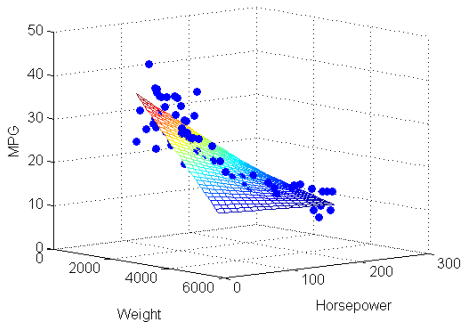

PROBLEM # 12.1 Explain each of the terms in the multiple regression model.

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p + e$$

PROBLEM # 12.2- For the multiple regression equation $\hat{y} = 100 + 20\,x_1 + 3x_2 + 120\,x_3$:

a. Identify the y-intercept and partial regression coefficients.
b. If x_1=12, x_2= 5, and x_3= 10, what is the estimated value of y?
c. If x_3 were to increase by 4, what change would be necessary in x_2 in order for the estimated value of y to remain unchanged?

PROBLEM # 12.3 During its manufacture, a product is subjected to four different tests, each giving a score of 200 to 800. An efficiency expert claims the fourth test to be unnecessary since its results can be predicted based on the first three. The following Minitab printout is an analysis of scores received by a sample of 12 units subjected to all four tests.

Regression Analysis
The regression equation is
TEST04= 12.0 + 0.274 TEST01 + 0.376 TEST02 + 0.326 TEST03

Predictor	**Coef**	**StDev**	**T**	**P**
Constant	11.98	80.50	0.15	0.885
TEST01	0.2745	0.1111	2.47	0.039
TEST02	0.37619	0.09858	3.82	0.005
TEST03	0.32648	0.08084	4.04	0.004

S=52.72 R-sq=87.2% R-Sq(adj) =82.4%

Analysis of Variance

Source	**DF**	**SS**	**MS**	**F**	**P**
Regression	3	151417	50472	18.16	0.001
Error	8	22231	2779		
Total	11	173648			

a. Interpret the multiple regression equation.
b. An individual unit from the assembly area received ratings of 350, 400, and 600 on the first three tests. Using the estimation equation in the printout, what is the unit's estimated score on test 4?

PROBLEM # 12.4- The owner of a large chain of health spas has selected eight of her smaller clubs for a test in which she varies the size of the newspaper ad and the amount of the initiation fee discount to see how this might affect the number of prospective members who visit each club during the following week. The results are shown in the table below.

Club	**New Visitors, y**	**Ad Column-Inches, x_1**	**Discount Amount, x_2**
1	**23**	**4**	**$100**
2	**30**	**7**	**20**
3	**20**	**3**	**40**

4	**26**	**6**	**25**
5	**20**	**2**	**50**
6	**18**	**5**	**30**
7	**17**	**4**	**25**
8	**31**	**8**	**80**

a. Determine the least squares multiple regression equation.
b. Interpret the y-intercept and partial regression coefficients.
c. What is the estimated number of new visitors to a club if the size of the ad is 5 column-inches and a $75 discount is offered?

	A	B	C	D	E	F	G
14	SUMMARY OUTPUT			Visitors	Col-Inches	Discount	
15	*Regression Statistics*			23	4	100	
16	Multiple R	0.8465		30	7	20	
17	R Square	0.7165		20	3	40	
18	Adjusted R Square	0.6031		26	6	25	
19	Standard Error	3.3749		20	2	50	
20	Observations	8		18	5	30	
21				17	4	25	
22	ANOVA			31	8	80	
23		*df*	*SS*	*MS*	*F*	*Significance F*	
24	Regression	2	143.924	71.962	6.318	0.043	
25	Residual	5	56.951	11.390			
26	Total	7	200.875				
27							
28		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
29	Intercept	10.687	3.875	2.758	0.040	0.726	20.648
30	Col-Inches	2.157	0.628	3.434	0.019	0.542	3.771
31	Discount	0.042	0.044	0.949	0.386	-0.071	0.154

