4.5. Two-way ANOVA

Example. (Moore/McCabe Example 13.3–4)

Red palm oil, due to its high content of vitamin A, is thought to reduce the occurence and severity of malarity for young children. To investigate whether this is indeed the case, a supplement will be prepared that contains either a placebo, a low dose, or a high dose of red palm oil. Because boys and girls may differ in exposure to malaria and the response to the red palm oil supplement, we consider a *two-way ANOVA*, that takes also gender into account. Suppose we recruit 75 boys and 75 girls to the study. We will then randomly assign 25 of each gender to each of the red palm oil levels:

	Gender		
Red palm oil	Girls	Boys	Total
Placebo	25	25	50
Low dose	25	25	50
High dose	25	25	50
Total	75	75	150

This example illustrates several advantages of two-way designs. The first is efficiency. With a two-way design, we can use the same tool to answer two questions: 1. Does red palm oil/ 2. Does gender have an impact upon prevalence of malaria? That is, we get two one-way ANOVAs for the price of one.

The second and more important advantage is that we may investigate interactions between factors. Suppose that boys and girls react indeed differently to the red palm oil supplement. That piece of information would not come out in one-way ANOVAs neither for red palm oil nor for gender.

The final advantage is increased power of the tests. Suppose we would run a one-way design and there are indeed differences between boys and girls. The one-way ANOVA would assign this variation to the *residual* (within groups) part of the model. In the two-way ANOVA, gender is included in the *fit* (between groups) part of the model, which increases the power of the tests.

Assumptions for Two-Way ANOVA

The assumptions for two-way ANOVA are the same as for one-way ANOVA, just that we have now 2 instead of only 1 factor:

We have independent random samples of size n_{ij} from each of $I \times J$ normal populations. The population means μ_{ij} may differ, but all populations have the same variance σ^2.

Let x_{ijk} represent the kth observation from the population having factor A at level i and factor B at level j. The statistical model is

$$X_{ijk} = \mu_{ij} + \epsilon_{ijk}, \qquad \text{where}$$

$i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, n_{ij};$ and $\epsilon_{ijk} \sim N(0, \sigma^2)$.

We estimate μ_{ij} by $\bar{x}_{ij} = \dfrac{1}{n_{ij}} \displaystyle\sum_{k=1}^{n_{ij}} x_{ijk},$

and σ^2 by (using $s_{ij}^2 = \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{ij})^2/(n_{ij}-1)$):

$$s^2 = \frac{\sum_{ij}(n_{ij}-1)s_{ij}^2}{N - IJ} = \frac{\text{SSE}}{\text{DFE}} = \text{MSE},$$

where DFE = observations–groups $= N - IJ$.

Main effects and interactions

We shall now explore in detail the FIT part of the model, represented by the population means μ_{ij}.

Because we have independent samples from each group, we can think of the problem initially as a one-way ANOVA with IJ groups, that is SST = SSM + SSE, where

$$\text{SSM} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{n_{ij}} (\bar{x}_{ij} - \bar{\bar{x}})^2$$

with $\quad \text{DFM} = \text{groups} - 1 = IJ - 1.$

In two-way ANOVA, the terms SSM and DFM are broken down into a main effect for A, a main effect for B, and an AB interaction as follows:

$$\text{SSM} = \text{SSA} + \text{SSB} + \text{SSAB}$$
$$\text{DFM} = \text{DFA} + \text{DFB} + \text{DFAB}$$

SSA represents variation among the means for the different levels of A. Because there are I such means, SSA has DFA= $I - 1$ degress of freedom. Similarly, DFB= $J - 1$. For DFAB we obtain by subtraction:

$$\begin{aligned} \text{DFAB} &= \text{DFM} - \text{DFA} - \text{DFB} \\ &= (IJ - 1) - (I - 1) - (J - 1) \\ &= (I - 1)(J - 1). \end{aligned}$$

The two-way ANOVA model

Taking the split up of the population means μ_{ij} into main effects and an interaction effect into account, the two-way ANOVA model may be restated as

$$X_{ijk} = \overbrace{\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}}^{\mu_{ij}} + \epsilon_{ijk},\ \epsilon_{ijk} \sim N(0, \sigma^2),$$

where μ is the overall mean; α_i is the effect of level i ($i = 1, \ldots, I$) of factor A; β_j is the effect of level j ($j = 1, \ldots, J$) of factor B; $(\alpha\beta)_{ij}$ is the interaction effect of levels i and j; and ϵ_{ijk} is the error associated with the kth data point from level i of factor A and level j of factor B.

