Chapter 15 – Multiple Regression

SECTION EXERCISES

SECTION 15.1

1.

a) $\hat{y} = 20,986.09 - 7483.10(2) + 93.84(1000) = \$99,859.89.$

b) The residual for the house that just sold for \$135,000 is \$135,000 - \$99,859.89 = \$35,140.11.

c) The house sold for more than the model predicted. The model underestimated the selling price.

2.

a) $\hat{y} = 28.4 + 11.37(15) + 2.91(20) = 257.15$ calories

b) The residual for this candy that has 227 calories per serving is 227 – 257.15 = –30.15 calories.

c) The candy has fewer calories that the model predicted. The model overestimated the calories in her candy.

SECTION 15.2

3.

a) $\widehat{USGross} = -22.9898 + 1.13442Budget + 24.9724Stars - 0.403296RunTime$

b) After allowing for the effects of *Stars* and *RunTime*, each additional million dollars in the budget for making the film yields about 1.13 million dollars in gross revenue.

4.

The manager is incorrectly interpreting the coefficient causally. The model says that longer films had smaller gross incomes (after allowing for budget and Stars), but it doesn't say that making a movie shorter will increase its gross. In fact, cutting arbitrarily would, for example, probably reduce the Star rating. Also, the coefficient is negative in the presence of the two other variables.

SECTION 15.3

5.

a) **Linearity**: The scatterplot shows that the relationship is reasonably linear; no curvature is evident in the scatterplot. The linearity condition is satisfied.

b) **Equal Spread:** The scattering of points seems to increase as the budget increases. The points are much more scattered (spread apart) to the right than to the left. Consequently, it appears that the equal spread condition is not satisfied.

c) **Normality:** A scatterplot of two variables gives us no information about the distribution of the residuals. Therefore we cannot determine if the normality assumption is satisfied.

6.

a) **Linearity**: A histogram gives us no information about the form of the relationship between two variables. Therefore, we cannot determine if the linearity condition is satisfied.

b) **Nearly Normal Condition:** The histogram shows that the distribution is unimodal and slightly skewed to the right. However, the presence of an outlier to the right is apparent in the histogram. This would violate the nearly normal condition.

c) **Equal Spread condition:** While a histogram shows the spread of a distribution, a histogram does not show if the spread is consistent for various values of *x*. We would need to see either a scatterplot of the data or a residual plot to determine if this were the case. Therefore, we cannot determine if the equal spread condition is satisfied.

SECTION 15.4

7.

a) The null hypothesis for testing the coefficient associated with *Stars* is: H_0: $\beta_{Stars} = 0$.

b) The *t*-statistic for this test is given by:

$$t_{n-k-1} = \frac{b_j - 0}{SE(b_j)}$$

$$t_{120-3-1} = \frac{24.9724}{5.884} = 4.24$$

c) The associated P-value is ≤ 0.0001.

d) We can reject the null hypothesis. There is sufficient evidence to suggest that the coefficient of *Stars* is not equal to zero. *Stars* is a significant predictor of a film's gross income.

8.

a) The null hypothesis for testing the coefficient associated with *Run Time* is: H_0: $\beta_{RunTime} = 0$.

b) The *t*-statistic for this test is given by:

$$t_{n-k-1} = \frac{b_j - 0}{SE(b_j)}$$

$$t_{120-3-1} = \frac{-0.403296}{0.2513} = -1.60$$

c) The *t*-statistic is negative because the regression coefficient for *Run Time* is negative.

d) The associated P-value is 0.1113.

e) We cannot reject the null hypothesis (at $\alpha = 0.05$). There is not sufficient evidence to suggest that the coefficient of *Run Time* is not equal to zero. *Run Time* is not a significant predictor of a film's gross income.

SECTION 15.5

9.

a) For this regression equation, $R^2 = 0.474$ or 47.4%. This tells us that 47.4% of the variation in the dependent variable *USGross* is explained by the regression model that includes the predictors *Budget, Run Time* and *Stars*.

b) The *Adjusted* R^2 value is slightly lower than the value of R^2. This is because the *Adjusted* R^2 is adjusted downward (imposing a "penalty") for each new predictor variable added to the model. Consequently, it allows the comparison of models with different numbers of predictor variables.

