ANOVA

Analysis of Variance









Lets say you want to find if the beverage that people drink affects their reaction time.

The mean of all three groups is the same.

Null hypothesis: H_0







Water	Juice	Coffee
29	28	25
30	29	28
31	27	29
31	30	27
29	29	29







Water	Juice	Coffee
10	11	12
12	14	13
18	19	17
24	23	25
36	38	37

Why the within-group spread matters

Both data sets ask the same question about the same three beverages.

First data set — readings tightly clustered

- Means 30.0, 28.6, 27.6. MS within = 1.70, MS between = 7.27
- $F = 7.27 / 1.70 = 4.28 > F(2,12) = 3.89 \rightarrow$ reject H_0 ($p = 0.04$)

Second data set — readings widely scattered

- Means 20.0, 21.0, 20.8. MS within = 109.9, MS between = 1.40
- $F = 1.40 / 109.9 = 0.01 \rightarrow$ fail to reject H_0 ($p = 0.99$)

A gap between group means only counts relative to the scatter within the groups. That ratio is exactly what F measures.

Figure out how much of the total variance comes from:
The variance between the groups
The variance within the groups

$$F = \frac{between\ group}{within\ group}$$

The larger this number is , it is more likely the groups have different means. (reject H_0)

Data

x_{11}	x_{21}	x_{31}	x_{41}
x_{12}	x_{22}	x_{32}	x_{k2}
x_{13}	x_{23}	x_{33}	
	x_{24}	x_{34}	
		x_{35}	

Sum of squares for variance between the samples

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

$$SS\ within = \sum(x_{1i} - \bar{x}_1)^2 + \sum(x_{2i} - \bar{x}_2)^2 + \cdots + \sum(x_{ki} - \bar{x}_k)^2$$

Mean Square between samples

$$MS\ between = \frac{SS\ between}{k-1}$$

$$MS\ within = \frac{SS\ within}{n-k}$$

SS for total variance

$$SS\ for\ total\ variance = \sum \left(x_{ij} - \bar{\bar{x}}\right)^2$$

$$SS\ for\ total\ variance = SS\ between + SS\ within$$

$$(n-1) = (k-1) + (n-k)$$

$$F - ratio = \frac{MS\ between}{MS\ within}$$

Table

Source of variation	Sum of Squares	Degrees of Freedom	Mean Square	F-ratio
Between Samples	SS between	k-1	SS between / (k – 1)	MS between / MS within
Within samples	SS within	n-k	SS within / (n – k)	
Total	$\Sigma (x_{ij} - \bar{x})^2$	n-1		

Per acre production data for three varieties of wheat are , each grown on 4 plots are shown here. State if the variety differences are significant.

Plot of land	A	B	C
1	6	5	5
2	7	5	4
3	3	3	3
4	8	7	4

Table 4(a): Critical Values of *F*-Distribution (at 5 per cent)

v_2 \ v_1	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	249.1	243.3
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	8.85	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	2.51	2.30

Results

The calculated F = 1.50 is less than the table value 4.26 at the 5% level with d.f. = (2, 9). We therefore fail to reject the null hypothesis: the variety differences are not significant.

Source of variation	Sum of Squares	Degrees of Freedom	Mean Square	F-ratio
Between Samples	8	2	**4**	1.5
Within samples	24	9	**2.67**	
Total	**32**	11		

$$F(2,9) = 4.26$$

What ANOVA assumes

- Independence — observations are independent of one another, within and between groups.
- Normality — values within each group are roughly normal. ANOVA is fairly robust to mild departures, especially with equal group sizes.
- Homogeneity of variance — the groups have similar variances (check with Levene's or Bartlett's test).

If an assumption fails

- Unequal variances → Welch's ANOVA.
- Strong non-normality, or ordinal data → Kruskal–Wallis, the non-parametric counterpart.

With only two groups, ANOVA is the two-sample t-test — in fact $F = t^2$.

A significant F — but which groups differ?

F tests one thing only: are all the group means equal? A significant F says at least one differs, not which one.

Post-hoc tests, run only after a significant F

- Tukey's HSD — compares every pair; the usual default.
- Bonferroni — simple and conservative: divide α by the number of comparisons.
- Dunnett — when every group is compared against one control.

Why not just run many t-tests?

Three groups need three t-tests. At α = 0.05 each, the chance of at least one false positive is about 14%, not 5%. Post-hoc tests hold the overall error rate at 5%.