Statistics 621 Waiver Exam

August 24, 2003

This is an **open-book, open-notes** exam. You are limited to 1 textbook and 2 pages of notes, please. No laptops or other computers are allowed.

You have **two hours** for the exam.

The **computer output** associated with one or more items should be considered an essential part of the questions.

The multiple-choice questions are equally weighted. Your score is the total number of correct answers given in questions that are scored. Some questions may be dropped and not counted as part of the overall score.

You should interpret the word “significant” in this exam as implying “statistically significant.”

Please also note the following when filling in the answer sheet.

- **– Fill in your name and student id number** on the answer form.
- **– Mark the “bubbles”** under your name and student id number on the form.
- – Choose **one *best* answer** by marking the item on the answer form.
- – Mark the answer form using only a **#2 pencil**. Erase all changes completely.

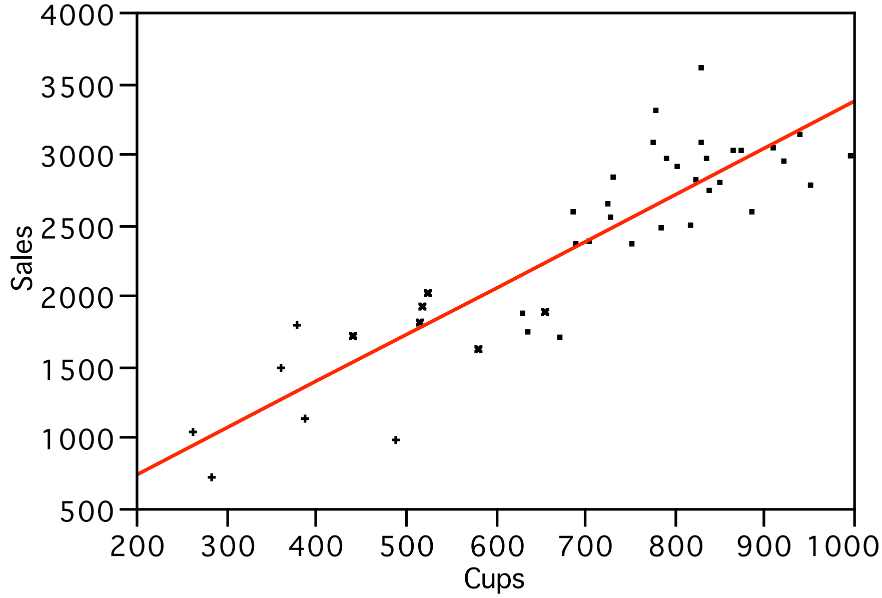
Turn in the solution page only; keep the test. Mark your copy of the exam in order to remember your choices so that you can check how well you did.

Solutions will be posted in WebCafe. A list of those passing the exam will be available from the MBA program office on the third floor of Huntsman Hall.

STOP

Do not turn the page until you are instructed.

(Questions 1–11) The owner of an independent coffee shop collected sales data over a six-week period. For each day, the owner recorded the total sales (*Sales*, in dollars) and the number of cups of coffee sold (*Cups*). The owner then obtained the following results.

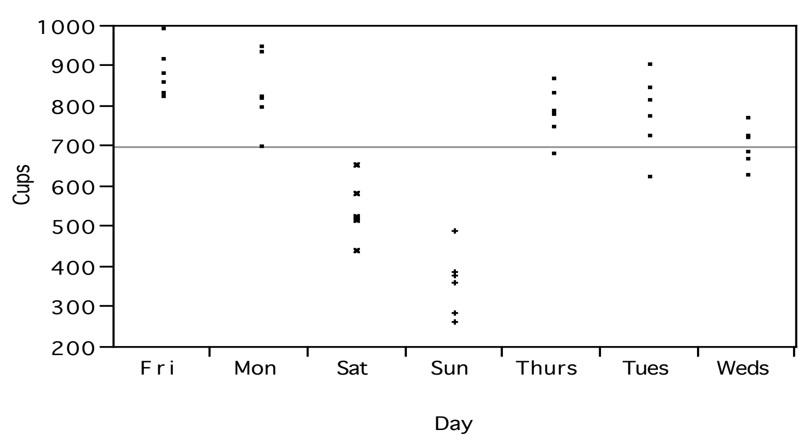


Summary of Fit

RSquare	0.789
Root Mean Square Error	305
Mean of Response	2394
Observations	42

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	79.9	195.6	0.41	0.6851
Cups	3.3	0.3	12.25	<.0001



(1) From the shown output, how much on average does a customer spend for a cup of coffee in this coffee shop?
(a) About \$0.80
(b) About \$2.40
(c) About \$3.30
(c) Less than \$2.00
(d) More than \$4.00

(2) The confidence interval for the intercept in the fitted model includes zero, implying that
(a) the model should be re-estimated without an intercept.
(b) the data are a sample from a population for which the intercept is zero.
(c) customers only purchase coffee in the shop.
(d) the data is consistent with samples from a population with intercept zero.
(e) None of the above.

(3) On a day similar to these, the shop sold 500 cups of coffee. Based on these results, we would *expect* the sales of the shop to be approximately
(a) \$80
(b) \$1500
(c) \$1650
(d) \$1730
(e) \$2394

(4) The manager of the shop hired a new group of employees. She used the results shown here to see how well these new employees did on their first day. The sales that day were \$300 lower than the model predicted. This difference of the data from the prediction
(a) is consistent with random variation and does not indicate a serious problem.
(b) is larger than typical random variation and indicates a serious problem.
(c) indicates that the model is no longer valid and should not be used for these new employees.
(d) implies that too few cups of coffee were sold that day for the model to be used.
(e) implies that too many cups of coffee were sold that day for the model to be used.

(5) Portions of the output show that many fewer cups of coffee are sold on the weekend days (Saturday and Sunday). Because fewer cups are sold on the weekend,
(a) the model should be fit separately on the weekend and weekdays.
(b) a categorical variable indicating the weekend should be added to the model.
(c) a categorical variable indicating the day of the week should be added to the model.
(d) the data are dependent, violating a key regression assumption.
(e) the confidence interval for the slope is shorter than if only weekdays were used in the model.

