Hypothesis testing

OBJECTIVES

Population mean

1. Describe the data and hypotheses for one-way ANOVA
2. Check the assumptions for one-way ANOVA
3. Perform a one-way ANOVA hypothesis test
4. Perform a one-way ANOVA test with technology
5.

one : μ z or t-distribution
two : μ_1, μ_2 t-distribution
more than two means : $\mu_1, \mu_2, \mu_3, \ldots, \mu_n$ f-distribution

OBJECTIVE 1
DESCRIBE THE DATA AND HYPOTHESES FOR ONE-WAY ANOVA

TESTING MORE THAN TWO POPULATION MEANS

We have already studied tests for a null hypothesis that two populations had equal means. There are situations where we want to compare three, four or more populations. In these situations, when appropriate assumptions are satisfied, we may test the hypothesis that all the population means are equal.

This is done with a method called analysis of variance (ANOVA).

Welds, used to join metal parts, are made by heating a powdered material called flux. In a study conducted by G. Fredrickson at the Colorado School of Mines, four welding fluxes, with differing chemical compositions, were prepared. The purpose of the study was to determine whether the welds made from the fluxes would have different degrees of hardness. Five welds using each flux were made, and the hardness of each was measured in Brinell units; the higher the number, the harder the weld.

We could think about the data that is generated by each welding flux as a different population. Furthermore, we may want to compare the mean performances of each welding flux. This may be a possible situation to apply an ANOVA. Data were gathered from the fluxes, along with sample means and standard deviations.

Flux	Sample Values	Sample Mean	Sample Standard Deviation
A	250 264 256 260 239	253.8	9.7570
B	263 254 267 265 267	263.2	5.4037
C	257 279 269 273 277	271.0	8.7178
D	253 258 262 264 273	262.0	7.4498

We observe from the table that flux C produced the largest sample mean and flux A produced the smallest. However, these results may change if the experiment was repeated and a different sample was selected.

The question we want to answer is this: **Can we conclude that there are differences in the population means among the four flux types?**

VARIABLES

In the previous example, we can identify a few variables. One variable is an outcome variable, hardness. The other variable is an explanatory variable, flux type.

The explanatory variable is qualitative, and is called a **factor**. The outcome variable is called the **response**. The different values for the factor are called **treatments**. Therefore, flux A, B, C, and D are treatments.

Since there is only one factor, which is flux type, this is called a **one-factor experiment**.

ONE-WAY ANOVA

Since the table involves only one factor, it is called **one-way table**, and the method of determining whether the population means differ is called **one-way ANOVA.**

Flux	Sample Values	Sample Mean	Sample Standard Deviation
A	250 264 256 260 239	253.8	9.7570
B	263 254 267 265 267	263.2	5.4037
C	257 279 269 273 277	271.0	8.7178
D	253 258 262 264 273	262.0	7.4498

OBJECTIVE 2
CHECK THE ASSUMPTIONS FOR A ONE-WAY ANOVA

Data for a one-way analysis of variance consist of samples from several populations. The methods of one-way ANOVA require that the populations satisfy the following assumptions:

1) We have independent simple random samples from three or more populations
2) Each of the populations must be approximately normal
3) The population most all have the same variance, θ^2.

BALANCED VS. UNBALANCED DESIGN

When creating a one-way analysis of variance, we have a choice between a balanced or unbalanced designs.

The **balanced design** is when the sample sizes are all the same. Although one-way analysis of variance can be used with both designs, the balanced design offers a big advantage. A balanced design is much *less sensitive* to violations of the assumption of equality of variance than an unbalanced one.

CHECKING THE ASSUMPTIONS

When the sample sizes are small, it is often the case that there is no good way to check the assumptions of normality and equality of variance. We are fortunate in that the methods of analysis of variance work well unless the assumptions are *severely* violated. One way to check for normality is with a dotplot. When the design is balanced, the method of ANOVA is not sensitive to the assumption of equal population variances.

When the design is unbalanced, we can check the assumption of equal population variances by computing the following quotient:

$$\frac{\text{largest sample standard deviation}}{\text{smallest sample standard deviation}}$$

If the result is less than ____2____, we conclude that the assumption of equal population variances is satisfied.