Note that since α_i, β_j and $(\alpha\beta)_{ij}$ are deviations from the overall mean μ, in the fixed effects model the sums of all these deviations are zero:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_{i,j} (\alpha\beta)_{ij} = 0.$$

The Hypothesis Tests in Two-Way ANOVA

Factor A main-effects test:

$$\begin{aligned} H_0&: \quad \alpha_i = 0 \text{ for all } i = 1, \ldots, I \\ H_1&: \quad \text{Not all } \alpha_i \text{ are zero} \\ F &= MSA/MSE \sim F(I-1, N-IJ). \end{aligned}$$

Factor B main-effects test:

$$\begin{aligned} H_0&: \quad \beta_j = 0 \text{ for all } j = 1, \ldots, J \\ H_1&: \quad \text{Not all } \beta_j \text{ are zero} \\ F &= MSB/MSE \sim F(J-1, N-IJ). \end{aligned}$$

Test for AB interactions:

$$\begin{aligned} H_0&: \quad (\alpha\beta)_{ij} = 0 \ \forall \ i{=}1,\ldots,I; \ j{=}1,\ldots,J \\ H_1&: \quad \text{Not all } (\alpha\beta)_{ij} \text{ are zero} \\ F &= MSAB/MSE \sim F((I-1)(J-1), N-IJ). \end{aligned}$$

Note: Finding a significant interaction effect substantially reduces the usefulness of factor A/B main effects, because then they apply only on average, but not individually for each level of the other factor.

Example: (Azcel, 4th edition.)

The brightness of films produced by 3 different manufacturers has been compared using 3 different development processes:

Kodak			Fuji			Agfa		
A	B	C	A	B	C	A	B	C
32	26	28	43	32	32	23	27	25
34	29	28	41	38	32	24	30	27
31	27	27	44	38	36	25	25	26
30	30	30	50	40	35	21	25	22
37	31	32	47	36	34	26	27	25

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Film Manufacturer	1363.378	2	681.6889	117.3078	1.7E-16	3.259446
Processing (A/B/C)	165.6444	2	82.82222	14.25239	2.76E-05	3.259446
Interaction	247.0222	4	61.75556	10.62715	8.63E-06	2.633532
Within	209.2	36	5.811111			
Total	1985.244	44				

All p-values are below 0.01%, so both manufacturer and development method have an impact.

There is also an interaction effect:

$F = 61.756/5.811 = 10.627$
Degrees of freedom:
$(3-1)(3-1) = 4$ and $45 - 3 \cdot 3 = 36$
Critical value: $F_{0.05}(4, 36) = 2.63$

