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

Date	: Monday, March 27th 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

_____

Part I. Theoretical questions: (20 points)

1. Suppose you want to perform a test to compare the mean GPA of all freshmen with the mean GPA of all sophomores in a college? What type of sampling is required for this test?
    a. Independent sampling with qualitative data
    b. Independent sampling with quantitative data
    c. Matched-pairs sampling with qualitative data
    d. Matched-pairs sampling with quantitative data

2. When calculating the standard error of $\bar{X}_1 - \bar{X}_2$, under what assumption do you pool the sample variances s_1^2 and s_2^2 ?
    a. Known population variances
    b. Unknown population variances that are assumed equal
    c. Unknown population variances that are assumed unequal
    d. All of the above

3. A particular personal trainer works primarily with track and field athletes. She believes that her clients run faster after going through her program for six weeks. How might she test that claim?
    a. A hypothesis test for $p_1 - p_2$
    b. A hypothesis test for $\mu_1 - \mu_2$
    c. A matched pairs hypothesis test for μ_D
    d. We are unable to conduct a hypothesis test since the samples would not be independent

4. What is the null hypothesis if we want to test the hypothesis that the mean score on Region 1 is higher than on Region 2?

a. $\mu_1 - \mu_2 = 0$

b. $\mu_1 - \mu_2 \neq 0$

c. $\mu_1 - \mu_2 > 0$

d. $\mu_1 - \mu_2 \leq 0$

5. You would like to determine if there is a higher incidence of smoking among women than among men in a neighborhood. Let men and women be represented by populations 1 and 2, respectively. The relevant hypotheses are constructed as _____.

a. $H_0 : (\mu_1 - \mu_2) \leq 0, H_A : (\mu_1 - \mu_2) > 0$

b. $H_0 : (\mu_1 - \mu_2) \geq 0, H_A : (\mu_1 - \mu_2) < 0$

c. $H_0 : (p_1 - p_2) \leq 0, H_A : (p_1 - p_2) > 0$

d. $H_0 : (p_1 - p_2) \geq 0, H_A : (p_1 - p_2) < 0$

6. For ANOVA two way with interaction, which is correct from the statement below:

a. Ho: there is interaction between Factor A and Factor B

H1: there is no interaction between Factor A and Factor B

b. F=MSE/MSA

c. The degree of freedom for source of variation from row is r-1

d. The degree of freedom for source of variation from error is $n_T - 1$

7. Identify the assumption that is not applicable for a one-way ANOVA test.

a. The populations are normally distributed.

b. The populations standard deviations are not equal.

c. The samples are selected independently.

d. The sample is drawn at random from each population.

8. Between-treatments variability is based on weighted sum of squared differences between the:

a. Population variances and the overall mean of the data set.

b. Sample means and the overall mean of the data set.

c. Sample variances and the overall mean of the data set.

d. Population means and the overall mean of the data set.

9. If the amount of between-treatments variability is significantly greater than the amount of within-treatments variability, then:

a. Reject the null hypothesis of equal population means.

b. Do not reject null hypothesis of equal population means.

c. Conclude that the ratio of between-treatments variability to within-treatments variability is significantly less than 1.

d. Perform further analysis using the two-way ANOVA with interaction test.

10. Which of the following is a feature of the F distribution?
    a. The F distribution depends on one degree of freedom.
    b. The F distribution is bell-shaped with values ranging from negative infinity to infinity.
    c. The F distribution becomes increasingly symmetric when the degrees of freedom increase.
    d. The F distribution is negatively skewed.

