FAKULTAS EKONOMI DAN BISNIS
UNIVERSITAS INDONESIA

FINAL EXAM
EVEN SEMESTER OF ACADEMIC YEAR 2016/2017
ADVANCED STATISTICS (ECEU601201)

Date	**: Monday, May 29th 2017**
Time	**: 180 Minute**
Lecturers	**: Sita Wardhani (Coordinator)**
	Maria Agriva/Diahhadi Setyonaluri
	Tika Arundina
	Wisam Rohilina
	Witri Indriyani/Uswatun Hasanah
	Uka Wikarya/Chotib
	Ainul Huda
Notes	**: - Closed Book**
	- You can use a Calculator
	- Never use the cell phone as calculator

PROBLEM A

1. Other things being equal, the greater the standard error of the slope coefficient, the
   a) larger the *t*-value for the slope coefficient
   b) smaller the *t*-value for the slope coefficient
   c) larger the intercept term
   d) smaller the intercept term
   e) larger the slope coefficient

2. Which of the following statistics is not constructed on the ranks of the data?
   a) The sign test
   b) The Mann–Whitney *U* test
   c) The Wilcoxon signed-rank test
   d) Spearman's rank correlation test
   e) The Kruskal–Wallis test

3. To construct the current 30-day moving average of stock prices, we
   a) Use the first 30 days of stock prices
   b) Use the most recent 30 days of stock prices
   c) Use any 30 days of stock prices
   d) Divide the most recent stock price by 30
   e) Multiply the most recent stock price by 30

4. Which of the following is the non-parametric counter part of the *t* test to compare the means of two independent populations?
   a) Chi-square goodness of fit test
   b) Chi-square test of independence
   c) Mann–Whitney *U* test

d) Wilcoxon signed-rank test
e) Friedman test

5. If the sales for a company exhibit constant growth over time, the best method for forecasting would be
a) A linear time trend
b) A log-linear time trend
c) A simple moving average
d) A first-order autoregressive model
e) A centered moving average

6. If the slope coefficient β is insignificant, it means that
a) the independent variable does a good job of explaining the dependent variable
b) the independent variable does not do a good job of explaining the dependent variable
c) the dependent variable does a good job of explaining the independent variable
d) the dependent variable does not do a good job of explaining the independent variable
e) none of the above

7. The sales of appliance manufacturers are tied closely to the status of the economy. If the economy is doing well, in general, sales are better. The sales for an appliance manufacturer time series would show a significant
a) cyclical component.
b) irregular component.
c) seasonal component.
d) trend component.
e) none of the above

8. In a dataset, if *x* and *y* have a strong negative correlation, then a scatter diagram would fit loosely around
a) a horizontal line
b) a vertical line
c) a line going down to the right
d) a line going up to the right
e) All of the above are possible.

9. The price of an item is graphed. Over time, there has been a general increase in price, possibly due to inflation. The time series component used to explain the long term increase is the
a) cyclical component.
b) irregular component.
c) seasonal component.
d) trend component.
e) none of the above

10. If the slope coefficient for a regression is 2.4 and the standard error of the slope coefficient is 0.8, then the *t*-value used to test H_0: β = 1 is

a) 0.8 / 2.4
b) 2.4 / 0.8
c) 2.4 / √0.8
d) (2.4-1) / √0.8
e) (2.4-1) / 0.8

11. The accompanying table shows the regression results when estimating $y = \beta_0 + \beta_1 x + \varepsilon$.
What is the value of the test statistic when testing whether *x* significantly influences *y*?

	Coefficients	*Standard Error*	*t stat*	*p-value*
Intercept	0.083	3.56		0.9822
x	1.417	0.63		0.0745

A. 0.66

B. 1.42

C. 1.96

D. 2.25

E 2.23

12. Which of the following is another name for nonparametric statistics?

a. Nominal statistics
b. Distribution-free statistics
c. Ratio statistics
d. Scale statistics
e. *t* statistics

13. A component of the time series model that results in the multi-period above-trend and below-trend behavior of a time series is

a) a trend component
b)a cyclical component
c)a seasonal component
d)an irregular component
e)none of the above

14. A car-rental company wanted to predict the annual operating cost (y) of its cars

using both the number of kilometers a car was driven and the size of the car (subcompact, compact, midsize, and full-size). To incorporate the effect of the size of the car, the following dummy variables were defined.