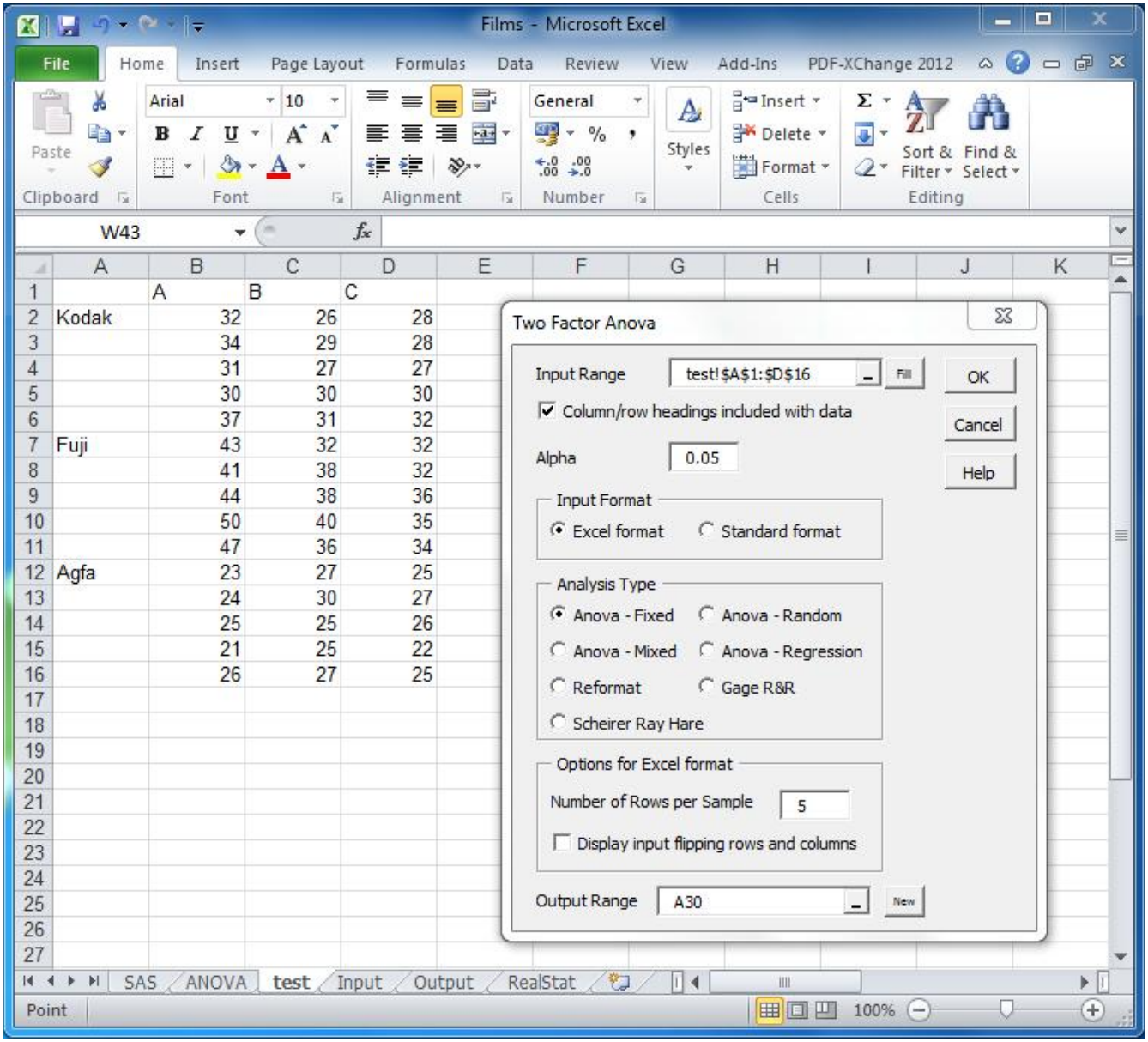
This implies that main effects must be checked for each level of the other factor separately.

