Uva Wellassa University of Sri Lanka
Faculty of Science and Technology
Department of Science and Technology
400 Level 1st Semester Examination – Jul./Aug. 2016
SCT 401-1 Data Handling and Statistics

Uva Wellassa University

Instructions to candidates.
Duration: 02 Hours
Number of questions: Four (04) Esssay Questions
Mark allocation: 100 mark
Answer All Questions.

1. A researcher investigated whether soil water availability, measured as soil matric potential (SMP, in kPa) is influenced by the amount of rainfall, measured as sliding thirty day totals (STDT, mm) at two sites, Sinharaja and Udawalawe.
   (The data are stored in a Minitab work sheet named "Site.mtw" given on your computer.)

Site	STDT	SMP	Site	STDT	SMP
Sinharaja	222.3	-10.25	Udawalawe	485.7	-11.03
Sinharaja	189.4	-10.30	Udawalawe	470.7	-18.90
Sinharaja	209.2	-26.58	Udawalawe	814	-5.51
Sinharaja	154.4	-31.18	Udawalawe	875.5	-8.78
Sinharaja	106.5	-28.76	Udawalawe	840.2	-5.47
Sinharaja	107.2	-61.67	Udawalawe	803.7	-10.31
Sinharaja	111	-12.00	Udawalawe	725.2	-10.32
Sinharaja	33.7	-40.83	Udawalawe	644	-6.22
Sinharaja	141.1	-6.60	Udawalawe	388	-5.77
Sinharaja	134.6	-8.63	Udawalawe	565.5	-6.41
Sinharaja	68.2	-24.85	Udawalawe	519.75	-12.10
Sinharaja	197.1	-6.46	Udawalawe	347.25	-6.27
Sinharaja	137.3	-5.58	Udawalawe	492	-5.60
Sinharaja	293.6	-5.81	Udawalawe	277	-10.72

a) Calculate the means, standard deviations and coefficient of variance (CV) of soil water availability and rainfall respectively for each site and comment of your results. (05 mark)

b) Use $\log_{10}$ transformation to transform soil water availability measurements for each site. SMP values are negative, so you will need to use absolute values of SMP for the log10 transformation. Calculate new values and store in new colunm in Minitab worksheet. **Use this transformed dataset for the rest parts of this question.** (03 mark)

c) Construct a scatterplot to examine the relationship between $\log_{10}$-transformed values of soil water availability and rainfall at each site. (04 mark)

d) Test for a correlation between $\log_{10}$-transformed values of soil water availability and rainfall at each site. Interpret your results. (04 mark)

(5)

(63)

e) Find the linear equation to predict the $\log_{10}$-transformed values of soil water availability as a function of rainfall at each site. (05 mark)

f) Test whether the fitted regression model is statistically significant by using ANOVA table with appropriate hypothesis, and explain what does it mean. (05 mark)

g) Using a simple linear regression analysis, compute the coefficient of determination for the relationship between $\log_{10}$-transformed values of soil water availability and rainfall at each site. For this regression, check the adequcy of all assumptions. (06 mark)

h) Use your equation to predict soil water availability values (in kPa) for rainfall measurements of 450 mm and 670 mm respectively. (03 mark)

2. An experiment was carried out to study the effect of two factors on the blood sugar level in rabbits. One factor was two types (I and II) of insulin and the other factor was the dose level ("low" or "high") at which each type of insulin was given. Four rabbits were used in the experiment, and each was given all four treatment combinations, in random order, separated by suitable intervals of time in the hope of avoiding carryover effects from one treatment to the next. Measurements of blood sugar level (mg/100cc) are given in the table.

