Introduction

Microsoft Excel features statistics in two places: in the list of available functions under Insert on the menu bar, and in Data Analysis under Tools. If you don't see a Data Analysis item, you need to install it from the set of Add-ins provided with Microsoft Office.

If you would like to go beyond Excel's capabilities, there are two popular choices. The Excel Add-in *StatTools* from Palisade (<u>www.palisade.com</u>) improves Excel's numerical accuracy and adds useful features, and *Resampling Stats in Excel* (<u>www.resamplingstats.com</u>) shows how to solve most of the problems here using straightforward simulation methods.

Note that the methods described here will work in almost any version of Excel.

Chapter One

This chapter reviews some simple graphics techniques.

Example 1.11

Making a bar chart (called a *histogram*) in Microsoft Excel is nearly automatic. For the case of Example 1.11, just list each category in its own cell in column A of a new worksheet. Then list the frequency number for each category in the adjacent cell of column B. Your worksheet should now look like Figure 1.11a.

\diamond	A	В
1	Academic Problems	7
2	Poor advising	3
3	Needed break	2
4	Economic reasons	11
5	Family	4
6	Attend other school	9
7	Personal	3
8	Other	3

Figure 1.11a

To make a bar chart (if the bars are up and down, Excel calls them Columns), click the Chart icon on the main Excel toolbar. You will be asked to pick a chart type (the plain chart choice at upper left will do nicely). Click Next at the bottom of the screen, and then specify a data range (here it's A1:B8). Click the option for data in columns, and then Next. In the preview of the chart that results, you can add the label, and add "frequency" as the y-axis label. Finally, to clean things up, go to the Legends option and click Display Legend off.

When you select Finish, you'll see the result in Figure 1.11b. If your chart doesn't look exactly like this, click on a corner of the frame and drag it until the size accommodates all the category labels.





Example 1.12

Dotplots are an early paper-and-pencil graphical method, and they don't appear as one of Excel's chart operations. So for this example, you use Excel as a blank sheet of graph paper.

Label cell A1 of a new worksheet as California, and cell B1 as Texas. Now using the numbers along the left as a guide, fill in the dots as asterisks. Figure 1.12 shows how this procedure looks for a large range of values.

This dotplot is of course vertical rather than horizontal, so you might want to print it out for sideways viewing.

0	A	B
32		*
33		
34		
35	*	***
36		
37	**	
38	-	
39		**
40	-	
41	*	
42		
43		
44	*	
45		
46	-	*
47	1	
48	-	
40		
50		
51		
52		
52		
54	2	
55	-	
56	-	
57		
58		
50		
60		
61		
62	-	
63	*	*
64	*	
65		*
66	**	
67	*	*
68	*	0.000
60		
70	*	
71		**
72	*	
72	*	
74	*	
74		

Figure 1.12

Chapter Two

This one simple simulation example is just the beginning of what can be done with simulation in statistics.

Example 2.3

There are several ways to approach this example, but they all start the same way. Pick any cell in a new worksheet, select Insert from the Excel main menu, then Function... from the Insert options, and then pick Math & Trig from the function groups, and finally RAND from the functions (Figure 2.3a).



Figure 2.3a

Now RAND returns a random number between 0 and 1, not 0 and 100. So a first step is to multiply this random number times 100. The next step is to use the function ROUND (also in Math & Trig) to round off the decimal places. A final refinement is to change the term RAND()*100 to RAND()*100 + 1.0. This makes it possible to reach the random number 100, despite rounding off, and to avoid getting zero as a bottle number. Typical results of using three cells with this formula appear in Figure 2.3b.

\diamond	A	В	C	D	E	F
1						
2	Three ra	andom n	umbers,	in the ra	ange 1 to	100
3						
4	72					
5	51					
6	36					
7						

Figure 2.3b

The final formula is thus =ROUND(RAND()*100 + 1.0).

There's still the nagging problem that for selecting bottles, you need three *different* random numbers, and sometimes this formula will produce duplicates in a set of three. In real statistical software, the random number functions available usually let you specify non-repeating random numbers. In this case, probably your best option is simply to discard sets with duplicates.

Chapter Three

Most graphics in Excel come directly from the Chart Wizard, featured as an icon on the menu bar.

Example 3.1

This example works out nicely in Excel. First, just enter the data from the Example 3.1 table as shown in Fig. 3.1a into a new worksheet. Now click the Chart Wizard icon on the Excel toolbar and pick the first type of Column chart. Excel automatically selects the data range and labels – you simply provide the title and label for the y-axis. Click Finish and you see the result in Fig. 3.1b.