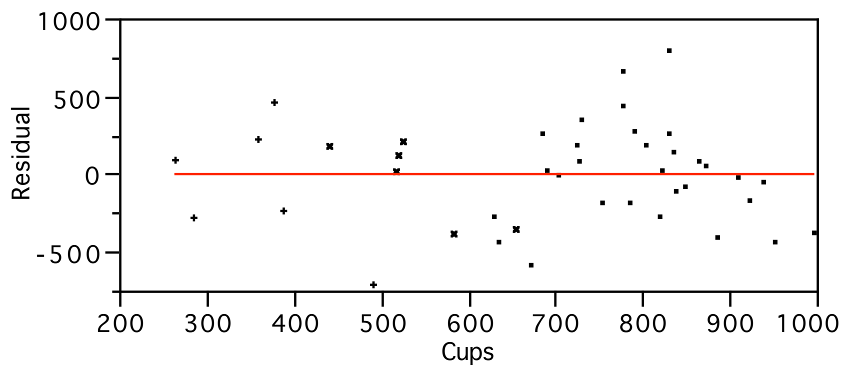
(6) If the shop were to start selling specialty coffees (such as cappuccino, which costs more per cup than typical coffee), then we can expect that a regression fit to the data that includes such specialty sales would have
(a) a smaller intercept.
(b) a larger intercept.
(c) a larger slope.
(d) a smaller slope.
(e) a smaller RMSE.

(7) On average with 95% confidence, sales on days with 700 cups exceed sales on days with 600 cups sold by about
(a) [$329, $331]
(b) [$300, $360]
(c) [$270, $390]
(d) [$-280, $635]
(e) [$400, $940]

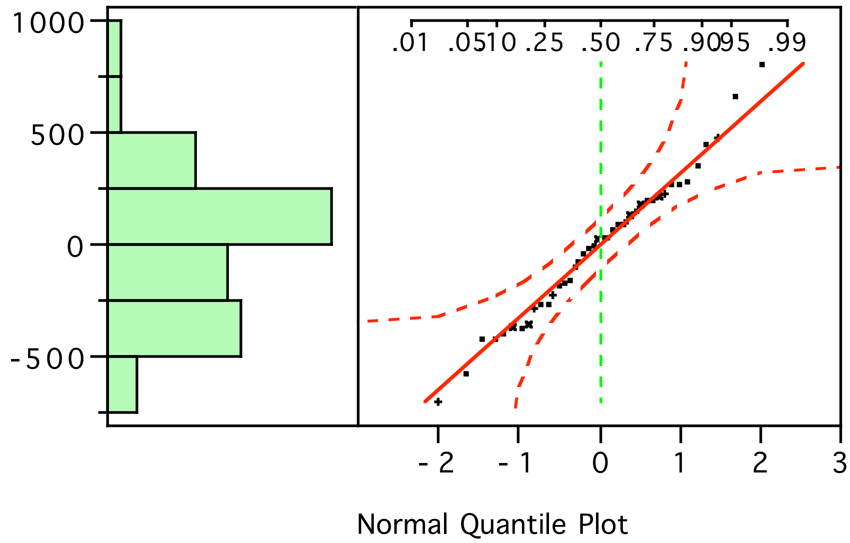
(8) An important check of the validity of the results that is not evident in the shown output is
(a) a plot of the fitted values on the number of cups.
(b) a time sequence plot of the residuals.
(c) a leverage plot of the fitted model.
(d) a plot of the residuals by day (i.e., Sunday, Monday,…,Saturday).
(e) not needed; the model description is complete.

(9) If the store were to expand the model by adding 3 more weeks of data (i.e., 21 more observations), then we can be sure that when fit to the larger collection of 42+21=63 observations that
(a) the R^2 of the model would be larger.
(b) the RMSE of the model would be larger.
(c) the confidence interval for the intercept would no longer include zero.
(d) the confidence interval for the slope would be shorter.
(e) nothing would change in a noticeable, predictable way.

(10) The following plot based on the fitted model indicates that
(a) the fitted model should be non-linear.
(b) the data are dependent.
(c) the variance of the error terms grows with the number of cups sold.
(d) the variation of the observed errors is not consistent with normality.
(e) the fit of the model appears OK.



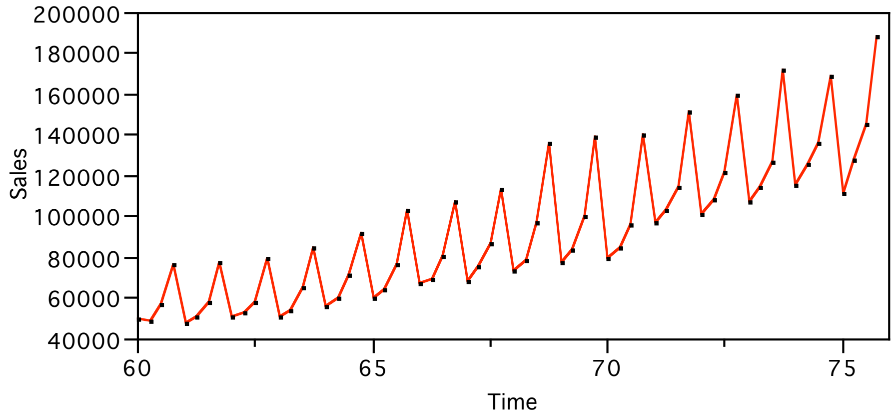
(11) The following plot of the residuals from the fitted model indicates that
(a) The underlying errors are autocorrelated.
(b) The underlying errors lack constant variance.
(c) The underlying errors are a sample from a normal distribution.
(d) The underlying errors are not a sample from a normal distribution.
(e) The results are consistent with models having normally distributed errors.

