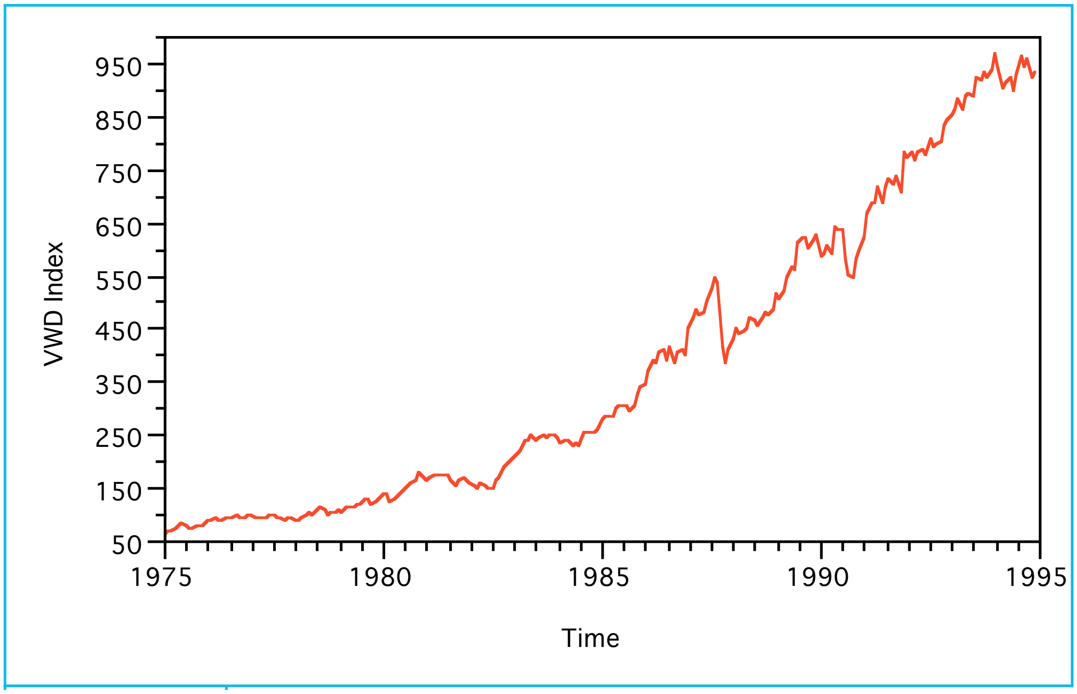
Statistics 621 Waiver Examination
August 26, 1996

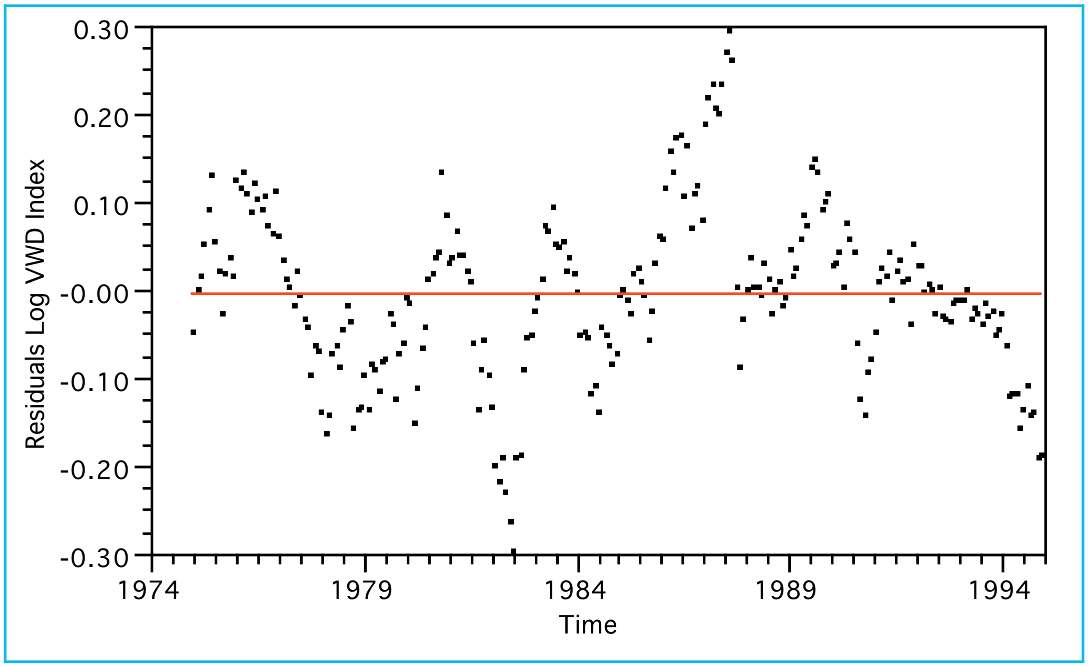
This is an open-book, open-notes exam. You have two hours to complete the exam. The computer output associated with one or more items should be considered an essential part of the question. The questions are equally weighted. The exam solutions sheets are scored electronically, so keep these issues in mind:

- *Be sure to use a #2 pencil*. If you do not have a pencil, we will supply you with one. *Do not* mark your solutions with a pen.
- Before starting, be sure to *fill in your name and student id* number.
- Choose the *one best possible answer*.
- The exam is scored by *counting the number of correct answers.*

