CHAPTER 8 EXCEL EXERCISES (EXCEL 2010)

For instructions on how to access the data files used in these exercises, see Appendix C.

The Sampling Distribution of Sample Proportions

Assume you have a very large population of items and that half of the items are defective and half are not. We'll use π to represent the proportion of defectives in the population, so that

$$\pi = .5 \qquad (N \text{ is large})$$

We can generate a large number of samples of size n from this population, compute the proportion of defectives in each of the samples— call it $\bar{p}$ — and show that for increasingly larger sample sizes, the distribution of sample proportions will look more and more normal.

We'll use 0s and 1s to indicate whether or not a particular item is defective: 0 = non-defective, 1 = defective.

Step 1: Generate random samples of size n from the population.

We'll begin with samples of size 4 (n = 4).

> Select a cell B1 on the worksheet. On the Excel ribbon, click the **Formulas** tab, then the ***fx*** button. Click on the down arrow to the right of the "or select a category" box. From the list of function categories, choose **Math & Trig**, then **RANDBETWEEN**. Click OK. In the box labeled **Bottom**, enter 0. In the box labeled **Top**, enter 1. Click **OK**. This will randomly place either a 0 or a 1 in cell B1. Now click on the marker in the lower right corner of the cell showing this first number (you should see a solid "cross" when you are over the marker) and drag the cell contents across the row until you've covered 4 cells (B1 to E1). Release the mouse button. You should see the four cells filled with 0s and 1s.

This represents our first sample of size 4.

Step 2: Compute $\bar{x}$, the mean of the sample.

> Select cell G1. On the Excel ribbon, click the **Formulas** tab, then the ***fx*** button. Click on the down arrow to the right of the "or select a category" box. From the list of function categories, choose **Statistical**, then **AVERAGE**. Click OK. In the box labeled **Number 1**, enter the range of your sample data: B1:E1. (You can just use the cursor to highlight the cells containing your sample data.) Click **OK**. The proportion of 1s in cells B1 through E1 should appear in cell G1. This is the sample proportion, $\bar{p}$, of defective items.

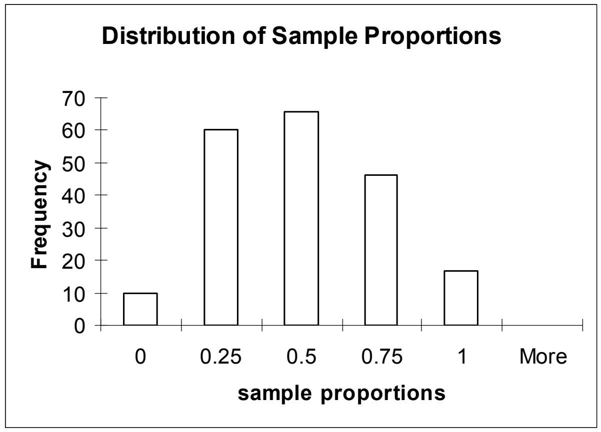
Step 3: Produce the Sampling Distribution of $\bar{x}$.

> Highlight cells B1 through G200. Click on **Home** tab on the ribbon at the top of the screen, In the **Editing** group at the far right of the expanded ribbon, choose **Fill,** then **Down**. The 200 entries in the G column that you create make up the $\bar{p}$ distribution.

Step 4: Show the Graph of the Sampling Distribution

> Select Cell J3 and type the label "Bin." In Cells J4 through J8, enter the values 0, .25, .5, .75, and 1. From the ribbon at the top of the screen, click the **Data** tab, then choose Data Analysis

from the Analysis group at the far right of the expanded ribbon. Choose **Histogram**. Click OK. In the **Input Range** box, enter the range of cells that contain your 200 sample means: G1:G200. In the **Bin Range** box, enter the range of the cells below the cell you labeled "Bin": J4:J8. Check the **Output Range** box. Click on the **Output Range** box itself and enter J18 as the cell range. Check the box marked **Chart Output**. Click **OK**. You should see the histogram for the sample proportions appear. "Clean it up" so that it looks roughly like the distribution shown below:



Step 5: Compute the Mean and Standard Deviation of the $\overline{p}$ distribution.

Select cell P5. On the ribbon at the top of the screen, click the **Formulas** tab, then the ***fx*** button. Click on the down arrow to the right of the "or select a category" box. From the list of function categories, choose **Statistical**, then **AVERAGE**. Click OK. In the box labeled **Number 1**, enter the range of cells containing the list of sample proportions: G1:G199. (You can just use the cursor to highlight the cells containing these values.) Click OK. The average of all the $\overline{p}$ values should appear in cell P5.

Now select cell P6. On the ribbon at the top of the screen, click the **Formulas** tab, then the ***fx*** button. Click on the down arrow to the right of the "or select a category" box. From the list of function categories, choose **Statistical**, then **STDEV.P**. Click OK. In the box labeled **Number 1**, enter the range of cells containing the list of sample proportions: G1:G199. (You can just use the cursor to highlight the cells containing these values.) Click **OK**. The standard deviation of the sample proportions distribution should appear in cell P6.

1. Follow the steps above for samples of size 4.

2. Adapt the steps above for samples of size 10. Use as your "Bin":

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Using EXCEL to Compute a Sample Proportion

3. A survey of 25 randomly selected American consumers was conducted. The table below shows responses to the question: "Do you have a positive image of American made cars?"

yes	no	no	yes	no
no	not sure	yes	no	yes
yes	yes	not sure	no	no
no	no	yes	no	yes
yes	no	not sure	yes	not sure

Determine the proportion of *yes* responses.

> Enter the data in cells A1 to E5 on a new worksheet. In cell G2, type the label "Count." In cell H2, enter **=Countif(A1:E5, "yes").** This will count the number of *yes* responses in the data set. In cell G3, type the label "Proportion." In cell H3, enter **=H2/25**. This will produce the sample proportion.