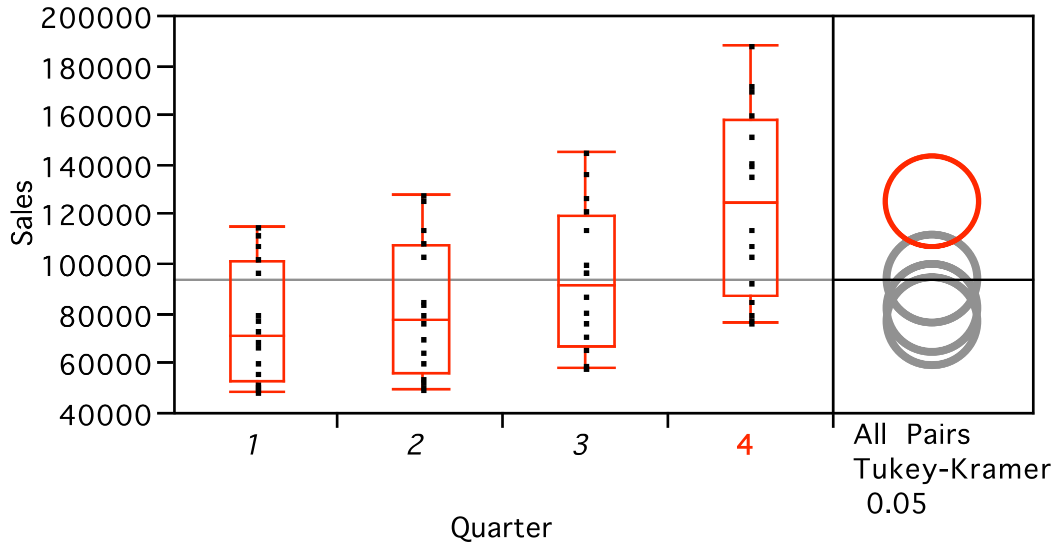


(12) The shown model uses only the total number of cups. If the owner also knows the size of the cups that were sold (*Cups* = *Small Cups* + *Medium Cups* + *Large Cups*), then the owner should
(a) add a categorical variable to the model indicating the cup size.
(b) add both a categorical variable for size as well as the associated interaction to the model.
(c) replace *Cups* in the model by *Small Cups*, *Medium Cups*, and *Large Cups*.
(d) add *Small Cups* and *Medium Cups* to the model.
(e) either "c" or "d" since these produce to the same fit.

(Questions 13–16) The Fields Department Store of Chicago tracked its quarterly sales (recorded in thousands of $'s) from the first quarter of 1960 though the last quarter of 1975. These 16 years of quarterly data give a sequence of 64 quarterly sales figures. For this analysis, the time period is recorded as a column labeled *Time* with values 60, 60.25, 60.5,…,75.75.

The following analysis builds a model for the sales over this time period and presents several alternative views of the data. Output for the modeling begins on the next page.





Means Comparisons

Mean[i]-Mean[j]	4	3	2	1
4	0	31113	43013	48373
3	-31113	0	11899	17260
2	-43013	-11899	0	5360
1	-48373	-17260	-5360	0

Abs(Dif)-LSD	4	3	2	1
4	-27688	3426	15325	20685
3	3426	-27688	-15788	-10428
2	15325	-15788	-27688	-22328
1	20685	-10428	-22328	-27688

Positive values show pairs of means that are significantly different.

Response Sales

RSquare	0.957
Root Mean Square Error	7390
Mean of Response	94063
Observations	64

Analysis of Variance

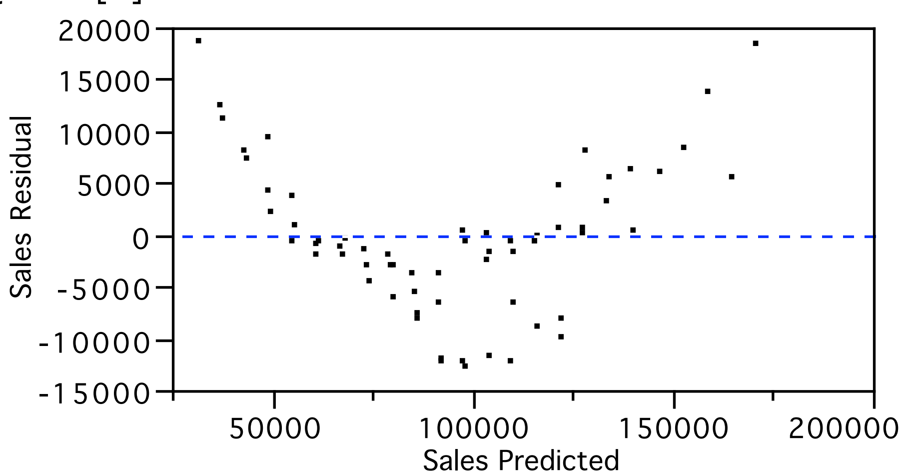
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	7.198e+10	1.8e+10	329.3560
Error	59	3223449773	54634742	**Prob > F**
C. Total	63	7.52e+10		<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Time	1	1	4.947e+10	905.5067	<.0001
Quarter	3	3	1.884e+10	114.9438	<.0001

Durbin-Watson	Number of Obs.	AutoCorrelation
1.233	64	0.2730

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-315311	13636	-23.12	<.0001
Time	6031	200	30.09	<.0001
Quarter[1]	-15486	1602	-9.67	<.0001
Quarter[2]	-11634	1601	-7.27	<.0001
Quarter[3]	-1242	1601	-0.78	0.4407
Quarter[4]	28363	1602	17.70	<.0001



(13) Accepting the fitted regression, in 5 years time, we should expect quarterly sales to grow on average with 95% confidence by
(a) $30,000,000 ± 400,000
(b) $30,000,000 ± 2,000,000
(c) $30,000,000 ± 14,800,000
(d) $30,000,000 ± 72,000,000
(e) $6,000,000 ± 400,000

(14) When adjusted for the modeled linear growth through time, the estimated *difference* in sales between quarter 1 and quarter 4 is about
(a) $48,400,000
(b) $43,800,000
(c) $28,400,000
(d) $20,700,000
(e) $15,500,000

(15) The fit of the shown multiple regression implies that when compared to the total sales in the last year (1975), total sales over the next 4 quarters will, on average, grow by
(a) $6,031,000
(b) $15,100,000
(c) $24,120,000
(d) $94,063,000
(e) About zero (i.e., the same as the total sales in 1975).

(16) If the shown model is used to predict sales in the next year, this output suggests that these predictions will
(a) Be rather close to the actual values because the R^2 of the model is so large.
(b) Be within about $7,390,000 of the actual sales for each quarter, with 95% confidence.
(c) Be within about $14,780,000 of the actual sales for each quarter, with 95% confidence.
(d) Grossly over-predict the sales in the coming year.
(e) Grossly under-predict the sales in the coming year.