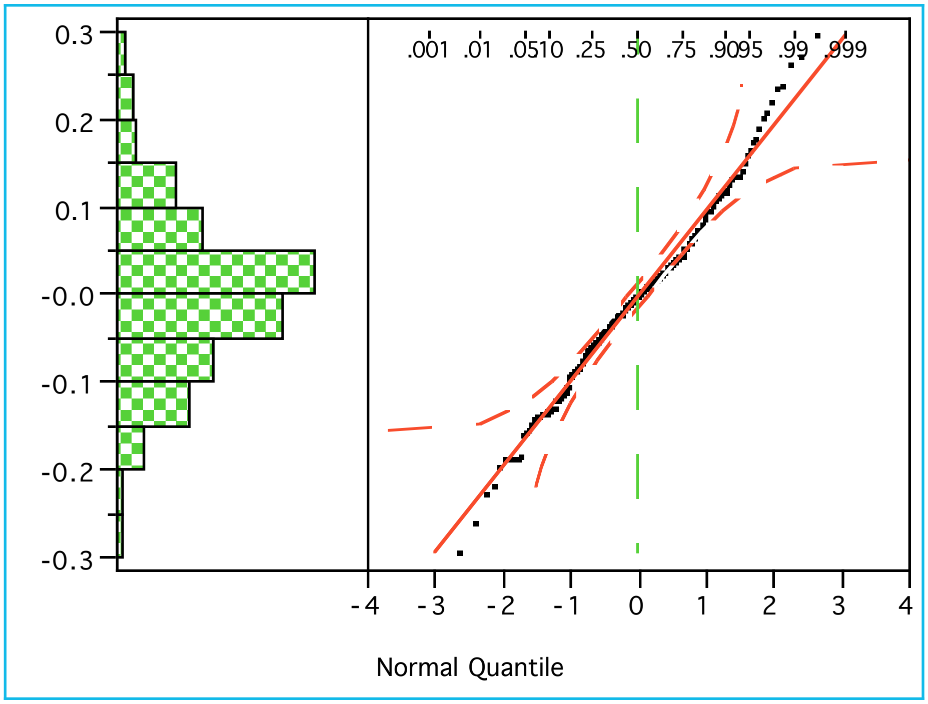
(Items 1-7) A study of recent properties of the stock market generates the following plot and associated regression output. The analysis covers the 20 years 1975-1994 and uses monthly data on an index of the performance of the stock market (labeled *VWD Index* in the output). The log used is the so-called "natural" log to base *e* (often denoted ln). The variable *Time* runs from 1975 through 1994. Residual plots follow on the next page.



```
                 Log VWD Index = -269.3 + 0.1385 Time

                 RSquare                              0.986
                 Root Mean Square Error               0.097
                 Observations                           240

    Term          Estimate   Std   Error   t  Ratio   Prob>|t|
     Intercept      -269.3           2.15   -125.2      <.0001
     Time           0.1385         0.0011    127.8      <.0001
```



Residuals Log VWD Index



(1) The fitted model implies that
(a) the stock index grew on average 13.8 units per month.
(b) the stock index grew 270 % over these years.
(c) the stock index grew at an average annual rate of about 13.8 % over these years.
(d) for each 1% increase in the length of time, the stock index grew 13.8 units.
(e) the elasticity of the stock index with respect to time is 13.8.

(2) Based on this model, a prediction interval for the index of January, 1995 (Time = 1995) would be approximately
(a) 6 ± 0.2
(b) [990, 1200]
(c) [890, 1340]
(d) [790, 1440]
(e) Need more information in order to obtain the prediction interval.

(3) Predictions of the next few months (early 1995) based on this model are likely to
(a) be too large.
(b) be too small.
(c) suffer from large error variation.
(d) be unreliable because the stock market is unpredictable.
(e) be identical to the last observed value.

(4) The residual plots indicate that the errors in the model
(a) meet the usual assumptions.
(b) are probably normally distributed.
(c) are likely autocorrelated.
(d) are probably heteroscedastic.
(e) cannot be interpreted since the data have been transformed.

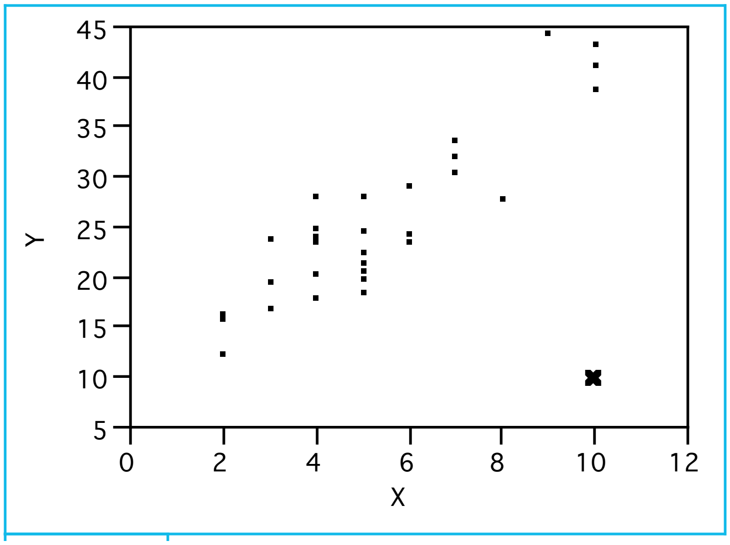
(5) The size of the R^2 summary statistic implies that the fitted model
(a) explains 98.6% of the variation in the stock index.
(b) explains 98.6% of the variation in the log of the stock index.
(c) explains significant variation.
(d) does not explain significant variation.
(e) cannot be interpreted because of autocorrelation.

(6) Assuming the needed assumptions, the same type of model was fit to a previous 20-year period, with least squares result

```
Log VWD Index = -260.625 + 0.1127 Time
```

The root MSE of this fit is comparable to that obtained for 1975-1994.
(a) These estimates are too far from those for 1975-1994. The second analysis is wrong.
(b) The coefficients of the fitted models are similar and do not differ significantly.
(c) One needs to add a dummy variable indicating the analysis period.
(d) The rate of growth of the market is significantly higher in the period 1975-1994.
(a) The two fits cannot be compared.

(7) The normal probability plot for the residuals from this model indicates that the
(a) model errors are normally distributed.
(b) model errors are not normally distributed, but the deviation is slight.
(c) model errors are not normally distributed, and thus the t-statistics are unreliable.
(d) reported p-values for the slope and intercept are too small.
(e) prediction intervals from this model are unreliable since the ± 2SD rule does not hold.

