Week 7 Cheat Sheet

Statistics and Data Analysis with Excel, Part 2

Charlie Nuttelman

Here, I provide the mathematical equations and some of the important Excel functions required to perform various calculations in Week 7 of the course. The headings represent the screencasts in which you will find those calculations and concepts. Not all screencasts are referenced below – just the ones that have complex mathematical formulas or Excel formulas that are tricky to use.

My equations and analysis follow that of Montgomery and Runger in: *Applied Statistics and Probability for Engineers, 7th edition*, Douglas C. Montgomery and George C. Runger, Wiley (2018). This is an excellent text for applied probability and statistics and highly recommended if you need a supplementary text for the course.

Single-Factor ANOVA (Parts 1, 2, and 3)

Q: What is one-way ANOVA (analysis of variance) used for?

A: When we have three or more groups that we wish to compare (single factor) the means of, then we can use one-way ANOVA. We can compare many groups, but there is only one factor that we are measuring/observing. If we wish to compare the means of two groups, then we use a simple "comparison of means" scheme, as we discussed previously in the course.

Basic data collection:

		Treatment (level)		
Observations	1	2	...	a
1	y_{11}	y_{21}	$\cdots$	y_{a1}
2	y_{12}	y_{22}	$\cdots$	y_{a2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{1n}	y_{2n}	$\cdots$	y_{an}

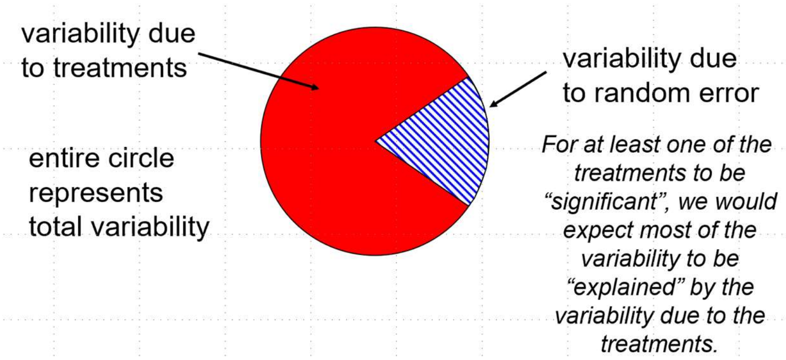
Terminology:

$\boldsymbol{a}$ = number of treatment levels (groups)

$\boldsymbol{n}$ = number of samples (observations) per group

$\boldsymbol{N}$ = total number of samples (observations) = $\boldsymbol{a \cdot n}$

Partitioning the sum of squares:



$$SS_T = SS_{Treatments} + SS_E$$

$y_{..}$ = sum of all y_{ij}

$$\boldsymbol{y_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}}$$

$\bar{y}_{..}$ = "grand average" of all y_{ij}

$$\bar{y}_{..} = \sum_{i=1}^{a}\sum_{j=1}^{n}\frac{y_{ij}}{N} = \sum_{i=1}^{a}\sum_{j=1}^{n}\frac{y_{ij}}{(a \cdot n)} = \frac{y_{..}}{N}$$

$y_{1.}$ = sum of all y_{ij} for first treatment group (i =1):

$$y_{1.} = \sum_{j=1}^{n} y_{1j}$$

Effects model:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad i = 1,2,\ldots,a \quad j = 1,2,\ldots,n$$

Computational formulas:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^2 - \frac{y_{..}^2}{N}$$

Note that SS_T can also be calculated via a “shortcut” method in Excel: **=VAR.S(y)*(a*n-1)**, where y represents all of the responses.

$$SS_{Treatments} = \frac{1}{n}\sum_{i=1}^{a} y_{i.}^2 - \frac{y_{..}^2}{N}$$

$$SS_E = SS_T - SS_{Treatments}$$

ANOVA Table (one-way, aka single factor):

Source	Sum of squares	Degrees of freedom	Mean square	F_0
Treatments	$SS_{Treatments}$	$a-1$	$MS_{Treatments}$	$\frac{MS_{Treatments}}{MS_E}$
Error	SS_E	$a(n-1)$	MS_E	
Total	SS_T	$an-1$		

(For single-factor analysis, only Factor A, Error and Total.)

Calculate mean squares:

$$MS_{Treatments} = \frac{SS_{Treatments}}{a-1}$$

$$MS_{Error} = \frac{SS_E}{an-1}$$

Statistical Test:

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1 : \tau_i \neq 0 \quad for\ at\,least\ one\ i$$

Test statistic:

$$F_0 = \frac{MS_{Treatments}}{MS_E} \in F_{a-1,N-a}$$

Reject H_0 (and accept H_1) if**:**

$$F_0 > F_{\alpha,a-1,N-a}$$

If we accept H_1, then the (simple) conclusion that we can make is: there is a significant difference between treatments of the groups. Usually, this is not very interesting because it does not answer the question “which treatments are different?”, but to answer this question we can use pairwise comparisons (see below).

