Section 10.1 - Single Factor ANOVA
(more than 2 population means: $\mu_1, \mu_2, \mu_3, \ldots, \mu_I$)

Single-factor ANOVA focuses on comparison of more than two population or treatment means. Let

I = the number of populations or treatments being compared (I>2)

μ_1 = the mean of population 1 or the true average response when treatment 1 is applied

⋮

μ_I = the mean of population I or the true average response when treatment I is applied

The relevant hypotheses are:

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

versus

H_a : at least two of the μ_i's are different

Example:

The article "Compression of Single-Wall Corrugated Shipping Containers Using Fixed and Floating Test Platens" (*J. Testing and Evaluation*, 1992: 318–320) describes an experiment in which several different types of boxes were compared with respect to compression strength (lb). Table 10.1 presents the results of a single-factor ANOVA experiment involving $I = 4$ types of boxes (the sample means and standard deviations are in good agreement with values given in the article).

6 sample values

Type of Box	Compression Strength (lb)						Sample Mean	Sample SD
1	655.5	788.3	734.3	721.4	679.1	699.4	713.00	46.55
2	789.2	772.5	786.9	686.1	732.1	774.8	756.93	40.34
3	737.1	639.0	696.3	671.7	717.2	727.1	698.07	37.20
4	535.1	628.7	542.4	559.0	586.9	520.0	562.02	39.87
					Grand mean =		682.50	

$I=4$

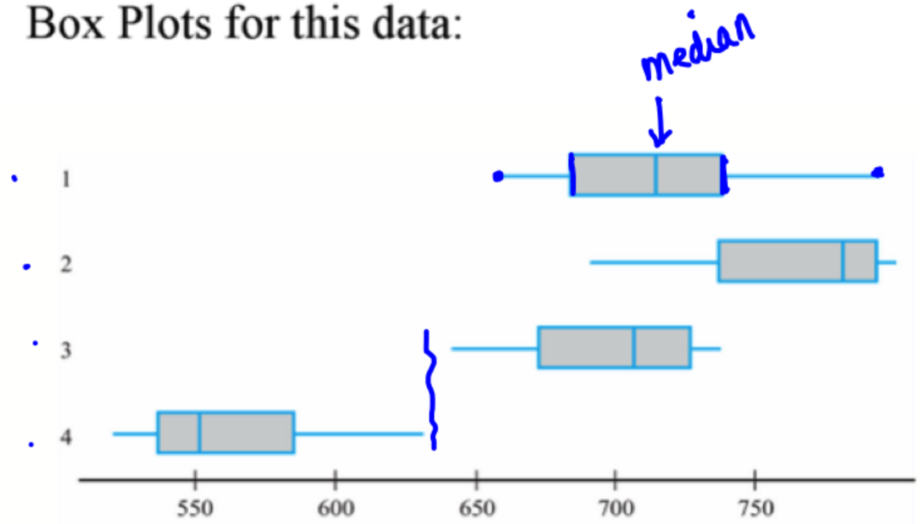
sum ÷ 4

Within each type, there are 6 levels

↖ J

$x_{2,3} = 786.9$

Box Plots for this data:



Notation:

$X_{i,j}$ = the random variable (rv) that denotes the jth measurement taken from the ith population, or the measurement taken on the jth experimental unit that receives the ith treatment

$x_{i,j}$ = the observed value of $X_{i,j}$ when the experiment is performed

$$\overline{X}_{i\cdot} = \frac{\sum_{j=1}^{J} X_{ij}}{J} \qquad i = 1, 2, \ldots, I$$

$$\overline{X}_{\cdot\cdot} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}}{IJ}$$

← grand mean

mean of all means

$$S_i^2 = \frac{\sum_{j=1}^{J} (X_{ij} - \overline{X}_{i\cdot})^2}{J - 1} \qquad i = 1, 2, \ldots, I$$

ASSUMPTIONS

The I population or treatment distributions are all normal with the same variance σ^2. That is, each X_{ij} is normally distributed with

$$E(X_{ij}) = \mu_i \quad V(X_{ij}) = \sigma^2$$

if H_0 is true

$E(MSTr) = E(MSE) = \sigma^2$

DEFINITION

Mean square for treatments is given by

$$\text{MSTr} = \frac{J}{I-1}[(\bar{X}_{1.} - \bar{X}_{..})^2 + (\bar{X}_{2.} - \bar{X}_{..})^2 + \cdots + (\bar{X}_{I.} - \bar{X}_{..})^2]$$

$$= \frac{J}{I-1}\sum_i (\bar{X}_{i.} - \bar{X}_{..})^2$$

and **mean square for error** is

$$\text{MSE} = \frac{S_1^2 + S_2^2 + \cdots + S_I^2}{I}$$

The test statistic for single-factor ANOVA is $F = \text{MSTr/MSE}$.