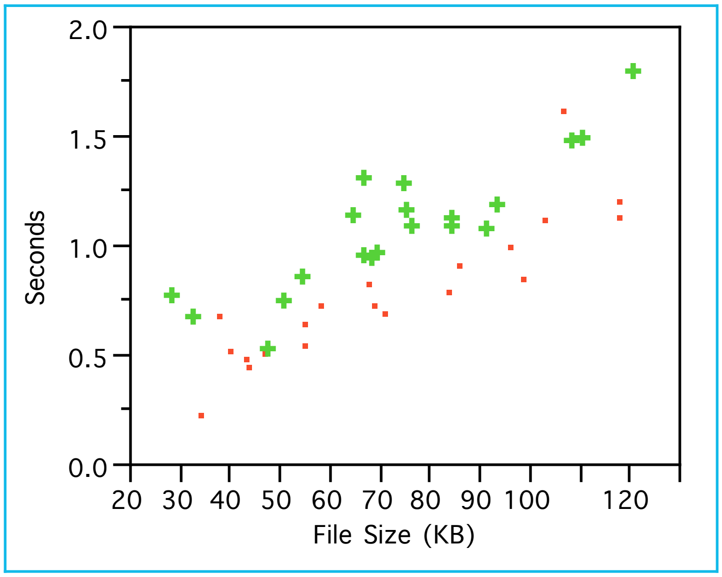


(8) Based on the data shown in the above plot, *removing* the point marked with an "x" from a least squares regression fit to this data would
(a) have no effect on the fit since this observation is leveraged, but not influential
(b) cause the slope to decrease and the intercept to increase.
(c) cause the slope to increase and the intercept to decrease.
(d) cause the R^2 and RMSE to decrease.
(e) cause the R^2 and RMSE to increase.

(Items 9- 15) A company has decided to increase its use of computers to exchange information among its worldwide offices. As part of the range of issues faced by the company, it must choose between two computer programs that actually make the data transfers. The company has run a test of the two programs labeled "a" (• in figures) and "b" (+ in figures). In the test, the company timed (in seconds) how long each program took to send 20 randomly chosen files of various sizes (measured in 1000's of bytes; a byte is one character).

The associated output is on the next page, with questions following.

Seconds By File Size (KB)



```
 Variable          Seconds  File  Size  (KB)
Seconds              1.000             0.810
File  Size  (KB)     0.810             1.000
```

Model ONE

```
                 Response:    Seconds
RSquare                                  0.817
Root Mean Square Error                   0.149
Observations                                40

 Term             Estimate  Std   Error   t  Ratio   Prob>|t|      VIF
Intercept           0.1808        0.0714      2.53     0.0157    0.000
File Size (KB)      0.0104        0.0009     11.25     <.0001    1.004
Program[a-b]       -0.1512        0.0236     -6.41     <.0001    1.004
```

Model TWO

```
                              Response:    Seconds
                          RSquare                            0.818
                          Root Mean Square Error             0.151

 Term                          Estimate  Std   Error   t  Ratio  Prob>|t|        VIF
Intercept                        0.1766        0.0729      2.42    0.0206       0.000
File Size (KB)                   0.0105        0.0009     11.08    <.0001       1.020
Program[a-b]                    -0.1068        0.0729     -1.46    0.1518       9.334
Program[a-b]*File Size   -0.0004 0.0009         -0.41    0.6848     9.315
```

(9) The fitted estimates for model ONE indicate that
(a) the two programs transfer about 100 characters per second.
(b) the startup time for each program is not significantly different from zero.
(c) the two programs take about 0.01 seconds to transfer 1000 characters (1 KB).
(d) the models need to be estimated using more than a total of 40 files.
(e) neither model can be used to predict the file transfer time of a new file.

(10) From the statistical output on the previous page, the company
(a) should use program "a" because it performs significantly better than program "b".
(b) should use program "b" because it performs significantly better than program "a".
(c) can adopt either since the observed differences are not significant.
(d) should use program "a" for transferring small files, and program "b" for large files.
(e) should use program "b" for transferring small files, and program "a" for large files.

(11) If two separate regressions were fit, one to the data for program "a" and one to the data for program "b", then
(a) the regression for program "a" would have much higher R^2 than for program "b".
(b) the regression for program "b" would have much higher R^2 than for program "a".
(c) the two models would not be reliable because the fits would ignore the dependence.
(d) the two models would yield significantly different root mean squared errors.
(e) the two fits would be virtually parallel.

(12) What residual plot would be particularly useful to check next in this example?
(a) Plots of the residuals on the fitted value for each fit.
(b) Leverage plot for the predictor `Program[a-b]*File Size`.
(c) Leverage plot for the categorical variable `Program[a-b]`.
(d) Comparison boxplots/quantile plots for the residuals grouped by program.
(e) Normal quantile plot.

(13) If one fit a constrained regression model which forces parallel slopes for the fits to the data for the two programs, then the difference in the intercepts would
(a) indicate that "a" experiences about 0.15 seconds less "overhead" when transferring a file.
(a) indicate that "a" experiences about 0.18 seconds more "overhead" when transferring a file.
(b) indicate that "a" experiences about 0.3 seconds less "overhead" when transferring a file.
(c) not be significantly different from zero.
(d) be unreliable since the value of zero is too far from the range of observed file sizes.

(14) The variance inflation factors in model ONE
(a) indicate that collinearity is a severe problem in this regression.
(b) indicate that collinearity is a moderate problem in this regression.
(c) simply confirm that the file sizes used to test each program are comparable.
(d) reveal that the correlation between each predictor and the response is near 1.
(e) are not relevant since the shown t-statistics are so large in absolute value.