\diamond	A	В	C	D
1	Perceived Risk	Smokers	Former	Non
2	Very Harmful	0.6	0.78	0.86
3	Somewhat	0.3	0.16	0.1
4	Not Too	0.07	0.04	0.03
5	Not At All	0.03	0.02	0.01

Figure 3.1a





Example 3.4

Once again, put he data for this example into an Excel table, like the one shown in Figure 3.4a. Drag-select the table, click the Chart Wizard icon, select Pie as the chart type, and select the simplest kind of pie. Then just click Next and then Finish, and Excel returns the chart shown in Figure 3.4b.

\diamond	A	В
1	Outcome	relative frequency
2	caught, 1st attempt	0.43
3	caught, 2nd attempt	0.17
4	caught, 3rd attempt	0.01
5	not caught	0.39
6		

Figure 3.4a



Figure 3.4b

Example 3.8

Stem-and-leaf charts don't have an automatic chart type in Excel. This is probably because stem-and-leaf charts are a World War I era statistical graphic originally designed for typewriters, and Excel can generate the equivalent bar chart from the same data instantly. Figure 3.8 shows the stem-and-leaf display of Figure 3.10 of the text, simply using a new Excel worksheet as a blank piece of paper. Note: you have to format the "leaf" cells as Text in Excel to get these large numbers to display correctly.

\diamond	A	В	
1	0	4	
2	1	134567889	
3	2	1223456666777889999	
4	3	011223334455666677777888899999	
5	4	11122222334444556666666777888889999	
6	5	00111222233455666667777888899	
7	6	01111244455666778	

Figure 3.8

Example 3.17

In Excel, histograms are easily produced with the Histogram function in the set of Data Analysis tools found, naturally, in the Tools menu (Figure 3.17a). (Note: the Data Analysis tools are provided as an Add-In that must be installed separately from the Microsoft Office CD.)



Figure 3.17a

First put the mercury data in column A of a new worksheet. In column B, put 10,20,30,40,50,60,70 as the "bin range". Now select Histogram from the Data Analysis options. In the Histogram dialog box, use A1:A40 as the input data range, B1:B7 as the bin range, and select new worksheet ply and Chart Output as the desired output format.

After clicking Finish, you'll see the output in Figure 3.17b. Well, almost. You'll see thinner columns, so double-click on any column, pick the options screen at far right in the formatting set, and set column gap to zero.



Figure 3.17b

Example 3.21

Scatterplots are a built-in function available from the Chart Wizard. Start with a new worksheet. In column A, put the label "soil pH" and the set of pH vales as x-data. In column B, put the label "dieback %" and the percentage values as y-data. Drag-select this data and click the Chart Wizard.

Excel generates (just click Finish) the chart seen in Figure 3.21 (identical to Figure 3.32b in your textbook). Excel does the labeling and axis-scaling automatically.



Figure 3.21

Chapter 4

This chapter covers basic descriptive statistics.

Example 4.3

In a new worksheet, simply list the 40 website-visit values in Column A. Select cell B1. Choose Insert from the main menu, and select Function. From the Function selections choose Statistical, and from the statistical functions pick AVERAGE. In the AVERAGE dialog box, put A1:A40 as the input (Number 1). Click finish to see the result 23.10.

For a slightly different perspective on the data set, choose the statistical function MEDIAN instead. You should see 13 as the answer, suggesting that the data set contains a few big numbers bringing up the mean.

Example 4.8

The functions STDEV, VAR, and so forth are all available as functions to insert, but for this problem there's a simpler method. The Data Analysis options under Tools in the main menu has a choice called Descriptive Statistics.

So type the data for this example into cells A1 through A7, and use the for input in the Descriptive Statistics dialog box. Click the Summary Statistics choice and then OK, and you see the results in Figure 4.8.

0	A	B
1	Column	1
2		
3	Mean	287.7142857
4	Standard Error	42.42961358
5	Median	270
6	Mode	#N/A
7	Standard Deviation	112.2582058
8	Sample Variance	12601.90476
9	Kurtosis	1.40978784
10	Skewness	0.992544847
11	Range	342
12	Minimum	155
13	Maximum	497
14	Sum	2014
15	Count	7

Figure 4.8

Example 4.11

Excel doesn't do boxplots. If you have three data sets, you can approximate one with the Chart Wizard choice Stock, variant Open-Hi-Lo-Close, but here we have only one data set.

But Excel does have a Quartile function as part of the statistical function set. If you arrange the data and use the Quartile function as shown in Figure 4.11, you at least have a basis for drawing a boxplot by hand.

\diamond	A	В	C	D	
1	0.553				
2	0.57				
3	0.576	0.553	=QUARTILE(A1:A20,0)	min	
4	0.601	0.606	=QUARTILE(A1:A20,1)	Q1	
5	0.606	0.641	=QUARTILE(A1:A20,2)	median	
6	0.606	0.6765	=QUARTILE(A1:A20,3)	Q3	
7	0.609	0.933	=QUARTILE(A1:A20,4)	max	
8	0.611				
9	0.615				
10	0.628				
11	0.654				
12	0.662				
13	0.668				
14	0.67				
15	0.672				
16	0.69				
17	0.693				
18	0.749				
19	0.884				
20	0.933				
21					

Figure 4.11

Chapter 5

Excel has a number of tools for describing and correlating (x,y) data.