4. Adapt the approach in Exercise 3 to calculate the proportion of *no* answers in the survey.

CHAPTER 9 EXCEL EXERCISES (EXCEL 2010)

For instructions on how to access the data files used in these exercises, see Appendix C.

1. The U.S. Department of Energy is sponsoring a competition to find alternative energy sources to power vehicles of the future. The prizes are multi-million grants for continued research. One prize will be awarded to the competitor who is first to show that his/her electric battery technology can power an automobile for an average of more than 500 miles before recharging is necessary.

Astro Technologies has entered a vehicle. For a sample of 50 trials, Astro shows the following results:

Miles Before Recharging for a Sample of 50 Trials

530	487	562	476	532
548	543	589	507	511
500	558	508	578	505
587	479	482	590	523
422	430	541	485	437
556	507	587	463	568
540	512	568	488	563
436	420	557	447	590
478	478	565	601	524
405	533	550	496	428

Use the sample results to test the following hypotheses:

H_0: $\mu \leq 500$ (The population average is no more than 500 miles.)
H_a: $\mu > 500$ (The population average is more than 500 miles.)

As you proceed,
a) Calculate the sample mean and the sample standard deviation.
b) Calculate the appropriate t_{stat} for the sample mean.
c) Determine the critical *t* score, t_c, if the significance level for the test is .05.
d) Calculate the *p-value* for the sample mean.

Based on the sample result, can you reject the $\mu \leq 500$ null hypothesis? Explain.

Enter the data in cells A1 through E10 of a new worksheet.

a) In cell A13, type "X-bar =". Select cell B13. On the Excel ribbon at the top of the screen, click the **Formulas** tab, then the ***fx*** button. Click on the down arrow to the right of the "or select a category" box. From the list of function categories, choose **Statistical**, then **AVERAGE**. Click OK. In the *Number 1* row of the box that appears, enter A1:E10. Click OK. In cell A14 type "s =". Select cell B14. At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **STDEV.S**. In the **Number 1** row of the box that appears, enter A1:E10 (or highlight the range of your data). Click OK.

b) In cell A16, type "Std Err =". Select cell B16, then enter =B14/SQRT(50). In cell A18, type "*tstat* =". Select cell B18 and enter =(B13-500)/B16.

c) In cell A19, type "*tc* =". Select cell B19. At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **T.INV.** Click OK. In the **Probability** row of the box that appears, enter .05. (This is the significance level of .05.) In the **Deg_freedom** row, enter 49. Click OK. (The *t* score produced here is "-" the critical *t* score for your test.)

d) In cell A21, type "*p-value* =". Select cell B21. At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **T.DIST.RT.** Click OK. In the **X** row of the box that appears, enter B18; in the **Deg_freedom** row enter 49. Click OK.

Your worksheet should look like the one below. (Notice we've left out the actual values for x-bar, s, etc.)

	A	B	C	D	E	F
1	530	487	562	476	532	
2	548	543	589	507	511	
3	500	558	508	578	505	
4	587	479	482	590	523	
5	422	430	541	485	437	
6	556	507	587	463	568	
7	540	512	568	488	563	
8	436	420	557	447	590	
9	478	478	565	601	524	
10	405	533	550	496	428	

11						
12						
13	x-bar =					
14	s =					
15						
16	std err =					
17						
18	*tstat =*					
19	*tc =*					
20						
21	*p-value =*					
22						

2. Repeat your work in Exercise 1 for the sample data set below:

Miles Before Recharging for a Sample of 50 Trials

525	547	482	481	527
495	538	584	502	506
543	553	503	573	500
477	474	582	585	518
417	425	536	480	432
502	551	582	458	563
535	507	563	473	558
431	415	552	442	585
473	473	560	596	519
400	538	545	491	423

3. DFC, a package delivery service, promises its customers that average delivery time for West Coast deliveries, using its standard rate, is no more than 18 hours. A random sample of DFC's West Coast deliveries is selected. Delivery times in the sample are shown below:

15.3	19.3	13.6	20.6	14.6
18.5	21.8	16.7	23.1	25.7
21.6	18.3	24.2	12.8	18.0
12.8	15.0	19.7	27.1	17.1
25.4	27.1	13.8	19.3	23.1
20.6	23.8	22.6	20.6	15.2
22.4	20.4	23.2	15.6	16.7
25.6	10.7	16.3	23.9	25.2
10.5	16.4	19.7	18.3	28.3
15.8	17.9	15.9	23.0	20.1

Following the steps in Exercise 1, use the sample results to test the following hypotheses:

H_0: $\mu \leq 18$ (The population average is no more than 18 hours.)
H_a: $\mu > 18$ (The population average is more than 18 hours.)

As you proceed, a) Calculate the sample mean and the sample standard deviation.

b) Calculate the appropriate t_{stat} for the sample mean.
c) Determine the critical t score, t_c, if the significance level for the test is .05.
d) Calculate the *p-value* for the sample mean.

Based on the sample result, can you reject the $\mu \leq 18$ null hypothesis? Explain.

CHAPTER 10 EXCEL EXERCISES (EXCEL 2010)

For instructions on how to access the data files used in these exercises, see Appendix C.