10.

a) To compute R^2 from the values in the table, we have

$$R^2 = \frac{SSR}{SST} = \frac{224995}{474794} = 0.474 \text{ or } 47.4\%$$

b) From the table, we see that $F = 34.8$.

c) The null hypothesis tested with the *F*-statistic is

$$H_0 : \beta_{Budget} = \beta_{Stars} = \beta_{RunTime} = 0$$

d) The P-value (see table) is very small. Therefore we can reject the null hypothesis and conclude that at least one slope coefficient is significantly different from zero (or that at least one predictor is significant in explaining *USGross*.

CHAPTER EXERCISES

11. Police salaries 2013.

a) **Linearity condition**: The scatterplots appear at least somewhat linear but there is a lot of scatter.
Randomization condition (Independence Assumption): States may not be a random sample but may be independent of each other.
Equal spread condition (Equal Variance Assumption): The scatterplot of Violent Crime vs Police Officer Wage looks less spread to the right but may just have fewer data points. Residual plots are not provided to analyze.
Nearly Normal condition (Normality Assumption): To check this condition, we will need to look at the residuals which are not provided for this example.

b) The R^2 of that regression would be $(0.051)^2 = 0.0026 = 0.26\%$.

12. Ticket prices.

a) **Linearity condition**: The first two scatterplots appear linear. The last one has a lot of scatter.
Randomization condition (Independence Assumption): These data are collected over time and may not be mutually independent. We should check for time-related patterns.
Equal spread condition (Equal Variance Assumption): The scatterplots show no tendency to thicken or spread.
Nearly Normal condition (Normality Assumption): To check this condition, we will need to look at the residuals which are not provided for this example.

b) The R^2 of that regression would be $(0.961)^2 = 0.9235 = 92.35\%$.

13. Police salaries, part 2.

a) The regression model:
$\widehat{Violent\ Crime} = 1370.22 + 0.795\,Police\,Officer\,Wage - 12.641\,Graduation\,Rate$

b) After allowing for the effects of *Graduation Rate*, states with higher *Police Officer Wages* have more *Violent Crime* at the rate of 0.7953 crimes per 100,000 for each dollar per hour of average wage.

c) $\widehat{Violent\ Crime} = 1370.22 + 0.795 * 20 - 12.641 * 70 = 501.25$ crimes per 100,000.

d) The prediction is not very good. The R^2 of that regression is only 22.4%.

14. Ticket prices, part 2.

a) The regression model:
$\widehat{Receipts} = -18.32 + 0.076\,Paid\ Attendance + 0.007\,\#\,Shows + 0.24\,Average\,Ticket\,Price$

b) After allowing for the effects of the # *Shows* and *Average Ticket Price*, each thousand customers account for about \$76,000 in receipts. That's about \$76 per customer, which is very close to the average ticket price.

c) $\widehat{Receipts} = -18.32 + 0.076 \times 200 + 0.007 \times 30 + 0.24 \times 70 = \$13.89\,\text{million}$

d) The prediction is very good. The R^2 of that regression is 99.9%.

15. Police salaries, part 3.

a) $0.221 = \dfrac{0.7947}{3.598}$

b) There are 50 states used in this model. The degrees of freedom are shown to be 47, which is equal to $n - k - 1 = 50 - 2 - 1 = 47$. There are two predictors.

c) The *t*-ratio is negative because the coefficient is negative meaning that *Graduation Rate* contributes negatively to the regression.

16. Ticket prices, part 3.

a) $126.7 = \dfrac{0.076}{0.0006}$

b) There are 78 weeks used in this model. The degrees of freedom are shown to be 74, which is equal to $n - k - 1 = 78 - 3 - 1 = 74$. There are three predictors.

c) The *t*-ratio is negative because the coefficient is negative, meaning that the intercept is negative.

17. Police salaries, part 4.

a) The hypotheses are: $H_0 : \beta_{Officer} = 0; H_A : \beta_{Officer} \neq 0$.

b) P = 0.8262 which is not small enough to reject the null hypothesis at $\alpha = 0.05$ and conclude that the coefficient is different from zero.

