Analysis-of-Variance Technique

appropriate methods for testing the hypothesis

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_k,$$
$$H_1\colon\ \text{At least two of the means are not equal.}$$

F-Test in ANOVA

The null hypothesis H_0 is rejected at the α-level of significance when

$$f > f_\alpha[k-1, k(n-1)].$$

Table 13.3: Analysis of Variance for the One-Way ANOVA

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	SSA	$k-1$	$s_1^2 = \dfrac{SSA}{k-1}$	$\dfrac{s_1^2}{s^2}$
Error	SSE	$k(n-1)$	$s^2 = \dfrac{SSE}{k(n-1)}$	
Total	SST	$kn-1$		

Example 13.1:

Table 13.1: Absorption of Moisture in Concrete Aggregates

Aggregate:	**1**	**2**	**3**	**4**	**5**	
	551	595	639	417	563	
	457	580	615	449	631	
	450	508	511	517	522	
	731	583	573	438	613	
	499	633	648	415	656	
	632	517	677	555	679	
Total	3320	3416	3663	2791	3664	16,854
Mean	553.33	569.33	610.50	465.17	610.67	561.80

Test the hypothesis $\mu_1 = \mu_2 = \cdots = \mu_5$ at the 0.05 level of significance for the data of Table 13.1 on absorption of moisture by various types of cement aggregates.

***Solution*:** The hypotheses are

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_5,$$
$$H_1\colon\ \text{At least two of the means are not equal.}$$
$$\alpha = 0.05.$$

Critical region: $f > 2.76$ with $v_1 = 4$ and $v_2 = 25$ degrees of freedom. sum-of-squares computations give

$$SST = 209{,}377, \quad SSA = 85{,}356,$$
$$SSE = 209{,}377 - 85{,}356 = 124{,}021.$$

Source	DF	Sum of Squares	Mean Square	F Value
Model	4	85356.4667	21339.1167	4.30
Error	25	124020.3333	4960.8133	
Corrected Total	29	209376.8000		

H_0 is rejected at the α-level of significance when

$$f > f_\alpha[k-1, k(n-1)].$$

Decision: Reject H_0 and conclude that the aggregates do not have the same mean

13.3 In an article "Shelf-Space Strategy in Retailing," published in *Proceedings: Southern Marketing Association*, the effect of shelf height on the supermarket sales of canned dog food is investigated. An experiment was conducted at a small supermarket for a period of 8 days on the sales of a single brand of dog food, referred to as Arf dog food, involving three levels of shelf height: knee level, waist level, and eye level. During each day, the shelf height of the canned dog food was randomly changed on three different occasions. The remaining sections of the gondola that housed the given brand were filled with a mixture of dog food brands that were both familiar and unfamiliar to customers in this particular geographic area. Sales, in hundreds of dollars, of Arf dog food per day for the three shelf heights are given. Based on the data, is there a significant difference in the average daily sales of this dog food based on shelf height? Use a 0.01 level of significance.

Shelf Height		
Knee Level	**Waist Level**	**Eye Level**
77	88	85
82	94	85
86	93	87
78	90	81
81	91	80
86	94	79
77	90	87
81	87	93

Solution:

The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_6,$$
$$H_1 : \text{At least two of the means are not equal.}$$

$\alpha = 0.05.$
Critical region: $f > 2.77$ with $v_1 = 5$ and $v_2 = 18$ degrees of freedom.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatment	5.34	5	1.07	0.31
Error	62.64	18	3.48	
Total	67.98	23		

Decision: The treatment means do not differ significantly.

13.7 It has been shown that the fertilizer magnesium ammonium phosphate, $MgNH_4PO_4$, is an effective sup-

Treatment

50 g/bu		100 g/bu		200 g/bu		400 g/bu	
13.2	12.4	16.0	12.6	7.8	14.4	21.0	14.8
12.8	17.2	14.8	13.0	20.0	15.8	19.1	15.8
13.0	14.0	14.0	23.6	17.0	27.0	18.0	26.0
14.2	21.6	14.0	17.0	19.6	18.0	21.1	22.0
15.0	20.0	22.2	24.4	20.2	23.2	25.0	18.2

Can we conclude at the 0.05 level of significance that different concentrations of $MgNH_4PO_4$ affect the average attained height of chrysanthemums? How much $MgNH_4PO_4$ appears to be best?

Solution:

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$,
H_1 : At least two of the means are not equal.

$\alpha = 0.05$.
Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Treatments	119.787	3	39.929	2.25
Error	638.248	36	17.729	
Total	758.035	39		

with P-value=0.0989.
Decision: Fail to reject H_0 at level $\alpha = 0.05$.

