Jack Rocloni
STAT 6058
Assignment 3
2/16/22

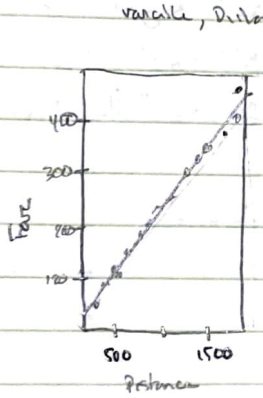
1.) Chp 3, Exercise 1 (A modern approach to Regression w/ R; pg 103)

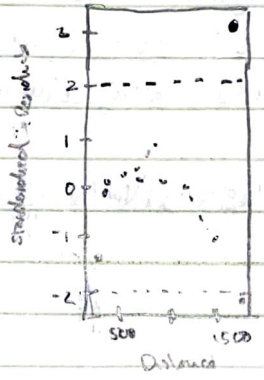
(1) The data file airfares.txt gives the one-way airfare (in US dollars) and distance (in miles) from city A to 17 other cities in the U.S. Interest centers on modeling airfare as a function of distance. The first model fit to the data was

$$\text{Fare} = \beta_0 + \beta_1 \text{Distance} + \epsilon$$

(a) Based on the output for the above model, a business analyst concluded the following:

The regression coefficient of the predictor variable, Distance is highly statistically significant and the model explains 99.4% of the variability in the y-variable, Fare. Thus the model is a highly effective model for both understanding the effects of Distance on Fare and for predicting future values given the value of the predictor variable, Distance.





Provide an detailed critique of these conclusions.

- First, I would notice that the standardized residuals do not seem to randomly fluctuate around 0. In fact they seem to be a quadratic function of X (they seem to have a parabolic shape to them). This indicates to me that we have not fit a correct model and should try improving the model by adding a quadratic term to our model.
- Second, we have two points (the two points where Distance > 1500) that can be classified as bad leverage points. I would first check to see if these two points are unusual or different in some way from the rest of the data. If we conclude that they are, I would remove them from our dataset and rerun our regression and see if the problems w/ the residuals (mentioned above) persists. If they are not different from the rest of the data, I would rerun the regression w/ the quadratic term included and see if those points are still bad leverage points.

(1) cont'd

(b) Does the ordinary straight line regression model seem to fit the data well? If not, carefully describe how the model can be improved.

- see discussion from part (a). I would try the following model

$$\text{Fare} = \beta_0 + \beta_1 \text{Distance} + \beta_2 (\text{Distance}^2) + e$$

see Stat 641 R.O.11 pg 56

2.) Explain in words why when we create confidence intervals and prediction intervals using a transformed response variable Y, we can simply take the inverse transformation of the endpoints to get a confidence or prediction interval in the original units of Y.

- Because in general $E[g(y)] \neq g(E(y))$. In other words, the Expected value of a function of Y is not, in general, equal to the function evaluated at the expected value of Y. Using a Taylor Series expansion of $g(y)$ about μ_y we obtain:

$$X = g(y) = g(\mu_y) + g'(\mu_y)(y - \mu_y) + \tfrac{1}{2} g''(\mu_y)(y - \mu_y)^2 + R$$

$$\mu_x = E[X] = E[g(y)] = g(\mu_y) + g'(\mu_y) E[y - \mu_y] + \tfrac{1}{2} g''(\mu_y) E[(y - \mu_y)^2] + E[R]$$

$$\approx g(\mu_y) + \tfrac{1}{2} g''(\mu_y) \text{Var}(Y) + E[R]$$

3.) Recall the model w/ two indicator variables from question 3 of the previous homework. Calculate the hat matrix (use software if you like; it might be faster by hand). Explain what that projection matrix does and why it makes sense, as if to someone who has taken one semester of statistics.