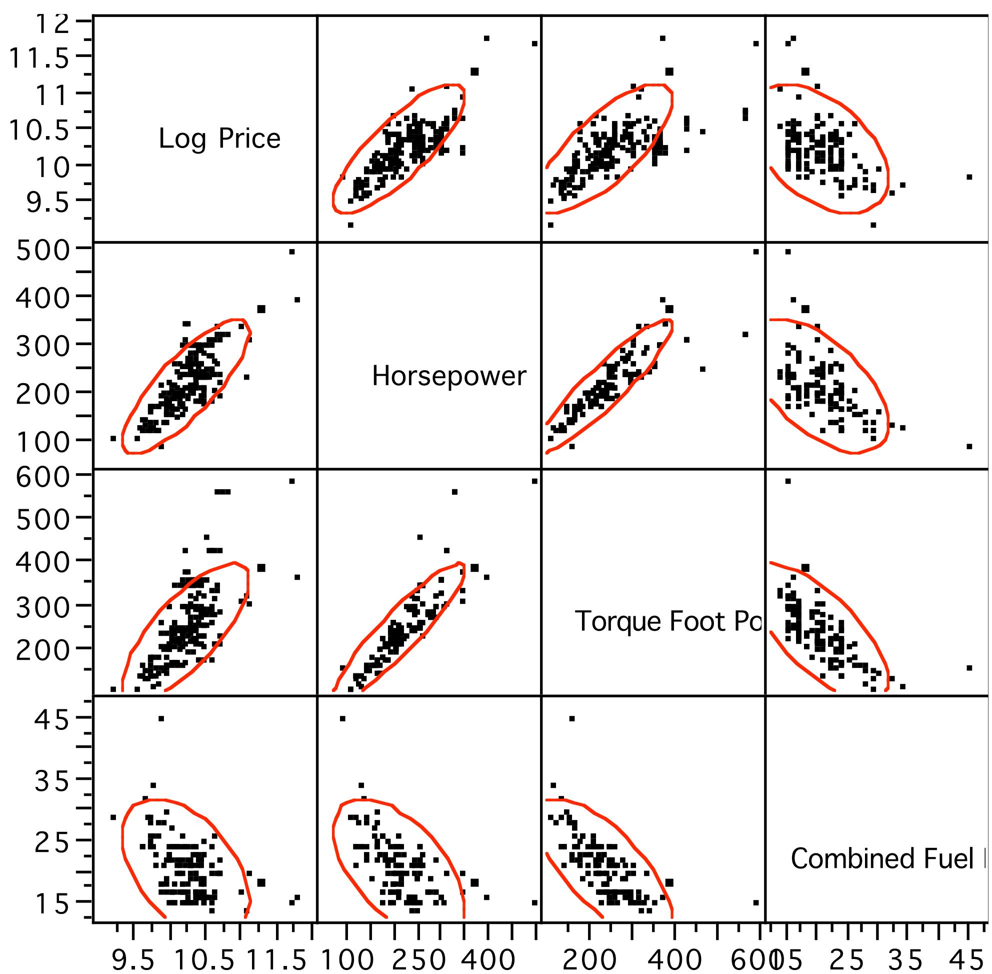
(17) It has been claimed that holiday sales of this department store (i.e., sales in Q4) have been growing more rapidly than sales during other quarters. With regard to this claim, the shown multiple regression
(a) Implies that sales in Quarter 4 are growing more rapidly than sales in other quarters.
(b) Implies that sales in Quarter 4 are growing at the same rate as sales in other quarters.
(c) Implies that sales in Quarter 4 are growing more slowly than sales in other quarters.
(d) Fits the data too well to allow for comparisons of the trends in different quarters.
(e) Lacks the interaction terms needed to estimate differences in rates of growth.

(18) The most useful step to take next that would improve this model the most would be to
(a) Add an interaction between the categorical variable (*Quarter*) and *Time*.
(b) Remove the insignificant indicator for Quarter 3.
(c) Check for collinearity using variance inflation factors.
(d) Re-express the response on a log scale and refit the model.
(e) Add a lagged value of the response to correct for the evident autocorrelation.

(Questions 19-30) An automobile manufacturer is trying to decide how to price a new vehicle it intends to introduce into the market. As part of its planning, it built a model for how the prices of competitors' cars depend on several performance characteristics, including the engine horsepower, the engine torque (large torque implies starting power at low speeds), the number of engine cylinders, and fuel economy (combined city and highway). The 200 cars used in the regression were models from rival companies that this manufacturer considers comparable (and competitive) with its new car. The analysis uses the natural log (i.e., base *e*) of the manufacturers suggested retail price (MSRP) for the cars included in this collection. Note that missing data for some of the cars reduced the total used in the multiple regression to only 146 out of the total of 200.

Correlations

	Log Price	Horsepower	Torque	Fuel Economy
Log Price	1.000	0.811	0.757	-0.464
Horsepower	0.811	1.000	0.917	-0.565
Torque	0.757	0.917	1.000	-0.702
Fuel Economy	-0.464	-0.565	-0.702	1.000



Cylinder Frequencies

Level	Count	Prob
10	5	0.025
12	1	0.005
4	43	0.215
5	2	0.010
6	80	0.400
8	69	0.345
Total	200	

Response Log Price

RSquare	0.676
Root Mean Square Error	0.213
Mean of Response	10.227
Observations	146

