ANOVA

Deepa Anwar

ANOVA
for comparing means between more than 2 groups

Hypotheses of One-Way ANOVA

- $H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_c$
  - All population means are equal
  - i.e., no treatment effect (no variation in means among groups)
- H_1 : Not all of the population means are the same
  - At least one population mean is different
  - i.e., there is a treatment effect
  - Does not mean that all population means are different (some pairs may be the same)

The F-distribution

- A ratio of variances follows an F-distribution:

$$\frac{\sigma^2_{between}}{\sigma^2_{within}} \sim F_{n,m}$$

- The F-test tests the hypothesis that two variances are equal.
- F will be close to 1 if sample variances are equal.

$$H_0 : \sigma^2_{between} = \sigma^2_{within}$$
$$H_a : \sigma^2_{between} \neq \sigma^2_{within}$$

ANOVA Table

Source of variation	d.f.	Sum of squares	Mean Sum of Squares	F-statistic	p-value
Between (k groups)	k-1	SSB (sum of squared deviations of group means from grand mean)	SSB/k-1	$\frac{SSB/(k-1)}{SSW/(nk-k)}$	Go to $F_{k-1,nk-k}$ chart
Within (n individuals per group)	nk-k	SSW (sum of squared deviations of observations from their group mean)	s^2=SSW/nk-k		
Total variation	nk-1	TSS (sum of squared deviations of observations from grand mean)			

TSS=SSB + SSW

Example

Treatment 1	Treatment 2	Treatment 3	Treatment 4
60 inches	50	48	47
67	52	49	67
42	43	50	54
67	67	55	67
56	67	56	68
62	59	61	65
64	67	61	65
59	64	60	56
72	63	59	60
71	65	64	65

Example

Step 1) calculate the sum of squares between groups:

Mean for group 1 = 62.0

Mean for group 2 = 59.7

Mean for group 3 = 56.3

Mean for group 4 = 61.4

Treatment 1	Treatment 2	Treatment 3	Treatment 4
60 inches	50	48	47
67	52	49	67
42	43	50	54
67	67	55	67
56	67	56	68
62	59	61	65
64	67	61	65
59	64	60	56
72	63	59	60
71	65	64	65

Grand mean= 59.85

SSB = $[(62-59.85)^2 + (59.7-59.85)^2 + (56.3-59.85)^2 + (61.4-59.85)^2]$ *xn per group*= 19.65*x10* = 196.5

Example

Step 2) calculate the sum of squares within groups:

$(60-62)^2+(67-62)^2+(42-62)^2+(67-62)^2+(56-62)^2+(62-62)^2+(64-62)^2+(59-62)^2+(72-62)^2+(71-62)^2+(50-59.7)^2+(52-59.7)^2+(43-59.7)^2+67-59.7)^2+(67-59.7)^2+(69-59.7)^2$…+….(sum of 40 squared deviations) = **2060.6**

Treatment 1	Treatment 2	Treatment 3	Treatment 4
60 inches	50	48	47
67	52	49	67
42	43	50	54
67	67	55	67
56	67	56	68
62	59	61	65
64	67	61	65
59	64	60	56
72	63	59	60
71	65	64	65

Step 3) Fill in the ANOVA table

Source of variation	d.f.	Sum of squares	Mean Sum of Squares	F-statistic	p-value
Between	3	196.5	65.5	1.14	.344
Within	36	2060.6	57.2	-	-
Total	39	2257.1	-	-	-

Step 3) Fill in the ANOVA table

Source of variation	d.f.	Sum of squares	Mean Sum of Squares	F-statistic	p-value
Between	3	196.5	65.5	1.14	.344
Within	36	2060.6	57.2	-	-
Total	39	2257.1	-	-	-

INTERPRETATION of ANOVA:

How much of the variance in height is explained by treatment group?

R^2=“Coefficient of Determination” = SSB/TSS = 196.5/2275.1=9%

Coefficient of Determination

$$R^2 = \frac{SSB}{SSB + SSE} = \frac{SSB}{SST}$$

The amount of variation in the outcome variable (dependent variable) that is explained by the predictor (independent variable).

Beyond one-way ANOVA

Often, you may want to test more than 1 treatment. ANOVA can accommodate more than 1 treatment or factor, so long as they are independent. Again, the variation partitions beautifully!

TSS = SSB1 + SSB2 + SSW