WRITTEN EXAM

Course code: ME 612

Course name: Survey of Statistical Methods

Date: May 31, 2019

Duration: 09.00 – 13.00

Total points: 5 x 10= 50 points

Note: Answer all questions.

1. Consider the following data for two variables X and Y.

X	Y
1	16
2	23
4	35
3	28
5	44
6	40
3	22
8	61
9	82

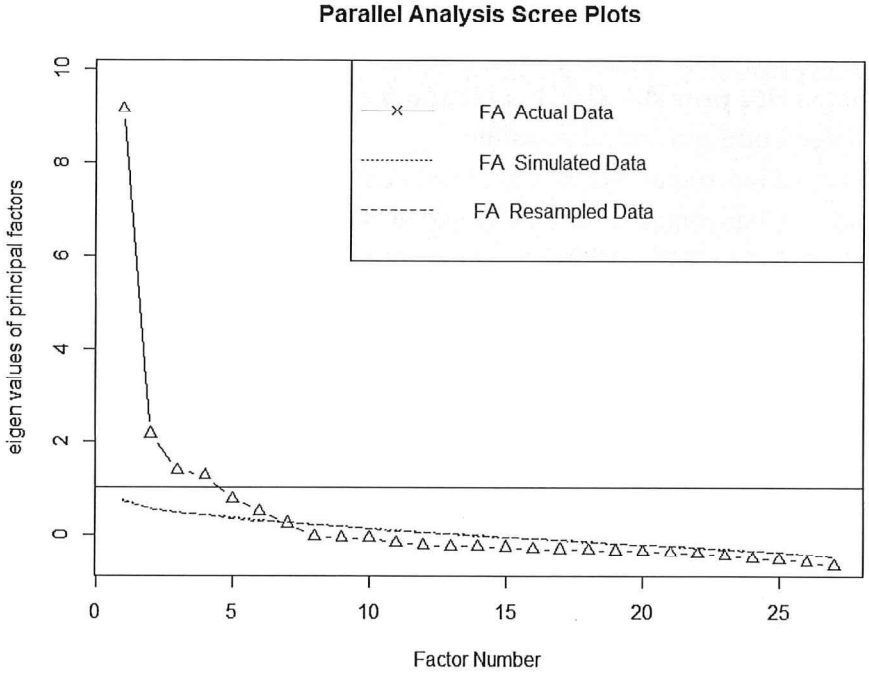
a. Sketch a scatterplot for (x,y). (2 points)
b. Compute the correlation coefficient (r) for (x,y). (3 points)
c. Compute the coefficients of the linear regression line, $y = b_0 + b_1 x$. (3 points)
d. What is the estimated value, y_p for x = 7? (2 points)

2. A research study was conducted to examine the differences between older and younger adults on perceived life satisfaction. A pilot study was conducted to examine this hypothesis. Ten older adults (over the age of 70) and ten younger adults (between 20 and 30) were given a life satisfaction test (known to have high reliability and validity). Scores on the measure range from 0 to 60 with high scores indicative of high life satisfaction; low scores indicative of low life satisfaction. The data are presented below.

Older Adults	**Younger Adults**
45	34
38	22
52	15
48	27
25	37
39	41
51	24
46	19
55	26
46	36

a. Compute means, standard deviation and variance for both groups. (3 points)
b. What would be the null and alternative hypothesis in this study? (1 points)
c. What significance level would you choose to test the null hypothesis? Why? (1 points)
d. What is the degree of freedom? (1 points)
e. Name the appropriate t-test for this study. (1 points)
f. What is your t-statistic value? Is there a significant difference between the two groups (consider t-critical value 2.101 for two-tailed test at 5% significance for the correct degree of freedom for this study)? (2 points)
g. If you have made an error, would it be a Type I or a Type II error? Explain briefly. (1 points)

3. In June 2018, the city of Kristiansand replaced all buses with hybrid-electric busses. Earlier in 2019, a researcher wanted to investigate the effects of perceived service quality and environmental performance of hybrid-electric bus on customer satisfaction. Furthermore, the researcher wanted to investigate the effect of customer satisfaction on life satisfaction of the bus users. Structural Equation Modelling (SEM) is an appropriate approach to such research problems where the variables of interest are rather abstract and needs multiple items to define them. As a prerequisite of SEM, exploratory factor analysis (EFA) using 27 items was performed and the scree plot are presented below:



a. When it comes to identifying the number of factors in the underlying data, what is the role of the researcher? Based on eigenvalue in the scree plot, how many factors will you suggest for further consideration? (4 points)

Now, the researcher decided to proceed with six factors for the underlying data. She conducted confirmatory factor analysis (CFA) to proceed with six factor structure. Measurement items and their corresponding factor loadings are presented below:

Constructs and their respective items	Standardized loadings
Tangible features of the buses (BT, alpha: 0.85, CR: 0.85)	-
1. Physical facilities of a bus are modern looking	0.707
2. Neat and clean inside a bus	0.703