Hypothesis Tests for the Difference between Two Population Means (Independent Samples)

1. Two large shipments of components have been received by your department—one from Supplier 1, one from supplier 2. You take a simple random sample of 10 of the components from each shipment and measure the breaking strength of each of the components in the two samples. Sample results are shown below:

Breaking Strength in Pounds

Sample from Supplier 1	Sample from Supplier 2
260	250
270	230
285	270
280	230
310	250
270	260
265	270
260	240
270	250
280	280

Use EXCEL to conduct a two-tailed t test testing the proposition that there is no difference in the average breaking strengths for the two shipments. Use a significance level of 5%. The hypotheses are

H_0: $\mu_1 = \mu_2$ or, equivalently, $\mu_1 - \mu_2 = 0$
H_a: $\mu_1 \neq \mu_2$ or, equivalently, $\mu_1 - \mu_2 \neq 0$

Enter the two columns of data, including labels, onto your worksheet. On the Excel ribbon at the top of the screen, click the **Data** tab, then choose **Data Analysis** (at the far right). Select **t-test: Two Sample assuming Equal Variances**. Click OK. In the box that appears, enter the cell range of the Shipment A data (including the label) in the space labeled **Variable 1 Range**, then enter the cell range of the Shipment B data (including the label) in the space labeled **Variable 2 Range**. In the **Hypothesized Mean Difference** space, enter "0". Check the **Labels** box. Make sure the **Alpha** space shows .05. Check the circle next to the **output range** label, then enter the cell location where you want to show the upper left-hand corner of the output table that EXCEL will produce. Click **OK**.

The output you produce should look similar to that shown below:

	A	B	C	D	E	F
1	Ship 1	Ship 2				
2	260	250		t-Test: Two-Sample Assuming Equal Variances		
3	270	230				
4	285	270			*Ship 1*	*Ship 2*
5	280	230		Mean	275	253
6	310	250		Variance	222.22	290
7	270	260		Observations	10	10
8	265	270		Pooled Variance	256.111	
9	260	240		Hypothesized Mean Differ	0	
10	270	250		df	18	
11	280	280		t Stat	3.073	
12				P(T<=t) one-tail	0.0032	
13				t Critical one-tail	1.734	
14				P(T<=t) two-tail	0.0064	
15				t Critical two-tail	2.101	
16						
17						

one-tailed *p-value* (→ P(T<=t) one-tail)

two-tailed *p-value** (→ P(T<=t) two-tail)

** Notice that the two-tailed p-value is twice the size of the one-tail p-value.*

Report your conclusion and explain. Don't just say "reject the null hypothesis" or "don't reject the null hypothesis." Express your decision and explain your reasoning in language that a non-statistician would understand.

2. Your engineers have developed two alternative production methods for manufacturing a new unit to be introduced onto the market. Method 1 is the less expensive of the two, but you're not sure if the quality of product is as good as the quality of product produced by the more expensive Method 2. You take a sample of 21 units produced by each of the two processes and test each to measure useful life (in hours). Results of the testing are shown below. Set up a one-tail hypothesis test to test the proposition that there is no difference in the average useful life for units produced by the two processes. Let $\alpha = 5\%$. Use the hypotheses

H_0: $\mu_1 \geq \mu_2 \rightarrow \mu_1 - \mu_2 \geq 0$

H_a: $\mu_1 < \mu_2 \rightarrow \mu_1 - \mu_2 < 0$

Method 1	Method 2
1540	1476
1528	1488
1464	1539
1522	1554
1433	1497
1513	1512
1537	1563
1478	1535
1560	1573
1472	1498
1436	1486

1404	1533
1529	1561
1537	1490
1422	1472
1457	1556
1510	1540
1426	1461
1463	1483
1528	1564
1507	1582

Report your conclusion and explain. Don't just say "reject the null hypothesis" or "don't reject the null hypothesis." Express your decision and explain your reasoning in language that a non-statistician would understand.

3. A shorter version of the *t* test for testing the difference between means involves the use of the TTEST function from the INSERT/ FUNCTION/ STATISTICAL menu. It produces a *p-value* for the one- or two-tail case. Try it on the data in Exercises 1 and 2.

> Select a cell near the data. At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **Ttest**. Click OK. In the **array 1** space on the box that appears, enter the range of cells containing the first sample data. In the **array 2** space, enter the cell range for the second sample data. In the **tails** space, enter 1 or 2, depending on whether you are conducting a one-tail or a two-tail test. In the **type** space, enter 2, which indicates you are assuming that the population standard deviations are the same (*i.e.*, you want to use the "pooled" sample standard deviations to estimate the population standard deviation). Click **OK**.

The number that appears in the cell you selected on the worksheet will be the *p*-value for the sample mean difference in your data. If it's smaller than the α you've chosen for the test, you can reject the "no difference" null hypothesis.

Hypothesis Tests for the Difference between Two Population Means (Matched Samples)

4. A major ready-to-assemble furniture company is testing two alternative designs for its new home media cabinet. The company is especially concerned about ease of assembly for the purchaser. Twenty-five consumers were randomly selected for the test. Each consumer was asked to assemble both new cabinet models. The order in which each consumer assembled the cabinets was randomly determined. The table below shows how long, in minutes, each assembly took for each consumer.

Conduct a two-tailed *t* test for *matched* samples, testing the null hypothesis that there is no difference in the average assembly times for the two models. Use a significance level of 5%.

Consumer	Model 1	Model 2
1	81.6	82.3
2	79.8	77.6
3	67.9	68.3
4	82.2	79.8
5	76.8	80.1

6	68.8	67.2
7	74.6	70.5
8	63.2	66.8
9	78.9	77.2
10	80.2	84.3
11	76.8	70.1
12	69.9	64.7
13	76.5	79.5
14	80.7	83.2
15	84.2	80.1
16	83.5	76.4
17	78.0	77.1
18	81.3	75.4
19	67.8	66.1
20	76.2	79.4
21	70.5	66.6
22	78.1	75.7
23	81.2	78.0
24	79.9	83.4
25	75.4	76.2

After entering the data on a new worksheet, click the Data Tab at the top of the screen, then select **Data Analysis** (at the far right). Select **t-test: Paired Two Sample for Means**. Proceed as in Exercise 1.

CHAPTER 11 EXCEL EXERCISES (EXCEL 2010)

For instructions on how to access the data files used in these exercises, see Appendix C.