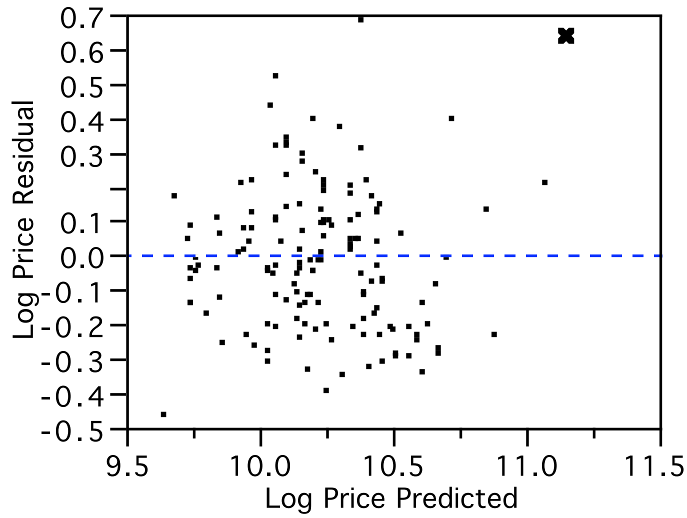
Analysis of Variance

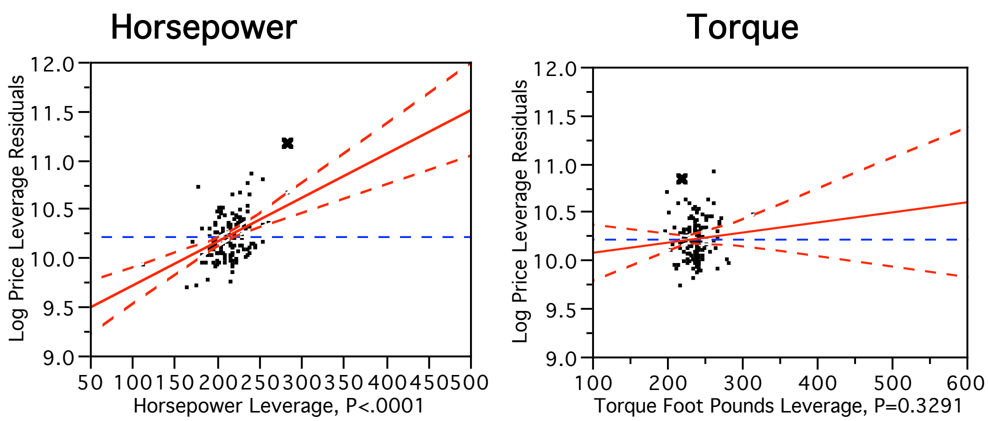
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	13.0316	1.86165	41.1903
Error	138	6.2371	0.04520	**Prob > F**
C. Total	145	19.2687		<.0001

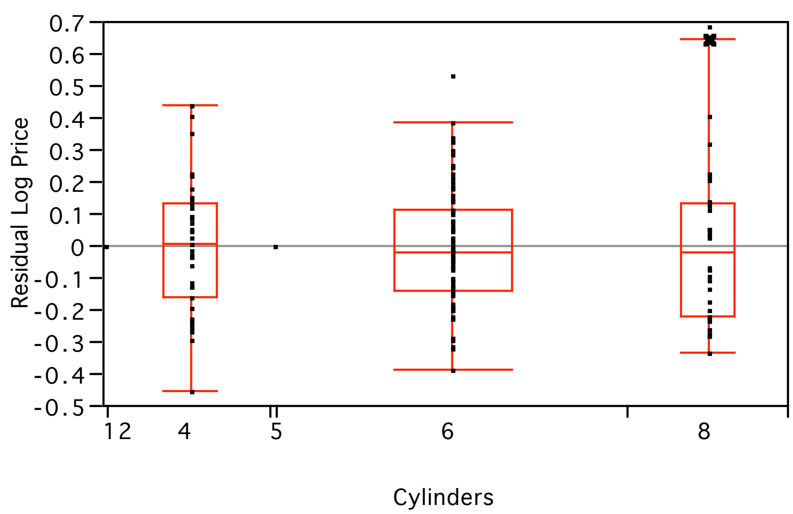
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Horsepower	1	1	1.310	28.982	<.0001
Torque	1	1	0.043	0.959	0.3291
Fuel Economy	1	1	0.008	0.188	0.6655
Cylinders	4	4	0.324	1.794	0.1334

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	9.05	0.2828	32.00	<.0001
Horsepower	0.0045	0.0008	5.38	<.0001
Torque	0.0011	0.0011	0.98	0.3291
Fuel Economy	0.0029	0.0066	0.43	0.6655
Cylinders[12]	-0.220	0.26	-0.86	0.3940
Cylinders[4]	-0.075	0.11	-0.65	0.5169
Cylinders[5]	0.480	0.19	2.57	0.0112
Cylinders[6]	-0.072	0.08	-0.89	0.3745
Cylinders[8]	-0.108	0.07	-1.52	0.1308


0.7
0.6
0.5
0.4
0.3
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
Log Price Residual
9.5
10.0
10.5
11.0
11.5
Log Price Predicted


Horsepower
Torque
Log Price Leverage Residuals
12.0
11.5
11.0
10.5
10.0
9.5
9.0
50 100 150 200 250 300 350 400 450 500
Horsepower Leverage, P<.0001
100 200 300 400 500 600
Torque Foot Pounds Leverage, P=0.3291


Residual Log Price
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
12
4
5
6
8
Cylinders

(19) Would a simple regression of *Log Price* on *Torque* (using these same 146 cars) explain significant amounts of the variation in the response?
(a) Yes, the correlation between *Log Price* on *Torque* is larger than zero.
(b) Yes, the F ratio for the simple regression is larger than 4.
(c) No, the multiple regression coefficient for *Torque* is not significantly different from zero.
(d) No, the correlation between *Log Price* on *Torque* is not far from zero.
(e) Cannot be deduced from the shown output.

(20) In addition to indicating the strength of the linear association between the response and the continuous predictors in this model, the scatterplot matrix shows
(a) The range of each of these predictors and the response.
(b) Outliers in the relationships among these columns.
(c) The presence of non-linear patterns.
(d) Collinearity among the predictors.
(e) All of the above.

(21) A correct interpretation of the R^2 of the fitted multiple regression is that
(a) The correlation between predicted values and response is 0.822.
(b) The variance of the residuals is 67.6% of the variance in log price.
(c) The multiple regression accurately predicts the log price for 67.6% of these cars.
(d) About 67.6% of these cars fall within 2 RMSE of the fitted model.
(e) All of the above.

(22) The predicted price for the new model based on this fitted regression is \$37,000. Based on the shown output, how high could the manufacturer price this new model without attracting undue attention (i.e., be within the range of most, say 95%, prices typical for a car with the features of its model)?
(a) \$37,000
(b) \$37,213
(c) \$37,416
(d) \$45,800
(e) \$56,650

(23) Does the number of cylinders used in the engine of the car have a significant effect on the (log of the) price of these cars?
(a) Yes, the dummy term Cylinders[5] is significant.
(b) No, some of the dummy terms, such as Cylinders[4], are not significant.
(c) Yes, the partial F-test for Cylinders is significant.
(d) No, the partial F-test for Cylinders is not significant.
(e) The variance of cars with different numbers of cylinders varies and cannot be tested.

(24) Which design change would be expected to exert the larger impact on price: raising horsepower from 120 to 140, or raising horsepower from 160 to 180? Assume that other features are not altered.
(a) Raising the horsepower from 120 to 140.
(b) Raising the horsepower from 160 to 180.
(c) Neither change would affect the price of the car noticeably.
(d) Both effects would be noticeable and of the same size.
(e) Cannot be deduced without knowing the other features of the car.