(15) Compared to the fit obtained by the simple regression of *Seconds* on *File Size*, the output shows that the amount of variation explained with two separate regressions

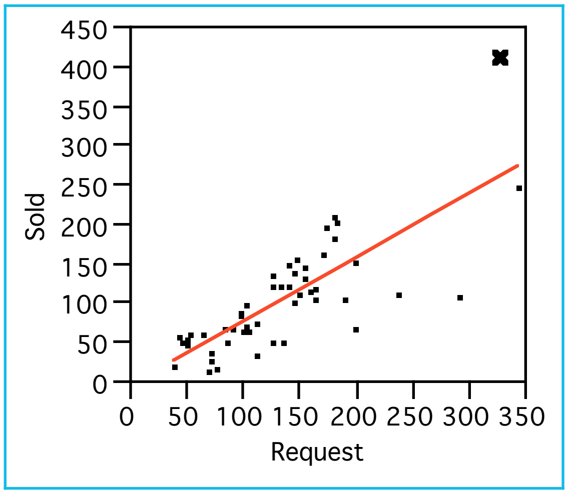
(a) is not significantly higher since the coefficients of Program[a-b] and Program[a-b]*File Size in model TWO are not significant.
(b) is significantly higher because the coefficient of Program[a-b] is significant in model ONE.
(c) is not significantly higher because the partial F-test is significant.
(d) is significantly higher because the partial F-test is significant.
(e) cannot be answered without further regression output.

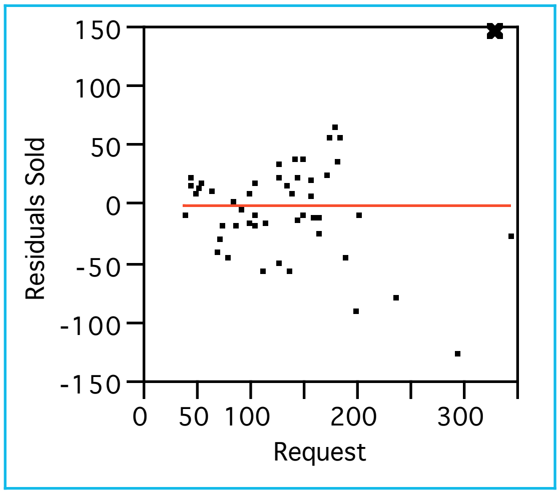
(Items 16-20) A university bookstore has decided to study the relationship between the number of books that a department requests for its courses and the number that are actually purchased. For each course that it offers, a department asks the bookstore to order a given number of copies. If the bookstore orders too many, it must return to the publisher (at the expense of the bookstore) any unsold books.

Data and questions follow. The data are the number of books requested (*Request*) and the number actually purchased (*Sold*) for a sample of 50 courses offered during one semester. The bookstore also considered the differences among orders for four departments (“a”, “b”, “c”, and “d”).

```
Model  ONE      Response:   Sold
                        RSquare                              0.617
                        Root Mean Square Error              43.276
                        Observations                            50

                 Term           Estimate   Std  Error   t  Ratio   Prob>|t|
                 Intercept        -2.022       13.779     -0.15     0.8840
                 Request           0.805        0.092      8.80     <.0001
```

```
Model  TWO      Response:   Sold

                        RSquare                              0.649
                        Root Mean Square Error              42.817

                 Term                 Estimate   Std  Error   t  Ratio   Prob>|t|
                 Intercept              -2.789        14.77     -0.19     0.8511
                 Request                 0.804         0.09      8.62     <.0001
                 Department[a-d]        17.756        10.69      1.66     0.1036
                 Department[b-d]         3.012        13.44      0.22     0.8236
                 Department[c-d]        -1.944         9.55     -0.20     0.8395

              Source        Nparm   DF   Sum  of  Squares    F  Ratio    Prob>F
              Request           1    1           136244.9      74.32    <.0001
              Department        3    3             7395.6       1.34    0.2718
```





(16) The negative intercept in model ONE

(a) implies that two books are returned to the publisher for the typical course.
(b) implies that a nonlinear relationship has been missed.
(c) implies that the bookstore needs to order two more books for the typical course.
(d) is not significantly different from zero and not a source of concern.
(e) is due to the omission of an important predictor from model ONE.

(17) The outlying value (marked “x” in the plots)

(a) is an influential outlier and should be excluded from this analysis.
(b) is indicative of a tendency to higher variation for larger orders.
(c) is leveraged, but not influential and thus can be ignored.
(d) would not be unusual if the appropriate nonlinear fit had been used.
(e) improves the ability of the model to make predictions at small request sizes.

(18) Model ONE suggests that

(a) textbook requests from departments are typically 25% larger than needed.
(b) textbook requests from departments are typically 80% larger than needed.
(c) there is no significant relationship between the number requested and number sold.
(d) most requests overestimate the number of books needed.
(e) most requests underestimate the number of books needed.

(19) Using the output for models ONE and TWO, it is clear that

(a) some departments are better than others at guessing the number of books required.
(b) without the value for department “d” the analysis is incomplete.
(c) the error variation differs substantially among the four departments.
(d) the slope for department “a” is significantly larger than those for other departments.
(e) there are no evident differences among the four departments used in the analysis.