X2	X3	X4	Car size
0	0	0	subcompact
1	0	0	compact
0	1	0	midsize
0	0	1	full size

The difference in mean annual operating costs between a full-size car and a subcompact car is

a) bo
b) b4
c) bo + b4
d) bo + b1X1
e) bo + b1X1 + b4

15. Consider the following simple linear regression model: $y = \beta_0 + \beta_1 x + \varepsilon$. When determining whether *x* significantly influences *y*, the null hypothesis takes the form

A. $H_0 : \beta_1 = 0$

B. $H_0 : \beta_1 = 1$

C. $H_0 : b_1 = 0$

D. $H_0 : b_1 = 1$

E. $H_o: b_1 \neq 0$

PROBLEM I.B

A firm wants to know whether there is a relationship between Employees scores of a certain test and their productivities in a unit of time. Data from 10 employees was collected and the result is shown by table below:

Employee	Test Score (X)	Productivity (Y)
A	65	30
B	70	25
C	76	35
D	75	40
E	80	38
F	78	42
G	83	48
H	84	50
I	85	55
J	90	45

As a researcher, you are asked to assist the firm's problem by solving these questions:

1. Find the correlation between test score and productivity of employees, by calculating the Spearman Rank Correlation Coefficient! (7 point)
2. What can you conclude from (1)? (3 point)
3. Test the answer you get from (2), to know whether the observed relationship is real or due to chance! (7 point)
4. What is your conclusion? (3 point)

PROBLEM II.B

A manager at an ice cream store is trying to determine how many customers to expect on any given day. Overall business has been relatively steady over the past several years, but the customer count seems to have ups and downs. He collects data over 30 days and records the number of customers, the high temperature (degrees Fahrenheit), and whether the day fell on a weekend (1 = weekends, 0 = otherwise). The results are presented in the following tables. (20 points)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.960
R Square	0.922
Adjusted R Square	0.916
Standard Error	35.347
Observations	30

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	397,672.9	...	...	0.000
Residual	...	...	...		
Total	29	431,407.4			

	Coefficients	*Standard Error*	*t Stat*	*p-value*
Intercept	-74.695	75.162	-0.994	0.329
Temperature	6.962	0.985	...	0.000
Weekend	201.871	15.513	13.013	0.000

a. Find the missing values in the tables. (4 points)
b. Formulate the estimated regression equation. (2 points)
c. Interpret:
   i. the Temperature coefficient. (2 points)
   ii. the R Square (2 point)
d. At the 5% significance level;
   i. Specify the null and alternative hypothesis to know whether Temperature and Weekend are jointly siginificant in explaining Customers? (1 points)
   ii. Test whether Temperature and Weekend are jointly siginificant in explaining Customers? (3 points)
   iii. Specify the null and alternative hypothesis to know whether Weekend individually siginificant in explaining Customers? (1 points)
   iv. Test your hypothesis whether Weekend individually siginificant in explaining Customers? (3 points)
e. How many customers should the manager expect on a Sunday with a forecasted temperature of 80^{o}? (2 points)

PROBLEM III.B

A farmer is concerned that a change in fertilizer to an organic variant might change his crop yield. He subdivides 6 lots and uses the old fertilizer on one half of each lot and the new fertilizer on the other half. The farmer wants to know whether the crop yields from using old and new fertilizer is different. The following table shows the results.

Lot	Crop yield using old fertilizer	Crop yield using new fertilizer
1	10	12
2	11	10
3	10	13
4	12	9
5	12	11
6	11	12

a. What method is best to use to test your hypothesis given that the data collected does not follow normality assumptions.
b. Specify the hypothesis to prove that the median between the crop yields using two different fertilizers is different.
c. At 5 % significance level, help the farmer to decide whether he has to consider to use the new fertilizer.

PROBLEM IV.B

A financial consultant was given the monthly return data on 20-years Treasury Bonds from 2010-2014. He was requested to estimate the linear trend model and the seasonally adjusted series. He has already calculated the ratio-to-moving average with 12-month moving average. The following data is the ratio to moving average:

Year	Jan	Feb	Mar	Apr	May	Jun
2010	-	-	-	-	-	-
2011	0.995976	0.993117	0.970329	1.000253	1.009119	1.075841
2012	0.937416	0.976706	0.957015	0.982663	1.024404	1.072297
2013	0.849253	0.944319	0.935644	0.955125	1.053135	1.11175
2014	1.050175	1.059832	1.077923	1.101966	1.010656	0.977924
Year	Jul	Aug	Sep	Oct	Nov	Dec
2010	1.048951	1.010778	0.980119	0.985127	0.958316	0.961287
2011	1.062799	1.033058	1.007721	1.014085	0.964909	0.97502
2012	1.068105	1.063172	1.026331	1.069712	1.036617	0.776794
2013	1.05521	1.025763	0.967761	0.959354	0.972384	1.01558
2014	-	-	-	-	-	-

a. As a student of FEB UI, you are asked to help him in computing the adjusted indices for the 12 months. (10points)
b. Intepret the monthly index for:
    a. May (2 points)
    b. November (2 points)
c. He then proceeded to find he following linear trend model for the data, $\hat{T}_t = 5.160547 - 0.02228t$,
   Use the trend and seasonal estimates to make forecasts for the first three month of 2015. (6 points)
   (Hint: Pay attention to the years in consideration)