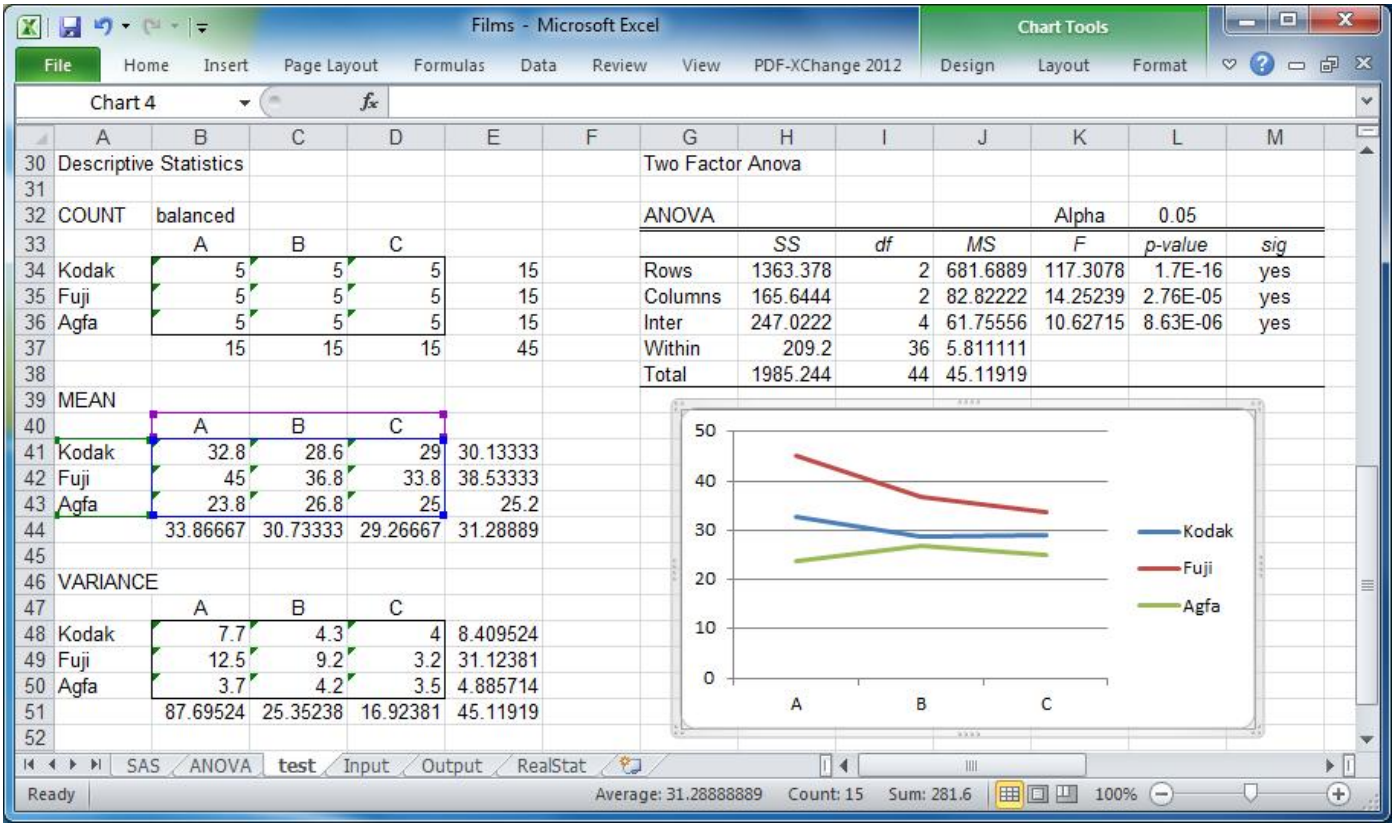
Two-way ANOVA in Excel

You can get the two-way ANOVA table either from Excel's Data Analysis tool via 'Anova: Two-Factor With Replication' or from the Real Statistics Two Factor ANOVA tool.

Doing it with Real Statistics has the advantage that you get the group specific means in such a way, that excel finds it easy to produce a line plot of those, which in the context of ANOVA is called an interaction plot.

Non-parallel lines in such a plot indicate that there is an interaction effect, which implies that the main effects hold only on average for each factor, but not for the factor levels individually, making it necessary to investigate the factor levels one by one. The p-value of the interaction effect in the ANOVA table tells whether interaction is present also out of sample.


Films - Microsoft Excel
Two Factor Anova
Input Range
test!$A$1:$D$16
Column/row headings included with data
Alpha
0.05
Input Format
Excel format
Standard format
Analysis Type
Anova - Fixed
Anova - Random
Anova - Mixed
Anova - Regression
Reformat
Gage R&R
Scheirer Ray Hare
Options for Excel format
Number of Rows per Sample
5
Display input flipping rows and columns
Output Range
A30
OK
Cancel
Help
SAS
ANOVA
test
Input
Output
RealStat


Films - Microsoft Excel
Chart Tools
Descriptive Statistics
COUNT balanced
MEAN
VARIANCE
Two Factor Anova
ANOVA
Alpha 0.05
Rows 1363.378 2 681.6889 117.3078 1.7E-16 yes
Columns 165.6444 2 82.82222 14.25239 2.76E-05 yes
Inter 247.0222 4 61.75556 10.62715 8.63E-06 yes
Within 209.2 36 5.811111
Total 1985.244 44 45.11919
Kodak
Fuji
Agfa
SAS
ANOVA
test
Input
Output
RealStat
Ready
Average: 31.28888889
Count: 15
Sum: 281.6

Two-Way ANOVA with 1 Observation/Cell

The case of one data point in every cell presents a problem in two-way ANOVA because then

$$\text{DFE} = \text{observations} - \text{groups} = 0.$$

However, if we may *assume that there are no interactions*, then SSAB is only due to error and contains no other information. In that case we may use SSAB and its associated degrees of freedom DFAB= $(I-1)(J-1)$ in place of SSE and its degrees of freedom.