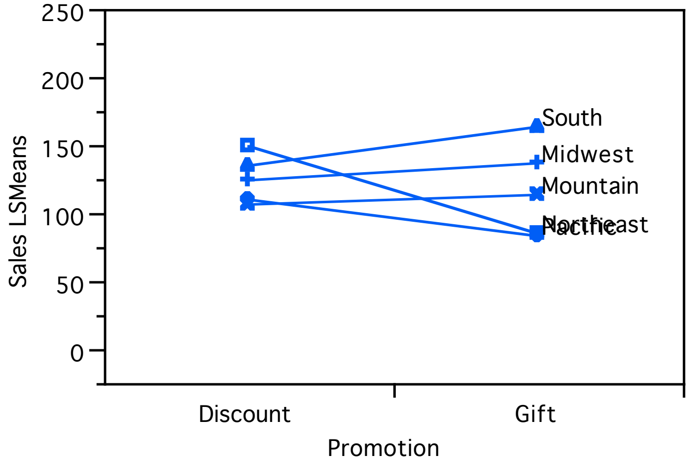
(20) This analysis would be improved by

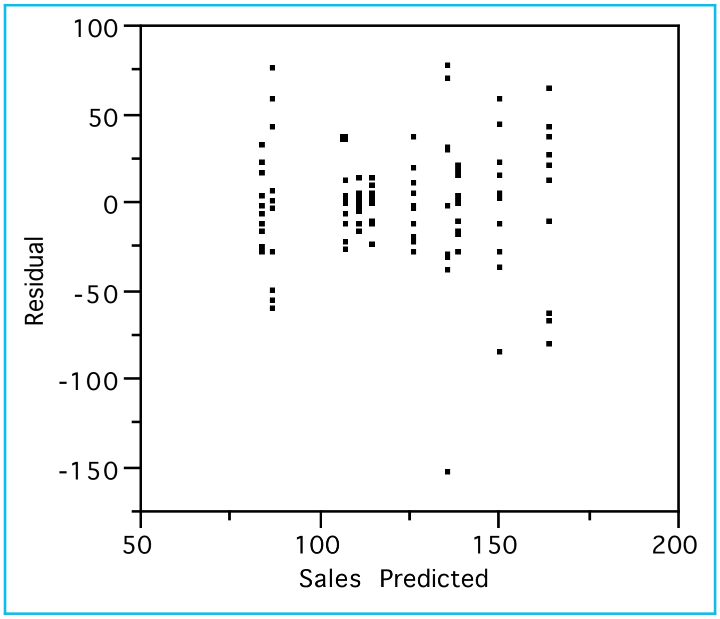
(a) removing the outlier (marked “x”) and repeating the regressions shown.
(b) adding information for other departments to increase the degrees of freedom.
(c) checking for autocorrelation.
(d) accommodating the unequal error variances via a weighted analysis.
(e) checking for collinearity.

(Items 21-26) A chain of retail stores uses two incentive programs to improve sales. One of the programs uses a “free” gift, the other is a special discount coupon. The chain has 5 national districts (northeast, south, midwest, mountains, and pacific). The results of some statistical analyses which compare the two types of incentives follow.

```
                              t-Test
                      Difference    t-Test     DF   Prob>|t|
          Estimate        8.5164     1.007     98     0.3163
          Std Error       8.4560


                         Response:    Sales
Term                                          Estimate  Std  Error  t  Ratio  Prob>|t|
Intercept                                       121.96        3.58     34.06     0.000
Region[Midwest-South]                            10.47        7.16      1.46     0.147
Region[Mountain-South]                          -11.09        7.16     -1.55     0.125
Region[Northeast-South]                          -3.28        7.16     -0.46     0.648
Region[Pacific-South]                           -24.34        7.16     -3.40     0.001
Promotion[Discount-Gift]                          4.26        3.58      1.19     0.237
Region[Midwest-South]*Promo[Discount-Gift]      -10.32        7.16     -1.44     0.153
Region[Mountain-South]*Promo[Discount-Gift]      -7.79        7.16     -1.09     0.280
Region[Northeast-South]*Promo[Discount-Gift]     27.57        7.16      3.85     0.000
Region[Pacific-South]*Promo[Discount-Gift]        9.14        7.16      1.28     0.205


                              Effect  Test
          Source           Nparm   DF  Sum  of  Squares   F  Ratio    Prob>F
         Region                4    4        32666.959     6.3707    0.0001
         Promotion             1    1         1813.238     1.4145    0.2374
         Region*Promotion      4    4        27143.436     5.2935    0.0007


                           Analysis  of  Variance
                 SourceDF  Sum  of  Squares  Mean  Square    F  Ratio
            Model       9          61623.63        6847.07     5.3412
            Error      90         115373.39        1281.93    Prob>F
            C Total    99         176997.02                    <.0001
```

Region*Promotion Profile Plot





(21) The initial two-sample t-test
(a) offers a concise summary of the main conclusions of the complete analysis.
(b) obscures important differences between the two incentive programs.
(c) reveals important differences between the two incentive programs.
(d) should be replaced by a paired t-test that compares the two incentives.
(e) comes from Stat 603 and I don't need to know that stuff.

(22) From this analysis, it would be appropriate to conclude that
(a) neither incentive program is effective.
(b) the incentive programs have approximately the same effect.
(c) the *Pacific* and *South* regions have significantly different levels of sales.
(d) the effectiveness of the incentive programs depends upon the region.
(e) the analysis is incomplete without further diagnostics, and none of these conclusions is appropriate.

(23) The partial F-test for region (Effect Test)
(a) indicates that regions obtain similar levels of sales.
(b) indicates that regions obtain significantly different levels of sales.
(c) indicates that regions obtain different levels of sales in spite of significant interaction.
(d) is equivalent to the t-test for each region.
(e) cannot be interpreted because of an evident lack of constant variation among the regions.

(24) In the *Mountain* region, the level of sales generated by the *Gift* program
(a) is not significantly different from the level of sales generated by the *Discount* program.
(b) is significantly greater than the level of sales generated by the *Discount* program.
(c) is significantly less than the level of sales generated by the *Discount* program.
(d) cannot be compared to the *Discount* program because of interaction.
(e) cannot be compared to the *Discount* program without further information.

(25) The highest level of sales has been achieved by
(a) the *Pacific* region using the *Gift* program.
(b) the *Pacific* region using the *Discount* program.