(25) The standard errors of the estimated coefficients for the dummy terms representing the number of cylinders differ considerably because
(a) The data include varying numbers of cars in each group identified by *Cylinder*.
(b) Collinearity among the predictors increased the variation for some categories.
(c) Leveraged outliers have affected some of the estimates more than others.
(d) Only one of the terms is significant, leaving the others as pure random error.
(e) The data do not follow a normal distribution.

(26) The best explanation for the apparent small effect of *Torque* and *Fuel Economy* on the prices of these cars is that
(a) Customers do not place much value on engine torque or fuel economy.
(b) These characteristics are redundant given the horsepower of the car.
(c) The model lacks interactions that would capture the effects of these predictors.
(d) The data are dependent and so the model has over-estimated the SE for these predictors.
(e) The sample size is too small to estimate the effects of these factors.

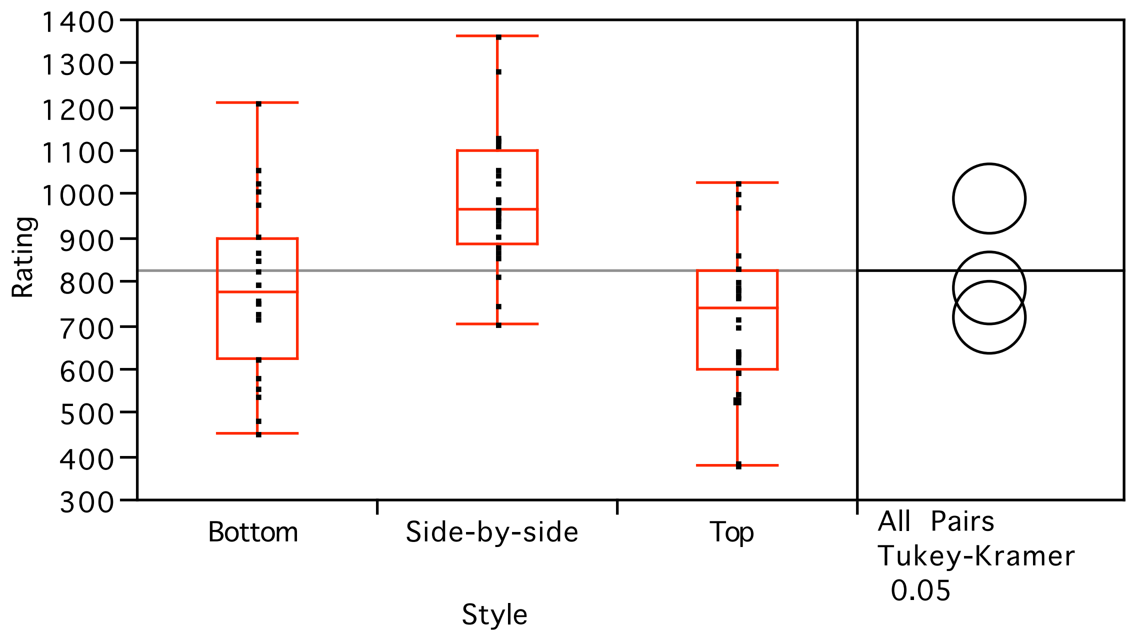
(27) The p-value for the estimated effect of *Torque*, if we accept the underlying assumptions for a multiple regression, is
(a) The probability that the population slope for *Torque* is zero.
(b) The probability that the population slope for *Torque* is not zero.
(c) One minus the probability that the population slope for *Torque* is not zero.
(d) The percentage of cars whose price is affected by *Torque*.
(e) The probability of a fitted slope this large in magnitude (or larger) if the true slope is zero.

(28) The non-zero intercept in the fitted model implies that that
(a) Cars with no horsepower, torque, fuel economy and cylinders only cost about $9.05.
(b) Cars with no horsepower, torque, fuel economy and cylinders only cost about $8,500.
(c) The continuous predictors should be re-expressed on a log scale and the model re-fit.
(d) Little, because the estimated intercept is not significantly far from zero.
(e) Little, because the estimated intercept is a large extrapolation from the fitted data.

(29) If the outlying car which is highlighted with an "x" in the plots were removed from the analysis and the model re-estimated, then
(a) The RMSE would increase.
(b) The slope for *Horsepower* would decrease.
(c) The slope for *Torque* would decrease.
(d) The slope for *Cylinders*[12] would not be estimable since only this car has 12 cylinders.
(e) None of the above.

(30) The leverage plot for *Torque* shows the points grouped into a small range along the x-axis because
(a) *Torque* does not have a significant effect in the model.
(b) the marginal variation in *Torque* is small compared to that in other predictors.
(c) the regression model includes a categorical predictor with many levels.
(d) *Torque* is collinear with other predictors in the model.
(e) the software has a programming error.

(Questions 31-40) A seller of large, household appliances conducted a marketing study to assess the attitudes of customers to features of refrigerators. The seller recruited 72 customers. As part of this study, each customer was shown a picture and description of a type of refrigerator. The customer was then asked to assign "an expected selling price", here called the *Rating*, to that type of refrigerator. The following analysis reports results associated with two attributes: the *Style* (as determined by freezer location; top, bottom, and side-by-side) and the *Efficiency* (of energy use over a year; low, med, high). Each of the 9 combinations of *Style* and *Efficiency* was seen by 8 customers. Several analyses of this data follow.



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Style	2	974125.3	487063	16.3248	<.0001
Error	69	2058665.8	29836		
C. Total	71	3032791.2			

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Bottom	24	780.99	35.26	710.65	851.3
Side-by-side	24	988.52	35.26	918.18	1058.9
Top	24	715.69	35.26	645.36	786.0