We can thus conduct the tests for the main effects by dividing MSA (MSB) by MSAB when testing for factor A (factor B) main effects. The resulting F-statistics has $I-1$ and $(I-1)(J-1)$ degrees of freedom for factor A, and $J-1$ and $(I-1)(J-1)$ degrees of freedom for factor B.

The fixed effects two-way ANOVA model with one observation per cell reduces then to:

$$X_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \qquad \epsilon_{ijk} \sim N(0, \sigma^2).$$

Randomized Complete Block Design

When we want to compare the means of k population means while controlling extraneous variables, a procedure known as *blocking* is used. A block (lohko) is a collection of k experimental units that are as nearly alike as possible with respect to the extraneous variables. (A block could be the same person trying k different products.)

Each treatment is randomly assigned to 1 unit within each block (random order in trying the products). Since the effect of the extraneous variables is controlled by matching like experimental units, any differences in response are attributed to treatment effects.

This *randomized complete block design* (also called *repeated measures ANOVA* or *one-way ANOVA with repetition*) may be regarded as a two-way ANOVA with one item per cell because the blocks may be viewed as one factor and the treatment levels as the other. In the randomized block design, however, we are only interested in the treatment levels and not in the blocks.

Because the randomized complete block design is just a two-way ANOVA with one observation per cell, we may test the null hypothesis

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_k.$$

with an F-test of the form

$$F = \frac{\text{MSTR}}{\text{MSE}} \sim F(k-1, (k-1)(b-1)),$$

where k denotes the number of treatments and b denotes the number of blocks.

Note that the randomized complete block design assumes that there is *no interaction between treatments and blocks!*

You get the randomized complete block design from Excel's Data Analysis tool via 'Annova: Two-Factor Without Replication' or from Real Statistics' 'One Repeated Measures ANOVA'. If you choose the Real Statistics tool, which allows you also to calculate contrasts and Tukey's HSD, make sure to put the treatments into columns and the blocks into rows. Blocks are then denoted as subjects and treatments as groups.

Example: (Cochran & Snedecor)
Below are percentages of soya beans failing to sprout, planted under 5 individually identical conditions (blocks) under 5 different treatments, with ANOVA output from Excel:

	Arasan	Spergon	Semesan	Fermate	None	
Block 1	2	4	3	9	8	
Block 2	6	10	5	7	10	
Block 3	7	9	9	5	12	
Block 4	11	8	10	5	13	
Block 5	5	10	6	3	11	
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	49.84	4	12.46	2.30	0.103	3.01
Columns	83.84	4	20.96	3.87	0.022	3.01
Error	86.56	16	5.41			
Total	220.24	24				

Degrees of freedom:

Treatment: $5 - 1 = 4$
Block: $5 - 1 = 4$
Error: $(5 - 1)(5 - 1) = 16$

There is a significant main effect for treatments, but not for blocks.

