## ECO 173: Applied Statistics

## Sample Final Exam

## A) Matched Pair Experiment

For hypothesis test comparing two population means using matched pairs experiment design, apply these following steps (you are required to memorize these steps and apply them in exact order in the exam):

Step 1: Calculate the differences, the mean of differences and the standard deviation of differences (if these are not provided already in the question)

Step 2: State the theory you are testing. Make sure to mention which is population 1 and which is population 2.
Step 3: State the null and alternative hypothesis, the sampling distribution of the test statistic, and whether it is a left tail/right tail/two-tail test.

Step 4: Compute the test statistic (t - statistic of differences).
Step 5: Identify the rejection region. Include a sketch of the sampling distribution.

## Step 6: State your decision

* use one of these formulas to compute variance of differences:

Here n is the number of pairs.

$$
s^{2}=\frac{\sum\left(x_{i}-\bar{x}\right)^{2}}{n-1} \quad s^{2}=\frac{\sum x_{i}^{2}-\frac{\left(\sum x_{i}\right)^{2}}{n}}{n-1}
$$

Problem 1: A state legislator wants to determine whether her voter's performance rating ( $0-100$ ) has changed from last year to this year. The following table shows the legislator's performance rating for the same 16 randomly selected voters for last year and this year.

| Voter | Rating (last <br> year) | Rating (this <br> year) |
| :---: | :---: | :---: |
| 1 | 60 | 56 |
| 2 | 54 | 48 |
| 3 | 78 | 70 |
| 4 | 84 | 60 |
| 5 | 91 | 85 |
| 6 | 25 | 40 |
| 7 | 50 | 40 |
| 8 | 65 | 55 |
| 9 | 68 | 80 |


| 10 | 81 | 75 |
| :---: | :---: | :---: |
| 11 | 75 | 78 |
| 12 | 45 | 50 |
| 13 | 62 | 50 |
| 14 | 79 | 85 |
| 15 | 58 | 53 |
| 16 | 63 | 60 |

a. What is the blocking variable used here to collect the sample data?
b. At $1 \%$ significance level, is there enough evidence to conclude that the legislator's performance rating has changed? Assume the performance ratings are normally distributed.

Problem 2: Police trainees were seated in a darkened room facing a projector screen and given a memory test. A matched pair random sample of 15 trainees who tool this test were then given a weeklong memory training course. The results are listed at follows:

| Police trainee | Memory score after <br> training | Memory score before <br> training |
| :---: | :---: | :---: |
| 1 | 6 | 6 |
| 2 | 8 | 5 |
| 3 | 6 | 6 |
| 4 | 7 | 5 |
| 5 | 9 | 7 |
| 6 | 8 | 5 |
| 7 | 9 | 4 |
| 8 | 6 | 6 |
| 9 | 7 | 7 |
| 10 | 5 | 8 |
| 11 | 9 | 4 |
| 12 | 8 | 5 |
| 13 | 6 | 4 |
| 14 | 8 | 6 |
| 15 | 6 | 7 |

Test at 5\% level of significance, whether the memory course improved the ability of the trainees to correctly identify license plates.

## B) Analysis of Variance (ANOVA)

For hypothesis tests comparing two or more population means using the ANOVA, apply these following steps (you are required to memorize these steps and apply them in exact order in your exam):

Step 1: State the theory and the null and alternative hypotheses. Mention the sampling distribution of the test statistic.

## Step 2: Calculate the F statistic:

1. Calculate SST
2. Calculate SSE
3. Calculate M ean Squares
4. Calculate the F statistic

Step 3: Identify the rejection region. Remember for one-way ANOVA, the rejection region is always $F \geq \boldsymbol{F}_{\alpha, k-1, n-k}$.

## Step 4: State your decision.

Problem 3: The National Transportation Safety Board (NTSB) wants to examine the safety of compact cars, midsize cars, and full-size cars. It collects a sample on the pressure applied to the driver's head during a crash test for each of the three treatments (cars types). Using the hypothetical data provided below, test using ANOVA whether the mean pressure applied to the driver's head during a crash test is equal for each types of car. Use $\alpha=5 \%$.

|  | Compact cars | Midsize cars | Full-size cars |
| :---: | :---: | :---: | :---: |
|  | 643 | 469 | 484 |
|  | 655 | 427 | 456 |
|  | 702 | 525 | 402 |
| $\bar{X}$ | 666.67 | 473.67 | 447.33 |
| S | 31.18 | 49.17 | 41.68 |

Present the data in an ANOVA table.