Fisher's Pairwise Comparison Test:

ANOVA determines that there are differences in factor (*aka* treatment) levels. The next step is to determine which levels are different.

$$H_0: \ \mu_i = \mu_j \qquad \forall \ i \neq j$$
$$H_1: \ \mu_i \neq \mu_j$$

Compute LSD (least significant difference)

$$LSD = t_{\alpha/2, N-a} \cdot \sqrt{2 \cdot MSE/n}$$

MS_E computed above.

Check to see if the difference in pairs of means are greater than the LSD:

$$\left| \overline{y}_{i\cdot} - \overline{y}_{j\cdot} \right| > LSD \qquad \forall \ i \neq j$$

If so, then there is a significant difference between μ_i and μ_j.

Two-Way ANOVA (Guided Workshop 7)

Q: What is two-way ANOVA (analysis of variance) used for?

A: Two-way (or two-factor) ANOVA is used to determine if there are significant differences in treatment levels of two factors, plus the interactions between those two factors .

Effects model:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases}$$

Terminology

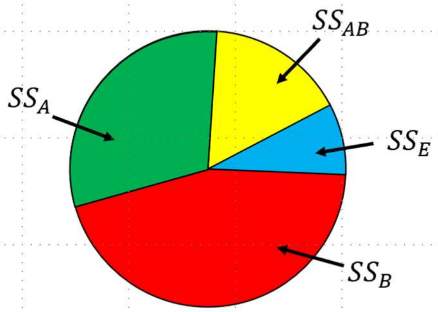
$\boldsymbol{a}$ = number of treatment levels (groups) of first factor

$\boldsymbol{b}$ = number of treatment levels (groups) of second factor

$\boldsymbol{n}$ = number of samples (observations) per group

$\boldsymbol{N}$ = total number of samples (observations) **=** $\boldsymbol{a \cdot b \cdot n}$

Partitioning of sum of squares



$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$SS_A = \sum_{i=1}^{a} \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}$$

$$SS_B = \sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$$

$$SS_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn} - SS_A - SS_B$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

$y_{...}$ = sum of all y_{ijk}

$$\boldsymbol{y_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}}$$

$\bar{y}_{...}$ = “grand average” of all y_{ijk}

$$\bar{y}_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} \frac{y_{ijk}}{N} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} \frac{y_{ijk}}{(a \cdot b \cdot n)} = \frac{y_{...}}{N}$$

$y_{i..}$ = sum of all y_{ijk} for i[th] treatment group of first level:

$$y_{i..} = \sum_{j}^{b}\sum_{k=1}^{n} y_{ijk}$$

$y_{.j.}$ = sum of all y_{ijk} for j[th] treatment group of second level:

$$y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk}$$

$y_{ij.}$ = sum of all y_{ijk} for i[th] treatment group of first level and j[th] treatment group of second level:

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk}$$

ANOVA Table (two-way):

Source	SS	dof	MS	F0	P-value
A treatments	SS_A	a-1	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MS_E}$	
B treatments	SS_B	b-1	$MS_B = \frac{SS_B}{b-1}$	$\frac{MS_B}{MS_E}$	
Interaction	SS_{AB}	(a–1)(b–1)	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_E}$	
Error	SS_E	ab(n-1)	$MS_E = \frac{SS_E}{ab(n-1)}$		
Total	SS_T	abn-1			

Statistical Tests:

Effect A:

$H_0: \tau_1 = \tau_2 = \cdots = \tau_a = 0$ (no main effect of factor A)

$H_1: \tau_i \neq 0$ for at least one i

Effect B:

$H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$ (no main effect of factor B)

$H_1: \beta_j \neq 0$ for at least one j

Interaction effect AB:

$H_0: (\tau\beta)_{11} = (\tau\beta)_{12} = \cdots = (\tau\beta)_{ab} = 0$ (no main effect of factor A)

$H_1: (\tau\beta)_{ij} \neq 0$ for at least one combination of i and j

Test statistic:

$$F_0 = \frac{MS_{Treatments}}{MS_E} \in F_{a-1,N-a}$$

Reject H_0 (and accept H_1) if**:**

$$F_0 > F_{\alpha,a-1,N-a}$$

If we accept H_1, then the (simple) conclusion that we can make is: there is a significant difference between treatments of the groups or interactions between the two groups. Usually, this is not very interesting because it does not answer the question “which treatments are different?”.