PROBLEM # 12.5 Pat Statsdud, a student ranking near the bottom of the statistics class, decided that a certain amount of studying could actually improve final grades. However, too much studying would not be warranted, since Pat's ambition was to ultimately graduate with the absolute minimum level of work. Pat was registered in a statistics course, which had only 3 weeks to go before the final exam, and where the final grade was determined in the following way:

Total mark= 20% (Assignment) + 30% (Midterm exam) + 50% (Final exam)

To determine how much work to do in the remaining 3 weeks, Pat needed to be able to predict the final exam mark on the basis of the assignment mark (worth 20 points) and the midterm mark (worth 30 points). Pat's marks on these were 12/20 and 14/30, respectively. Accordingly, Pat undertook the following analysis. The final exam mark, assignment mark, and midterm test mark for 30 students who took the statistics course last year were collected.

a. Determine the regression equation.
b. What is the standard error of estimate? Briefly describe how you interpret this statistic.
c. What is the coefficient of determination? What does this statistic tell you?
d. Test the validity of the model.
e. Interpret each of the coefficients.
f. Can Pat infer that the assignment mark is linearly related to the final grade in this model?

g. Can Pat infer that the midterm mark is linearly related to the final grade in this model?

h. Predict Pat's final exam mark.

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.8734				
5	R Square					
6	Adjusted R Square	0.7453				
7	Standard Error	3.75				
8	Observations	30				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	2	1223.2			0.0000
13	Residual	27		14.08		
14	Total	29	1603.4			
15						
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
17	Intercept	13.01	3.53		0.0010	
18	Assignment	0.194	0.200		0.3417	
19	Midterm	1.11	0.122		0.0000	

PROBLEM # 12.6 The marketing manager for a chain of hardware stores needed more information about the effectiveness of the three types of advertising that the chain used. These are localized direct mailing (in which flyers describing sales and featured products are distributed to homes in the area surrounding a store), newspaper advertising, and local television advertisements. To determine which type is most effective, the manager collected 1 week's data from 100 randomly selected stores. For each store, the following variables were recorded:

- Weekly gross sales
- Weekly expenditure on direct mailing
- Weekly expenditures on newspaper advertising
- Weekly expenditures on television commercials
- All variables were recorded in thousands of dollars.

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.4419				
5	R Square	0.1953				
6	Adjusted R Square	0.0803				
7	Standard Error	2.59				
8	Observations					
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression		34.10			0.1979
13	Residual		140.56			
14	Total	24				
15						
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
17	Intercept	12.31	4.70	2.62	0.0160	
18	Direct	0.570	1.72		0.7437	
19	Newspaper	3.32		2.16	0.0427	
20	Television		1.96	0.37	0.7123	

a. Find the regression equation.
b. What is the coefficient of determination? What do these statistics tell you about the regression equation?
c. What does the standard error of estimate tell you about the regression model?
d. Test the validity of the model.
e. Which independent variables are linearly related to weekly gross sales in this model? Explain.
f. Compute the 95% interval of the week's gross sales if a local store spends $800 on direct mailing, $1,200 on newspaper advertisements, and $ 2,000 on television commercials. (Optional)
g. Calculate the 95% interval of the mean weekly gross sales for all stores that spend $800 on direct mailing, $ 1,200 on newspaper advertising, and $2,000 on television commercials. (Optional)
h. Discuss the difference between the two intervals found in Parts f and g.

PROBLEM # 12.7 What is multicollinearity, and how can it adversely affect multiple regression analysis? How can we tell whether multicollinearity is present?

There are several clues to the presence of multicollinearity:

(1) an _____ variable known to be an important _____ _____ _____.

(2) a partial regression coefficient exhibits the _____; and/or,

(3) when an _____, the partial regression coefficients for the other variables _____ _____.

PROBLEM # 12.8 In simple linear regression, the regression equation is a straight line. In multiple regressions, what geometric form is taken by the regression equation when there are two independent variables? When there are three or more independent variables?

When there are two independent variables, the regression equation can be thought of in terms of a_____.