13.18 The following data are values of pressure (psi) in a torsion spring for several settings of the angle between the legs of the spring in a free position:

Angle (°)					
67	**71**	**75**		**79**	**83**
83	84	86	87	89	90
85	85	87	87	90	92
	85	88	88	90	
	86	88	88	91	
	86	88	89		
	87	90			

Compute a one-way analysis of variance for this experiment and state your conclusion concerning the effect of angle on the pressure in the spring. (From C. R. Hicks, *Fundamental Concepts in the Design of Experiments*, Holt, Rinehart and Winston, New York, 1973.)

Solution:

The hypotheses are

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_5,$$
H_1 : At least two of the means are not equal.

Computation:

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed f
Angles	99.024	4	24.756	21.40
Error	23.136	20	1.157	
Total	122.160	24		

with P-value< 0.0001.

Decision: Reject H_0.

There is a significant difference in mean pressure

Practice: Neil A Wesis

In Exercises 16.24–16.29, we have provided data from independent simple random samples from several populations. In each case, determine the following items.

a. SSTR b. MSTR c. SSE d. MSE e. F

16.24

Sample 1	Sample 2	Sample 3
1	10	4
9	4	16
	8	10
	6	
	2	

16.25

Sample 1	Sample 2	Sample 3
8	2	4
4	1	3
6	3	6
		3

In Exercises 16.42–16.47, we provide data from independent simple random samples from several populations. In each case,

a. compute SST, SSTR, and SSE by using the computing formulas given in Formula 16.1 on page 726.

b. compare your results in part (a) for SSTR and SSE with those in Exercises 16.24–16.29, where you employed the defining formulas.

c. construct a one-way ANOVA table.

16.44

Sample 1	Sample 2	Sample 3	Sample 4
6	9	4	8
3	5	4	4
3	7	2	6
	8	2	
	6	3	

16.45

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
7	5	6	3	7
4	9	7	7	9
5	4	5	7	11
4		4	4	
		8	4	

One-Way ANOVA Tables

To organize and summarize the quantities required for performing a one-way analysis of variance, we use a **one-way ANOVA table.** The general format of such a table is as shown in Table 16.4.

Source	df	SS	$MS = SS/\text{df}$	F-statistic
Treatment	$k-1$	$SSTR$	$MSTR = \dfrac{SSTR}{k-1}$	$F = \dfrac{MSTR}{MSE}$
Error	$n-k$	SSE	$MSE = \dfrac{SSE}{n-k}$	
Total	$n-1$	SST		

In one-way ANOVA, we measure the variation among the sample means by a weighted average of their squared deviations about the mean, $\bar{x}$, of all the sample data. That measure of variation is called the **treatment mean square, *MSTR*,** and is defined as

$$MSTR = \frac{SSTR}{k-1},$$

where k denotes the number of populations being sampled and

$$SSTR = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_k(\bar{x}_k - \bar{x})^2.$$

The quantity ***SSTR*** is called the **treatment sum of squares.**

Next we consider the measure of variation within the samples. This measure is the pooled estimate of the common population variance, σ^2. It is called the **error mean square, *MSE*,** and is defined as

$$MSE = \frac{SSE}{n-k},$$

where n denotes the total number of observations and

$$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2.$$

The quantity ***SSE*** is called the **error sum of squares.**[†‡]

Sums of Squares in One-Way ANOVA

For a one-way ANOVA of k population means, the defining and computing formulas for the three sums of squares are as follows.

Sum of squares	Defining formula	Computing formula
Total, SST	$\Sigma(x_i - \bar{x})^2$	$\Sigma x_i^2 - (\Sigma x_i)^2/n$
Treatment, $SSTR$	$\Sigma n_j(\bar{x}_j - \bar{x})^2$	$\Sigma(T_j^2/n_j) - (\Sigma x_i)^2/n$
Error, SSE	$\Sigma(n_j - 1)s_j^2$	$SST - SSTR$

In this table, we used the notation

$$n = \text{total number of observations}$$
$$\bar{x} = \text{mean of all } n \text{ observations;}$$

and, for $j = 1, 2, \ldots, k$,

$$n_j = \text{size of sample from Population } j$$
$$\bar{x}_j = \text{mean of sample from Population } j$$
$$s_j^2 = \text{variance of sample from Population } j$$
$$T_j = \text{sum of sample data from Population } j.$$

Note that summations involving subscript is are over all n observations; those involving subscript js are over the k populations.