Example 5.4

There are two ways to calculate a correlation in Excel. The Insert Function sequence gives you the option CORREL in the function category Statistical. There's also an option called Correlation in Data Analysis under Tools.

For simplicity, let's just use CORREL. Type the mare weights in A1 through A15, and the foal weights in B1 through B15. Then use these arrays as input to the function CORREL as shown in Figure 5.4. You should get a result that convinces you there's very little correlation at all.

A	B	C	D	E	F	G
556	129	CODDE				
638	119	CORREL				
588	132	Array1	a1:a15	<u></u>	556;638;588;551	
550	123.5	Array2	b1:b15	(A) = {	129;119;132;12:	
580	112					1.5
642	113.5			= 0	.00134831	
568	95	Returns the correlati	on coefficient betwe	en two data sets.		
642	104	Array2	is a second cell ran	ge of values. The values	ies should be numbers	
556	104	, and pe	names, arrays, or r	eferences that contain	numbers.	
616	93.5					
549	108.5	Pormula result	t = 0.00134831	(Cancel OK)
504	95			_		C
515	117.5					
551	128					
594	127.5					
	A 556 638 588 550 580 642 568 642 556 616 549 504 515 551 551 594	A B 556 129 638 119 588 132 550 123.5 580 112 642 113.5 568 95 642 104 556 104 616 93.5 549 108.5 504 95 515 117.5 551 128 594 127.5	A B C 556 129 CORREL 638 119 Array1 588 132 Array2 550 123.5 Array2 580 112 Array2 642 113.5 Array2 568 95 Returns the correlation 642 104 Array2 556 104 Array2 556 104 Formula resu 554 108.5 Formula resu 504 95 117.5 551 117.5 551 554 128 594 127.5	A B C D 556 129 CORREL al:a15 638 119 Array1 al:a15 588 132 Array2 bl:b15 580 112 Array2 bl:b15 642 113.5 S68 95 642 104 Array2 is a second cell ran names, arrays, or r 616 93.5 S49 108.5 S15 515 117.5 Formula result = 0.00134831 504 95 128 S94 127.5	A B C D E 556 129 CORREL CORREL CORREL 638 119 Array1 a1:a15 A = { 588 132 Array2 b1:b15 A = { 580 112 Array2 b1:b15 Image: Correlation coefficient between two data sets. 642 104 Array2 is a second cell range of values. The value names, arrays, or references that contain names, arrays, or references tha	A B C D E F 556 129 CORREL CORREL

Figure 5.4

Example 5.6

Excel's regression option under Data Analysis Tools handles this type of problem automatically. Type the mean times (x) in column A, with the label "mean time" in A1, and survival rates (y) in column B, with label "survival rate" in B1. Call up the regression dialog box, and enter A1:A 6 as the x-array and B1:B6 as the y-array. Check the boxes for and for line-fit plot and click OK.

The output is a complete regression study. If you scroll to the right of the data output, you'll find the regression plot, which after a bit of scaling (click on the plot and drag one of the handles) looks like Figure 5.6. If you click on the chart, the choice Chart appears in the main menu, and one of the options under Chart is Add Trendline... (choose Linear in the dialog box that appears).





Example 5.10

Again, the Regression tool in Data Analysis has the option for residual data. Fill in the label "resonance" and data in column A, and the label "acceleration" and the acceleration data in column B.

Enter this data in the Regression dialog box. Check Labels, Residuals, and Residual Plot. There's a large set of data output (Figure 5.10a) plot will be found to the right of the data output (Figure 5.10b). If you like, you can delete the last data point and compare results.

0	A	Б	C	
2				
3	Regress	sion Statistics		
4	Multiple R	0.775000393		
5	R Square	0.60062561		
6	Adjusted R Squa	0.567344411		
7	Standard Error	1.139752833		
8	Observations	14		
9				
10	ANOVA			
11		df	SS	
12	Regression	1	23.44370461	23.
13	Residual	12	15.58843825	1.2
14	Total	13	39.03214286	
15				
16		Coefficients	Standard Error	t
17	Intercept	42.37454497	1.840342438	23.
18	resonance	-0.064520998	0.015187932	-4.
19				
20				
21				
22	RESIDUAL OUT	PUT		
23				
24	Observation	Predicted acceleration	Residuals	
25	1	35.59984018	0.400159815	

Figure 5.10a



Figure 5.10b

Chapter Six

Excel also has reasonable capabilities for simulation, which helps in probability problems.