Rabbit	**Type of insulin**			
	1		2	
	Dose level		**Dose level**	
	Low (1)	High (2)	Low (1)	High (2)
A	88	59	76	63
B	89	58	74	65
C	86	61	79	62
D	85	59	72	68

a) Calculate the mean responses for the four treatment combinations, and illustrate these in a graph. Is there any suggestion of interaction between the two factors? (05 mark)

b) What is the suitable model to describe the above situation (describe each and every parameter in the model)? (05 mark)

c) Construct the analysis of variance (ANOVA). What is your decision based on ANOVA (Check validity of all possible hypotheses) (08 mark)

d) What is the R^2 and explain it? (03 mark)

e) Write a brief report on the results of the experiment by considering following Minitab output. (04 mark)

Output 01: Grouping Information Using Tukey Method

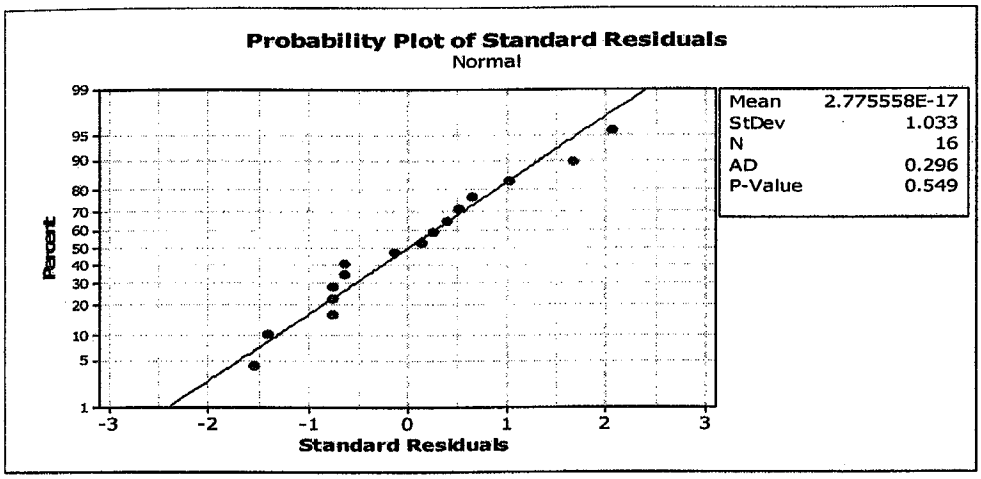
Grouping Information Using Tukey Method and 95.0% Confidence

Dose	Insulin type	N	Mean	Grouping
1	1	4	87.00	A
1	2	4	75.25	B
2	2	4	64.50	C
2	1	4	59.25	C

Means that do not share a letter are significantly different.

f) The normal probability plot of standardized residuals of above analysis is given below. What comments can you make using Graph 01 (Write a suitable hypothesis)? (04 Mark)

Graph 01: Normal Probability Plot of Standardized Residuals



g) Another researcher suggests that it would save time if, instead, four different groups, each of four rabbits, were used in this type of experiment. Each treatment combination would be given to one of the groups, chosen at random. Briefly diccuss the advantages and disadvantages of conducting an experiment in this way. (06 mark)

3. The following data are measures of atmospheric SO2 concentrations, in parts per billion (ppb), over a two-year study period. SO2 concentrations were measured at 12 randomly selected sampling sites in the year 2014 and then re-measured at the same locations in the following year, 2015.

		SO_2 concentrations (ppb)											
Site		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Year	**2014**	400	20	24	95	228	116	65	112	35	45	81	197
	2015	345	8	29	81	204	140	36	75	47	5	65	187

Noting any assumptions made and taking into account the paired nature of the data, carry out a suitable non-parametric test to check whether SO_2 concentrations in the year 2014 is differ from year 2015. Briefly comment on your results. (12 mark)

4. An experiment was conducted in a laboratory using tissue cultured plants to evaluate the growth of plants under five (5) different levels of hormone concentrations. Each concentration was repeated ten (10) times.

a) Identify the most suitable experimental design (model) for this experiment giving justifications. (05 mark)

b) Prepare the lay-out of the experimental design you proposed above elaborating your randomization procedure. (04 mark)

c) Construct the partial ANOVA giving sources of variations and degrees of freedom only. (04 mark)

d) What are the advantages and disadvantages of the experimental design you proposed? (05 mark)