18. Ticket prices, part 4.

a) The hypotheses are: $H_0 : \beta_{\#\,Shows} = 0; H_A : \beta_{\#\,Shows} \neq 0$.

b) P = 0.116 which is too large to reject the null hypothesis at α= 0.05. The coefficient may be zero. Practically speaking, the # *Shows* doesn't contribute to this regression model.

c) The coefficient of # *Shows* reports the relationship after allowing for the effects of *Paid Attendance* and *Average Ticket Price*. The scatterplot and correlation were only concerned with the relationship between *Receipts* and # *Shows* (two variables).

19. Police salaries, part 5. This is a causal interpretation, which is not supported by regression. For example, among states with high graduation rates, it may be that those with higher violent crime rates spend more to hire police officers, or states with higher costs of living must pay more to attract qualified police officers but also have higher crime rates.

20. Ticket prices, part 5. This is a causal interpretation, which is not supported by regression. Also, it attempts to interpret a coefficient without taking account of the other variables in the model. For example, the number of paid attendees is surely related to the number of shows. If there were fewer shows, there would be fewer attendees.

21. Police salaries, part 6.
Constant Variance Condition (Equal Spread): met by the residuals vs. predicted plot.
Nearly Normal Condition: met by the Normal probability plot.

22. Ticket prices, part 6.
Constant Variance Condition (Equal Spread): met by the residuals vs. predicted plot.
Nearly Normal Condition: met by the Normal probability plot
Independence Assumption: There doesn't appear to be very much autocorrelation when the residuals are plotted against time.

23. Real estate prices.

a) Incorrect: Doesn't mention other predictors; suggests direct relationship between only two variables: *Age* and *Price*.

b) Correct

c) Incorrect: Can't predict *x* from *y*

d) Incorrect interpretation of R^2 (this model accounts for 92% of the of the variability in *Price*)

24. Wine prices.

a) Incorrect: Doesn't mention other predictors; suggests direct relationship between only two variables: *Age* and *Price*.

b) Incorrect interpretation of R^2 (this model accounts for 92% of the of the variability in *Price*)

c) Incorrect: Doesn't mention other predictors; suggests direct relationship between only two variables: *Tasting Score* and *Price*.

d) Correct

25. Appliance sales.

a) Incorrect: This is likely to be extrapolation since it is unlikely that they observed any data points with no advertising of any kind.

b) Incorrect: Suggests a perfect relationship

c) Incorrect: Can't predict one explanatory variable (x) from another

d) Correct

26. Wine prices, part 2.

a) Incorrect: Doesn't mention other predictors

b) Correct

c) Incorrect: Can't predict one explanatory variable (x) from another

d) Incorrect: Can't predict x from y

27. Cost of pollution.

a) The negative sign of the coefficient for ln(*number of employees*) means that for businesses that have the same amount of sales, those with more employees spend less per employee on pollution abatement on average. The sign of the coefficient for ln(*sales*) is positive. This means that for businesses with the same number of employees, those with larger sales spend more on pollution abatement on average.

b) The logarithms mean that the effects become less severe (in dollar terms) as companies get larger either in *Sales* or in *Number of Employees*.

28. OECD economic regulations.

a) No, it says that *after allowing for the effects of all the other predictors in the model,* the effect of more regulation on GDP is negative.

b) The F is clearly significant. We can be confident that the regression coefficients aren't all zero.

c) It makes sense that 1988 GDP is a good predictor of 1998 GDP. All the other predictors are only helpful after taking into consideration this one variable.

29. Home prices.

a) $\widehat{Price} = -152,037 + 9530\,Baths + 139.87\,Area$

b) R^2 =71.1%

c) For houses with the same number of bathrooms, each square foot of area is associated with an increase of $139.87 in the price of the house, on average.

d) The regression model says that for houses of the same size, there is no evidence that those with more bathrooms are priced higher. It says nothing about what would actually happen if a bathroom were added to a house.

30. Home prices, part 2. The residuals are right skewed, and the residuals versus fitted values plot shows a possible outlier, which may be the cause of the skewness in the other residuals. The outlier should be examined and either corrected or set aside and the regression recomputed. In addition, there is a clear pattern (negative linear before the outlier) in the residual vs. fitted plot.

