13.5: Understanding the ANOVA Table

When running Analysis of Variance, the data is usually organized into a special ANOVA table, especially when using computer software.

Source of Variation	Sum of Squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F
Factor (Between)	SS_{Factor}	k-1	$MS_{Factor} = SS_{Factor}/k-1$	$F = MS_{Factor}/MS_{Error}$
Error (Within)	SS_{Error}	n-k	$MS_{Error} = SS_{Error}/n-k$	
Total	SS_{Total}	n-1		

Sum of Squares: The total variability of the numeric data being compared is broken into the variability between groups ($\mathrm{SS}_{\text{Factor}}$) and the variability within groups ($\mathrm{SS}_{\text{Error}}$). These formulas are the most tedious part of the calculation. T_c represents the sum of the data in each population and n_c represents the sample size of each population. These formulas represent the numerator of the variance formula.

$$\mathrm{SS}_{\text{Total}} = \Sigma\left(X^2\right) - \frac{(\Sigma X)^2}{n}$$

$$\mathrm{SS}_{\text{Factor}} = \Sigma\left(\frac{T_c^2}{n_c}\right) - \frac{(\Sigma X)^2}{n}$$

$$\mathrm{SS}_{\text{Error}} = \mathrm{SS}_{\text{Total}} - \mathrm{SS}_{\text{Factor}}$$

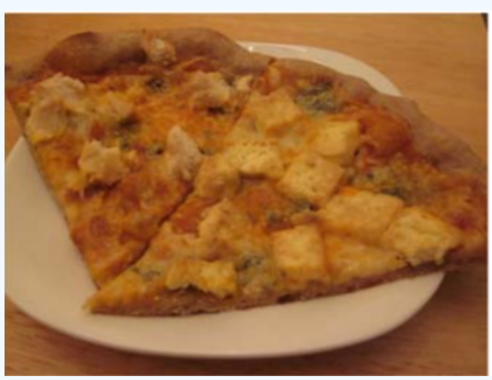
Degrees of freedom: The total degrees of freedom is also partitioned into the Factor and Error components.

Mean Square: This represents calculation of the variance by dividing Sum of Squares by the appropriate degrees of freedom.

F: This is the test statistic for ANOVA: the ratio of two sample variances (mean squares) that are both estimating the same population value has an F distribution. Computer software will then calculate the p- value to be used in testing the Null Hypothesis that all populations have the same mean.

Example: Party Pizza

Party Pizza specializes in meals for students. Hsieh Li, President, recently developed a new tofu pizza.



Before making it a part of the regular menu she decides to test it in several of her restaurants. She would like to know if there is a difference in the mean number of tofu pizzas sold per day at the Cupertino, San Jose, and Santa Clara pizzerias. Data will be collected for five days at each location.

At the .05 significance level can Hsieh Li conclude that there is a difference in the mean number of tofu pizzas sold per day at the three pizzerias?

Solution

Design

Response: tofu pizzas sold

Factor: location of restaurant

Levels: $k=3$ (Cupertino, San Jose, Santa Clara)

Research Hypotheses:

H_o: There is no difference in mean tofu pizzas sold due to location of restaurant.

H_a: There is a difference in mean tofu pizzas sold due to location of restaurant

H_o: $\mu_1=\mu_2=\mu_3$ (Mean sales same at all restaurants)

H_a: At least μ_i is different (Means sales not the same at all restaurants)

We will assume the population variances are equal $\sigma_1^2=\sigma_2^2=\sigma_3^2$, so the model will be **One Factor ANOVA**. This model is appropriate if the distribution of the sample means is approximately Normal from the Central Limit Theorem.

Type I error would be to reject the Null Hypothesis and claim mean sales are different, when they actually are the same. The test will be run at a level of significance (α) of 5%.

The test statistic from the table will be $\mathrm{F}=\dfrac{\mathrm{MS}_{\text{Factor}}}{\mathrm{MS}_{\text{Error}}}$. The degrees of freedom for numerator will be 3-1=2,and the degrees of freedom for denominator will be 13-3=10. (The total sample size turned out to be only 13, not 15 as planned).

Critical Value for F at α of 5% with $\mathrm{df}_{\text{num}}=2$ and $\mathrm{df}_{\text{den}}=10$ is 4.10. Reject H_o if F >4.10. We will also run this test using the p-value method with statistical software, such as Minitab.

Data/Results

	Cupertino	San Jose	Santa Clara	Total
	13	10	18	
	12	12	16	
	14	13	17	
	12	11	17	
			17	
T	51	46	85	182
n	4	4	5	13
Means	12.75	11.5	17	14
Σ^2	653	534	1447	2634

$$SS_{Total}=2634-\frac{182^2}{13}=86$$

$$SS_{Factor}=2624.25-\frac{182^2}{13}=76.25$$

$$\mathrm{SS}_{\mathrm{Error}}=86-76.25=9.75$$

$\mathrm{F}=38.125/0.975=39.10$ which is more than the critical value of 4.10, so reject H_o. Also from the Minitab output, p-value = 0.000 < 0.05 which also supports rejecting H_o.

One-way ANOVA: Cupertino, San Jose, Santa Clara

```
Source  DF      SS      MS      F      P
Factor   2  76.250  38.125  39.10  0.000
Error   10   9.750   0.975
Total   12  86.000

S = 0.9874   R-Sq = 88.66%   R-Sq(adj) = 86.40%


                               Individual 95% CIs For Mean Based on
                               Pooled StDev
Level        N    Mean  StDev  --------+---------+---------+---------+-
Cupertino    4  12.750  0.957        (-----*----)
San Jose     4  11.500  1.291  (----*-----)
Santa Clara  5  17.000  0.707                             (----*----)
                               --------+---------+---------+---------+-
                                    12.0      14.0      16.0      18.0
```

Conclusion

There is a difference in the mean number of tofu pizzas sold at the three locations.