1. Use EXCEL to conduct a simple regression analysis for the training/productivity example that was the focal point for our chapter discussion:

(x) Training (hours)	(y) Productivity (units/hour)
2	6
3	8
4	10
5	9

Enter the training/productivity data (including the "training" and "productivity" column labels) in two adjacent columns on your worksheet. On the Excel ribbon at the top of the screen, click the **Data** tab, then choose **Data Analysis** (at the far right). **Se**lect **REGRESSION**. Click OK. In the "**Input Y Range**" box, enter the cell range for the *y* (productivity) data on your worksheet (including the cell in which you entered the "productivity" label). In the "**Input X Range**" box, enter the cell range for the *x* (training) data. Check "**labels**" (since you have included labels in your data ranges). Check the circle for "**output range**", click on the box, and enter the first cell in which you want the output to appear. To produce a table of $(y-\hat{y})$ error terms, check "**residuals**". When you click **OK**, you should see output like that shown in the chapter.

2. The following survey data has been obtained from interviewing a sample of 20 students on campus:

Student	weekly study time	weekly work hrs	SAT score	HS GPA	Coll GPA
1	14.5	10	920	3.2	3.22
2	6	14	1080	2.9	2.81
3	17.5	20	830	3.1	2.65
4	20	5	1340	3.0	3.2
5	22.5	6	1220	3.6	3.77
6	20	12	1130	2.5	1.92
7	5	18	1010	2.3	2.13
8	16.5	8	1170	3.6	3.1
9	32	0	930	4	3.66
10	12	10	1030	2.5	2.87
11	23	5	860	2.9	3.25
12	22	35	760	2.4	1.76
13	16	30	830	3	2.45
14	10	18	1070	2.7	2.68
15	8.5	25	1120	2.5	2.41
16	2.5	20	960	2.4	2.18
17	15	5	1190	3.7	3.56
18	17.5	20	1230	3.1	3.62
19	12	14	940	2.8	2.44
20	30	0	1080	2.1	2.95

Repeat the EXCEL steps in Exercise 1, but use College GPA as your dependent variable (*y*) and hours of work as your independent variable (*x*). From your table of output, report:

a) The intercept and slope of the estimated regression line.
b) The coefficient of determination, r^2.
c) The correlation coefficient, r.
d) The explained variation (SSR), the unexplained variation (SSE), and the total variation (SST) for the dependent variable, College GPA.
e) The t_{stat} (*t* statistic) value to test a $\beta = 0$ null hypothesis.
f) The probability that a sample slope (*b*) as far (or farther) from 0 as the one you've generated would occur randomly if the population slope, β, is equal to 0. (This is the *p-value* for the sample slope.)
g) If a significance level of 5% is used, what can you conclude about the relationship between college GPA and work time? Is it statistically significant? Explain.

3. Repeat the EXCEL steps in Exercise 2, but this time use College GPA as your dependent variable (y) and High School GPA as your independent variable(x). From your table of output, report:

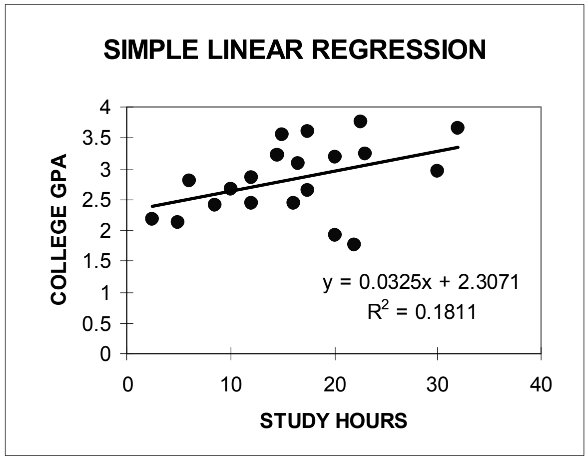
a) The intercept and slope of the estimated regression line.
b) The coefficient of determination, r^2.
c) The correlation coefficient, r.
d) The explained variation (SSR), the unexplained variation (SSE), and the total variation (SST) for the dependent variable, College GPA.
e) The t_{stat} (t statistic) value to test a β=0 null hypothesis.
f) The probability that a sample slope (b) as far (or farther) from 0 as the one you've generated would occur randomly if the population slope, β, is equal to 0. (This is the *p-value* for the sample slope.)
g) If a significance level of 5% is used, what can we conclude about the relationship between college GPA and High School GPA? Is it statistically significant? Explain.

It's possible to produce a smaller but more visual version of simple regression analysis using the Chart Wizard to draw a "scatter diagram" and then inserting a least squares line into the scatter diagram. The exercises below illustrate the approach.

4. Conduct a simple linear regression analysis using College GPA from the data set in Exercise 2 as the dependent variable (y) and Study Time (from that same data set) as the independent variable (x).

> Copy the Study Time column and the College GPA column onto another section of the worksheet. Be sure to place the data in adjacent columns, with the Study Time column to the left of the College GPA column. Use your mouse to select the range for these columns (including the labels). The range of cells containing the data should now be black. Click on the **Insert** tab on the Excel ribbon at the top of the screen. From the **Charts** group in the expanded ribbon, choose **Scatter**, then click the graph in the upper left corner of the choices presented (scatter with only markers).
>
> Be sure the chart is highlighted, then choose the **Layout** tab at the top of the screen. Use options from the **Labels, Axes and Background** groups to make your chart look like the one shown below. To insert a least squares line (TREND LINE) into the chart, click on the **Layout** tab, then on the expanded ribbon, choose **Analysis**, **Trendline** and **Linear**. Again click on the Layout tab, then choose Analysis, Trendline, and **More Trendline Options**. Check **DISPLAY EQUATION ON CHART** and **DISPLAY R2-VALUE ON CHART**. Click CLOSE.