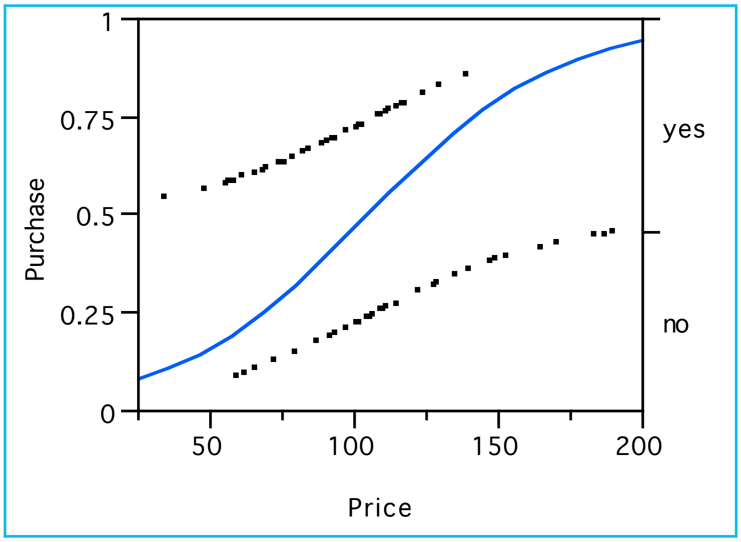
(c) the *South* region using the *Gift* program.
(d) the *South* region using the *Discount* program.
(e) some region, but the specific region cannot be identified without a table of means.

(26) The residual plot for this analysis shows
(a) that the errors for the model are not normally distributed.
(b) that the errors for the model clearly lack constant variance.
(c) that outliers have distorted the analysis and made it misleading.
(d) that the data are dependent and the analysis made unreliable.
(e) no meaningful violation of assumptions.

(Items 27- 31) A maker of men's dress shirts conducted a small marketing study to gain insights into how it should price a new shirt. Customers were sampled from among those making purchases of men's clothing at certain "premium" department stores. Each sampled customer was shown the new shirt, and asked "Would you consider purchasing such a shirt if the price is $xx?". The offered price varied. Some of the shirts are made domestically and others internationally, but the products are essentially the same. No visible differences are apparent. Nonetheless, when the shirt was shown to the customer, it was described as either "imported" or "domestic" in origin. The price and origin of manufacture were chosen randomly so that comparable prices were shown for the two groups.

Questions follow the output. The output has results for two fitted models.

Response: Purchase



```
Term            Estimate  Std  Error  ChiSquare  Prob>ChiSq
Intercept         -3.147      0.9487      11.00      0.0009
Price              0.030      0.0093      10.52      0.0012
```

```
Term                            Estimate  Std  Error  ChiSquare  Prob>ChiSq
Intercept                         -3.232      0.9614      11.30      0.0008
Price                              0.031      0.0093      10.74      0.0010
Origin[Domestic-Import]  -0.171   0.2533        0.45     0.5002
```

(27) From the logistic regression of *Purchase* on *Price*, it follows that
(a) the constant is negative and implies that this model is not reliable since probabilities cannot be negative.
(b) the fitted model implies that the probability of purchase decreases 1% for each increase of $3.15.
(c) the exponential of the slope is not significantly different from one and thus *Price* does not affect the probability of saying "yes".
(d) half of the customers in the survey who were offered the shirts at $105 said "yes".
(e) the probability of saying "yes" when the shirt costs $105 is about 0.5.

(28) How does price affect the probability of customers saying "yes" they would be interested in purchasing the shirt at the offered price?
(a) It has no effect.
(b) It has no statistically significant effect.
(c) The higher the price, the higher the probability of saying "yes".
(d) The higher the price, the lower the probability of saying "yes".
(e) For those shown imported shirts, the higher the price, the higher the probability of saying "yes".

(29) The coefficient of `Origin[Domestic-Import]`
(a) implies that the probability of "yes" is lower for those shown imported shirts.
(b) implies that the probability of "yes" is higher for those shown imported shirts.
(c) indicates that *Origin* has no significant effect on the probability of saying "yes".
(d) indicates that *Origin* only has an effect for those shown the imported shirt.
(e) is negative due to collinearity with the price of the shirt.

(30) A store used these results to determine how to adjust the price of shirts in order to obtain higher sales. Assuming that the store is now selling the domestic version for $100, approximately what price should the store charge to increase the *odds of purchase* by 20%? (The store is making the assumption that "saying yes" is the same as purchasing the shirt.)
(a) $50
(b) $80
(c) $94
(d) $106
(e) $120

(31) These results rely heavily upon the assumption that
(a) the offered prices are normally distributed.
(b) each customer is shown both a domestic and imported shirt.
(c) the effect of price is similar for both domestic and imported shirts.
(d) the errors in the fitted probabilities have constant variance.
(e) the sample size is large enough to have power to isolate a significant price effect.

(Items 32-40) In order to help her clients determine the price at which their house is likely to sell, a realtor gathered a sample of 150 purchase transactions in her area during a recent three month period. A multiple regression analysis was done to glean the important results from the data. For each house, the data used in the regression are:

Price - Sale price in 1000's of dollars (the response)
Sq Feet - Size of the house in 1000's of square feet.
Bath Rms - Number of bathrooms (as a count with $^1/_2$'s allowed).
Lot Size - Size of the lot in 1000's of square feet.
Median Inc - Median income of the surrounding "neighborhood" in 1,000's of $'s.
Fireplace - "yes" or "no", depending whether the house has a working fireplace

The initial regression output follows.

```
                    Response:    Price
                RSquare                        0.740
                Root Mean Square Error          38.2
                Mean of Response             213.852
                Observations                     150
Term               Estimate  Std  Error  t  Ratio  Prob>|t|        VIF
Intercept            43.615      10.857     4.02    <.0001        0.00
Sq Feet              21.034       3.380     6.22    <.0001        3.36
Bath Rms             12.622       5.571     2.27    0.0250        2.21
Lot Size              3.872       1.151     3.36    0.0010        1.42
Median Inc            0.522       0.083     6.27    <.0001        1.30
Fireplace[no-yes] -1.540  3.559    0.43   0.6627      1.30
       Source     DF   Sum  of  Squares  Mean  Square   F  Ratio
       Model       5          598473.2         119695   82.0276
       Error     144          210124.7           1459    Prob>F
       C Total   149          808597.8                   <.0001
```

(32) Does this regression model explain significant variation in the prices of houses?

(a) Yes, because the R^2 implies that $^3/_4$ of the variation is modeled.
(b) Yes, because most of the t-ratios are significant.
(c) Yes, because the overall F-ratio is significant.
(d) No, too many other important factors have been ignored.
(e) No, the residual variation is too large to be effective for prediction.