Example 6.10 Devore/Peck, 6.32 in Peck/Olsen/Devore

Type the formula =IF(RAND()>0.5,"B","G") into cell A1 and copy it down through cell A20. You get a string of B's and G's corresponding to boys and girls. The only problem here is that RAND recalculates every time you click anywhere. So Copy the range A1:A20, and (from the Edit menu) choose the Pate Special: Values option to paste a static set of B's and G's into B1:B20.

Now look down column B and pull out the substrings that end in B (see Figure 6.10). Actually, you know in advance that for a big enough starting string (for example, 100 B/G instead of just 20) the number of B's and G's will be nearly equal, since the IF formula generates them with equal probability.

\diamond	A	В	C	D	E	F	G	Н		J	K	L	M	Ν	0	
1	В	В	В													
2	G	В		В												
3	G	В			В											
4	В	G				G										
5	G	G				G										
6	G	В				В										
7	G	В					в									
8	G	В						В								
9	В	G							G							
10	В	В							в							
11	G	G								G						
12	G	G								G						
13	G	В								в						
14	В	В									В					
15	G	В										В				
16	В	В											В			
17	В	В												в		
18	В	G													G	
19	G	G													G	
20	В	В													В	

Figure 6.10

Chapter Seven

All common distribution are available as functions in the statistics category.

Example 7.13 Devore/Peck, 7.25 in Peck/Olsen/Devore

The function NORMDIST from the statistical functions category can be used for this example. Insert function NORMDIST in cell A1, and use the values 94 for x, 100 for mean, 6 for standard deviation, and 1-true for cumulative (Figure 7.13). This gives the cumulative value from 0 to 94 cm.

Do the same in cell A2, except using 112 as the x-value. This gives the cumulative value from 0 to 112. Now subtract A1 from A2. This gives you the value of the shaded area shown in Fig 7.24 of the text.

x	94	<u> </u>
Mean	100	A = 100
Standard_dev	6	a = 6
Cumulative	1	A = TRUE
turns the normal c	imulative distribution	of or the specified mean and standard
Consulation		TTIR FUTTILI, TURNER, TRACTORISTICS, TURNER, TRACT

Figure 7.13

Chapter Eight

There are no examples for Chapter Eight.

Chapter Nine

For confidence intervals in Excel, you have to do a bit of set-up work yourself.

Example 9.2

This example simply uses the functions AVERAGE and MEDIAN. Enter the data in cells A1 to A20 of a new worksheet, and just perform the calculation shown in Figure 9.2. There is also a function TRIMMEAN, but in this case it's simpler just to use AVERAGE and the right part of the array.

\diamond	A	В	C	
1	4			
2	5			
3	5			
4	5.25	AVERAGE(A1:A20)=	7.075	
5	5.5			
6	6.25	MEDIAN(A1:A20)=	7.125	
7	6.25			
8	6.5	trimmed mean=		
9	6.5	AVERAGE(A3:A18)=	7.03125	
10	7			
11	7.25			
12	7.75			
13	8			
14	8			
15	8			
16	8.25			
17	8.5			
18	8.5			
19	9.5			
20	10.5			

Figure 9.2

Example 9.5

This is just a straightforward calculation with the given values, as shown in Figure 9.5.

\diamond	A	В	C
1	critical value for z (90%)	1.645	
2	n	750	
3	р	0.167	
4			
5	z*SQRT(p(1-p)/n)	0.02240351	
6			
7	p+0.022	0.189	
8			
9	p-0.022	0.145	
10			

Figure 9.5

Example 9.9

Once again, the simplest way to perform this calculation in Excel is to plug the numbers into the textbook formula, since Excel doesn't have a built-in function that's convenient for this example. Figure 9.9 shows a suggested layout for the problem.

\diamond	A	В	C	
1	0.95	AVERAGE	0.9255	
2	0.85	STDEV(A1:A20)	0.0809467	
3	0.92			
4	0.95	2.86*s/SQRT(n)	0.05176667	
5	0.93			
6	0.86	AVERAGE + d	0.97726667	
7	1	AVERAGE - d	0.87373333	
8	0.92			
9	0.85			
10	0.81			
11	0.78			
12	0.93			
13	0.93			
14	1.05			
15	0.93			
16	1.06			
17	1.06			
18	0.96			
19	0.81			
20	0.96			



Chapter 10

Hypothesis testing on a single variable uses a variety of functions, since the Data Analysis tools don't feature an all-in-one solution

Example 10.11

Most of this calculation just follows the text, as shown in Figure 10.11. For calculating the P-value given the value z=7.14, use the Statistical function NORMSDIST. For the value of z here, NORMSDIST returns the value 1, meaning that there is essentially no area to the right of such a large z value, so P-value is zero.