Statistical Table

I. Critical Values of the Wilcoxon Signed Rank Test

ONE-TAIL	α = .05	α = .025	α = .01	α = .005
TWO-TAIL	**α = .10**	**α = .05**	**α = .02**	**α = .01**
n		*(Lower, Upper)*		
5	0,15	—,—	—,—	—,—
6	2,19	0,21	—,—	—,—
7	3,25	2,26	0,28	—,—
8	5,31	3,33	1,35	0,36
9	8,37	5,40	3,42	1,44
10	10,45	8,47	5,50	3,52

II. Wilcoxon Rank Sum Table

		Lower Tail						Upper Tail					
		prob						prob					
n_A	n_B	.005	.01	.025	.05	.10	.20	.20	.10	.05	.025	.01	.005
4	4			10	11	13	14	22	23	25	26		
	5		10	11	12	14	15	25	26	28	29	30	
	6	10	11	12	13	15	17	27	29	31	32	33	34
	7	10	11	13	14	16	18	30	32	34	35	37	38
	8	11	12	14	15	17	20	32	35	37	38	40	41
	9	11	13	14	16	19	21	35	37	40	42	43	45
	10	12	13	15	17	20	23	37	40	43	45	47	48
	11	12	14	16	18	21	24	40	43	46	48	50	52
	12	13	15	17	19	22	26	42	46	49	51	53	55
5	5	15	16	17	19	20	22	33	35	36	38	39	40
	6	16	17	18	20	22	24	36	38	40	42	43	44
	7	16	18	20	21	23	26	39	42	44	45	47	49
	8	17	19	21	23	25	28	42	45	47	49	51	53
	9	18	20	22	24	27	30	45	48	51	53	55	57
	10	19	21	23	26	28	32	48	52	54	57	59	61
	11	20	22	24	27	30	34	51	55	58	61	63	65
	12	21	23	26	28	32	36	54	58	62	64	67	69
6	6	23	24	26	28	30	33	45	48	50	52	54	55
	7	24	25	27	29	32	35	49	52	55	57	59	60
	8	25	27	29	31	34	37	53	56	59	61	63	65
	9	26	28	31	33	36	40	56	60	63	65	68	70
	10	27	29	32	35	38	42	60	64	67	70	73	75
	11	28	30	34	37	40	44	64	68	71	74	78	80
	12	30	32	35	38	42	47	67	72	76	79	82	84
7	7	32	34	36	39	41	45	60	64	66	69	71	73
	8	34	35	38	41	44	48	64	68	71	74	77	78
	9	35	37	40	43	46	50	69	73	76	79	82	84
	10	37	39	42	45	49	53	73	77	81	84	87	89
	11	38	40	44	47	51	56	77	82	86	89	93	95
	12	40	42	46	49	54	59	81	86	91	94	98	100
8	8	43	45	49	51	55	59	77	81	85	87	91	93
	9	45	47	51	54	58	62	82	86	90	93	97	99
	10	47	49	53	56	60	65	87	92	96	99	103	105
	11	49	51	55	59	63	69	91	97	101	105	109	111
	12	51	53	58	62	66	72	96	102	106	110	115	117
9	9	56	59	62	66	70	75	96	101	105	109	112	115
	10	58	61	65	69	73	78	102	107	111	115	119	122
	11	61	63	68	72	76	82	107	113	117	121	126	128
	12	63	66	71	75	80	86	112	118	123	127	132	135
10	10	71	74	78	82	87	93	117	123	128	132	136	139
	11	73	77	81	86	91	97	123	129	134	139	143	147
	12	76	79	84	89	94	101	129	136	141	146	151	154
11	11	87	91	96	100	106	112	141	147	153	157	162	166
	12	90	94	99	104	110	117	147	154	160	165	170	174
12	12	105	109	115	120	127	134	166	173	180	185	191	195

III. Upper Critical Values for the Spearman Rank – Correlation Coefficient

	Nominal α					
n	0.10	0.05	0.025	0.01	0.005	0.001
4	1.000	1.000	-	-	-	-
5	0.800	0.900	1.000	1.000	-	-
6	0.657	0.829	0.886	0.943	1.000	-
7	0.571	0.714	0.786	0.893	0.929	1.000
8	0.524	0.643	0.738	0.833	0.881	0.952
9	0.483	0.600	0.700	0.783	0.833	0.917
10	0.455	0.564	0.648	0.745	0.794	0.879
11	0.427	0.536	0.618	0.709	0.755	0.845
12	0.406	0.503	0.587	0.678	0.727	0.818

FORMULA

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$H = \left(\frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right) - 3(n+1)$$

$$R^2 = 1 - \frac{SSE}{SST}$$

$$t_{df} = \frac{b_j - \beta_{j0}}{S_{b_j}}$$

$$F_{(df_1, df_2)} = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE}$$

$$Ratio - to - moving\ average = \frac{y_t}{\bar{y}_t}$$

$$Unadjusted\ Seasonal\ Index = average\ of\ \frac{y_t}{\bar{y}_t}\ for\ each\ season$$

$$Adjusted\ Seasonal\ Index = Unadjusted\ seasonal\ index \times multiplier$$

$$Seasonally\ Adjusted\ Series = \frac{y}{\bar{S}}$$