Std Error uses a pooled estimate of error variance

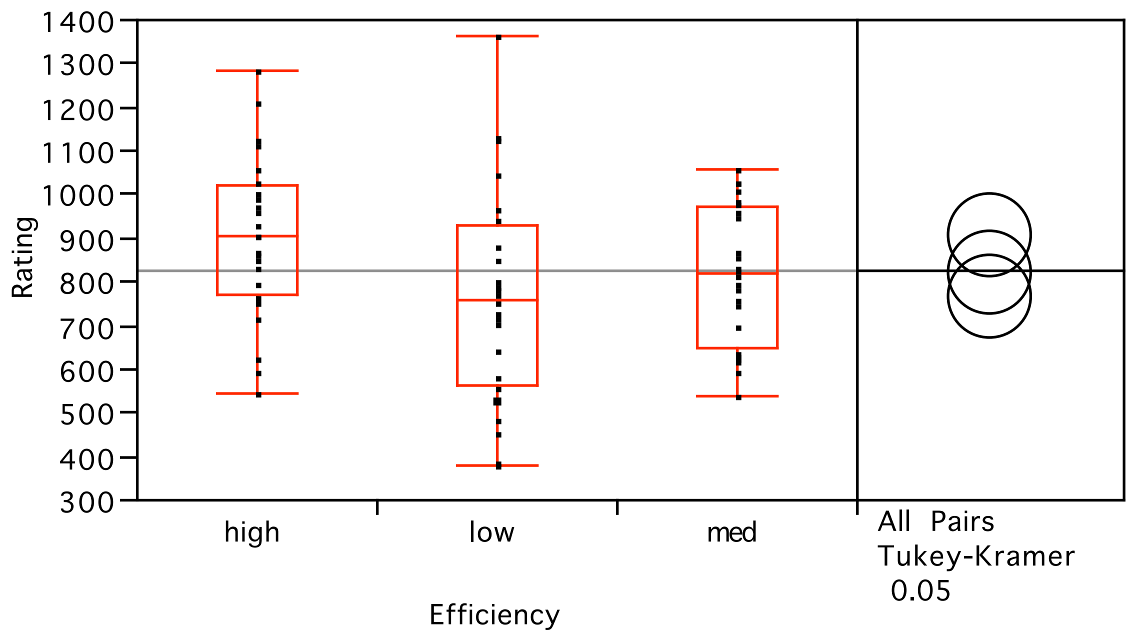
Means Comparisons

Mean[i]-Mean[j]	Side-by-side	Bottom	Top
Side-by-side	0.00	207.53	272.82
Bottom	-207.53	0.00	65.29
Top	-272.82	-65.29	0.00

Comparisons for all pairs using Tukey-Kramer HSD

Abs(Dif)-LSD	**Side-by-side**	**Bottom**	**Top**
Side-by-side	-119.44	88.09	153.39
Bottom	88.09	-119.44	-54.14
Top	153.39	-54.14	-119.44

Positive values show pairs of means that are significantly different.



Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**	**Prob > F**
Efficiency	2	234236.2	117118	2.8876	0.0625
Error	69	2798555.0	40559		
C. Total	71	3032791.2			

Level	**Number**	**Mean**	**Std Error**	**Lower 95%**	**Upper 95%**
high	24	903.13	41.11	821.12	985.14
low	24	764.74	41.11	682.73	846.75
med	24	817.33	41.11	735.32	899.34

Means Comparisons

Mean[i]-Mean[j]	**high**	**med**	**low**
high	0.00	85.80	138.39
med	-85.80	0.00	52.59
low	-138.39	-52.59	0.00

Comparisons for all pairs using Tukey-Kramer HSD

Abs(Dif)-LSD	high	med	low
high	-139.26	-53.45	-0.87
med	-53.45	-139.26	-86.67
low	-0.87	-86.67	-139.26

Positive values show pairs of means that are significantly different.

Response Rating

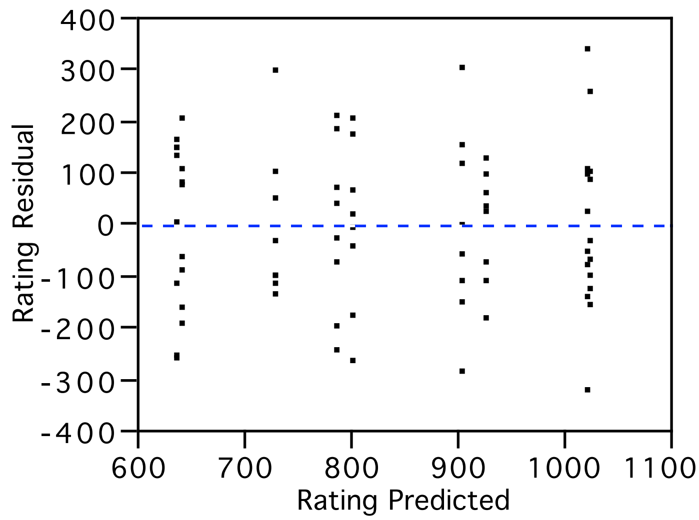
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	1395846	174481	6.7151
Error	63	1636945	25983	Prob > F
C. Total	71	3032791		<.0001

Effect Tests

Source	DF	Sum of Squares	F Ratio	Prob > F
Style	2	974125	18.75	<.0001
Efficiency	2	234236	4.51	0.0148
Efficiency*Style	4	187485	1.80	0.1392

Term	Estimate	SE	t Ratio	Prob>|t|
Intercept	828.4	19.0	43.61	<.0001
Style[Bottom]	-47.4	26.9	-1.76	0.0824
Style[Side-by-side]	160.1	26.9	5.96	<.0001
Style[Top]	-112.7	26.9	-4.20	<.0001
Efficiency[high]	74.7	26.9	2.78	0.0071
Efficiency[low]	-63.7	26.9	-2.37	0.0209
Efficiency[med]	-11.1	26.9	-0.41	0.6816
Efficiency[high]*Style[Bottom]	46.3	38.0	1.22	0.2278
Efficiency[high]*Style[Side-by-side]	-41.5	38.0	-1.09	0.2793
Efficiency[high]*Style[Top]	-4.8	38.0	-0.13	0.8997
Efficiency[low]*Style[Bottom]	-76.8	38.0	-2.02	0.0476
Efficiency[low]*Style[Side-by-side]	94.9	38.0	2.50	0.0151
Efficiency[low]*Style[Top]	-18.2	38.0	-0.48	0.6338
Efficiency[med]*Style[Bottom]	30.5	38.0	0.80	0.4254
Efficiency[med]*Style[Side-by-side]	-53.5	38.0	-1.41	0.1642
Efficiency[med]*Style[Top]	23.0	38.0	0.61	0.5472



Level	Efficiency*Style Least Sq Mean	Std Error
high,Bottom	902.0	57.0
high,Side-by-side	1021.8	57.0
high,Top	785.6	57.0
low,Bottom	640.6	57.0
low,Side-by-side	1019.8	57.0
low,Top	633.8	57.0
med,Bottom	800.4	57.0
med,Side-by-side	924.0	57.0
med,Top	727.6	57.0

(31) Should the manufacturer conclude that the efficiency of the refrigerator has a significant impact on the *Rating* assigned to the design by customers?
(a) Yes, the appropriate F-test for *Efficiency* is significant.
(b) No, the appropriate F-test for *Efficiency* is not significant.
(c) No, the Tukey-Kramer comparisons do not find a significant difference.
(d) No, the ratings when grouped by *Efficiency* are too variable and overlap considerably.
(e) The presence of so many interactions makes it impossible to tell.