Do the cleanup necessary to make your chart look like the example below. (Right click on various elements in the chart and select appropriate options. As an alternative, be sure the chart is highlighted, then choose the **Layout** tab at the top of the screen. Use options from the **Labels, Axes and Background** groups to make changes in the appearance of the chart.)



5. Repeat the steps in Exercise 4, but this time use College GPA as your dependent variable (y) and SAT score as your independent variable (x).

CHAPTER 12 EXCEL EXERCISES (EXCEL 2010)

For instructions on how to access the data files used in these exercises, see Appendix C.

F Distribution

1. In an F distribution with numerator degrees of freedom = 4 and denominator degrees of freedom = 18, what percentage of the values are greater than

a) 1.0? b) 3.6? c) 7.9?

At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **F.DIST.RT**. Click OK. In the table that appears, enter the given value of F (e.g., 1.0) in the top box, labeled "x". In the second box, enter the numerator degrees of freedom; in the bottom box, enter the denominator degrees of freedom. Click **OK**.

2. In an F distribution with numerator degrees of freedom = 5 and denominator degrees of freedom = 29,

a) .40 of the values are greater than ____?
b) .10 of the values are greater than ____?
c) .025 of the values are greater than ____?
d) .05 of the values are greater than ____?

At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **F.INV.RT**. In the table that appears, enter the given probability (e.g., .40) in the top box. In the second box, enter the numerator degrees of freedom; in the bottom box, enter the denominator degrees of freedom. Click **OK**.

Multiple Regression

A survey of 20 of XYZ Inc.'s online customers was conducted to try to establish a connection between various customer characteristics—age, annual income, years of education, credit card debt, and average hours spent online per week—and the average amount that the customer spends each time he/she makes an online purchase on the company's website. Results are shown in the table below:

AGE	INCOME ($000)	EDUCATION	CREDIT CARD DEBT	HRS. ONLINE	AVERAGE PURCHASE
26	31	16	1300	25	120
43	46	12	200	32	58
64	56.5	14	450	10	45
34	52	18	1800	5	230
25	36	16	2300	28	110
28	25	12	4500	40	85
41	46.8	8	120	22	42
32	35.4	16	540	16	36
35	46	12	350	42	75
26	23.7	10	900	35	80
57	72	14	1280	20	25
48	51	16	1800	10	40
36	24.5	20	100	0.1	140
18	20	14	4600	36	30
31	27.7	12	2500	18	85
36	42.3	10	3000	25	60
61	89	16	1200	28	32
44	52.3	16	5500	14	56
38	39.8	12	1600	18	78
27	22.5	14	450	22	100

3. With **Average Purchase** amount as the dependent variable (y) and **Age** as the only independent variable (x), use EXCEL's linear regression option (DATA/DATA ANALYSIS/REGRESSION) to determine whether the sample data indicates a useful linear relationship between the two variables at the "population" level. Specifically, determine whether can reject a "$\beta = 0$" null hypothesis? Use a 5% significance level.

Enter the **Average Purchase** column of values and the **Age** column of values on your worksheet, being sure to place the columns next to one another. At the top of the screen, click the **Data** tab, then click **Data Analysis** (at the far right of the expanded ribbon). From the list of tools, choose **REGRESSION**. Click OK. On the wizard that appears, enter the location of the y (e.g., average purchase) values, including the column label, in the top box. Enter the location of the x values (e.g., Age) in the second box. Check the **labels** box. Click on **output range** then click on the **output range** box. In the **output range** box, enter the cell location that will mark the upper lefthand corner of the output display. Click OK. You should see the full regression output, including the ANOVA table.

a) Use the appropriate *t test* to support your conclusion, comparing the sample *t value* (*t stat*) to the critical *t value* (t_c) for a two-tailed test at the 5% significance level (FORMULAS/ INSERT FUNCTION/ STATISTICAL/ T.INV.2T).

b) Use the appropriate *F*-test to support your conclusion, comparing the sample *F* value to the critical *F* value (FORMULAS/ INSERT FUNCTION/ STATISTICAL/ F.INV.RT).
c) Use the appropriate *p-value* to support your conclusion.

4. With **Average Purchase** as the dependent variable (y), and **Age** (x_1) and **Income** (x_2) as the independent variables, use EXCEL's linear regression option to determine whether the sample data shows a statistically significant relationship between the two independent variables and the dependent variable, at the 5% significance level.
a) Use the appropriate *F* test to support your conclusion, comparing the sample *F* value to the critical *F* value.
b) Report the appropriate *p-value* and compare it to the significance level of 5% to support your conclusion.

5. Determine the correlation between **Age** and **Income**. What does this tell you about the appropriateness of using both variables together in a multiple regression model for this situation?

> At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **CORREL**. Enter the cell locations for the first variable in the top box and the cell locations for the second variable in the second box. Click **OK**.

6. With **Average Purchase** as the dependent variable (y), and **Age** (x_1) and **Credit Card Debt** (x_2) and **Online Hours** (x_3) as the independent variables, use EXCEL's linear regression option to determine whether the sample data shows a statistically significant relationship between the set of three independent variables and the dependent variable at the 5% significance level.
a) Use the appropriate *F*-test to support your conclusion, comparing the sample *F* value to the critical *F* value (from the *F* table).
b) Report the appropriate *p-value* to support your conclusion.

7. Given your results in Exercise 6, conduct the appropriate *t test*s to determine which of the three independent variable coefficients are significant at the 5% significant level.
a) In each case, compare the sample *t*-score from the EXCEL printout to the critical *t*-score from the t-table and report the results.
b) In each case, compare the appropriate *p-value* to the significance level (5%) and report your conclusion.