OBJECTIVE 3
PERFORM A ONE-WAY HYPOTHESIS TEST

Let's suppose that we have I samples, each from a different population. The table below presents a summary of the various notation for conducting an ANOVA test.

Sample	Population Mean	Sample Mean	Sample Standard Deviation	Sample Size
1	μ_1	$\bar{x}_1$	s_1	n_1
2	μ_2	$\bar{x}_2$	s_2	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I	μ_I	$\bar{x}_I$	s_I	n_I

N is the total number of items in all samples combined.
$\bar{\bar{x}}$ is the grand mean, the average of all the items in all samples combined.

EXAMPLE: For the data in the table below, find I, n_1, n_2, $\cdots$, $n_I, \bar{x}_3$, s_2, N, $\bar{\bar{x}}$.

Flux	Sample Values	Sample Mean	Sample Standard Deviation
A	250 264 256 260 239	253.8	9.7570
B	263 254 267 265 267	263.2	5.4037
C	257 279 269 273 277	271.0	8.7178
D	253 258 262 264 273	262.0	7.4498

SOLUTION: There are 4 samples, so $I = 4$. Each sample contains 5 observations, so $n_1 = n_2 = n_3 = n_4 = 5$. The quantity $\bar{x}_3$ is the sample mean of third sample. This value is $\bar{x} = 271.0$, The total number of observations is $N = 20$. Finally, the grand mean is found by averaging all 20 observations; the grand mean $\bar{\bar{x}} = 262.5$, and $s_2 = 5.4037$

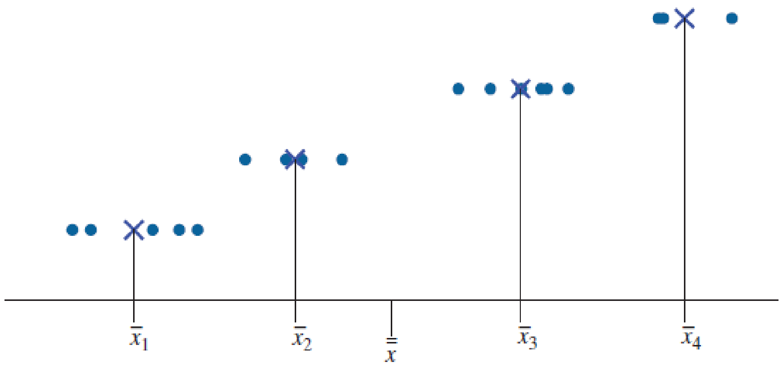
Hypotheses for One-Way ANOVA

Recall that the question that we want to answer is this: Can we conclude that there are differences in the population means among the four flux types? In other words, are the population means all the same or are two or more of the means different?

Therefore, the hypotheses for one-way ANOVA are:

$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_I$

H_1: Two or more of the μ_I are different



The figure illustrates several hypothetical samples from different treatments, along with their sample means and the sample grand mean. The sample means are spread out around the sample grand mean. One-way ANOVA provides a way to measure this spread. If the sample means are highly spread out, then it is likely that the treatment means are different, and we will reject H_0.

TREATMENT SUM OF SQUARES ($SSTr$)

The spread of the sample means around the sample grand mean, $\bar{\bar{x}}$, is measured by a quantity called the **treatment sum of squares** ($SSTr$).

$$SSTr = n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \dots + n_I(\bar{x}_I - \bar{\bar{x}})^2$$

Notice that each term in the $SSTr$ involves the distance from the sample mean to the sample grand mean. Also, each squared distance is multiplied by the sample size corresponding to its sample mean, so that the means for the larger samples count more.

The $SSTr$ provides an indication of how different the treatment means are from each other. If the $SSTr$ is large, then the sample means are spread out widely, and it is reasonable to conclude that the treatment means differ and to reject H_0. If, on the other hand, the $SSTr$ is small, then the sample means are all close to each other, so it is plausible that the treatment means are equal.

EXAMPLE: For the data in the table below, compute $SSTr$.

Flux	Sample Values	Sample Mean	Sample Standard Deviation
A	250 264 256 260 239	253.8	9.7570
B	263 254 267 265 267	263.2	5.4037
C	257 279 269 273 277	271.0	8.7178
D	253 258 262 264 273	262.0	7.4498

SOLUTION: The sample means are $\bar{x}_1 = 253.8$, $\bar{x}_2 = 263.2$, $\bar{x}_3 = 271.0$ and $\bar{x}_4 = 262.0$.

