1-Way
ANOVA
Laboratory

One-Way Analysis of Variance (ANOVA): Comparing More than Two Means

OVERVIEW

In this lab, you will be using a one-way analysis of variance (ANOVA) to compare **more than two population means**.

OBJECTIVES

By the end of this laboratory, you will be able to:

- Perform and interpret a one-way analysis of variance using *Minitab*
- Perform and interpret a multiple comparison procedure to find which groups are significantly different from each other.

EQUIPMENT

- PC with *Minitab*
- Computer diskette to save files

BACKGROUND MATERIAL

Statistical Terms and Topics

- Response variable or dependent variable
- Factors
- Quantitative factors
- Qualitative factors
- Factor levels
- Treatments
- Experimental unit
- Designed experiment
- Observational experiment
- Sum of Squares for Treatment (SST)
- Sum of Squares for Error (SSE)
- Mean Square for Treatments (MST)
- Mean Square for Error (MSE)
- "F" test statistic
- Multiple comparison procedure
- Tukey's Method

Scenario

Forty students (10 freshmen, 10 sophomores, 10 juniors, and 10 seniors) were asked to report their GPA to one decimal point.

Exercise

INSTRUCTIONS

1. How do your believe that the data from the groups should relate. **Hint: Do you have any concrete evidence that the average of the data from any one group should be specifically larger or smaller than the average of the data from any other group.* Write out your hypothesis below.

2. Now formulate the hypothesis using statistical notation.
   The **null hypothesis** is denoted H_o.
   The **alternative / research hypothesis** is denoted H_a.
   The general forms of the null and alternative hypotheses are:

   H_o: $\mu_1 = \mu_2 = \ldots \mu_k$
   H_a: at least two of the k treatments differ

   Now write the specific null and alternative hypothesis for the present scenario.

3. The test statistic for ANOVA is the F statistic. The formula to find F is:

$$F = \frac{MeanSquareTreatment(MST)}{MeanSquareError(MSE)}$$

Look back at you definitions for MST and MSE and write out why this makes sense. (*Hint: It may help to draw a diagram of the groups.)

____
____
____

DATA

Freshman	Sophomores	Juniors	Seniors
2.5	2.3	3.5	2.3
2.2	1.6	2.9	2.5
1.4	2.8	1.6	2.0
2.6	3.2	3.3	3.1
2.4	2.4	3.6	2.7
1.4	3.4	3.1	3.9
2.4	2.4	3.2	3.0
2.6	3.4	2.6	3.9
2.3	3.0	3.7	3.8
2.9	3.4	3.8	3.8

COMPUTER EXERCISE

1. Enter the data into the columns and name them **Fresh**, **Soph**, **Junior**, and **Senior**.
2. Run descriptive statistics on the sets of data.

   What are the means and standard deviations for each set of data?
   Fresh: mean _____ standard deviation _____
   Soph: mean _____ standard deviation _____
   Junior: mean _____ standard deviation _____
   Senior: mean _____ standard deviation _____

3. The assumptions for ANOVA are listed below. For each assumption, indicate whether or not you think the assumption is reasonable for this scenario.
   1. Samples are selected randomly and independently from the respective populations. _______________
   2. All *k* population probability distributions are normal. ____________
   3. The *k* population variances are equal. ___________

4. First, choose a maximum value of α that you are willing to tolerate.

α = _____

5. Perform the ANOVA on *Minitab*.

    - Go to **Stat> ANOVA>One-way (Unstacked)**
    - Select each of the four columns
    - Click **Ok**

    F test statistics = _____ *p*-value = _____ df = _____

6. Interpret the results of the one-way ANOVA.
    If the observed significance level (*p*-value) of the F-test statistic is less than the chosen value of α, reject the null hypothesis in favor of the alternative. Otherwise, do not reject the null hypothesis.
    *Note that rejecting the null hypothesis only lets you know that at least one mean is significantly different. You don't know which mean is significantly different, or how many means are different from each other.
    Based on the *p*-value and α, what do you conclude about the test?
    ____

7. Go back to step number 5 and look at the means and standard deviations of the three sets of data. Also, look back over your answer to number 3. Based on all of this information, which group(*s*) look as if they may be significantly different from each other? Why?
    ____
    ____
    ____

8. [***Multiple comparison procedures*** are higher-level statistical processes that are used to determine whether the means are different for all possible pairs of the factors. These multiple comparisons are used as a follow-up when significant differences are detected in population means.]
    Since there was a significant difference found between at least two population means for the ANOVA test, you will be using a multiple comparison procedure to determine which of the population means were significantly different from each other. **Tukey's multiple comparison test** is used when the sample sizes of the treatments are equal. Therefore, we will be using Tukey's test.
    **Note: there are other tests to use if the sample sizes of the treatments are equal*.

9. Perform the Tukey test.
    To use a multiple comparison procedure in *Minitab*, the data must be stacked.
    First, stack the data

- Go to **MANIP > STACK**
- Select the **columns** that you want to stack.
- Mark **Store Stacked Data in Column of worksheet**.
- Check **Use variable names in subscript columns**.
- Click **OK**.
- Name the response column “GPA”, and the factor column “Factor”.

Next, run the ANOVA on the stacked data.

- Go to **Stat** > **ANOVA** > **One-way**
- Select **GPA** for Response
- Select **Factor** for Factor
- Click **Comparison** Button
- Select **Tukey**
- Enter **5** in the family error
- Click **OK**

Look at the confidence intervals for each of the pairs of factors. Which pair(s) contain a significant difference? (*Note: Recall that if an interval does not contain zero, there is a statistically significant difference between the corresponding means. If the interval does contain zero, the difference between the means is not statistically significant.)

$\mu_1 - \mu_2 =$ ______ $\mu_2 - \mu_3 =$ ______ $\mu_1 - \mu_3 =$ ______ $\mu_3 - \mu_4 =$ ______
$\mu_1 - \mu_4 =$ ______ $\mu_2 - \mu_4 =$ ______

Family error rate = _____

10. Use complete sentences to state the conclusions of this analysis.

Application to Psychology

The power of the test is stronger if the number of data points is the same in each factor.

Ethics Application

Keep the identities of the individuals separate from the data, which is in this case the GPA's