\diamond	A	В	C
1			
2	pi	0.5	
3			
4	n	310	
5	count	217	
6	p=count/n	0.7	
7			
8	z=(p-pi)/SQRT(pi*(1-pi)/n)	7.04273	
9			
10	NORMSDIST(z)	1	
11			

Figure 10.11

Example 10.14

Use the formula for t to calculate its value from the data (Excel will give you the AVERAGE and the STDEV). The function TDIST gives the area, but since it doesn't accept negative values, so use

0.0669316

as x in the dialog box, and 1-TDIST(x) = P-value as the area (Figure 10.14)

~	A		-		U		E	
1	66	AVERAGE=		74.8				
2	70							
3	75	HYP. VALUE		75				
4	88							
5	69	STDEV=	9.4	44927981				
6	89	C5						
7	71	N=		10				
8	71							
9	63	t=	-0	.0669316				
10	86							
11			-(-c	:9,9,1)				
12		_						
13		TDIST						
14			x	-c9		A =	0.066931612	
15		Deg_free	dom	9		À =	9	
16			ails	1			21	
17						<u> </u>		
18						10	0 474040725	
19		NV-3 10 1450				-	0.474049725	
20		Returns the Stu	Returns the Student's t-distribution.					
21		1	Tails specifies the number of distribution tails to return: one-tailed distribution = 1; two-tailed distribution = 2.					
22								
23								
24		Formula	i resul	t = 0.474049725			Cancel OK	
25								

Figure 10.14

Chapter Eleven

The problems in this chapter are handled fairly automatically by Data Analysis routines.

Example 11.2

Excel's Data Analysis option under Tools has a test called t-Test: Paired Two Sample For Means that matches this example. Type the data sets in column A and column B, call up the t-test, and fill in the data as shown in Figure 112.a.

0	A	В	C	D		E	F
1	Never	Used					
2	0.82	0.94		t-Test: Pa	ired Two S	Sample for M	Means
;	0.94	1.09	loc	117			1
	0.96	0.97			¥		
	1.31	0.98	Vari	able 1 range:	al:all		
	0.94	1.14	Vari	able 2 range:	b1:b11		(A)
	1.21	0.85	Нур	othesized mean di	ifference:	0	
	1.26	1.3		abels			
	1.09	0.89	Alph	ia:	0.05		
)	1.13	0.87			0.05		-
2	1.14	1.01	Ou	tput Options			
2			00	Output range:	1		(A)
3				vew worksheet plv	e		
4			0	lew workbook			
5				ten norkbook			
5			(He	alp		Cancel	OK
7				<u></u>		Canter	

Figure 11.2a

When you click OK, you get the results shown in Figure 11.2b. The P-value shown is 0.137224136, slightly different from the textbook answer and the MINITAB answer. In this case, the answer depends strongly on the number of digits carried along during the calculation.

\diamond	А	В	C
1	t-Test: Paired Two Sample for		
2			
3		Never	Used
4	Mean	1.08	1.004
5	Variance	0.025511111	0.01933778
6	Observations	10	10
7	Pearson Correlation	0.049525026	
8	Hypothesized Mean Difference	0	
9	df	9	
10	t Stat	1.163750101	
11	P(T<=t) one-tail	0.137224136	
12	t Critical one-tail	1.833113856	
13	P(T<=t) two-tail	0.274448272	
14	t Critical two-tail	2.262158887	
15			

Figure 11.2b

Example 11.4

Again, there's a "canned" test in Excel for this example, t-Test: Two Sample Assuming Unequal Variances. Type the data in columns A and B, use 0 for the hypothesized difference and click OK.

You can use the elements of output to compute a confidence interval using the formula in the text, as shown in Figure 11.4. The formula for the difference element (the amount plus or minus the mean) is shown in the formula bar at top.

000		wt114.xls			
A	В	С			
t-Test: Two-Sample Assuming Unequal Variances					
8	talking	counting			
Mean	108.75	102.25			
Variance	22.5	29.0714286			
Observations	8	8			
Hypothesized Mean Di	ffere 0				
df	14				
t Stat	2.56008137				
P(T<=t) one-tail	0.01133482				
t Critical one-tail	1.76130925				
2 P(T<=t) two-tail	0.02266964				
3 t Critical two-tail	2.1447886				
4					
5 mean1-mean2=	6.5				
6					
7 delta=	5.44557921				
8					

Figure 11.4

Example 11.8

Excel doesn't have a t-based confidence interval, so you have to get a value for t-crit from the table in your textbook. The rest of the work is just the arithmetic shown in Figure 11.8.