The sample site are $n_1 = n_2 = n_3 = n_4 = 5$. The sample grand mean $\bar{\bar{x}} = 262.5$

$$SSTr = 5(253.8-262.5)^2 + 5(263.2-262.5)^2 + 5(271.0-262.5)^2 + 5(262.0-262.5)^2$$
$$= 743.4$$

ERROR SUM OF SQUARES (SSE)

Previous we stated that if the $SSTr$ is large, then the sample means are spread out widely, and it is reasonable to conclude that the treatment means differ and reject H_0. How do we determine what is large?

To determine whether $SSTr$ is large enough, we compare it to the **error sum of squares** (SSE). SSE measures the spread in the individual sample points around their respective sample means. This spread is measured by combining the sample variances.

$$SSE = (n_1-1)s_1^2 + (n_2-1)s_2^2 + \dots + (n_I-1)s_I^2$$

The size of SSE depends only on the standard deviations of the samples, and is not affected by the location of treatment means relative to one another.

EXAMPLE: For the data in the table below, compute SSE.

Flux	Sample Values	Sample Mean	Sample Standard Deviation
A	250 264 256 260 239	253.8	9.7570
B	263 254 267 265 267	263.2	5.4037
C	257 279 269 273 277	271.0	8.7178
D	253 258 262 264 273	262.0	7.4498

SOLUTION: The sample site are $n_1 = n_2 = n_3 = n_4 = 5$

The sample standard deviations are $s_1 = 9.7570$, $s_2 = 5.4037$, $s_3 = 8.7178$, $s_4 = 7.4498$

$SSE = (5-1)9.7570^2 + (5-1)5.4037^2 + (5-1)8.7178^2 + (5-1)7.4498^2$

$= 1023.6$

DEGREES OF FREEDOM

The $SSTr$ and SSE have different degrees of freedom.

The $SSTr$ involves the deviations of the I sample means around the grand mean, so

Degrees of freedom for $SSTr$ = $I-1$ where I is the number of samples

The SSE involves the I sample variances. The degrees of freedom for the sample variances are $n_1 - 1,\ n_2 - 1,\ \ldots,\ n_I - 1$. The degrees of freedom for SSE is the sum of these, so

Degrees of freedom for SSE = $N - I$ where N is the total number in all the samples combine and I is the number of samples

MEAN SQUARES

If we divide either of the sum of squares ($SSTr$ or SSE) by its degrees of freedom, the resulting quantity is called a **mean square**.

Treatment mean square is denoted $MSTr$.

$$MSTr = \frac{SSTr}{I-1}$$

Error mean square is denoted MSE.

$$MSE = \frac{SSE}{N-I}$$

TEST STATISTIC

$\boldsymbol{MSE}$ measures how spread out the samples are around their own means. It is an estimate of σ^2, which is the variance of each of the populations.

$\boldsymbol{MSTr}$ measures how spread out the samples are around the grand mean, which is affected by σ^2, and how spread out the population means are.

If H_0 is true, and all populations have the same mean, then $MSTr$ is an estimate of σ^2, just as MSE is. When the population means differ, $MSTr$ tends to be larger than σ^2, and thus larger than MSE. The test statistic for testing H_0: $\mu_1 = \cdots = \mu_I$ is $\boxed{F = \frac{MSTr}{MSE}}$

When H_0 is true, the numerator and denominator of F are the same size, so value of $F = \frac{MSTr}{MSE}$ tends to be near 1. When H_0 is false, the $MSTr$ tends to be larger, so value of $F = \frac{MSTr}{MSE}$ tends to be greater than 1.

When H_0 is true, the test statistic has an F distribution with $I - 1$ and $N - I$ degrees of freedom, denoted by $F_{I-1,\,N-I}$.

F TEST FOR ONE-WAY ANOVA

To test H_0: $\mu_1 = \mu_2 \cdots = \mu_I$ versus H_1: Two or more of the u_i are different:

Step 1. Check that the assumptions are satisfied.

Step 2. Choose a level of significance α.

Step 3. Compute $SSTr$ and SSE.