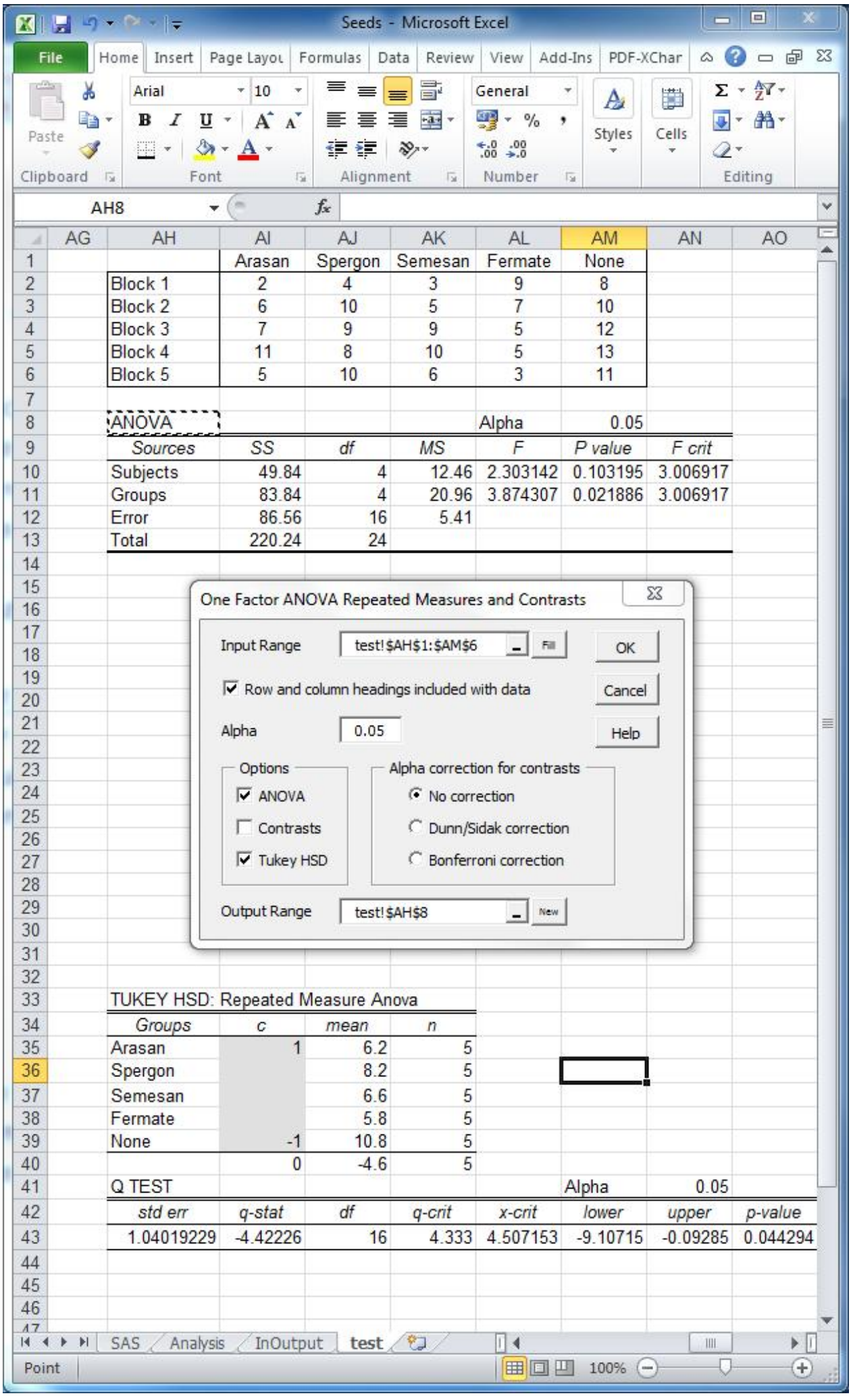
Further Two-Way Analysis

Post hoc tests and contrasts work in much the same way as in one-way ANOVA. We only need to replace MSE=MSW and DFE=DFW from one-way ANOVA with the new MSE and DFE for two-way ANOVA. (Recall that MSE is our estimator s^2 for the variance σ^2.)

Example: (soya beans continued.)
Consider comparing all pairs of treatments using Tukey's HSD. The minimum honestly significant difference at $\alpha = 5\%$ is

$$\begin{aligned} HSD &= q_\alpha(k, \text{DFE}) \cdot \sqrt{\frac{\text{MSE}}{n}} \\ &= q_{0.05}(5, 16) \cdot \sqrt{\frac{5.41}{5}} \\ &= 4.33 \cdot 1.04 = 4.5 \end{aligned}$$

Only the differences of the pairs Arasan-None and Fermate-None exceed that value and are therefore significant.