if H_0 false

$E(MSTr) > E(MSE) = \sigma^2$

$$F = \frac{MSTr}{MSE}$$

THEOREM

Let $F = \text{MSTr/MSE}$ be the test statistic in a single-factor ANOVA problem involving I populations or treatments with a random sample of J observations from each one. When H_0 is true and the basic assumptions of this section are satisfied, F has an F distribution with $\nu_1 = I - 1$ and $\nu_2 = I(J-1)$. With f denoting the computed value of F, the rejection region $f \geq F_{\alpha, I-1, I(J-1)}$ then specifies a test with significance level α. Refer to Section 9.5 to see how P-value information for F tests is obtained.

$df = I-1, I(J-1)$

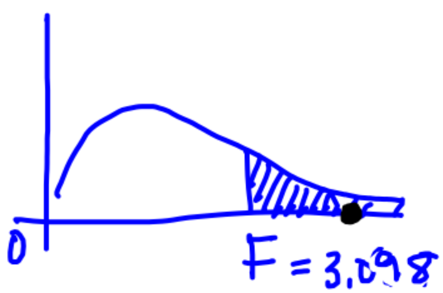
$\alpha = .05$

Now lets perform the hypothesis test for the example above:

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

H_a: at least one pair of μ_i's differs

$I = 4 \qquad J = 6$



$I - 1 = 3$

$I(J-1) = 20$

$F_{3,20}$

$qf(.95, 3, 20) = 3.098$

$$MSTr = \frac{J}{I-1}\left[\sum (\text{sample mean} - \text{grand})^2\right]$$

$$= 2\left[(713.0 - 682.5)^2 + (756.93 - 682.5)^2 + \dots\right]$$

$$\approx 42{,}455.86$$

$$MSE = \frac{46.55^2 + 40.34^2 + \dots}{4}$$

$$\approx 1691.92$$

$$f = \frac{MSTr}{MSE} = \frac{4245586}{1691.92} = 25.09$$

pvalue: $p(f > 25.09) = FCDF(25.09, 999999, 3, 20) = 5.5 \times 10^{-7}$

$1 - pf(25.09, 3, 20)$

Reject H_0

DEFINITION

The **total sum of squares (SST), treatment sum of squares (SSTr),** and **error sum of squares (SSE)** are given by

$$\text{SST} = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}x_{ij}^2 - \frac{1}{IJ}x_{..}^2$$

$$\text{SSTr} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{x}_{i.} - \bar{x}_{..})^2 = \frac{1}{J}\sum_{i=1}^{I}x_{i.}^2 - \frac{1}{IJ}x_{..}^2$$

$$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{i.})^2 \quad \text{where } x_{i.} = \sum_{j=1}^{J}x_{ij} \quad x_{..} = \sum_{i=1}^{I}\sum_{j=1}^{J}x_{ij}$$

Fundamental Identity

$$\text{SST} = \text{SSTr} + \text{SSE} \tag{10.1}$$

$$\text{MSTr} = \frac{\text{SSTr}}{I-1} \qquad \text{MSE} = \frac{\text{SSE}}{I(J-1)} \qquad F = \frac{\text{MSTr}}{\text{MSE}} \tag{10.3}$$

Table 10.2 An ANOVA Table

Source of Variation	df	Sum of Squares	Mean Square	f
Treatments	$I - 1$	SSTr	$MSTr = SSTr/(I - 1)$	MSTr/MSE
Error	$I(J - 1)$	SSE	$MSE = SSE/[I(J - 1)]$	
Total	$IJ - 1$	SST		

Another Example:

The article "Influence of Contamination and Cleaning on Bond Strength to Modified Zirconia" (*Dental Materials*, 2009: 1541–1550) reported on an experiment in which 50 zirconium-oxide disks were divided into five groups of 10 each. Then a different contamination/cleaning protocol was used for each group. The following summary data on shear bond strength (MPa) appeared in the article:

Treatment:	1	2	3	4	5	
Sample mean	10.5	14.8	15.7	16.0	21.6	Grand mean = 15.7
Sample sd	4.5	6.8	6.5	6.7	6.0	

$I = 5$

$J = 10$

Anova table	df	Sum of Squares	Mean Square	f
Treatments	4	SSTr	MSTr = 156.875	
Error	45	SSE	MSE = 37.926	
Total	49	SST		

$f = 156.875 / 37.926 \approx 4.14$

Fcdf(4.14, 99999, 4, 45) = .006 < .05 Reject H_0

1 − pf(4.14, 4, 45)