Recall Chp 3 Notes slide 13: hat matrix $\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = Hy$

Recall from Assign 2, (#3):

- $X = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{matrix} \} m \\ \} n-m \end{matrix}$

- $(X'X)^{-1} = \begin{bmatrix} \frac{1}{m} & 0 \\ 0 & \frac{1}{n-m} \end{bmatrix}$

- $X(X'X)^{-1} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/m & 0 \\ 0 & 1/(n-m) \end{bmatrix} = \begin{bmatrix} 1/m & 0 \\ \vdots & \vdots \\ 1/m & 0 \\ 0 & 1/(n-m) \\ \vdots & \vdots \\ 0 & 1/(n-m) \end{bmatrix}$ (m rows, n-m rows)

  $n \times 2 \quad 2 \times 2 \quad = \quad n \times 2$

  Q: Is there a name for this type of matrix

- $X(X'X)^{-1}X' = \begin{bmatrix} 1/m & 0 \\ 1/m & 0 \\ 0 & 1/(n-m) \\ 0 & 1/(n-m) \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 1/m & \cdots & 1/m & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 1/m & \cdots & 1/m & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1/(n-m) & \cdots & 1/(n-m) \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1/(n-m) & \cdots & 1/(n-m) \end{bmatrix}$ (m rows, n-m rows)

  $n \times 2 \quad 2 \times n \quad = \quad n \times n$

note: $Hy = \begin{bmatrix} 1/m & \cdots & 1/m & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 1/m & \cdots & 1/m & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1/(n-m) & \cdots & 1/(n-m) \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1/(n-m) & \cdots & 1/(n-m) \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_m \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \frac{1}{m}\sum_{i=1}^{m} y_i \\ \vdots \\ \frac{1}{m}\sum_{i=1}^{m} y_i \\ \frac{1}{n-m}\sum_{i=m+1}^{n} y_i \\ \vdots \\ \frac{1}{n-m}\sum_{i=m+1}^{n} y_i \end{bmatrix}$

$\hat{y} = Hy$, the hat matrix projects y onto the $E[y|X]$. This makes sense b/c for given values of our predictor variables, our best guess for y would be the average value of y when $X = x$

Q: If the above correct $\forall\, H$, or $\forall\, X$ where X is a design matrix?
I believe this is the case, as our regression model is defined as $E[y|X]$.

see book pg 154

4) For the simple linear regression model in the case that our assumption is met that the errors are iid w/ var σ^2:

(a) Show that the formula for the vector of the residuals $\hat{e}$ can be expressed compactly using the hat matrix:

$$(I - H)\underline{y}.$$

- $\underline{\hat{y}} = X\hat{\beta} + \hat{e} = X(X'X)^{-1}X'\underline{y} + \hat{e} = H\underline{y} + \hat{e}$

$$\Rightarrow \boxed{\hat{e} = \underline{y} - H\underline{y} = (I - H)\underline{y}}$$

see chp 2 notes slide 37 →

(b) Show that the covariance matrix of the residuals is therefore equal to

$$(I - H)\Sigma(I - H)'.$$

Where Σ is the covariance matrix of the errors, show that the covariance matrix of the residuals reduces to $(I - H)\sigma^2$. Please show that H is idempotent. That is, that $HH = H$

- $HH = X(X'X)^{-1}X'X(X'X)^{-1}X' = XI(X'X)^{-1}X' = X(X'X)^{-1}X' = H$

- $\text{cov}(\hat{e}) = \text{cov}((I - H)\underline{y})$ (by (a)) $= (I - H)\,\text{cov}(\underline{y})(I - H)'$ (we we to assume $(I-H)$ is constant in this example? (i.e. $\text{cov}(\hat{e})$ is really $\text{cov}(\hat{e} \mid X)$)

see chp 2 notes slide 37

$$= (I - H)\,\text{var}(\underline{y})(I - H)' = (I - H)\Sigma(I - H)' = (I - H)\sigma^2 I(I - H)$$

$$= \sigma^2(I - H)(I - H)' = \sigma^2(I - H)(I' - H')$$

$$= \sigma^2(II' - IH' - HI' + HH')$$

$$= \sigma^2(I - H' - H + H) \quad (HH' = H \text{ by idempotent property})$$

* NOTE: $H' = (X(X'X)^{-1}X')' = (X')'(X(X'X)^{-1})' = X((X'X)^{-1})'X' = X(X'X)^{-1}X' = H$

* recall: $(AB)' = B'A'$

$(A')' = A$

$$\boxed{\text{cov}(\hat{e}) = \sigma^2(I - H)}$$

(c) conclude that $\text{cov}(\hat{e}_i, \hat{e}_j) = -h_{ij}\sigma^2$, $i \neq j$