\diamond	A	С	D	
1	- 5	AVERAGE(A1:A8)=	-13.625	
2	-17			
3	-23	STDEV(A1:A8)=	8.27970841	
4	-22			
5	-17	X+TCRIT*STDEV/SQRT(n)	-6.687254	
6	-4			
7	-18	X-TCRIT*STDEV/SQRT(n)	-20.562746	
8	-3			
9				

Figure 11.8

Example 11.10

Excel does have a prepared test for this case, but it expects you to have the actual primary data to place in two arrays, rather than the summary values p_1 and p_2 as given in this example.

So we just perform the calculation in Figure 11.10 to get a z-value, and use the function NORMSDIST to get a value for P. The returned value 1 says that all the area of thee distribution is to the right of this z-value, so P = 0.

\diamond	A	В	C
1	p1	0.763	
2	n1	80	
3	p2	0.4	
4	n2	80	
5	рс	0.5815	
6	z test stat=	4.65386767	
7			
8	NORMSDIST(z)=	0.99999837	
9			



Chapter 12

In Excel, there's almost no work for you to do on chi-square calculation.

Example 12.7

The Excel function CHITEST returns a probability directly from the data arrays. Type in the data from Example 12.7 as shown in Figure 12.7. Then call the function CHITEST with the arrays for "Actual" and "Expected" as shown. The answer is slightly different from the textbook answer because the Excel calculation carries more significant digits.

0	A	В	C	D	E
1	observed	fem	male	row total	
2	never	210	350	560	
3	some	190	320	510	
4	always	400	530	930	
5	clm total	800	1200	2000	
6					
7	expected	224	336	560	
8		204	306	510	
9		372	558	930	
10		800	1200	2000	
11					
12			CHITEST=	=CHITEST(B	2:C4,B7:C9)
13					
14		CHITEST			
15		Actual_range	B2:C4	(≜) = {210),350;190,321
16		Expected range	B7.C9	A = {224	1 336-204 30
17			01.03		
18					7700/07
19				= 0.03	(299697
20		Returns the test for independence: the value from the chi-squared distribution for the			
21		Actual_range is the range of data that contains observations to test against			ons to test against
22			expected values.		
23					
24		Formula resul	t = 0.037399697	Ca	ncel OK
25					

Figure 12.7

Chapter 13

Simple regression in Excel is handled by the Data Analysis tool called, appropriately enough, Regression.

Example 13.2

Type the x data in column A and the y data in column B of a new worksheet. Fill in these arrays in the Regression dialog box, check Line Plot, and click OK.

In the output, you'll find slope and intercept of the line fit down the page (Figure 13.2a) and a fitted line plot on the right side of the sheet (Figure 13.2b).

	Coefficients
Intercept	-1163.45
x	245.15

Figure 13.2a



x Line Fit Plot

Figure 13.2b

Example 13.4

As in Example 13.2, enter the data in columns A and B, but also check the boxes for residual output in the Regression dialog box.

Output is shown in Figure 13.4. Note that s_b is given in the table as 0.591, the standard error of the intercept, so you can use this to calculate a confidence interval.

\diamond	A	В	C	D
1	SUMMARY O	UTPUT		
2				
3	Regression	Statistics		
4	Multiple R	0.79619214		
5	R Square	0.63392193		
6	Adjusted R S	0.59324658		
7	Standard Err	2.18828632		
8	Observations	11		
9				
10	ANOVA			
11		df	SS	MS
12	Regression	1	74.6298994	74.6298994
13	Residual	9	43.0973733	4.78859703
14	Total	10	117.727273	
15				
16		Coefficients	tandard Erro	t Stat
17	Intercept	88.7956487	5.75010819	15.4424309
18	X Variable 1	-2.3335102	0.59109519	-3.947774
19				

Figure 13.4

Example 13.6

Again, enter the data in the first two columns of a new worksheet.

In the Regression dialog box, check Residuals, Standardized Residuals, and Normal Probability Plot.

The residuals (actual and standardized) are shown in Figure 13.6a and the normal probability plot in Figure 136.b. Note that the standardized residual values are in some cases a bit different from those in the text.

21					
22	RESIDUAL O	UTPUT			
23					
24	Observation	Predicted y	Residuals	Standard Residuals	
25	1	110.9600998	-40.9601	-0.824303383	
26	2	196.2468828	-46.246883	-0.930697488	
27	3	196.2468828	63.7531172	1.283002495	
28	4	217.5685786	12.4314214	0.250176704	
29	5	217.5685786	37.4314214	0.753290337	
30	6	238.8902743	-73.890274	-1.487008173	
31	7	238.8902743	-13.890274	-0.279535455	
32	8	260.2119701	79.7880299	1.605697823	
33	9	281.5336658	23.4663342	0.472249305	
34	10	281.5336658	53.4663342	1.075985664	
35	11	302.8553616	-12.855362	-0.258708307	
36	12	302.8553616	37.1446384	0.747518958	
37	13	324.1770574	-99.177057	-1.995893184	
38	14	324.1770574	-24.177057	-0.486552286	
39	15	388.1421446	-8.1421446	-0.163856959	
40	16	388.1421446	11.8578554	0.238633947	

Figure 13.6a





Figure 13.6b

Example 13.11

Enter the data in the first two columns of a new worksheet. For such a large amount of data, you'll want to proofread your entries several times.

