Analysis of Variance
(ANOVA)

ANOVA

- Analysis of Variance (ANOVA) - An **inferential statistical** technique designed to test for significant relationships between two variables in **two or more samples**.
- The logic is the **same** as in t-tests, just extended to independent variables with **two or more** samples.

Understanding Analysis of Variance

- **One-way ANOVA –** An analysis of variance procedure using **one dependent** and **one independent variable**.
- ANOVAs examine the **differences between samples**, as well as the **differences within a single sample**.

The Structure of Hypothesis Testing with ANOVA

Assumptions:

- (1) Independent **random samples** are used. Our choice of sample members from one population has **no effect** on the choice of members from subsequent populations.
- (2) The dependent variable is measured at the **interval-ratio** level. Some researchers, however, do apply ANOVA to **ordinal level** measurements.

The Structure of Hypothesis Testing with ANOVA

Assumptions:

- (3) The population is **normally distributed**. Though we generally cannot confirm whether the populations are normal, we must **assume** that the population is **normally distributed** in order to continue with the analysis.
- (4) The population **variances are equal**.

Stating the Research and Null Hypotheses

H_1: At least one mean is different from the others.

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

The Structure of Hypothesis Testing with ANOVA

Between-Group Sum of Squares

This tells us the differences **between** the groups

$$SSB = \sum N_k(\bar{Y}_k - \bar{Y})^2$$

N_k = the number of cases in a sample (k represents the number of different samples)

$\bar{Y}$ = the overall mean

$\bar{Y}_k$ = the mean of a sample

The Structure of Hypothesis Testing with ANOVA

Within-Group Sum of Squares

This tells us the variations **within** our groups; it also tells us the amount of **unexplained** variance.

$$SSW = \sum (Y_i - \bar{Y}_k)^2$$

N_k = the number of cases in a sample (k represents the number of different samples)

Y_i =each individual score in a sample

$\bar{Y}_k$ = the mean of a sample

Alternative Formula for Calculating the

Within-Group Sum of Squares

$$\sum Y_i^2 - \sum \frac{(\sum Y_k)^2}{n_k}$$

where

n_k = the total of each sample

Y_i^2 =the squared scores from each sample

$\sum Y_k$ = the sum of the scores of each sample

The Structure of Hypothesis Testing with ANOVA

Total Sum of Squares

$$SST = SSB + SSW$$

The Structure of Hypothesis Testing with ANOVA

Mean Square Between

An estimate of the **between-group variance** obtained by **dividing** the between-group sum of squares by its **degrees of freedom**.

Mean square between = SSB/dfb

where dfb = degrees of freedom between

dfb = k – 1
k = number of categories

The Structure of Hypothesis Testing with ANOVA

Mean Square Within

An estimate of the **within-group variance** obtained by **dividing** the within-group sum of squares by its degrees of freedom.

Mean square between = SSW/dfw

where dfw = degrees of freedom within

dfw = N – k

N = total number of cases

k = number of categories

The F Statistic

The **ratio** of **between-group variance** to **within-group variance**

$$F = \frac{Mean Square Between}{Mean Square Within} = \frac{\frac{SSB}{dfb}}{\frac{SSW}{dfw}}$$

Definitions

- **F** *ratio* (**F** *statistic*) – Used in an analysis of variance, the F statistic represents the **ratio** of **between-group variance** to **within-group variance**
- **F** *obtained* – The test statistic **computed** by the ratio for between-group to within-group variance.
- **F** *critical* – The *F* score associated with a particular **alpha level** and **degrees of freedom**. This *F* score marks the beginning of the **region of rejection** for our null hypothesis.

Educational Attainment (Measured in Years) for Four GSS 2002 Groups

White Males $n_1=6$	Black Males $n_2=4$	White Females $n_3=6$	Black Females $n_4=5$
16	16	16	14
18	12	12	10
14	11	14	12
14	14	14	13
16		11	11
16		11	

White Males		Black Males		White Females		Black Females	
$n_1 = 6$		$n_2 = 4$		$n_3 = 6$		$n_4 = 5$	
Y_1	Y_1^2	Y_2	Y_2^2	Y_3	Y_3^2	Y_4	Y_4^2
16	256	16	256	16	256	14	196
18	324	12	144	12	144	10	100
14	196	11	121	14	196	12	144
14	196	14	196	14	196	13	169
16	256			11	121	11	121
16	256			11	121		
94	1484	53	717	78	1034	60	730

The F Statistic

$$F = \frac{MeanSquareBetween}{MeanSquareWithin} = \frac{\frac{SSB}{dfb}}{\frac{SSW}{dfw}}$$

$$F = \frac{13.7145}{3.298823529}$$

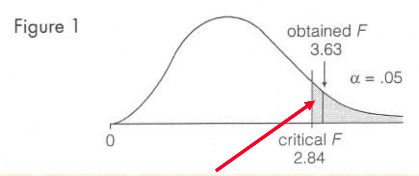
$\alpha = .05$

dfb → df_1

dfw → df_2

df_2	1	2	3	4	5	6	8	12	24	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	243.9	249.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.41	19.45	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.74	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.91	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.68	4.53	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.00	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.57	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.28	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.07	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.91	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.79	2.61	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.69	2.50	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.77	2.60	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.53	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.48	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.42	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.38	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.34	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.31	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.28	2.08	1.84

Example: Obtained vs. Critical F



Since the obtained F is beyond the critical F value, we reject the Null hypothesis of no difference