8. Given your results in Exercise 7, re-run the model, eliminating as one of the independent variables **Credit Card Debt,** the variable whose coefficient proved not to be statistically significant at the 5% significance level.
a) Use the appropriate *F*-test to determine if there is evidence of a useful linear relationship between the two remaining independent variables **Age** and **Hours Online**, and the dependent variable **Average Purchase Amount**.
b) Report the appropriate *p-value* to support your conclusion.
c) Are both the independent variable coefficients significant at the 5% significance level? Explain.

9. Use the EXCEL table results in Exercise 6 (where the model involved 3 independent variables, AGE, CREDIT CARD DEBT and HOURS ONLINE) and Part 8 (where the model used 2 independent variables, AGE and HOURS ONLINE) to compare the two models with respect to the
a) r^2 values in each case.
b) adjusted r^2 values in each case.
c) Comment on each of the comparisons.

10. Now include **all five** independent variables in your model, with **Average Purchase** as the dependent variable (y), and use EXCEL's linear regression option to determine whether the sample data shows a statistically significant relationship between the set of 5 independent variables and the dependent variable at the 5% significance level.

a) Use the appropriate F-test to support your conclusion, comparing the sample F value to the critical F value (from the F table).

b) Report the appropriate *p-value* to support your conclusion.

c) Are any of the independent variable coefficients significant at the 5% significant level? Explain.

Qualitative Variables

Below are results of a study of 25 similar sized retail outlets that feature your company's main product. The study is intended to establish a connection between average monthly product sales and a set of independent variables that include the store price of the major competing product, product positioning within the store, and local area promotional expenditures.

Sales (units)	Competing Price	Promo Exp ($00s)	In-Store Location	d_1	d_2
1340	5.78	36	R		
2350	6.55	51	*F*		
1720	5.55	38	C		
1200	4.89	32	R		
1560	5.3	36	C		
2300	6.29	44	*F*		
890	4.45	32	R		
1360	6.35	44	C		
2370	5.75	52	*F*		
1430	5.45	37	C		
1280	4.97	43	R		
940	4.58	25	R		
1690	5.99	30	C		
1560	5.58	34	C		
1600	5.39	33	*F*		
2100	6.66	56	*F*		
2650	5.72	38	C		
480	4.99	29	R		
1380	5.69	31	C		
1430	6.55	41	C		
2130	6.89	37	*F*		
2720	6.29	50	C		
1490	5.42	36	C		
2100	6.38	51	*F*		
1500	5.18	35	C		

11. In-store product placement locations were classified as Front (*F*), Center (C), and Rear (R). Use dummy variables d_1 and d_2 to represent the in-store location variable, with the following code:

	d_1	d_2
F	0	0
C	1	0
R	0	1

Enter the appropriate 0s and 1s in the d_1 and d_2 columns of the table.

12. Use EXCEL's linear regression option to determine which, if any, of the variables appears to be a significant factor influencing sales (at the 5% significance level).

13. Discuss the results of your analysis in Exercise 12, especially with regard to the categorical variable in-store location.

14. Re-run your model, this time using only Advertising and In-store Location as the independent variables. Re-code the location variable using just one dummy (d) and the following scheme:

Front or Center	$d = 0$
Rear	$d = 1$

CHAPTER 13 EXCEL EXERCISES (EXCEL 2010)

For instructions on how to access the data used in these exercises, see Appendix C.

The *F* Distribution

1. Use the proper EXCEL function to produce the following chi-square probabilities:

a) $P(F \geq 2.85)$ $df_1 = 5$, $df_2 = 15$ b) $P(F \geq 7.38)$ $df_1 = 4$, $df_2 = 25$

c) $P(F \geq 3.68)$ $df_1 = 5$, $df_2 = 150$

> On the Excel ribbon, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then choose **F.DIST.RT**. In the screen that appears, insert the desired value for *F* (*e.g.*, 2.85) or its cell location on your worksheet (*e.g.*, B4); in the second box enter the degrees of freedom in the numerator; in the third box enter the degrees of freedom in the denominator. This should produce the proper "greater than or equal to" probability.

2. Use the F.INV.RT function to fill in the following blanks:

For an *F* distribution with $df_1 = 4$, $df_2 = 27$,

a) 5% of the values will be greater than or equal to _________.
b) 1% of the values will be greater than or equal to _________.
c) 7% of the values will be greater than or equal to _________.

> At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **F.INV.RT**. Enter the desired probability (*i.e.,* percentage) in the first box, then the numerator degrees of freedom in the second box; in the third box, enter the denominator degrees of freedom. The result shown will be the "≥" *F* value you're looking for.

One-Way ANOVA

3. You are testing units from three different batches of product produced in your manufacturing operation at different times during the day to determine if the average diameter of the units is the same for all three batches. Results from a test of 15 randomly selected units from each batch are provided below:

Batch A Sample Diameters (mm)	Batch B Sample Diameters (mm)	Batch C Sample Diameters (mm)
14.33	14.73	14.26
15.10	14.88	14.30
14.68	15.28	14.78
14.52	14.92	14.41
14.97	14.89	14.56
15.03	14.93	14.03
14.54	15.27	14.14
14.65	14.15	14.75
15.17	15.38	14.47
14.29	14.99	14.89
14.56	15.26	14.52
14.33	14.83	14.13
14.93	14.72	14.83
15.08	15.19	14.11
14.73	14.98	14.43

Use one-way analysis of variance to determine whether you can conclude that the average diameter in at least one of the three batches is different from the others. Use a significance level of 5%.

> Enter the three columns of data on a worksheet. On the Excel ribbon, click the **Data** tab, then click **Data Analysis** (at the far right of the expanded ribbon). From the list of tools, choose **ANOV**A: **SINGLE FACTOR**. Click OK. Enter the range for your data in the **Input Range** box. If you've included the column labels in the range, check the **labels in first row** box. Insert the desired significance (alpha) level. Click **output range** and enter the cell location on your worksheet where you want to show the output table. Click **OK**. You should see the ANOVA table, together with a table of summary measures for each of the three data columns, in the location you chose on the worksheet.