$\forall i, j: i \neq j$, $I_{ij} = 0$ $\therefore$ $\text{cov}(\hat{e}_i, \hat{e}_j) = \sigma^2(I_{ij} - h_{ij}) =$

$$= \sigma^2(0 - h_{ij}) = -h_{ij}\sigma^2$$

$$\boxed{\text{cov}(\hat{e}_i, \hat{e}_j) = -h_{ij}\sigma^2, \; i \neq j}$$

5) For the simple linear regression model, show that the hat matrix H has the following properties:

(a) H is symmetric ($A^T = A$)

$$H' = (X(X'X)^{-1}X')' = (X')'((X(X'X)^{-1})' = X((X'X)^{-1})'X' = X(X'X)^{-1}X' = H.$$

[NOTE: we are using the following properties of matrices in the above:
- $(AB)' = B'A'$
- $(A^{-1})' = (A')^{-1}$]

(b) $0 \le h_{ii} \le 1$, where h_{ii} is the i^{th} diagonal entry of the hat matrix (Hint: First show that $h_{ii} \ge h_{ii}^2$ and note $h_{ii} = \sum_{j=1}^{n} h_{ij}^2$)

- $h_{ii} = \sum_{j=1}^{n} h_{ij}^2$.

$$h_{ii} = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2$$

[note $h_{ij}^2 \ge 0 \Rightarrow \sum h_{ij}^2 > 0$]

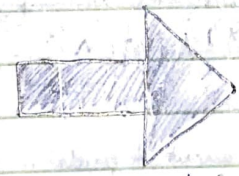
$$h_{ii} = h_{ii}^2 + \sum_{j \neq i} h_{jj}^2 \ge h_{ii}^2$$

$$h_{ii} \ge h_{ii}^2$$

- $h_{ii} \ge h_{ii}^2 \Rightarrow h_{ii} - h_{ii}^2 > 0 \iff h_{ii}(1 - h_{ii}) \ge 0 \Rightarrow \underline{0 \le h_{ii} \le 1}$

(c) The off-diagonals of the hat matrix are found by the formula

$$h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{\text{SSX}}$$



5.) (c) We know from previous work: $(X'X)^{-1} = \frac{1}{SXX}\begin{bmatrix} \frac{SXX}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$

$$(X'X)^{-1}X' = \frac{1}{SXX}\begin{bmatrix} \frac{SXX}{n} + \bar{x}^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix}$$

$$= \frac{1}{SXX}\begin{bmatrix} \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_1 & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_j & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_n \\ -\bar{x} + x_1 & \cdots & -\bar{x} + x_j & \cdots & -\bar{x} + x_n \end{bmatrix}$$

$$X(X'X)^{-1}X' = \frac{1}{SXX}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}\begin{bmatrix} \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_1 & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_j & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_n \\ -\bar{x} + x_1 & \cdots & -\bar{x} + x_j & \cdots & -\bar{x} + x_n \end{bmatrix}$$

$$= \frac{1}{SXX}\begin{bmatrix} \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_1 - \bar{x}x_1 + x_1^2 & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_j - \bar{x}x_1 + x_1x_j & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_n - \bar{x}x_1 + x_1x_n \\ \vdots & & \vdots & & \vdots \\ \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_1 - \bar{x}x_i + x_ix_1 & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_j - \bar{x}x_i + x_ix_j & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_n - \bar{x}x_i + x_ix_n \\ \vdots & & \vdots & & \vdots \\ \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_1 - \bar{x}x_n + x_nx_1 & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_j - \bar{x}x_n + x_nx_j & \cdots & \frac{SXX}{n} + \bar{x}^2 - \bar{x}x_n - \bar{x}x_n + x_n^2 \end{bmatrix}$$

$$h_{ij} = \frac{1}{SXX}\left(\frac{SXX}{n} + \bar{x}^2 - \bar{x}x_j - \bar{x}x_i + x_ix_j\right) = \frac{1}{n} + \frac{1}{SXX}\left(\bar{x}^2 - \bar{x}x_j - \bar{x}x_i + x_ix_j\right)$$

$$= \frac{1}{n} + \frac{1}{SXX}\left((x_i - \bar{x})(x_j - \bar{x})\right) = \boxed{\frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{SXX} = h_{ij}}$$