(32) The fitted model (with its usual assumptions) implies that 2/3 of customers would rate a high efficiency, side-by-side refrigerator in the range (*Rating* is on a $ scale)
(a) \$1022 ± 57
(b) \$787 ± 38
(c) \$1022 ± 161
(d) \$975 ± 57
(e) \$975 ± 114

(33) Do customers value efficiency more (i.e., assign a higher rating) in some styles than others?
- (a) No, the Tukey-Kramer comparisons of *Rating* based on *Efficiency* are not significant.
- (b) Yes, the Tukey-Kramer comparisons of *Rating* based on *Style* are significant.
- (c) Yes, the interaction of *Style* with *Efficiency* is not significant.
- (d) No, the interaction of *Style* with *Efficiency* is not significant.
- (e) These data are too collinear in order to identify such effects.

(34) If the shown two-way analysis of variance were treated as a multiple regression, then the R^2 of the multiple regression would be
- (a) 0.08
- (b) 0.17
- (c) 0.32
- (d) 0.46
- (e) Cannot be recovered from the shown output.

(35) If the interaction component is removed from the shown two-way analysis of variance, then the F-ratio for *Efficiency* would
- (a) Be unchanged.
- (b) Would decrease and be significant.
- (c) Would increase and be significant.
- (d) Would decrease and be insignificant.
- (e) Cannot be determined from the shown output.

(36) The vertical columns of points shown in the plot of the residuals indicates that
- (a) The data violate the assumption of independence.
- (b) The data violate the assumption of constant variance.
- (c) The data violate the assumption of normality.
- (d) Some of the 9 predicted values of the model are similar.
- (e) The data have very little collinearity and so are widely dispersed along the x-axis.

(37) In order to check the assumption of normality, we should view a
- (a) Normal quantile plot of the ratings.
- (b) Normal quantile plot of the residuals.
- (c) Side-by-side boxplots of the residuals.
- (d) Plot of the actual ratings on the fitted ratings.
- (e) Scatterplot matrix of the values for each combination of *Efficiency* and *Style*.

(38) Is further analysis needed in order to determine which combinations of *Efficiency* and *Style* have significantly higher assigned ratings?
- (a) Yes, we need a Tukey-Kramer comparison of the averages for the 9 combinations.
- (b) Yes, we need a Hsu comparison of the averages for the 9 combinations.
- (c) Yes, we need a one-way analysis of variance for the 9 combinations.
- (d) No, we can obtain this result from the estimated coefficients in the fitted model.
- (e) No, we can tell from the effect tests which combinations have largest ratings.

(39) A profile plot of the two-way analysis of variance would show
(a) A curved trend indicating the need to transform the response.
(b) Exactly parallel lines.
(c) Roughly parallel lines that never cross.
(d) Roughly parallel lines with a slight crossing.
(e) Significantly crossing lines with little hint of parallel trends.

(40) Suppose that it was later learned that rather than having each of the 72 customers rate a combination, to reduce costs, only 8 customers each rated all 9 combinations. This discovery implies that the shown analysis
(a) Violates the assumption of independence.
(b) Violates the assumption of equal variance.
(c) Violates the assumption of normality.
(d) Is contaminated by confounding and a new experiment is warranted.
(e) Is adequate as offered and no changes are warranted.

(Questions 41-44) A moderate size firm tracked its total spending for labor and capital over the last 20 quarters (5 years). It has used the following 3 regressions to summarize the relationship of these expenditures to its total production. *Labor* and *Capital* are measured in thousands of dollars, and the production output is measured in millions of units produced during a quarter.

Model A: Fit of Log Production By Log Capital

RSquare	0.79
Root Mean Square Error	0.076

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.50	0.32	13.93	<.0001
Log Capital	0.45	0.06	8.19	<.0001

Model B: Fit of Log Production By Log Labor

RSquare	0.34
Root Mean Square Error	0.134

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.60	1.49	1.75	0.0971
Log Labor	0.71	0.25	2.85	0.0069

Model C: Fit of Log Production By Log Capital & Log Labor

RSquare	0.841
Root Mean Square Error	0.068
Mean of Response	7.14

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.85	0.75	3.79	0.0015
Log Labor	0.34	0.14	2.62	0.0298
Log Capital	0.42	0.05	8.41	<.0001

(41) If the company increases its capital expenditures by 2% and holds the level of labor expenditures at the current level in the next quarter, it can expect to see
(a) a 0.42% increase in production.
(b) a 0.45% increase in production.
(c) a 0.84% increase in production.
(d) a 0.90% increase in production.
(e) Production increase by about 840,000 units.

(42) If the company has the ability to increase labor by 1%, increase capital spending by 1%, or increase both labor and capital by 1/2 of 1%, these results indicate that it can expect the largest expected increase by
(a) Allocating all of the increase to labor.
(b) Allocating all of the increase to capital.
(c) Allocating half of the increase to labor, half to capital.
(d) Hiring a statistician to analyze these trends rather than going it alone.
(e) Regression does not allow the comparison of fits between models with different predictors.

(43) The addition of the log of labor to a model containing only the log of capital results in
(a) A slight, but insignificant improvement in the predictive ability of the model.
(b) A significant improvement in the predictive ability of the model.
(c) Too much collinearity to assess added benefits of spending on labor.
(d) Dependence that lowers the overall F-ratio of the fitted multiple regression.
(e) Collinearity that reduces our ability to measure accurately the impact of capital spending.

(44) It has been claimed that the elasticity of production with respect to labor given current spending is significantly less than 0.5. Based on Model C, we can conclude with regard to this claim that:
(a) The elasticity for labor is significantly less than 0.5.
(b) The elasticity for labor is less than 0.5, but not by a significant margin.
(c) The elasticity for labor is significantly more than 0.5.
(d) The elasticity for labor is more than 0.5, but not by a significant margin.
(e) Spending levels for labor have grown too much recently address this issue.