31. Secretary performance.

a) The regression equation:

$$\widehat{Salary} = 9.788 + 0.110\,Service + 0.053\,Education + 0.071\,Test\ Score + 0.004\,Typing\ wpm + 0.065\,Dictation\ wpm$$

$$\widehat{Salary} = 9.788 + 0.110\times 120 + 0.053\times 9 + 0.071 * 50 + 0.004\times 60 + 0.065\times 30 = 29.205$$

b) $29,205

c) The t-value is 0.013 with 24 df and a P-value = 0.9897 (two-tailed), which is not significant at $\alpha = 0.05$.

d) You could take out the explanatory variable *typing speed* since it is not significant.

e) *Age* is likely to be collinear with several of the other predictors already in the model. For example, secretaries with longer terms of *Service* will naturally also be older.

32. Wal-Mart revenue.

a) The regression equation:

$$\widehat{Revenue} = 87.0089 + 0.0001 Retail\ Sales + 0.000011 Personal\ Consumption - 0.345 CPI$$

b) After allowing for the effects of the other predictors in the model, a change of 1 point in the CPI is associated with a decrease of 0.345 billion dollars on average in Wal-Mart revenue. Possibly higher prices (increased CPI) lead customers to shop less.

c) The Normal probability plot looks reasonably straight, so the Nearly Normal condition is met. With a P-value of 0.007, it is very unlikely that the true coefficient is zero.

33. Gross domestic product.

a) This model explains less than 4% of the variation in *GDP per Capita*. The P-value is not particularly low.

b) Because more education is general associated with a higher standard of living, it is not surprising that the simple association between *Primary Completion Rate* and GDP is positive.

c) The coefficient now is measuring the association between *GDP/Capita* and *Primary Completion Rate* after account for the two other predictors.

34. Lobster industry 2012, revisited.

a) $\widehat{LogValue} = 0.856 + 0.563 Traps - 0.000044 Fishers + 0.00381 Pounds / Trap$

b) Residuals show no pattern and have equal spread. Normal probability plot is straight. There is a question whether values from year to year are mutually independent.

c) After allowing for the number of *Traps* and *pounds/trap*, the *LogValue* of the lobster catch decreases by 0.000044 per *Fisher*. This doesn't mean that a smaller number of fishers would lead to a more valuable harvest. It is likely that *Traps* and *Fishers* are correlated, affecting the value and meaning of their coefficients.

d) The hypotheses are: $H_0 : \beta_{lbs/trap} = 0; H_A : \beta_{lbs/trap} \neq 0$; P-value = 0.0114 is below the common alpha level of 0.05, so we can reject the null hypothesis. However, this is not strong evidence that pounds/trap is an important predictor of the harvest value.

35. Lobster industry 2012, part 2.

a) $\widehat{EstimatedPrice/lb} = 1.094 + 1.236 Traps(M) - 0.000149 Fishers - 0.0180 Pounds / Trap$

b) Residuals show greater spread on the right and a possible outlier on the high end. We might wonder if the values from year to year are mutually independent. We should interpret the model with caution.

c) The hypotheses are: $H_0 : \beta_{pounds/trap} = 0; H_A : \beta_{pounds/trap} \neq 0$; P-value = 0.0011. It appears that *Pounds/Trap* does contribute to the model.

d) No, we can't draw causal conclusions from a regression. A change in *pounds/trap* would likely affect other variables in the model.

e) The adjusted R^2 accounts for the number of predictors and says to prefer the more complex model.

36. HDI.

a) $\widehat{HDI} = 0.09 + 0.01376 ExpectedYearsofSchooling + 0.00333 LifeExpectancy - 0.00012 MaternalMortality$
$+0.01686 MeanYrsSchool + 0.000745 PopUrban + 0.000000802 GDP / Capita + 0.00046 CellPhones / 100$

b) No. There appear to be two outliers with HDI higher than predicted. They make the residuals non-Normal.

c) The hypotheses are: $H_0 : \beta_{YrsSchool} = 0; H_A : \beta_{YrsSchool} \neq 0$; *t*-statistic 8 is quite large, so we can reject the null hypothesis.

d) The Normality assumption is violated. Most likely the standard deviation of the residuals is inflated. That would tend to make the t-ratios smaller. This one is so large that we probably can feel safe in rejecting the null hypothesis anyway.