## Problem 4:

The following ANOVA table, based on information obtained for three samples selected from three independent populations that are normally distributed with equal variances, has a few missing values.

| Source of Variation | Degrees of Freedom | Sum of Squares | Mean <br> Square | Value of the Test Statistic |
| :---: | :---: | :---: | :---: | :---: |
| Between | 2 |  | 19.2813 | $F=\square=$ |
| Within | 89.3677 |  |  |  |
| Total | 12 |  |  |  |

a. Find the missing values and complete the ANOVA table.
b. Using $\alpha=.01$, what is your conclusion for the test with the null hypothesis that the means of the three populations are all equal against the alternative hypothesis that the means of the three populations are not all equal?

## C) Simple Linear Regression

Problem 5: Answer the questions using the data below:

## Annual Bonus and Years of Experience

The annual bonuses ( $\$ 1,000$ s) of six employees with different years of experience were recorded as follows. We wish to determine the straight-line relationship between annual bonus and years of experience.

| Years of experience $\boldsymbol{x}$ | 1 | 2 | 3 | 4 | 5 | 6 |
| :--- | :--- | :--- | :--- | :--- | :---: | :---: |
| Annual bonus $y$ | 6 | 1 | 9 | 5 | 17 | 12 |

a) Does years of experience depend on annual bonus or does annual bonus depend on driving experience? Do you expect a positive or a negative relationship between these two variables?
b) Compute $\mathrm{SS}_{x x}, \mathrm{SS}_{y y}$, and $\mathrm{SS}_{\mathrm{xy}}$.
c) Find the least squares regression line by choosing appropriate dependent and independent variables based on your answer in part a.
d) Interpret the meaning of the values of $a$ and $b$ calculated in part $c$.
e) Plot the scatter diagram and the regression line. (use a graphing paper)
f) Calculate $r$ and $r^{2}$ and explain what they mean.
g) Predict annual bonus for an employee with 10 years of experience.
h) Compute the standard deviation of errors.
i) Construct a $90 \%$ confidence interval for B.

Problem 6: The values of $y$ and their corresponding values of $y$ are shown in the table below

| $x$ | 0 | 1 | 2 | 3 | 4 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| $y$ | 2 | 3 | 5 | 4 | 6 |

a) Compute $\mathrm{SS}_{x x}, \mathrm{SS}_{y y}$, and $\mathrm{SS}_{x y}$.
b) Find the least squares regression line by choosing appropriate dependent and independent variables based on your answer in part a.
c) Interpret the meaning of the values of $a$ and $b$ calculated in part $c$.
d) Plot the scatter diagram and the regression line. (use a graphing paper)
e) Calculate $r$ and $r^{2}$ and explain what they mean.
f) Predict $y$ for $x=7$.
h) Compute the standard deviation of errors.
i) Construct a 95\% confidence interval for B.
D) M ultiple Regression M odel
(see handout)
Multiple Choice Questions:

1. Suppose that in an ANOVA analysis of distribution of income across four regions in Bangladesh, the value of the observed F-statistic appeared larger than the critical value. Based on that information, the following conclusion could be made:
a. The mean incomes of the four regions are exactly equal to each other.
b. The mean incomes of the four regions are not equal to each other.
c. There is not enough evidence to prove that the incomes are the same.
d. There is no statistically significant difference among mean incomes.
2. One-way ANOVA is applied to independent samples taken from three normally distributed populations with equal variances. Which of the following is the null hypothesis for this procedure?
a. $\mu_{1}+\mu_{2}+\mu_{3}=0$
b. $\mu_{1}+\mu_{2}+\mu_{3} \neq 0$
c. $\mu_{1}=\mu_{2}=\mu_{3}$
d. $\mu_{1}=\mu_{2}=\mu_{3}=0$
3. In a completely randomized design for ANOVA, the numerator and denominator degrees of freedom are 4 and $\mathbf{2 5}$, respectively. The total number of observations must equal:
a. 24
b. 25
c. 29
d. 30

Make sure to be well-versed in the different theories of the four topics of your final exam. Follow handouts and slides closely.