2. (a) Finally the text states that "Plots of a small amount of correlation present in the standard residuals, even if the errors are independent" Comment on when the covariances of the residuals are close to zero, for a fixed sample size. Why does it make sense that the covariances are close to zero in those situations?

$$\text{cov}(e_i, e_j) = -h_{ij}\sigma^2 = -\frac{\sigma^2}{n} - \frac{\sigma^2(x_i - \bar{x})(x_j - \bar{x})}{SXX}$$

$$= -\frac{\sigma^2}{n} - \frac{\sigma^2(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

is x_i is also $\text{var}(e)$ yes. $\text{var}(e|x=i) = \text{var}(y|x=x)$ ↕ true

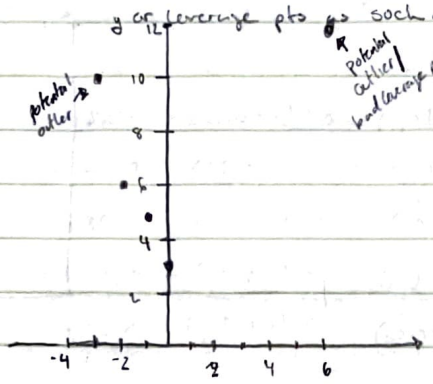
- For a fixed n the $\text{cov}(e_i, e_j)$ will be small when $\text{var}(y|x_i = x) = \sigma^2$ is small.

or if x_i is close to $\bar{x}$ or if x_j is close to $\bar{x}$.

$\sum (x_i - \bar{x})^2$

6.) Under the simple linear regression model, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ & the following data are recorded: $\underline{x} = [-3, -2, -1, 0, 6]$, $\underline{y} = [10, 6, 5, 3, 12]$. Show your work and do the following calculations by hand.

(a) First create a quick sketch of the scatter plot of the data. Label any potential outliers or leverage pts as such. Write out your design matrix.



$$X = \begin{bmatrix} 1 & -3 \\ 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 6 \end{bmatrix}$$

(b) using $\hat{y} = 7.2 + 0.5x$, calculate $\hat{\varepsilon}$

$\hat{\underline{\varepsilon}}' = [4.3, -0.2, -1.7, -4.2, 1.8]$

(c) Compute the leverage for each observation. Use the rule $h_{ii} > 4/n$ to identify potential leverage pts. Are there any points of high leverage "good" or "bad".

[Recall (Chp 3 notes Slide 22): • a bad leverage pt is a leverage pt. whose standardized residual falls outside the interval $[-2, 2]$
• a good leverage pt is a leverage pt. whose standardized residual falls w/in the interval $[-2, 2]$]

• leverage: $h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SXX}$

• leverages: $[0.38, 0.23, 0.12, 0.23, 0.82]$

• cutoff for high leverage: $\frac{4}{5} = 0.80$

• $(x, y) = (6, 12)$ is a high leverage pt.

• standardized residual for $(6, 12)$: $r_i = \frac{\hat{\varepsilon}_i}{s\sqrt{1 - h_{ii}}}$

$s = \sqrt{\frac{1}{3}\sum_{i=1}^{5} (e_i - \bar{e})^2} = 3.755$

I know this is a bad question, but how do we get s. Is that not the correct way?
- Nevermind. we wanted s of the residuals, I was calculating s_y before.

$r_i = \frac{1.8}{3.755\sqrt{1 - 0.82}} = 1.1192.2$

• yes, our 5th observation seems to be a "bad" leverage pt."