3. Good condition of air cooling and heating system	0.663
4. Good working condition of bus audio system	0.696
5. Visibility of a complete set of safety equipment (i.e. glass breaking device, emergency door, etc.) with instruction signs	0.722
6. Suitable for special need users (i.e. users with wheel chair, baby stroller, heavy luggage, etc.)	0.699
Bus drivers' quality (BD, alpha: 0.86, CR: 0.86)	-
1. Bus driver with good driving skills	0.766
2. Good appearance of bus driver (i.e. neat, clean and meets uniform standards)	0.766
3. Friendly, helpful and polite customer service of driver	0.803
4. Effective and correct emergency management	0.772
Empathy of bus service provider (EMP, alpha: 0.83, CR: 0.83)	-
1. Punctual departure and arrival schedule	0.709
2. Availability of bus routes within and outside city	0.736
3. Availability of bus routes between city and airport	0.632
4. Provision of compensation scheme in cases of loss or hazard	0.745
5. Attention paid when passengers are boarding on-and-off a bus	0.688
Customer satisfaction (CS, alpha: 0.94, CR: 0.94)	-
1. My satisfaction with this bus service has increased.	0.911
2. My impression of this bus service has improved	0.956
3. I now have a more positive attitude towards this bus service	0.891
Perceived environmental performance (EP, alpha: 0.92, CR: 0.92)	-
1. This hybrid electric bus service is more environmentally friendly	0.911
2. This hybrid electric bus service reduces CO2 emission from road transport	0.982
3. This hybrid electric bus service reduces noise pollution in comparison to diesel bus	0.779
4. Using hybrid electric bus service, I contribute to the betterment of global environment	0.787
Life satisfaction (LS, alpha: 0.85 CR: 0.91)	-
1. In most ways my life is close to my ideal.	0.851
2. The conditions of my life are excellent.	0.926
3. I am satisfied with life.	0.856
4. So far, I have gotten the important things I want in life.	0.804
5. If I could live my life over again, I would change almost nothing.	0.601

b. Calculate average variance extracted (AVE) for each factor. Do the AVEs meet the recommended level? (6 points)

4. (i) A teacher used a series of problems in a class that came from a variety of sources. After each set of problems, the students evaluated it in terms of usefulness, with 1 meaning very helpful and 5 meaning useless. The teacher wondered if the material from a prestigious school was better than the rest. He ran a regression using as the dependent variable the average student rating of the set of problems (remember, higher numbers mean less useful) and as an independent variable whether or not the problems came from the prestigious school (0 if from an ordinary school, 1 if from the prestigious school). Below are his results.

Variable	Coefficient	std error	t-statistic
constant	2.285	.034	67.071
Prestige?	.214	.057	3.780
$R^2 = .212$ $n > 40$			

a) What was the average rating of the lessons from the ordinary schools? (1 point)
b) What was the average rating of the lessons from the prestigious school? (1 point)
c) Was the expectation of the teacher confirmed? (1 point)
d) Suppose the claim was that the lessons from the prestigious school were just like the other lessons and that any differences are due to random chance. Does random chance look like a good explanation of the differences in the quality of the lessons as perceived by students? What number do we use to answer this? (2 points)

(Hint: This is a problem of comparing whether or not two means are the same. Here it is done with regression. It can also be done without regression using a two-sample t-test. The results will be the same regardless of which method is used. Use of a zero-one coding is common when we have an off-on situation. Variables with this coding are called dummy variables.)

(ii) Below are the results from a regression trying to predict the asking price of Cadillacs based on their mileage (measured in thousands of miles). (These data were taken from an issue of the *Chicago Tribune* a few years ago.)

R Square	.603			
Adjusted R Square	.591			
Variable	Regression Coefficient	Std. Error	T	Significance
Constant	26303.415	1928.098	13.642	.000
miles	-226.465	32.478	-6.973	.000

a) How successful is our attempt to explain the prices of these cars? (1 point)
b) If we have a Caddy that has 10,000 miles on it, what would we predict for its price? (2 points)
c) The level of significance for miles .000. What is the hypothesis being tested? (1 point)
d) There might be a problem with the regression. Miles and age tend to go together, with older cars having more miles. Perhaps we are capturing some of the effects of age when we include only miles. How do you think we could fix this problem? (1 point)

5. An investor is interested in forecasting bitcoin price. She learned that autoregressive models are particularly useful to model such data. However, stationarity is a prerequisite to most autoregressive models. She first ran augmented dicky-fuller test on the original time series to check for stationarity and found non-stationarity of data. Consider the bitcoin price (USD per bitcoin) for last 10 days:
(data collected from coindesk)

Date	USD per bitcoin
5/18/2019	7361.26
5/19/2019	7285.42
5/20/2019	8221.27
5/21/2019	7984.51
5/22/2019	7961.52
5/23/2019	7723.89
5/24/2019	7871.38
5/25/2019	7990.92
5/26/2019	8090.05
5/27/2019	8653.15

a) Which data transformation should be adopted before further stationarity check? (2 points)
b) Calculate the log transformed time series for the given time series. (2 points)
c) Calculate first differenced time series of the log transformed data. (2 points)
d) Assume that the researcher wants to get back to the original data. Calculate the original log transformed data from the first differenced log transformed data. (2 points)
e) Calculate the original USD per bitcoin time series from the log transformed data. (2 points)

Relevant Equations for ME 612 Exam

The following equation sheet will be provided during the final exam. Equations for basic calculations such as mean, variance, standard deviation, respective degrees of freedom for chosen test will not be provided.

$$\mu = \bar{X} \pm \left(t * \frac{sd}{\sqrt{n}}\right)$$

$$t = \frac{\bar{X} - \mu}{\frac{sd}{\sqrt{n}}}$$

$$t = \frac{\overline{(X_A} - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\left(\frac{sd_A^2}{n_A} + \frac{sd_B^2}{n_B}\right)}}$$

$$x^2 = \sum \frac{(O - E)^2}{E}$$

$$Cov(x, y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}) * (y_i - \bar{y})$$

$$r = \frac{Cov(x, y)}{sd_x * sd_y}$$

$$\beta = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2}$$

Consider the following critical values for t-test and chi-square test:

Two-tail t-distribution table:

Df/alpha	0.10	0.05	0.01
20	1.725	2.086	2.845
30	1.697	2.042	2.75
Infinity	1.645	1.96	2.576

Chi-squared distribution table:

Df/alpha	0.10	0.05	0.01
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345