4. A random sample of 50 adult consumers from each of three geographic regions of the country—East Coast, Central States, and West Coast—was selected and asked to fill out a questionnaire intended to test awareness of consumer protection laws currently in effect. Sample test scores based on questionnaire responses are shown below. Follow the procedure outlined in Exercise 3 to produce an ANOVA table from which you can determine whether there would be a difference in average test scores for the populations of consumers represented. Use a significance level of 5%.

EAST COAST	CENTRAL	WEST COAST
65	73	87
52	56	65
82	43	87
56	78	46
76	90	51
54	72	49
87	46	78
53	87	76
47	52	57
46	34	82
83	87	44
48	85	56
56	56	87
68	68	76
79	79	65
42	42	62
66	65	78
78	57	59
62	68	46
50	88	67
87	68	87
73	76	65
59	70	76
44	65	78
57	43	54
91	44	65
75	56	78
33	52	65
62	43	48
68	46	72
44	78	43
63	72	56
57	38	87
68	54	56
72	67	58
85	89	68
34	78	55
60	84	81
78	65	57
62	66	88
56	49	54
47	51	65
89	76	62
51	38	41
34	58	52
58	75	58
73	61	42
67	83	68
48	47	91
79	64	48

The Chi-Square Distribution

5. Use the proper EXCEL function to produce the following chi-square probabilities:

a) $P(\chi^2 \geq 5.331)$, $df = 5$ b) $P(\chi^2 \geq 14.678)$, $df = 10$ c) $P(\chi^2 \geq 22.653)$, $df = 14$

> At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **CHISQ.DIST.RT**. In the screen that appears, insert the desired value for χ^2 (for example 5.331) or its cell location on your worksheet (for example, B4); in the second box enter the degrees of freedom for the distribution. Click OK. This should produce the proper "greater than or equal to" probability.

6. Use the CHIINV function from the INSERT menu to fill in the following blanks:

For a chi-square distribution with $df = 9$,

a) 5% of the values will be greater than or equal to ________.
b) 1% of the values will be greater than or equal to _________.
c) 12% of the values will be greater than or equal to ________.

> At the top of the screen, click the **Formulas** tab, then the ***fx*** button. From the list of function categories, choose **Statistical**, then **CHISQ.INV.RT**. Enter the probability (that is, the percentage) in the first box, then the degrees of freedom in the second box. Click OK. The result shown will be the "≥" χ^2 value you're looking for.

Tests of Independence

7. Below is an extension of Exercise 44 at the end of the Chapter. It shows observed frequencies for airline customers who were involved in a year-long trial of four rewards programs. Passengers were asked to give their impression of the program in which they participated. Possible passenger responses were 'highly favorable', 'moderately favorable', 'moderately unfavorable', and 'highly unfavorable.'

OBSERVED FREQUENCIES
Response

Program	Highly Favorable	Moderately Favorable	Moderately Unfavorable	Highly Unfavorable	TOTALS
A	90	30	20	10	150
B	85	35	25	5	150
C	92	38	10	10	150
D	88	42	5	15	150
TOTALS	355	145	60	40	600

Use a chi-square test of independence to test the proposition that response is independent of program (that is, customer response would be the same for all four programs), using a significance level of 5%.

Open the **Contingency Table** file (See Appendix C). If you are asked whether you want to enable macros, click **enable macros**. In the box next to **number of rows** in the table, enter 4. In the box next to **number of columns** in the table, enter 4. Click the **clear table** button. Click the **table setup** button.

Enter the 16 cells of the data table above. (Don't include the labels or the totals.) Return to the **Contingency Table** worksheet. Paste the 16 cells that you copied from the Exercise 7 worksheet into the cells of the blank table. Click the **Go** button. You should see three tables appear on the Contingency Table work sheet: observed frequencies, expected frequencies, and a table of the 16 values that make up χ^2_{stat}.

Below the tables is χ^2_{stat}, plus the appropriate degrees of freedom for the analysis, two critical chi-square (χ_c^2) values—one for a significance level of .05 and one for a significance level of .01—and the *p*-value for χ^2_{stat}.

*Note: For a more "hands-on" approach using the CHISQ.TEST function from the **formulas/fx/statistical** menu, you might try adapting the "Chi-square Template" that's been included with the EXCEL Data (See Appendix C).*

8. A survey involving a random sample of 200 adults (18 years and older) in your local area was recently conducted. Partial results from the survey are given below:

Sex	Income Group	Highest Education Level	Age Group	Do You own Stocks?	Economy?	Better Off?
M	A	COLLGRAD	18-24	Y	OPT	Y
M	A	HS	35-49	Y	OPT	Y
F	B	<HS	50+	N	PESS	N
M	C	SOMECOLL	35-49	Y	NEUT	Y
M	B	COLLGRAD	25-34	N	NEUT	Y
F	D	SOMECOLL	50+	N	OPT	N
F	B	HS	18-24	N	PESS	N

Key: M-Male, F-Female; A-under$30K, B-$30K-$60K, C-$60K-$85K, D-over $85K;
<HS-No High School Diploma, HS-High School Diploma, SOMECOLL-some college, COLLGRAD-College Degree, GRAD-Graduate Degree (Masters, PhD, etc.);

In addition to recording demographic information (age, sex, annual income, education level), the survey asked participants three primary questions:

1) Do you own any stock?
2) Are you optimistic, pessimistic, or neutral about the economy over the next five years?
3) Are you economically better off today than you were five years ago?

Using the complete table of survey results, produce a pivot (contingency) table showing age group in the rows of the table and view of the economy in the columns. Show the frequency of responses in each of the table's cells.

