration	-.162[a]	-1.401	.166	-.171	.988
	schooling	-.231[a]	-1.920	.059	-.232	.889

a. Predictors in the Model: (Constant), knowledge

b. Dependent Variable: glucose

Regression 2

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	fatalism		Forward (Criterion: Probabilit y-of-F-to-e nter <= .050)
2	schooling		Forward (Criterion: Probabilit y-of-F-to-e nter <= .050)

a. Dependent Variable: knowledge

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	851.617	1	851.617	20.983	.000[a]
	Residual	2678.662	66	40.586		
	Total	3530.279	67			
2	Regression	1013.941	2	506.970	13.096	.000[b]
	Residual	2516.339	65	38.713		
	Total	3530.279	67			

a. Predictors: (Constant), fatalism

b. Predictors: (Constant), fatalism, schooling

c. Dependent Variable: knowledge

Model Summary[c]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.491[a]	.241	.230	6.371
2	.536[b]	.287	.265	6.222

a. Predictors: (Constant), fatalism

b. Predictors: (Constant), fatalism, schooling

c. Dependent Variable: knowledge

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	47.846	2.825		16.935	.000			
	fatalism	-.654	.143	-.491	-4.581	.000	-.491	-.491	-.491
2	(Constant)	44.982	3.094		14.541	.000			
	fatalism	-.578	.144	-.434	-4.007	.000	-.491	-.445	-.420
	schooling	3.220	1.572	.222	2.048	.045	.333	.246	.214

a. Dependent Variable: knowledge

Excluded Variables[c]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	duration	.027[a]	.240	.811	.030	.924
	schooling	.222[a]	2.048	.045	.246	.934
2	duration	.073[b]	.653	.516	.081	.890

a. Predictors in the Model: (Constant), fatalism

b. Predictors in the Model: (Constant), fatalism, schooling

c. Dependent Variable: knowledge

Regression 3

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	duration	.	Forward (Criterion: Probabilit y-of-F-to-e nter <= .050)

a. Dependent Variable: fatalism

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.275[a]	.076	.062	5.279

a. Predictors: (Constant), duration

b. Dependent Variable: fatalism

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	150.740	1	150.740	5.409	.023[a]
	Residual	1839.202	66	27.867		
	Total	1989.941	67			

a. Predictors: (Constant), duration

b. Dependent Variable: fatalism

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	16.805	1.151		14.602	.000			
	duration	1.787E-02	.008	.275	2.326	.023	.275	.275	.275

a. Dependent Variable: fatalism

Excluded Variables[b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	schooling	-.200[a]	-1.661	.102	-.202	.937

a. Predictors in the Model: (Constant), duration

b. Dependent Variable: fatalism