Run the Regression routine from Data Analysis tools. You might note that the results in Figure 13.11a are very slightly different from the MINITAB results in the text – MINITAB is a bit "more right" than Excel.

\diamond	A	В	C
1	SUMMARY OUTPUT	Γ	
2			
3	Regression S	statistics	
4	Multiple R	0.87608976	
5	R Square	0.767533268	
6	Adjusted R Square	0.761998346	
7	Standard Error	1.364978208	
8	Observations	44	
9			
10	ANOVA		
11		df	SS
12	Regression	1	258.3670487
13	Residual	42	78.25295132
14	Total	43	336.62
15			
16		Coefficients	Standard Error
17	Intercept	0.716918657	1.288887838
18	х	0.961294224	0.081632574
19			
20			
21			

Figure 13.11a

The confidence interval calculation is shown in Figure 13.11b, and it gives a confidence interval quite close to that using MINITAB figures. In reality, given the precision of the original data, a confidence interval of (14.8, 15.5) would be acceptable.

16		Coefficients
17	Intercept	0.716918657
18	x	0.961294224
19		
20		
21		
22		
23	a+b*15=	15.13633202
24		
25	s(a+b15)=	0.211709722
26		
27	a+b15+tcrit*s	15.49200436
28		
29	a+b15-tcrit*s	14.78065969
20		

Figure 13.11b

Chapter Fourteen

There is just one trick in Excel in moving from simple regression to multiple regression, which is putting all the x-variables in an array.

Example 14.6

Arrange the data for this example in three columns, A and B for the x's and C for the y's. Now when you call up the Regression dialog box, enter C1:C14 as the y-array and A1:B14 as the x-array.

		-	÷.
		В	C
1	SUMMART OUTPUT		
2			
3	Regression St	atistics	
4	Multiple R	0.97389281	
5	R Square	0.9484672	
6	Adjusted R Square	0.93816064	
7	Standard Error	4.37937466	
8	Observations	13	
9			
10	ANOVA		
11		df	SS
12	Regression	2	3529.903083
13	Residual	10	191.7892244
14	Total	12	3721.692308
15			
16		Coefficients	Standard Error
17	Intercept	-7.3506598	3.484667762
18	x1	0.11273292	0.029690829
19	x2	0.34900031	0.071306428
20			

Figure 14.6

Figure 14.6 shows that Excel understood from the x-array that this was a multiple regression problem. Compare the results to the MINITAB output in the text – in this case there's an exact match.

Example 14.17

This multiple regression problem requires a bit of thinking about convenient handling of the x-array. One way to handle things is shown in Figure 14.17, in which all the data is entered in its own space, and data to be tested as a model is transferred to an "active site". To test this two-facto model, enter A1:B10 as the x-array in the Regression dialog box.

\diamond	A	В	C	D	E	F	G	Н		
1	x1	x4			у	x1	x2	хЗ	x4	
2	2166600	2.01			4.89	2166600	30	4	2.01	
3	751658	3.54			3.49	751658	40	2	3.54	
4	2422650	3.63			4.33	2422650	28	3	3.63	
5	224573	4.65			8.24	224573	25	2	4.65	
6	3917800	1.71			5.1	3917800	26	4	1.71	
7	2866526	2.27			2.79	2866526	35	4	2.27	
8	1698161	3.12			5.89	1698161	28	3	3.12	
9	1046260	4.77			6.38	1046260	33	4	4.77	
10	1108828	7.56			5.25	1108828	28	4	7.56	

Figure 14.17

For this data, using x_1 and x_4 , the output shows an adjusted R^2 of zero, indicating a pretty poor model. If you try all other x pairs and sets of three, you will find x_1 , x_2 , and x_3 as the best predictors, although the R^2 for this set is still pretty low by modeling standards.

Example 14.18

This example actually calls for statistics software that can automatically test all combinations of variables. Here, trying all combinations by hand in Excel would take weeks.

But you can test the 6-variable choice in the text. Enter the data in a y-column and four xcolumns, and then compose columns for $(x1^2)$, $(x2^2)$, $(x1^*x3)$, $(x1^*x4)$, and $(x2^*x3)$. When you call Regression and use these 5 composite variables plus x2 as the x-array for input, you get the output shown in Figure 1418. Note that the coefficients agree with the MINITAB output in the text.

