

Stella Maris College
Department of Economics

Data Analysis Using SPSS Software
End Semester Examinations – November 2015

Section A

(Max Marks = 100)

Answer any 5 Questions (5 x 8 Marks = 40 Marks)

1. Assume that a marketing manager wishes to compare five different package designs. He is interested in knowing which is the most preferred one so that the same can be introduced in the market. A random of 200 consumers gives the following picture. Test using appropriate statistic whether the consumer preferences for designs show any significant difference 5% and 1% level of Significance. **(8 Marks)**

Package Design	A	B	C	D	E
Colour Preference	36	52	40	35	37

2. Consider the gain in weight of 19 female rats (given below) between 28 and 84 days after birth. 12 were fed on a high protein diet and 7 on a low protein diet. Test with appropriate statistical tool whether the rats would all benefit (in terms of weight gain) from a high protein diet in comparison with those on a low protein diet. **(8 Marks)**

High Protein Diet	134	146	104	119	124	161	107	83	113	129	97	123
Low Protein Diet	70	118	101	85	107	132	94					

3. From the data given below (Colour.xls), find out whether there is any significant association between gender and colour preference. Test whether the colour preference is same for both male and female at 1% & 5% level of significance using appropriate statistical test. **(8 Marks)**

Colour	Gender		Total
	Male	Female	
Red	30	40	70
Blue	50	20	70
Green	40	20	60
Total	120	80	200

4. The amount of lactic acid in the blood was examined for 10 men, before and after a strenuous exercise, with the results in the following table. (a) Test if exercise changes the level of lactic acid in blood. Use significance level $\alpha=0.01$. (b) Find a 95% CI for the mean change in the blood lactose level. **(8 Marks)**

Before: 15, 16, 13, 13, 17, 20, 13, 16, 14, 18

After: 33, 20, 30, 35, 40, 37, 18, 26, 21, 19

5. The scores of various measures of VCR qualities are given on a 5 point scale in the data set 'Cluster.sav.' 5 measures of picture quality, 3 measures of reception quality, 3 measures of audio quality, 1 measure of ease of programming, 1 measure of no. of events events, 1 measure of feature programming, 3 measures of remote control and 3 measures of extra qualities are given on a 5 point scale. Data is available in 'Cluster.sav.' Cluster television shows into homogenous groups based on viewer characteristics using cluster analysis in SPSS. (8 Marks)

'Cluster.sav' (Double Click to get full data set in Excel)

brand	price	pictur1	pictur2	pictur3	pictur4	pictur5
SONNY	520	5	5	5	5	5
ANGLER	535	5	5	5	5	5
MITTENSU	515	5	5	5	5	5
SINGBO	470	5	5	5	5	5
WHACKAC	525	5	5	5	5	5
SILVERMC	370	4	4	4	4	4
EXPERTSI	430	4	4	4	4	4
FROMSHE	505	5	5	5	5	5
POTASON	450	3	3	3	3	3

6. Assume that a marketer wishes to compare five different package designs. He is interested in knowing which the most preferred one is so that the same can be introduced in the market. A random sample of 200 consumers gives the following picture. Does the consumer preference for the designs show any significant difference? (note: Data may be converted to a form in which SPSS accepts for carrying out a particular statistical test) (8 Marks)

Package Design	A	B	C	D	E	Total
Preference by consumers	36	52	40	35	37	200

Section B

Answer any 3 Questions (3 x 20 Marks = 60 Marks)

1. Build a data file in SPSS using the data given in 'Swimming.docs' specifying the nature & characteristics of each of the variables and answer the corresponding questions. (20 Marks)
- Recode all non-numeric variables to numeric as specified and give appropriate value labels:
Gender: F = 1, M = 2 ; Region: N = 1, S = 2, E = 3, W = 4 ; Food_Type: Veg = 1, NV = 2
 - Generate frequency distribution table for the variables 'Region' and 'Food_Type'.
 - What is the average swim time and how dispersed are the values of swim time from mean swim time
 - Draw appropriate graph/chart to show the share of each region (N,S,E & W) in total.
 - Compute Natural Log of Income variable and name the corresponding variable as 'Ln_Income'.
 - Generate a new variable 'Time_Catg' and categorize the 'Swim_time" into three categories as given: **Fast Swimmers:** [Lowest till 32.00 = 1]; **Medium Swimmers:** [32.10 to 34.50]; **Slow Swimmers:** [Above 34.50]
 - Create a bivariate table for the variables 'Time_Catg' and 'Gender'

2. (a) use 'Swimming.docs' data set and bring out correlation matrix among the variables 'Height', 'Weight', 'Swim_Time', and 'Age'.

(b) Run a regression analysis and bring out the direction and magnitude of relationship of the variables Age, Height & weight on Swim_Time. Test if the observed relationships are statistically significant at 1% level of significance. (20 Marks)

'Swimming.Doc' (Double Click to get full data set in Excel)

Sl. No.	Name	Age	Gender	Region	Height	Weight
1	Sara	23	F	S	152	66
2	Jason	20	M	N	187	53
3	Jones	32	F	N	135	83
4	Donna	25	F	S	139	43
5	Philip	26	M	S	130	57
6	Hanna	23	F	W	154	73
7	Sam	22	M	E	178	64
8	Ben	23	M	W	130	77
9	Abby	25	F	W	173	45

3. (a) use 'Swimming.docs' data set given above and create dummies for the variables Gender & Region. Regress income on Region dummies and interpret the same. Test whether there is a statistically significant difference in income because of the difference in region they belong at 5% and 1% level of significance.

(b) Add one more categorical variable viz., gender to the model [as asked in 3 (a)] to your set of independent variables. Find out the interaction effect of gender and region on Income. (20 Marks)

4. A super market that has a chain of stores is concerned about its service quality reputation perceived by its customers. The Table below shows the perceived service quality with regards to politeness of the staff. The number in each cell of the table is percentage of people who have said that the staffs are polite. Perform a two way ANOVA and draw your inferences about the population means of the politeness corresponding to the days, as well as the stores. (20 Marks)

Day/Store	A	B	C	D	E
Monday	79	81	74	77	66
Tuesday	78	86	89	97	86
Wednesday	81	87	84	94	82
Thursday	80	83	81	88	83
Friday	70	74	77	89	68

5. Run a factor analysis for the data available in the file **'Factor1.sav'**. and interpret all the essential components in the output. What are the underlying dimensions of our standardized scores for the given variables?
(20 Marks)

'Factor.sav' (Double Click to get full data set in Excel)

item13	item14	item15	item16	item17	item18	item19
5	5	5	5	5	5	5
5	4	4	4	5	3	4
5	4	5	5	5	5	5
4	4	4	4	4	2	4
5	5	4	5	4	5	5
5	5	4	4	5	5	4
5	5	5	5	5	4	5
5	5	4	4	5	5	5
5	5	5	5	5	4	4