11. The chi-square test of a contingency table is a test of independence for:
    a. A single qualitative variables
    b. Two qualitative variables
    c. Two quantitative variables
    d. Three or more quantitative variables

12. For the chi-square test of a contingency table, the expected cell frequencies are found as:
    a. The row total multiplied by the column total dividend by the sample size
    b. The observed cell frequency
    c. $(r - 1)(c - 1)$
    d. $(r)(c)$

13. The chi-square test of a contingency table is valid when the expected cell frequencies are:
    a. Equal to 0
    b. More than 0 but less than 5
    c. At least 5
    d. Negative

14. A χ^2_{df} distribution tends to the _____, as the degrees of freedom increase.
    a. F distribution
    b. Uniform distribution
    c. Student's t distribution
    d. Normal distribution

15. A fund manager wants to know if it equally likely that the Dow Jones Industrial average will go up each day of the week. For each day of the week, the fund manager observes the following number of days when the Dow Jones Industrial average goes up.

Day of Week	Day	Observed
1	Monday	192
2	Tuesday	189
3	Wednesday	202
4	Thursday	199
5	Friday	218

For the goodness-of-fit test, what are the degrees of freedom for the chi-squared test statistic?

a. 4
b. 5
c. 6
d. 7

Part 2. Essay questions (@20 points):

Problem 1

A cereal manufacturer, KOWKOW KRUNCH is considering three alternatives box colors –red, yellow, and blue. To check whether such a consideration has any effect on sales, the boxes are sent to some stores, where each store receive each of the three box colors. After a week, check is made on the number of sales in each store. The results of sales from each color of boxes are shown in the following table.

Red	Yellow	Blue
7	15	11
12	12	7
11	15	6
15	8	9
9		7
14		

The manufacturer thus want to test whether the population mean sales level are the same for all three box colors. Perform the hypothesis test to answer the manufacturers question, by:

1. Constructing the null and alternative hypothesis **(5 points)**
2. Helping KOWKOW KRUNCH test their hypothesis, Use $\alpha=0.01$ **(10 points)**
3. What is the conclusion? **(5 points)**

Problem 2

Delta Food Grocery has two grocery stores located in Bije District. One store is located at Usman Street and the other at Sinda Street and each is run by a different manager. Each manager claims that her store's layout maximizes the amount of impulsive customers. Both managers surveyed a sample of their customers and asked them how much more they spent than they had planned to, in other words, how much did they spend on impulse (spending more than they had planned). The following table shows the sample data collected from the two stores.

Usman Street (IDR thousand)	Sinda Street (IDR thousand)
15.78	15.19
17.73	18.22
10.61	12.87
15.79	17.79
13.45	13.96
12.85	13.74
	12.47
	10.83
$S_1 = 2.5529$	$S_2 = 2.5675$

Upper-level management at Delta Foods Grocery wants to know if there is a difference in the mean amounts of impulsive consumption between the two stores. Perform the statistical analysis through:
1. Constructing the null and alternative hypothesis **(5 points)**
2. Help the two managers test their hypothesis, Use $\alpha=0.05$ **(10 points)**
3. What is the conclusion? **(5 points)**

Problem 3

A university has six colleges and takes a poll to gauge student support for a tuition increase. The university wants to insure each college is represented fairly. The below table shows the observed number students that participate in the poll from each college and the actual proportion of students in each college.

College	Obsereved	Actual proportion of students in each college
1	457	0.2
2	206	0.08
3	301	0.13
4	792	0.29
5	336	0.15
6	373	0.15

