ANSWERS CHAPTER 14

THINK IT OVER

think it over

TIO 14.1: This is fundamental to understanding ANOVA. Before answering the specific question of between-sample influences, think about what three separate but overlapping distributions tell us. If the means are far apart then the null hypothesis, which states they are the same, cannot be true and therefore is rejected. Technically we say the between-treatments estimate will overestimate the population variance because the variance of the between-treatments is large. So, if you want visual confirmation of accepting or rejecting the null hypothesis, plot your samples and see where the means are in relation to one another.

If the samples are independent, an assumption for ANOVA, then no they cannot influence one another.

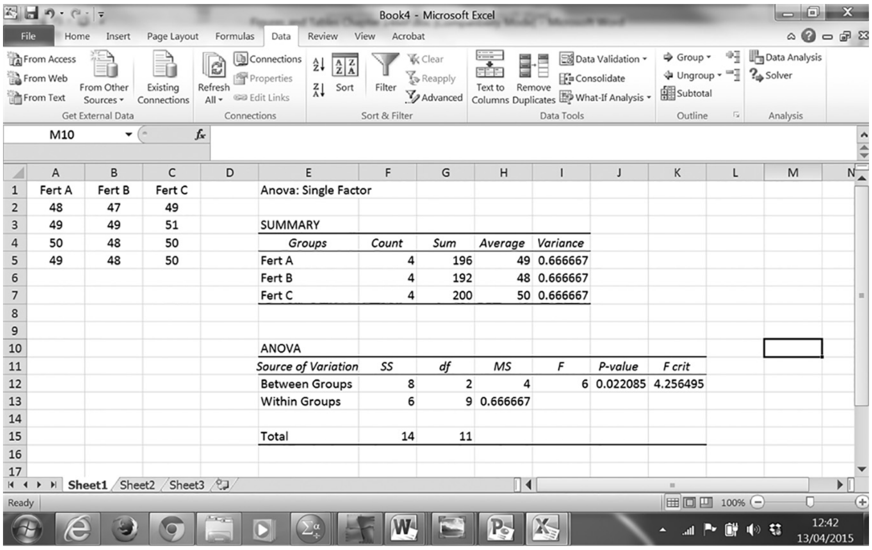
TIO 14.2: No answer required.

TIO 14.3: If F has a value of 0.002 then the fraction $F = \frac{MS_M}{MS_R}$ tells us that the value of the mean square of the model, which is the variation explained by the model, is much less than the mean square of the residuals (unsystematic errors that the model cannot account for). This can be interpreted to mean the model is not very good and the results are probably due to chance. If $MS_M = MS_R$, i.e. $F = 1$, then the result is unclear and therefore it is unlikely the model is useful. If you look at the critical values of F tables, you will see a value of 1 occurs when both values for the degrees of freedom are infinite. An F value of 50 could be interpreted to mean the model is significant, but this needs to be checked by comparing the critical value for your chosen significance level with the one given in the tables.

EXERCISES

1. (a) SS_T is the same, SS_M is higher (22.467 compared with 1.389) and SS_R is higher (96.567 compared with 17.847). The SS_R value says that there is considerably more variation due to extraneous factors if experience is not taken into account.
   (b) If I hadn't run the model with experience as a cofactor, I would recommend including other factors such as experience.
   (c) Experience is a significant factor.
   (d) The F value indicates that the model with experience included is better than the one with it left out.