(33) If the variables *Price, Sq Feet, Lot Size* and *Median Inc* were expressed in dollars and square feet rather than thousands of dollars and thousands of square feet, then

(a) the R^2 and RMSE would increase.
(b) the R^2 and RMSE would remain the same.
(c) the t-ratios of *Bath Rms* and *Fireplace* would become 1000 times larger.
(d) the coefficients of *Bath Rms* and *Fireplace* would become 1000 times larger.
(e) nothing in the output would change.

(34) The coefficient of *Bath Rms* implies that

(a) converting space within a house into a bathroom adds about $12,600 to its selling price.
(b) extending the house with a bathroom adds about $12,600 to its selling price.
(c) the number of bathrooms is not an important predictor of selling price.
(d) the bathrooms in the typical house in this dataset are worth $12,600 apiece.
(e) the bathrooms in the typical house in this dataset are worth $12,600 altogether.

(35) The realtor believes a client is asking too much for his house. She would like him to lower his asking price by $10,000. Is this regression likely to be able to help her convince him that his price should be lower?

(a) Yes, the model fit explains significant overall variation.
(b) Yes, because only 26% of the total variation is not explained.
(c) No, because 26% of the total variation is not explained.
(d) No, because the model includes an insignificant predictor.
(e) No, because the estimated error standard deviation is almost $40,000.

(36) In evaluating the collection of five slopes, allowing an overall 5% chance for error, we conclude

(a) all but *Fireplace* are significantly different from zero.
(b) all but *Fireplace* and *Bath Rms* are significantly different from zero.
(c) all the coefficients are significantly different from zero since R^2 is so large.
(d) all the coefficients are significantly different from zero since the overall F is so large.
(e) *Median Inc* has the smallest impact on prices.

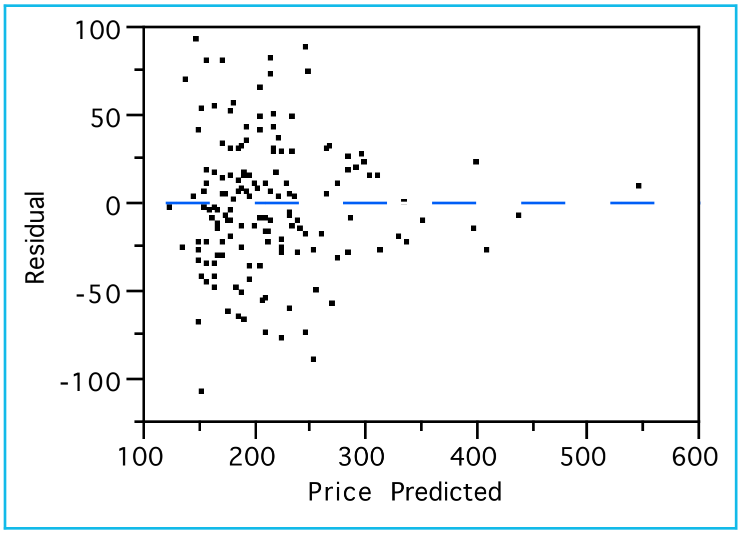
(37) A contractor is considering using this model to estimate how to cost jobs by estimating the cost he will charge as a percentage of the estimated change to the value of the house. For example, he interprets the model to predict a $21,000 increase in the selling price of the house for a 1000 square foot addition. This use of the model is

(a) questionable since we do not know corr(*Price*, *Sq Feet*).
(b) questionable since we do not know the nature of the space being added.
(c) questionable since the contractor has not obtained the associated prediction interval.
(d) questionable because the VIF indicates that collinearity has made the coefficient for *Sq Feet* insignificant.
(e) valid.

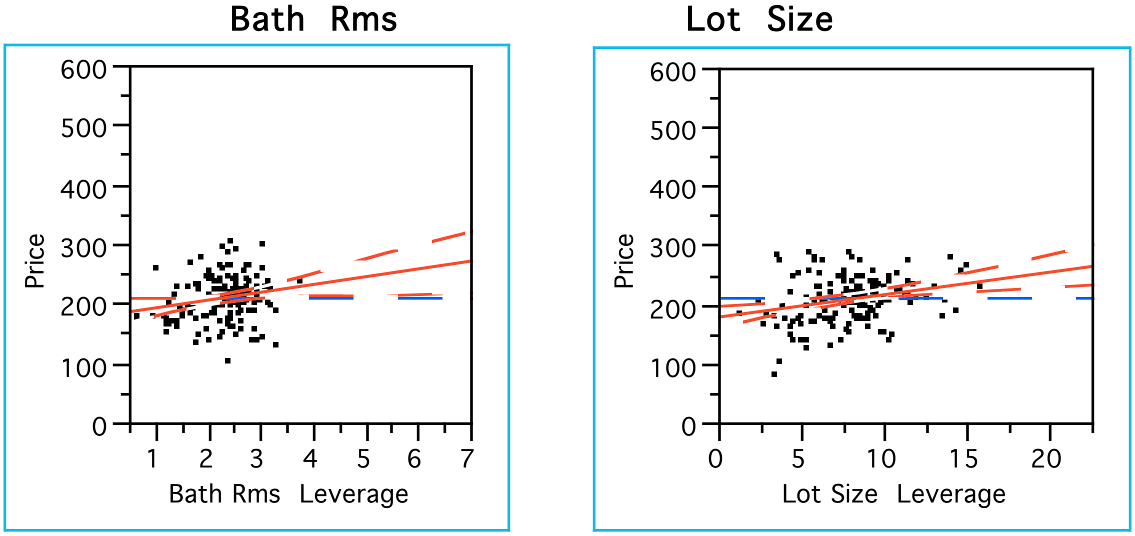
(38) Would it be possible for the two-sample t-statistic which compares the average selling price for houses with and without a fireplace to be significant?