37. Wal-Mart revenue, part 2.

a) $\widehat{PBE} = 87.0 - 0.345\,CPI + 0.000011\,Personal\ Consumption + 0.0001\,Retail\ Sales$

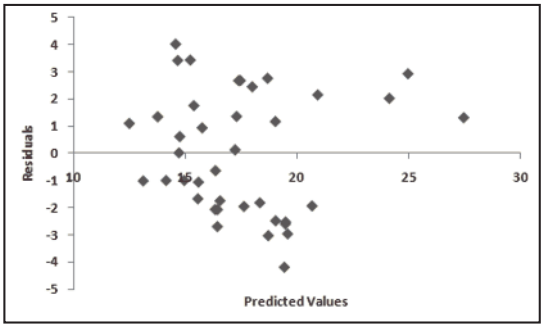
	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	87.00892605	33.59897163	2.589631	0.013908
CPI	-0.344795233	0.120335014	-2.86529	0.007002
Personal Consumption	1.10842E-05	4.40271E-06	2.51759	0.016546
Retail Sales Index	0.000103152	1.54563E-05	6.67378	1.01E-07

Regression Statistics	
Multiple R	0.816425064
R Square	0.666549886
Adjusted R Square	0.637968448
Standard Error	2.326701861
Observations	39

b) $R^2 = 66.7\%$ and all *t*-ratios are significant. It looks like these variables can account for much of the variation in Wal-Mart revenue.

38. Wal-Mart revenue, part 3.

a) The plot does not show any pattern or spread. There are a few high values to the right that could be considered outliers.



b) The December results correspond to the high values. Performing the regression analysis again without the four December values:

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-22.6410169	35.98903632	-0.62911	0.533887
CPI	0.04756868	0.129728836	0.366678	0.71635
Personal Consumption	7.8443E-07	4.27905E-06	0.183319	0.855741
Retail Sales Index	1.3382E-05	2.26191E-05	0.591614	0.558399

Regression Statistics	
Multiple R	0.64980747
R Square	0.42224975
Adjusted R Square	0.36633844
Standard Error	1.87418242
Observations	35

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	79.58196867	26.52732	7.552134	0.000623298
Residual	31	108.8893521	3.51256		
Total	34	188.4713207			

c) Without the December values, none of these variables is obviously different from zero. None of the *t*-ratios or P-values are significant. The F-statistic of 7.55 is highly significant with a P-value of < 0.001 indicating that the slope coefficients are not zero. The regression analysis as a whole doesn't provide much insight into the Wal-mart revenues. It appears that the model of the previous exercise was only about holiday sales.

39. Clinical trials.

a) $\widehat{Logit(Drop)} = -0.4419 - 0.0379\,Age + 0.0468 HDRS$

b) To find the predicted log odds (logit) of the probability that a 30-year-old patient with an HDRS score of 30 will drop out of the study: set *Age* = 30 and *HDRS* = 30 in the estimated regression equation: – 0.4419 – 0.0379 (30) – 0.0468(30) = –0.1749.

c) The predicted dropout probability of that patient is

$$\hat{p} = \frac{1}{1+e^{-(-0.1749)}} = \frac{1}{1+1.19} = 0.4564$$

d) To find the predicted log odds (logit) of the probability that a 60-year-old patient with an HDRS score of 8 will drop out of the study: set *Age* = 60 and *HDRS* = 8 in the estimated regression equation: – 0.4419 – 0.0379 (60) + 0.0468(8) = –2.342.

e) The associated predicted probability is

$$\hat{p} = \frac{1}{1+e^{-(-2.342)}} = \frac{1}{1+10.42} = 0.0877$$

40. Cost of higher education.

a) $\widehat{Logit(Type)} = -13.1461 + 0.08455 Top10\% + 0.000259\,\$ / Student$

The Outlier Condition is satisfied because there are not outliers in either predictor, but this is not a sample, but the top 25 colleges and universities in the U.S. It may be used for predicting, but the inference is not clear.

b) Yes; the P-value is < 0.05.

c) Yes; the P-value is < 0.05.