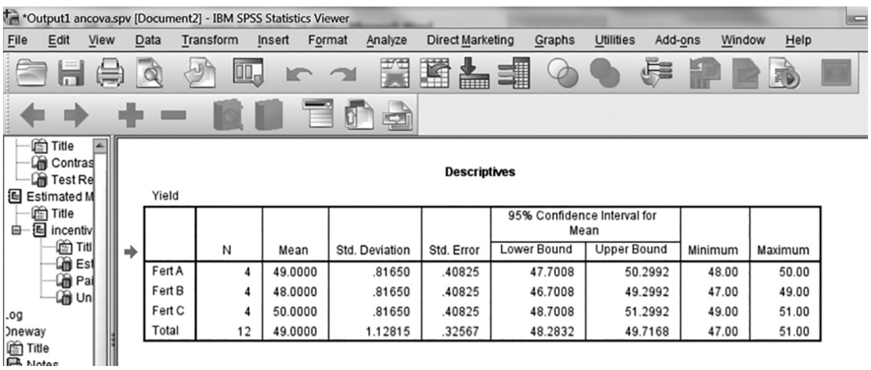
2. Excel output



	A	B	C	D	E	F	G	H	I	J	K
1	Fert A	Fert B	Fert C		Anova: Single Factor						
2	48	47	49								
3	49	49	51		SUMMARY						
4	50	48	50		*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	49	48	50		Fert A	4	196	49	0.666667		
6					Fert B	4	192	48	0.666667		
7					Fert C	4	200	50	0.666667		
8											
9											
10					ANOVA						
11					*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
12					Between Groups	8	2	4	6	0.022085	4.256495
13					Within Groups	6	9	0.666667			
14											
15					Total	14	11				

(a) One-way ANOVA since we are interested in analysing the effects of one independent variable, i.e. fertiliser.
(b) A: 49, B: 48, C: 50.
(c) (49 + 48 + 50)/3 = 49 kg per greenhouse.
(d) 14.
(e) 8.
(f) 6.

SPSS output



Descriptives

Yield

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Fert A	4	49.0000	.81650	.40825	47.7008	50.2992	48.00	50.00
Fert B	4	48.0000	.81650	.40825	46.7008	49.2992	47.00	49.00
Fert C	4	50.0000	.81650	.40825	48.7008	51.2992	49.00	51.00
Total	12	49.0000	1.12815	.32567	48.2832	49.7168	47.00	51.00

ANOVA

Yield

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.000	2	4.000	6.000	.022
Within Groups	6.000	9	.667		
Total	14.000	11			

3.

Between-Subjects Factors

		Value Label	N
Machine	1.00	Machine A	10
	2.00	Machine B	10
	3.00	Machine C	10
	4.00	Machine D	10
Shift	1.00	Shift 1	20
	2.00	Shift 2	20

Descriptive Statistics

Dependent Variable: Output

Machine	Shift	Mean	Std. Deviation	N
Machine A	Shift 1	4.8000	.83666	5
	Shift 2	6.0000	1.58114	5
	Total	5.4000	1.34990	10
Machine B	Shift 1	8.2000	1.30384	5
	Shift 2	8.8000	1.92354	5
	Total	8.5000	1.58114	10
Machine C	Shift 1	6.4000	1.67332	5
	Shift 2	6.2000	1.92354	5
	Total	6.3000	1.70294	10
Machine D	Shift 1	5.6000	1.51658	5
	Shift 2	7.6000	1.94936	5
	Total	6.6000	1.95505	10
Total	Shift 1	6.2500	1.80278	20
	Shift 2	7.1500	2.05900	20
	Total	6.7000	1.96377	40

Tests of Between-Subjects Effects

Dependent Variable: Output

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	65.600[a]	7	9.371	3.536	.006
Intercept	1795.600	1	1795.600	677.585	.000
Machine	51.000	3	17.000	6.415	.002
Shift	8.100	1	8.100	3.057	.090
Machine * Shift	6.500	3	2.167	.818	.494
Error	84.800	32	2.650		
Total	1946.000	40			
Corrected Total	150.400	39			

a. R Squared = .436 (Adjusted R Squared = .313)

(a) The *F* value for the interaction Machine * Shift, has a value of 0.818 which indicates the interaction is not significant and therefore the null hypothesis cannot be rejected. However, if you look at the *F* value for Machine, 6.415, this says that the machines are not equally effective at the 0.05 level, therefore we can reject the null hypothesis that the machines have equal means.

(b) The *F* value for Shift is 3.057, therefore the null hypothesis cannot be rejected, i.e. there is no significant difference between shifts.

(c) Overall, the analysis tells us that there is a difference between the machines and this difference is the same across the two shifts. In other words, the increase in rejected items is not due to operator error but to machine deficiency.