Enter the data on your worksheet, including the labels at the top of each column (be sure each of the labels occupies just one cell). Click the **Insert** tab at the top of the screen, then select **Pivot Table** (at the left end of the expanded ribbon). Check the circle next to "select a table or range". Enter the range of the data (including labels) that you entered on the worksheet (*e.g.,* A5:E25) This range may already be automatically entered for you. Check the box for **existing worksheet**, then click on the cell on your worksheet where you want to show your table. Click **OK**. From the "choose fields to add" on the form that appears, use the mouse to drag the **age** label to the section marked **Row Labels**. Similarly, drag the **economy** label to the **Column Labels** box. Drag the **age** label (again) **Σ Values** box. You should see the table appear. You can right click on the table and experiment with some of the options shown.

9. Using the table you produced in EXCEL Exercise 8, follow the directions in EXCEL Exercise 7 to conduct a chi-square test of independence. Use the test to determine if you can reject the null hypothesis that age and view of the economy are independent factors. Use a significance level of 1%.

10. Repeat the procedures you used in Exercise 8 and Exercise 9, but this time determine if the survey provides sufficient sample evidence to reject the null hypothesis that income and view of the economy are independent factors. Use a significance level of 5%.

11. Repeat the procedures to determine if the survey provides sufficient sample evidence to reject the null hypothesis that education and stock ownership are independent factors. Use a significance level of 5%.

CHAPTER 14 EXCEL EXERCISES (EXCEL 2010)

Download the "Quality Control.xls" file and the EXCEL data for Chapter 14 from www.lulu.com/content/320281.

1. Wilson Toys ships party balloons to wholesalers throughout the country. The company uses an acceptance sampling plan to check on the percentage of balloons in the shipment that will burst under normal pressure. Wilson selects a sample of 50 balloons from each shipment. The shipment is sent out only if there are no more than two balloons in the sample that burst when tested.

a) Show the OC curve for the plan.
b) Show the AOQ curve for the plan.
c) Report the producer's risk (α) and consumer's risk (β) for this sampling plan if Wilson sets its AQL at 1% and its LTPD at 10%.
d) According to the AOQ curve here, what is the "worst" average outgoing quality (highest % defective) possible under this sampling plan? (This is the AOQL value.)

Open the Quality Control workbook. Click on the "*OC-AOQ*" tab at the bottom of the worksheet. Enter the sample size and acceptance number in the appropriate cells on the worksheet.

*Note: If you want to change the scale on the *y*-axis of either chart, right click on the axis, then choose "format axis". Choose the "scale" tab and then enter your desired minimum and maximum values for the scale, together with the distance you want between markers along the axis. When finished, click OK.

2. You have just re-engineered your assembly line and are preparing a *P-chart* to monitor its performance. During the first few days of operation, you begin monitoring the process by taking 20 samples of size 300. The number of defective items in each sample is shown below.

Sample #	Number of Defectives	Sample #	Number of Defectives
1	4	11	9
2	8	12	7
3	6	13	4
4	7	14	5
5	11	15	6
6	9	16	7
7	6	17	8
8	8	18	9
9	12	19	11
10	3	20	12

a) Use these sample results to set up the appropriate *P-chart*. Report the LCL, UCL and center line values.
b) Identify any points which indicate that the process is out of statistical control. Assuming you can eliminate the cause behind these points, remove the violating point(s) from the list of 20 samples and then use the remaining values plus one or more of the following points (as needed to total 20 points) to construct a new chart: 6, 1 ,8 , 9, 10, 11, 6.
c) Report the LCL, UCL and center line values on the revised chart. Does your chart now show that the process is in statistical control?

Open the Quality Control workbook. Click on the "*p-Charts*" tab at the bottom of the worksheet. Enter the sample data in the blue section of the worksheet. Enter the sample size in the yellow sample size cell.

*Note: If you want to change the scale on the *y*-axis of the chart, right click on the axis and then choose "format axis". Choose the "scale" tab and then enter your desired minimum and maximum values for the scale, together with the distance you want between markers along the axis. When finished, click OK.

3. Martell Industrial Solvents produces a thermoplastic polyurethane elastomer (TPE) used for bonding water insoluble synthetic organic polymers. One of the key ingredients is methyl ethyl ketone (MEK). Martell wants to control the amount of MEK added to each 5-liter container of TPE produced by its mixing process. Below are sample results based on 20 recent samples, each consisting of five containers. Use these sample results to construct a proper $\overline{X}$ - and *R*-chart.

Sample Values
(milliliters of MEK added to each container)

	1	2	3	4	5
1	27.25	29.75	29.50	29.75	28.25
2	29.00	29.50	29.00	28.50	30.50
3	30.50	29.75	29.25	30.50	30.75
4	29.75	29.50	29.50	29.25	29.25
5	28.75	29.50	30.75	30.75	31.25
6	31.25	31.00	29.50	27.75	28.50
7	29.75	29.75	29.50	29.00	28.50
8	29.00	29.00	29.50	29.00	29.50
9	29.50	29.50	30.00	30.00	31.75
10	29.75	30.25	29.75	30.25	30.50
11	29.50	29.50	30.25	29.00	30.00
12	29.50	29.75	29.50	31.50	31.75
13	29.75	30.25	29.75	30.25	30.50

14	28.75	29.50	30.75	30.75	31.25
15	31.25	31.00	29.50	27.75	28.50
16	29.50	29.75	29.50	30.00	31.25
17	27.75	29.75	29.50	29.75	28.25
18	29.50	29.50	29.25	28.50	30.50
19	30.50	29.75	29.25	30.75	30.75
20	28.50	29.50	29.50	29.25	32.75

a) Report the LCL, UCL and center line for each of the charts.
b) Is the process in control? Explain.

Open the Quality Control workbook. Click on the "*Xbar R Charts*" tab at the bottom of the worksheet. Enter the sample data in the blue section of the worksheet. Enter the proper factor values in the yellow cells.