(a) No.
(b) No, because the coefficient for fireplace is not significant.
(c) Yes, because the coefficient for fireplace is not significant.
(d) Yes, because there may be enough collinearity.
(e) Yes, because the observed difference in price from the regression is over $3,000.

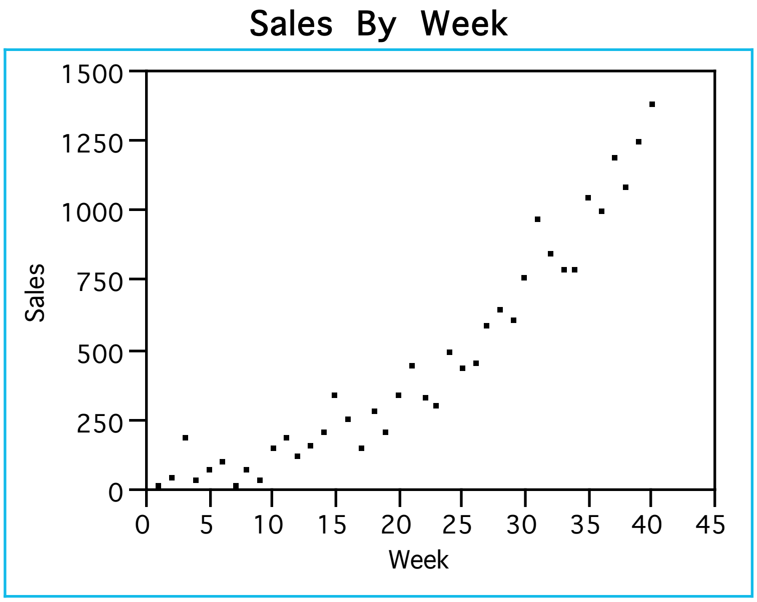
(39) The following residual plot implies that
(a) the underlying model errors lack constant variance.
(b) some houses cost a lot more than the others, but otherwise no problem is indicated.
(c) collinearity has distorted the analysis and we must drop the correlated factors.
(d) the underlying model errors are not normally distributed.
(e) the measurements are dependent and the associated statistics unreliable.



(40) The following leverage plots imply that
(a) there is some weak collinearity that reduces the significance of these factors.
(b) *Lot Size* is a less important factor in the regression than *Bath Rms.*
(c) several leveraged observations are determining the slope for *Lot Size.*
(d) neither factor is useful in the equation and both should be excluded.
(e) since *Bath Rms* takes on only discrete values, something is wrong in the code.



(Items 41-43) In the rush to satisfy impatient superiors, an analyst constructed the following simple regression equation to model how sales of a new product are growing as a function of time. Time in this problem is the week since the product was introduced.



```
FIRST                    Sales = -176.42 + 31.47 Week
                       RSquare                              0.880
                       Root Mean Square Error              137.96
                       Observations                            40
                Term          Estimate  Std  Error  t  Ratio   Prob>|t|
                Intercept      -176.42       44.46     -3.97     0.0003
                Week             31.47        1.89     16.66     <.0001

                  Durbin-Watson   Number of Obs.   AutoCorrelation
                      0.6200367               40            0.6025
```

(41) The Durbin-Watson statistic for this regression equation
(a) indicates that the fitted model suffers from autocorrelation in the underlying error process.
(b) indicates that the fitted model does not have a problem satisfying the usual assumptions.
(c) is not appropriate in simple regression models.
(d) is so small because the model does not capture the nonlinear trend in the data.
(e) indicates that the sample size is too small to build a predictive model.

(42) The analyst also built the following alternative simple regression model for this data.

```
SECOND                   Log(Sales) = 3.87332 + 0.08844 Week
                       RSquare                              0.876
                       Root Mean Square Error               0.393
                       Observations                            40

                Term          Estimate  Std  Error  t  Ratio   Prob>|t|
                Intercept       3.8733      0.1267     30.56     <.0001
                Week            0.0884      0.0054     16.42     <.0001
```

Comparing this model (labeled *SECOND*) to the model in question #41,
(a) since the R^2's of these models are comparable, either will fit the data equally well.
(b) since the slopes of both models have comparable t-ratios, the two fits are similar.
(c) since its RMSE is lower, the second model will predict better.
(d) both variables in the second model need to be expressed on a log scale.
(e) the R^2's are not comparable since the models use different dependent variables.

(43) Another analyst constructed a different model for the data.

```
THIRD              Sales = 95.0428 - 7.30591 Week + 0.94587 Week^2

                    RSquare                          0.964
                    Root Mean Square Error          76.382

               Term          Estimate  Std  Error  t  Ratio  Prob>|t|
               Intercept        95.04       38.12       2.49    0.0173
               Week             -7.31        4.29      -1.70    0.0968
               Week^2            0.95        0.10       9.33    <.0001
```

Comparing this model to the initial model offered in question #41, the third model
(a) explains significantly more variation since the R^2 is so much larger.
(b) explains significantly more variation since the t-ratio for *Week*^{2} is significant.
(c) is comparable in fit to the first model and would yield similar predictions.
(d) cannot be compared since this model has two predictors, and the first just one.
(e) is defective since the coefficient of *Week* is not significant, suggesting no time trend.

(44) Which of the following statements about collinearity is **not** true?
(a) Collinearity is present in a model which has small t-ratios but a large overall F-statistic.
(b) VIF's increase as the degree of collinearity increases.
(c) Collinearity implies that the lines in a profile plot are not parallel.
(d) The interpretation of regression coefficients is problematic in the presence of substantial collinearity.
(e) Severe collinearity does not invalidate the interpretation of regression prediction intervals.

(45) Which of the following statements about interaction is **not** true?
(a) Interaction of two predictors in a regression implies that the slope of one depends upon the level of the other predictor.
(b) Cross-product terms are used to incorporate interactions into regression models.
(c) The interaction terms in an analysis of covariance represent the differences among the slopes fitted to the separate groups.
(d) When interaction is present in a regression model, it implies that at least one predictor is redundant.
(e) Profile plots are used to judge visually interaction in an anova.