Step 4. Compute $MSTr$ and MSE.

Step 5. Compute the test statistic: $F = \frac{MSTr}{MSE}$

Step 6. Find the critical value by using Table A.5 with $I - 1$ and $N - I$ degrees of freedom. If the value of the test statistic is greater than or equal to the critical value, reject H_0. Otherwise do not reject H_0. Use the p-value to reject or do not reject H_0.

Step 7. State a conclusion.

EXAMPLE: For the data below, test the null hypothesis that all the means are equal. Use the $\alpha = 0.05$ level of significance. What do you conclude?

Flux	Sample Values	Sample Mean	Sample Standard Deviation
A	250 264 256 260 239	253.8	9.7570
B	263 254 267 265 267	263.2	5.4037
C	257 279 269 273 277	271.0	8.7178
D	253 258 262 264 273	262.0	7.4498

SOLUTION: Previously, we have already checked the assumptions of normality. The design is balanced, so the procedure is not very sensitive to violations of all the assumption of equal variances. In any event, the largest sample standard deviation is less than twice as large as the smallest. The assumptions are satisfied.

We $SSTr = 743.4$ and $SSE = 1023.6$. We have $I = 4$ samples and $N = 20$ observations in all the samples taken together.

Thus, $MSTr = \frac{SSTr}{I-1} = \frac{743.4}{4-1} = 247.8$ and

$$MSE = \frac{SSE}{N-I} = \frac{1023.6}{20-1} = 63.975$$

The test statistic is $F = \frac{MSTr}{MSE} = \frac{247.8}{63.975} = 3.8734$

using TI-84, ANOVA(L1, L2, L3, L4)

p-value = 0.0294

p-value < α, we reject H_0

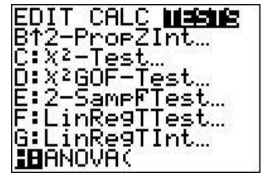
we conclude that the mean hardness is not the same for all fluxes. There are at least two fluxes with different means.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: two or more μ_i are different

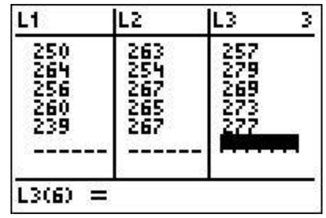
HYPOTHESIS TESTING ON THE TI-84 PLUS

The **ANOVA** command will perform a one-way ANOVA hypothesis test. This command is accessed by pressing **STAT** and highlighting the **TESTS** menu.

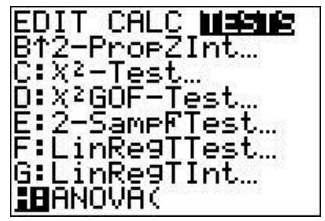


Test the null hypothesis that all the means are equal. Use the α = 0.05 level of significance. What do you conclude?

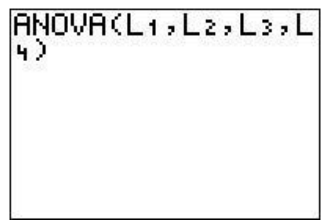
Solution:
We enter the values from flux A into **L1**, the values from flux B into **L2**, the values from flux C into **L3**, and the values from flux D into **L4**.



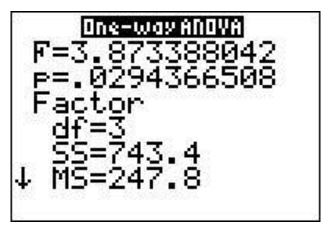
Press **STAT** and highlight the **TESTS** menu. Select **ANOVA** and press **ENTER**.



Enter **L1, comma, L2, comma, L3, comma, L4**, and then press **Enter.**



The *P*-value is 0.029 which is less than α = 0.05. We reject the null hypothesis and conclude that the mean hardness is not the same for all fluxes. There are at least two fluxes with different means.



YOU SHOULD KNOW …

- The hypotheses for one-way ANOVA
- The variables involved in a one-way ANOVA
- The assumptions for a one-way ANOVA
- The advantages of a balanced design
- How to check that the assumptions are satisfied for one-way ANOVA
- The notation involved in a one-way ANOVA
- How to perform the F-test for one-way ANOVA
- How to find the various degrees of freedom