	AG	AH	AI	AJ	AK	AL	AM	AN	AO
1			Arasan	Spergon	Semesan	Fermate	None		
2		Block 1	2	4	3	9	8		
3		Block 2	6	10	5	7	10		
4		Block 3	7	9	9	5	12		
5		Block 4	11	8	10	5	13		
6		Block 5	5	10	6	3	11		
7									
8		ANOVA				Alpha	0.05		
9		*Sources*	*SS*	*df*	*MS*	*F*	*P value*	*F crit*	
10		Subjects	49.84	4	12.46	2.303142	0.103195	3.006917	
11		Groups	83.84	4	20.96	3.874307	0.021886	3.006917	
12		Error	86.56	16	5.41				
13		Total	220.24	24					
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									
28									
29									
30									
31									
32									
33		TUKEY HSD: Repeated Measure Anova							
34		*Groups*	*c*	*mean*	*n*				
35		Arasan	1	6.2	5				
36		Spergon		8.2	5				
37		Semesan		6.6	5				
38		Fermate		5.8	5				
39		None	-1	10.8	5				
40			0	-4.6	5				
41		Q TEST					Alpha	0.05	
42		*std err*	*q-stat*	*df*	*q-crit*	*x-crit*	*lower*	*upper*	*p-value*
43		1.04019229	-4.42226	16	4.333	4.507153	-9.10715	-0.09285	0.044294
44									
45									
46									

Example: (film processing continued.)

Consider testing the following two contrasts:
$H_0(1)\colon \mu_A = \frac{1}{2}(\mu_B + \mu_C)$, $\quad H_0(2)\colon \mu_B = \mu_C$.

	A	B	C	$\sum \lambda_i^2$
$\bar{X}_i$	33.867	30.733	29.267	
$\lambda_i(1)$	1	-0.5	-0.5	1.5
$\lambda_i(2)$	0	1	-1	2

The contrast values and standard errors are:

$L_1 = 3.867$, $s(L_1) = \sqrt{5.811 \cdot 1.5/15} = 0.762$

$L_2 = 1.467$, $s(L_2) = \sqrt{5.811 \cdot 2/15} = 0.880$

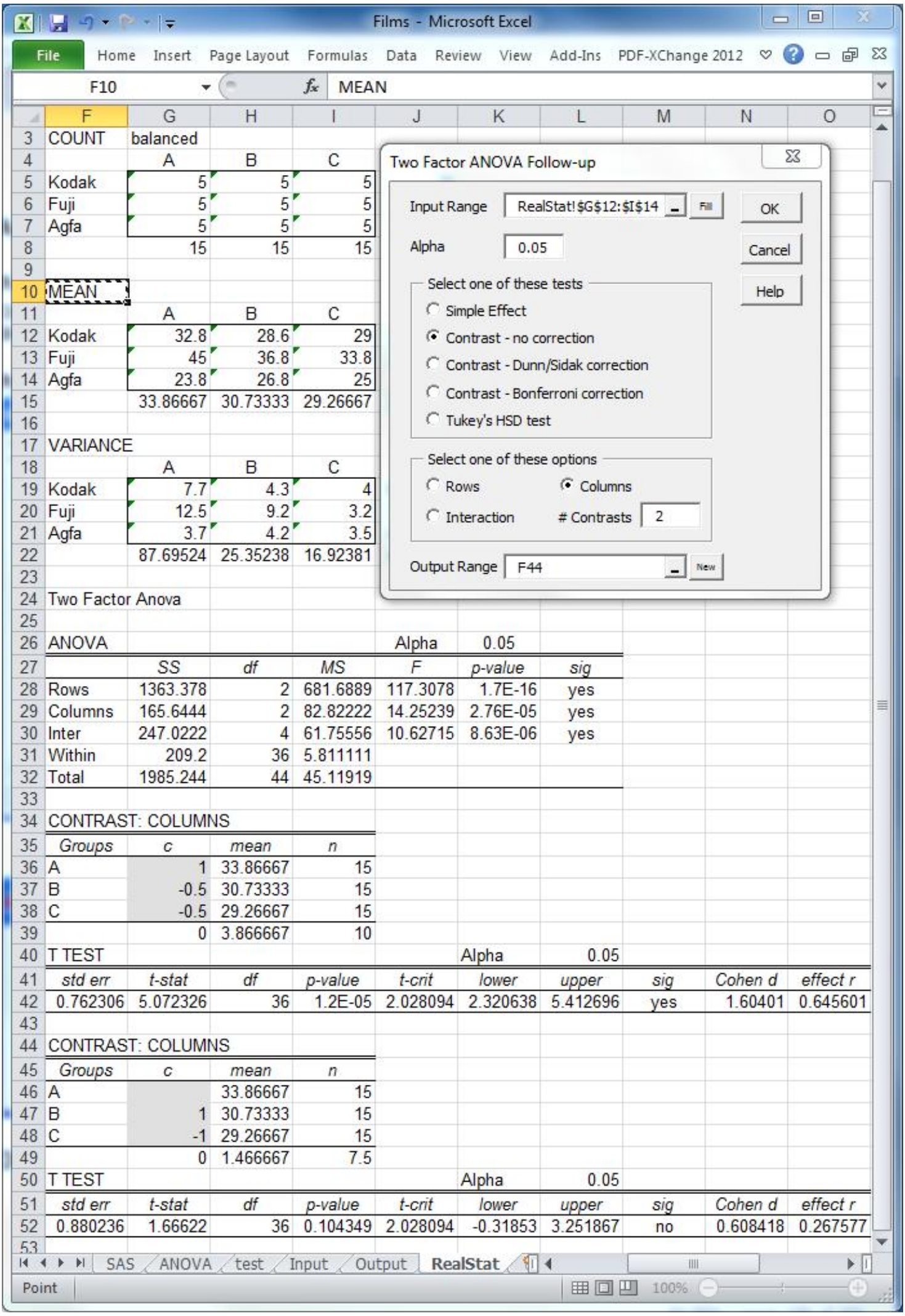
after inserting MSW=5.811 and n=15 into

$$s(L) = \sqrt{\text{MSE} \sum \lambda_i^2 / n}.$$

The t-statistics for the two contrasts are:

$$t_1 = \frac{L_1}{s(L_1)} = 5.07 \quad \text{and} \quad t_2 = \frac{L_2}{s(L_2)} = 1.67$$

with $df = N - IJ = 45 - 3 \cdot 3 = 36$, such that $t_{0.025}(36) = 2.03$; $p_1 = 0.000$, $p_2 = 0.104$. So we reject $H_0(1)$ and accept $H_0(2)$.



	F	G	H	I	J	K	L	M	N	O
3	COUNT	balanced								
4		A	B	C						
5	Kodak	5	5	5						
6	Fuji	5	5	5						
7	Agfa	5	5	5						
8		15	15	15						
9										
10	MEAN									
11		A	B	C						
12	Kodak	32.8	28.6	29						
13	Fuji	45	36.8	33.8						
14	Agfa	23.8	26.8	25						
15		33.86667	30.73333	29.26667						
16										
17	VARIANCE									
18		A	B	C						
19	Kodak	7.7	4.3	4						
20	Fuji	12.5	9.2	3.2						
21	Agfa	3.7	4.2	3.5						
22		87.69524	25.35238	16.92381						
23										
24	Two Factor Anova									
25										
26	ANOVA				Alpha	0.05				
27		*SS*	*df*	*MS*	*F*	*p-value*	*sig*			
28	Rows	1363.378	2	681.6889	117.3078	1.7E-16	yes			
29	Columns	165.6444	2	82.82222	14.25239	2.76E-05	yes			
30	Inter	247.0222	4	61.75556	10.62715	8.63E-06	yes			
31	Within	209.2	36	5.811111						
32	Total	1985.244	44	45.11919						
33										
34	CONTRAST: COLUMNS									
35	*Groups*	*c*	*mean*	*n*						
36	A	1	33.86667	15						
37	B	-0.5	30.73333	15						
38	C	-0.5	29.26667	15						
39		0	3.866667	10						
40	T TEST					Alpha	0.05			
41	*std err*	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*Cohen d*	*effect r*
42	0.762306	5.072326	36	1.2E-05	2.028094	2.320638	5.412696	yes	1.60401	0.645601
43										
44	CONTRAST: COLUMNS									
45	*Groups*	*c*	*mean*	*n*						
46	A		33.86667	15						
47	B	1	30.73333	15						
48	C	-1	29.26667	15						
49		0	1.466667	7.5						
50	T TEST					Alpha	0.05			
51	*std err*	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*Cohen d*	*effect r*
52	0.880236	1.66622	36	0.104349	2.028094	-0.31853	3.251867	no	0.608418	0.267577
53										