\diamond	A	В	C
1	SUMMARY O	JTPUT	
2			
3	Regression	Statistics	
4	Multiple R	0.94071862	
5	R Square	0.88495152	
6	Adjusted R S	0.85493887	
7	Standard Err	0.53336655	
8	Observations	30	
9			
10	ANOVA		
11		df	SS
12	Regression	6	50.3289627
13	Residual	23	6.54303728
14	Total	29	56.872
15			
16		Coefficients	tandard Errol
17	Intercept	-1.2183495	0.45403276
18	x2	0.95991429	0.14416067
19	(x1^2)	-0.037276	0.00975326
20	(x2^2)	-0.0389469	0.00669591
21	(x1*x3)	0.00368402	0.00075011
22	(x1*x4)	0.01925046	0.00595758
23	(x2*x3)	-0.0012827	0.00067215

Figure 14.18

Chapter Fifteen

The Data Analysis tools have three different ANOVA routines.

Example 15.4

This example is a perfect match to the Excel routine Single-Factor ANOVA in the Data Analysis tools. With four columns of data plus headings, the input array is A1:D21.

The results are shown in Figure 15.4, and agree with Table 15.3 of the text.

\diamond	A	В	C	D	E	F	
1	Anova: Single Factor						
2							
3	SUMMARY						
4	Groups	Count	Sum	Average	Variance		
5	рор	20	50	2.5	0.68421053		
6	rock	20	53	2.65	0.97631579		
7	hard	20	56	2.8	0.90526316		
8	metal	20	71	3.55	0.78684211		
9							
10							
11	ANOVA						
12	Source of Variation	SS	df	MS	F	P-value	
13	Between Groups	13.05	3	4.35	5.1899529	0.00257751	
14	Within Groups	63.7	76	0.83815789			
15							
16	Total	76.75	79				

Figure 15.4

Example 15.6

Just enter the data as shown in Figure 15.6a, and the Single-Factor ANOVA tool returns the results in Figure 15.6b.

\$	A	В	C	D
1	control	1 g/kg	2 g/kg	4 g/kg
2	88.6	63	44.9	31
3	73.2	53.9	59.5	39.6
4	91.4	69.2	40.2	45.3
5	68	50.1	56.3	25.2
6	75.2	71.5	38.7	22.7
7				

Figure 15.6

0	A	В	С	D	E	F	
1	Anova: Single Fac	tor					
2							
3	SUMMARY						
4	Groups	Count	Sum	Average	Variance		
5	control	5	396.4	79.28	103.652		
6	1 g/kg	5	307.7	61.54	87.313		
7	2 g/kg	5	239.6	47.92	89.512		
8	4 g/kg	5	163.8	32.76	91.373		
9							
10							
11	ANOVA						
12	ource of Variatior	SS	df	MS	F	P-value	
13	Between Groups	5882.3575	3	1960.78583	21.0922236	8.3249E-06	
14	Within Groups	1487.4	16	92.9625			
15							
16	Total	7369.7575	19				
17							

Figure 15.6b

Example 15.10

This is another case where Excel does the work. Enter data as in Figure 15.10a. Now choose Anova: Two-Factor Without Replication from Data Analysis tools and enter the array as input. The ANOVA output is shown in Figure 15.10b.

\diamond	A	В	С	D	E	F	G	Н		J
1		1	2	3	4	5	6	7	8	9
2	Α	12	10	7	7	8	9	8	7	9
3	В	15	14	14	11	11	11	12	11	13
4	С	12	13	13	10	8	11	12	8	10
5	D	10	12	9	9	7	10	11	7	8
6										

Figure 15.10a

19						
20	ANOVA					
21	Source of Variation	SS	df	MS	F	P-value
22	Rows	81.1944444	3	27.0648148	22.3556405	3.9346E-07
23	Columns	66.5	8	8.3125	6.86615679	0.00010609
24	Error	29.0555556	24	1.21064815		
25						
26	Total	176.75	35			
27						

Figure 15.10b

Example 15.14

Once again, use Anova; Two-Factor Without Replication. The output is shown in Figure 15.14.

\diamond	A	В	C	D	E	F
1	Anova: Two-Factor Without Replication					
2						
3	SUMMARY	Count	Sum	Average	Variance	
4		3	163	54.3333333	70.3333333	
5	factor A	3	152	50.6666667	6.33333333	
6	coating	3	142	47.3333333	6.33333333	
7		3	146	48.6666667	24.3333333	
8						
9	1	4	215	53.75	52.9166667	
10	2	4	188	47	13.3333333	
11	3	4	200	50	2.66666667	
12						
13						
14	ANOVA					
15	Source of Variation	SS	df	MS	F	P-value
16	Rows	83.5833333	3	27.8611111	1.35723951	0.34221547
17	Columns	91.5	2	45.75	2.22868742	0.18887989
18	Error	123.166667	6	20.5277778		
19						
20	Total	298.25	11			

Figure 15.14