Based on the number of students participating in the poll, can you say that each college is represented fairly? To answer the question;
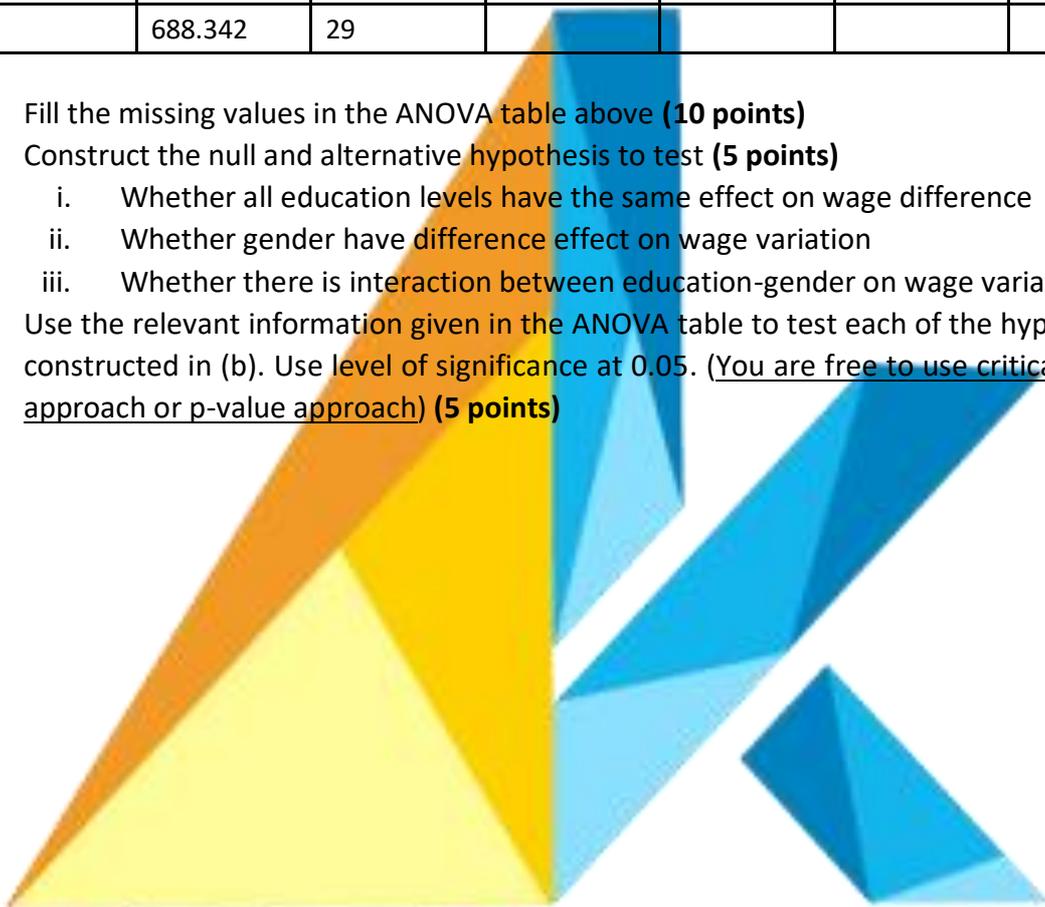1. Construct the null and alternative hypothesis **(5 points)**
2. Test the hypothesis, Use $\alpha=0.05$ **(10 points)**
3. What is the conclusion? **(5 points)**

Problem 4

A researcher is conducting a study on labor wage variation determinants in PT XYZ. He suspects that labor's education and gender might cause wage to differ. He also expects that there is coincidence effect between education and gender to affect wage variation. In order to validate his prior expectation, he then conducts survey, with three replications for each pair of factors and use ANOVA method to analyze the result. In the study, the first factor (education) were divided into 5 level of educations; (1)PhD, (2)Graduate, (3)Undergraduate, (4)Diploma and (5)High school or below. The following table shows partial result from ANOVA analysis at 5% level of significance.

Source of Variation	SS	df	MS	F	P-value	F crit
Level of education	683.383	……….	……….	……….	2.27E-27	……….
Sex/gender	2.408	……….	……….	……….	2.88E-06	……….
Education X sex		……….	0.34575	……….	0.0026	……….
Error	1.167	……….	……….			
Total	688.342	29				

a. Fill the missing values in the ANOVA table above **(10 points)**
b. Construct the null and alternative hypothesis to test **(5 points)**
   i. Whether all education levels have the same effect on wage difference
   ii. Whether gender have difference effect on wage variation
   iii. Whether there is interaction between education-gender on wage variation
c. Use the relevant information given in the ANOVA table to test each of the hypothesis constructed in (b). Use level of significance at 0.05. (You are free to use critical value approach or p-value approach) **(5 points)**