41. Motorcycles.

The scatterplot of *MSRP* versus *Wheelbase* indicates that this relationship is not linear. While there is a positive relationship between *Wheelbase* and *MSRP*, a curved pattern is evident. The scatterplot of *MSRP* versus *Displacement* indicates that this relationship is positive and linear. The scatterplot of *MSRP* versus *Bore* indicates that this relationship is also positive and linear. Based on these plots, it appears that both *Displacement* and *Bore* would be better predictors of *MSRP* than *Wheelbase.*

42. Motorcycles, part 2.

a) The hypotheses are: $H_0 : \beta_{Bore} = 0$; $H_A : \beta_{Bore} \neq 0$; P-value = 0.0108 which is greater than 0.05, so we fail to reject the null hypothesis.

b) Although *Bore* might be individually significant in predicting *MSRP*, in the multiple regression, after allowing for the effects of *Displacement*, it doesn't add enough to the model to have a coefficient that is clearly different from zero.

43. Motorcycles, part 3.

a) Yes, with an $R^2 = 90.9\%$ says that most of the variability of *MSRP* is accounted for by this model.

b) No, in a regression model, you can't predict an explanatory variable from the response variable.

44. Demographics.

a) The only model that seems to do poorly is the one that omits *murder.* The other three are hard to choose among.

$$\widehat{Life\,\exp} = 70.1421 - 0.238597(Murder) + 0.039059(HSgrad) + 0.000095(Income)$$

$$R^2 = 66.4\%$$

$$\widehat{Life\,\exp} = 69.7354 - 0.258132(Murder) + 0.051791(HSgrad) + 0.253982(Illiteracy)$$

$$R^2 = 66.8\%$$

$$\widehat{Life\,\exp} = 71.1638 - 0.273632(Murder) + 0.000381(Income) + 0.036869(Illiteracy)$$

$$R^2 = 63.7\%$$

b) Each of the models has at least one coefficient with a large P-value. This predictor variable could be omitted to simplify the model without degrading it too much.

c) No. Regression models cannot be interpreted that way. Association is not the same thing as causation.

d) Plots of the residual highlight Hawaii, Alaska, and Utah as possible outliers. These seem to the principal violations of the assumptions and conditions.

45. Burger King nutrition.

a) With an R^2 = 100%, the model should make excellent predictions.

b) The value of *s*, 3.140 calories, is very small compared to the initial standard variation of *calories*. This means that the model fits the data quite well, leaving very little variation unaccounted for.

c) No, the residuals are not all 0. Indeed, we know that their standard deviation is $s = 3.140$ calories. They are very small compared with the original values. The true value of R^2 was likely rounded up to 100%.

46. Health expenditures.

a) $\widehat{Expenditures} = 0.1994 + 0.232\,ExpectedYrsofSchooling + 0.051\,InternetUsers/100\,people$

b) Residuals show no pattern and have equal spread. There is one fairly high residual and one fairly low residual, but neither seem to be large outliers. Normal probability plot is fairly straight. Assuming that the errors are independent, the conditions are met.

c) The hypotheses are: $H_0: \beta_{Years} = 0$; $H_A: \beta_{Years} \neq 0$; the *t*-value is 2.81 (93df) and a P-value = 0.0006; we reject the null hypothesis of 0 slope.

d) No. The model says nothing about causality. It says that accounting for *Expected Year of Schooling*, higher numbers of Internet users are generally associated with higher health expenditures.

Ethics in Action

Kenneth's Ethical Dilemma: With all 5 independent variables included, gender shows no significant effect on sales performance and Kenneth wants to eliminate it from the model. Nicole reminds him that women had a history of being offered lower starting base salaries and when that is removed, gender is significant. When all variables are together, the detailed effects are confounded and gender is not an issue in sales performance.

Undesirable Consequences: Eliminating gender as a predictor of sales performance may mask the effects of gender and actually give an incorrect conclusion.