When there are three or more independent variables, the regression equation becomes a mathematical entity called a _____;

_____

_____

_____

_____

PROBLEM # 12.9 What assumptions are required in using the multiple regression models?

In terms of the residual component of the model, the assumptions underlying multiple regression are:

1. For any _____ for the _____, the population of residuals will be _____ with a mean of _____ and a standard deviation of σ.

2. The standard deviation of the error terms is the _____regardless of the combination of values taken on by the independent variables.

3. The error terms are statistically _____.

PROBLEM # 12.10- The following data have been reported for a sample of 10 major U.S. zoological parks:

City	Y=Budget	X_1= Attendance	X_2= Acres	X_3= Number of Species
1	$19.5 million	0.6 million	210	271
2	40.0	2.0	216	400
3	11.9	0.4	70	377
4	14.0	1.0	125	277
5	11.6	1.5	55	721
6	22.2	1.3	80	400
7	20.5	1.3	42	437
8	26.0	2.5	91	759
9	17.0	0.9	125	270
10	14.6	1.1	92	260

a. Determine the least squares multiple regression equation.

	G	H	I	J	K	L	M
1			Budget	Attend	Acres	Species	
2			19.5	0.6	210	271	
3	SUMMARY OUTPUT		40.0	2.0	216	400	
4			11.9	0.4	70	377	
5	*Regression Statistics*		14.0	1.0	125	277	
6	Multiple R	0.8817	11.6	1.5	55	721	
7	R Square	0.7774	22.2	1.3	80	400	
8	Adjusted R Square	0.6662	20.5	1.3	42	437	
9	Standard Error	4.9139	26.0	2.5	91	759	
10	Observations	10	17.0	0.9	125	270	
11			14.6	1.1	92	260	
12	ANOVA						
13		*df*	*SS*	*MS*	*F*	*Significance F*	
14	Regression	3	506.06	168.69	6.99	0.022	
15	Residual	6	144.88	24.15			
16	Total	9	650.94				
17							
18		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
19	Intercept	4.31855	6.600	0.654	0.537	-11.8314	20.4685
20	Attend	11.95568	4.142	2.887	0.028	1.8209	22.0904
21	Acres	0.06115	0.033	1.829	0.117	-0.0206	0.1429
22	Species	-0.01538	0.016	-0.984	0.363	-0.0536	0.0229

b. Interpret the y-intercept and partial regression coefficients.
c. What is the estimated budget for a zoological park that draws an annual attendance of 2.0 million, occupies 150 acres, and has 600 species

Understanding the Basics: Suggested Problems from the Book.

In **Bold** are the Suggested Problems, in Green are the problems on Connect and the book.

	Chapter 14								
14.1	The Multiple Regression Model and the Least Squares Point Estimates								
14.2	Rsquare and Adjusted RSquare	**14.01**	**14.02**	**14.03**	**14.04**	**14.05**	14.06	14.07	
14.3	Model Assumptions and the Standard Error								
14.4	The Overall F Test	**14.08**	**14.09**	**14.10**	**14.11**	**14.12**	**14.13**	**14.14**	**14.15**
14.5	Testing the Significance of an Independent Variable	**14.16**	14.17	**14.18**	**14.19**	**14.20**	**14.21**	**14.22**	
	Supplementary	14.63	14.64						

This statistical workbook is compiled from the following books:

- Keller, G. (2012). *Statistics for management and economics*. Mason: Cengage Learning.
- McClave, J. T., Benson, G. P., & Sincich, T. (2008). *Statistics for Business and Economics*. New Jersey: Prentice Hall.
- Weiers, R. M. (2011). *Introduction to Business Statistics*. Mason: Cengage Learning.
- (GMAC), F. t. (Ed.). (2005). *GMAT -Quantitative Review.* Oxford, UK: Blackwell.

Bowerman, B. L., O'Connell, R. T., Murphree, E., Huchendorf, S. C., & Porter, D. C. (2003). *Business statistics in practice*(pp. 728-730). New York: McGraw-Hill/Irwin.