AnSc 5403
Biometry

Lecture Notes Supplement

I. Using Microsoft Excel® for statistical AOV and simple linear regression analysis

A. Although most students will find that the statistical software like the Statistical Analysis System (SAS; SAS Inst., Inc., Cary, NC) provides powerful tools for statistical analyses, more commonly available programs can be very useful

B. We will briefly examine the capabilities provided by Microsoft Excel® for analysis of variance of CRD and RCBD data, as well as simple linear regression

II. Example 1 - Analysis of data from a CRD with Microsoft Excel®

A. We will use data from Snedecor and Cochran (1967; p. 259) in which the grams of fat absorbed by batches of doughnuts (six batches of 24 doughnuts each per treatment) was measured:

Fat 1	Fat 2	Fat 3	Fat 4
64	78	75	55
72	91	93	66
68	97	78	49
77	82	71	64
56	85	63	70
95	77	76	68

B. Steps for analysis in Microsoft Excel®

1. Input these data into a worksheet exactly as shown in the table above

2. From the main menu, select “Tools” and then select “Data analysis”

   a. Excel® is not always loaded with the data analysis features, so you might need to go back to the “Add-Remove” programs feature in Windows and install this add-in.

   b. A new input box will appear with a listing of various data analysis tools

3. Select “Anova: Single factor”

   a. A new input box will appear with the options for the single-factor AOV

4. The first required action is to input the range of data to be analyzed

   a. Use your mouse or the shift and arrow keys to highlight the entire data range, including the data labels

5. Next, check whether the data are in rows or columns (columns in our case)

6. Click the option "Labels in first row"

7. The alpha option can be used to enter the level at which you want to evaluate critical values for the F statistic. The default value is 0.05.

8. Specify the output range. You may enter the reference for the upper-left cell of the output table and put the AOV output anywhere you want it. Excel® will automatically determine the size of the output area and display a message if the output table will replace existing data or extend beyond the bounds of the worksheet. The default option is a new worksheet ply.

9. Click "OK" and the output that follows is generated:

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Fat 1	6	432	72	178
Fat 2	6	510	85	60.4
Fat 3	6	456	76	97.6
Fat 4	6	372	62	67.6

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1636.5	3	545.5	5.406342914	0.006875948	3.098392654
Within Groups	2018	20	100.9			
Total	3654.5	23				

a. Notice that the AOV results are presented in a form similar to the one we have used, and a convenient summary of the means and variances for each treatment group are presented above the AOV table

III. Example 2 - Analysis of data from a RCBD with Microsoft Excel®

A. We will use data from Snedecor and Cochran (1967; p. 300) in which the number of failures out of 100 planted soybean seeds from five varieties were recorded in five replications (blocks):

	Varieties				
Block	Check	Arasan	Spergon	Semesan, Jr.	Fermate
1	8	2	4	3	9
2	10	6	10	5	7
3	12	7	9	9	5
4	13	11	8	10	5
5	11	5	10	6	3

B. Steps for analysis in Microsoft Excel®

1. As before, input these data into a worksheet exactly as shown in the table above (except for the "Varieties" heading)

2. From the main menu, select "Tools" and then select "Data analysis"

3. Select "Anova: Two-Factor Without Replication"

   a. As with the single-factor AOV, an input box will appear with the options

4. As before, input the range of data to be analyzed, including the data labels

5. Click the option "Labels"

6. As before, the alpha option can be used if desired.

7. As before, specify the output range or use the default of a new worksheet ply.

8. Click "OK" and the output that follows is generated:

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
1	5	26	5.2	9.7
2	5	38	7.6	5.3
3	5	42	8.4	6.8
4	5	47	9.4	9.3
5	5	35	7	11.5
Check	5	54	10.8	3.7
Arasan	5	31	6.2	10.7
Spergon	5	41	8.2	6.2
Semesan, Jr.	5	33	6.6	8.3
Fermate	5	29	5.8	5.2

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	49.84	4	12.46	2.303142329	0.10319528	3.006917382
Columns	83.84	4	20.96	3.874306839	0.021886163	3.006917382
Error	86.56	16	5.41			
Total	220.24	24				

9. As with the CRD, notice that the AOV results are presented in a form similar to the one we have used, and a summary of the means and variances for each block and variety are presented above the AOV table

IV. Example 3 – Simple linear regression analysis with Microsoft Excel®

A. We will use data from Snedecor and Cochran (1967; p. 150) in which the percentage of wormy fruit was regressed on the size of the apple crop on a tree:

Tree	Crop size, hundreds (X)	Wormy, % (Y)
1	8	59
2	6	58
3	11	56
4	22	53
5	14	50
6	17	45
7	18	43
8	24	42
9	19	39
10	23	38
11	26	30
12	40	27

B. Steps for analysis in Microsoft Excel®

1. As before, input these data into a worksheet exactly as shown in the table above, including the headings

2. From the main menu, select “Tools” and then select “Data analysis”

3. Select “Regression”

   a. As before, an input box will appear with the options

4. Input the range for the X data (including labels) to be analyzed, then click in the Y range box, and input the Y data (including labels)

5. Click the option “Labels”

6. The default confidence level (used to set confidence limits on the regression estimates in this case) is 95%, but this can be changed if desired

7. As before, specify the output range or use the default of a new worksheet ply

8. You can ask for several very useful plots (residual plots and line plots) by checking various boxes

9. Click “OK” and the output that follows is generated:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8808547
R Square	0.7759049
Adjusted R Square	0.7534954
Standard Error	5.2330121
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	Significance F
Regression	1	948.1558442	948.15584	34.623921	0.000154307
Residual	10	273.8441558	27.384416		
Total	11	1222			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	64.246753	3.602904977	17.831931	6.569E-09	56.21898072	72.27452577
Crop size, hundreds	-1.012987	0.172153459	-5.8842095	0.0001543	-1.396568822	-0.629405204

10. This output is a bit different than what we have discussed, in that it presents the results of regression in the context of an AOV table

    a. This format is very useful because the value for $s_{y.x}$ can be readily obtained from the AOV table by taking the square root of the mean square error (Residual MS) in this table

    b. The coefficients for the intercept and slope are given in the last two lines, along with their standard errors.