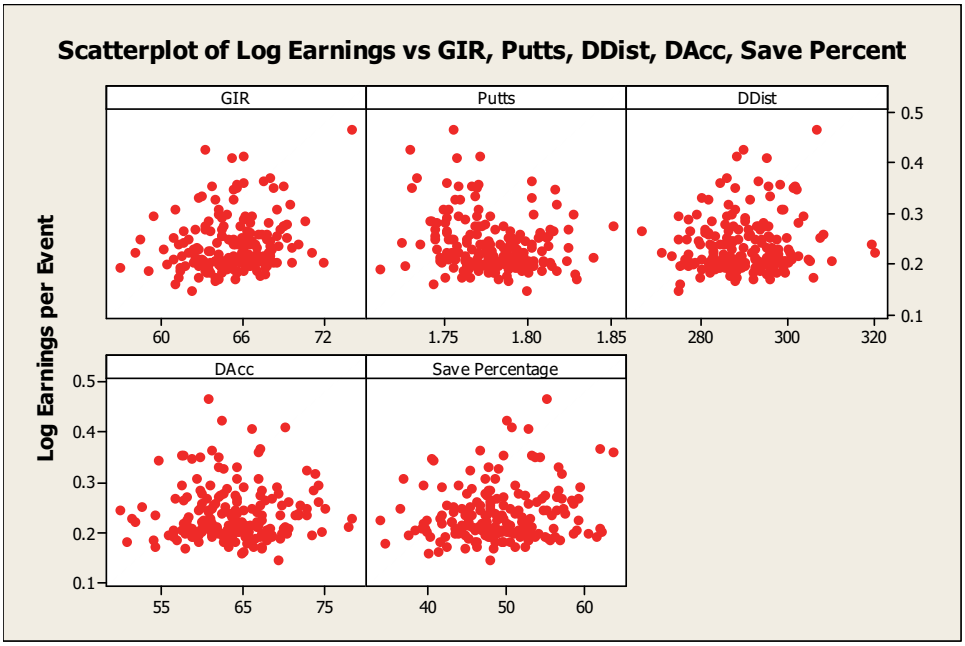
Ethical Solution: Kenneth needs to listen to Nicole's logic and her recollection of history regarding women and lower starting base salaries. Because the starting base salaries were inequitable, it is confounded with gender and starting base salaries and should be eliminated from the model until the data reflect the court order adjustment.

For further information on the official American Statistical Association's Ethical Guidelines, visit:
http://www.amstat.org/about/ethicalguidelines.cfm
The Ethical Guidelines address important ethical considerations regarding professionalism and responsibilities.

Brief Case – Golf Success

Report:

Of the potential variables considered for predicting golfers' success (measured in log earnings per event), the best model includes two significant independent variables: *GIR* and *Putts*. *GIR* stands for "Greens in Regulation" and is defined as the percentage of holes played in which the ball is on the green with two or more strokes left for par. The variable *Putts* is the average number of putts per hole in which the green was reached in regulation. In the scatterplots of *Log Earnings* versus all potential independent variables (shown below), weak linear relationships are observed (*Log Earnings* has a weak positive linear association with *GIR* and a weak negative linear association with *Putts*). The model is $\widehat{LogEarningsperEvent} = 0.828 + 0.00497\ GIR - 0.515\ Putts$. The model is significant with a moderate explanatory power ($R^2 = 37.2\%$). This model explains less than 40% of the variability in golfers' success. Examination of the residuals plotted against fitted values indicates that the equal spread condition is reasonably satisfied, but the histogram of residuals is skewed right indicating problems with the nearly normal condition even though a log transformation of earnings is used.



General Regression Analysis: Log$/Event versus GREENS IN REG., PUTT AVG.

Regression Equation

Log$/Event = 11.6983 + 0.0641226 GREENS IN REG. - 6.31982 PUTT AVG.

Coefficients

Term	Coef	SE Coef	T	P
Constant	11.6983	1.37020	8.53767	0.000
GREENS IN REG.	0.0641	0.00862	7.44206	0.000
PUTT AVG.	-6.3198	0.76064	-8.30860	0.000

Summary of Model

S = 0.329649 R-Sq = 37.20% R-Sq(adj) = 36.50%
PRESS = 20.0679 R-Sq(pred) = 34.85%