6) (cont'd)

(d) compute the variances of the residuals (assume the variance of the errors simply to be σ^2)

$$\text{var}(\hat{e}_i) = \sigma^2 (1 - h_{ii})$$

[note: we know from (5) that for the simple linear regression model

$$h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{SXX} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \quad]$$

- for our case: $h_{11} = \frac{1}{5} + \frac{(-3-0)^2}{50}$ $\quad \text{var}(\hat{e}_1) = \sigma^2(1 - 0.38) = 0.62\sigma^2$

$h_{22} = \frac{1}{5} + \frac{(-2-0)^2}{50}$ $\Rightarrow$ $\text{var}(\hat{e}_2) = \sigma^2(1 - 0.28) = 0.72\sigma^2$

$h_{33} = \frac{1}{5} + \frac{(-1-0)^2}{50}$ $\quad \text{var}(\hat{e}_3) = \sigma^2(1 - 0.22) = 0.78\sigma^2$

$h_{44} = \frac{1}{5} + \frac{(0-0)^2}{50}$ $\quad \text{var}(\hat{e}_4) = \sigma^2(1 - 0.20) = 0.80\sigma^2$

$h_{55} = \frac{1}{5} + \frac{(6-0)^2}{50}$ $\quad \text{var}(\hat{e}_5) = \sigma^2(1 - 0.92) = 0.08\sigma^2$

(e) compute the standardized residuals:

(recall H.D. pg 20): $r_i = \frac{\hat{e}_i}{s\sqrt{1 - h_{ii}}}$

$$\underline{r}' = \left[1.454,\ -0.063,\ -0.513,\ -1.251,\ 1.695\right]$$

- This conflicts a bit w/ (b) b/c pt 1 has the highest residual (calculated in (6)), but point 5 has the largest standardized residual. This occurs b/c pt 5 has a lower variance than pt 1.

(f) comment on why the pt w/ the largest leverage in this dataset had the smallest variance.

- Variance is given by $\sigma^2(1 - h_{ii})$ where $0 \le h_{ii} \le 1$ is our leverage.

Thus, pts w/ higher leverage will tend to have lower variances.

7) When Y has both mean & variance equal to μ, we showed in the notes that the appropriate transformation of Y for stabilizing the variance is the square root of the transformation. Now if that Y has mean μ and variance μ^2. Show the appropriate transformation of Y for stabilizing variance is the log transformation. (Q7 Chp 3, pg 112 in textbook).

[* see chp 3 notes pgs 45;46]

- Let $f(y) = \ln(y)$, $E[Y] = \mu$, $\text{var}[Y] = \mu^2$
- $\text{var}(f(y)) = (f'(E[Y]))^2 \text{var}(Y)$
- $f'(\mu) = \frac{1}{\mu} \Rightarrow (f'(\mu))^2 = \frac{1}{\mu^2}$

$\Rightarrow \text{var}(f(y)) = \frac{1}{\mu^2} \cdot \mu^2 = 1$ ← constant wrt μ.

8.) Download the dataset called company.csv from canvas. The data contains a systematic sample (every 10th company, we'll take this as randomly selected) for the forbes 500 list. The variables of interest are Sales & Assets of the companies. As w/ many financial datasets, many of these variables are skewed. Your job is to choose appropriate power transformations s.t. the relationship between Assets (response var) & sales (EV) are approximately linear.

(a) The scatter plot seems to suggest a log transformation might be appropriate.

- First weakness of the simple linear regression model we fit is that the residuals aren't normally distributed. This can be seen clearly in the Q-Q plot of the standardized residuals.
- We also have a few outliers in the data, specifically the 16th, 48th observations have standardized residuals > 2.
- Additionally we have 5 observations that seem to be highly influential (16, 40, 43, 48, 54) when looking at cooks vs cooks Distances. All these points are above the cutoff $4/(n-2)$ to be considered influential.
- Finally, we see that we have non-constant error variance.

see stat4140.97.47 for box cox code

(b) Choose an appropriate transformation for Sales. Explain how you made your choice. Include plots if applicable

- Taking log(sales) seems to be an appropriate transformation. After a box cox transformation and find the appropriate power to be -0.068. However a 95% CI for the appropriate power was $(-0.235, 0.095)$. Since the CI contained 0, used the log-transform. Furthermore, a shapiro-wilks test on the log transformed sales data gives a p-value > 0.05 indicating we cannot conclude the log transformed data does not have a normal distribution.

8.) (c) Choose an appropriate transformation for assets, and again explain how you made your choice. Because an inverse response plot in this example is messy, you can (1) just fit a regression model of Assets vs. the transformed model of sales, then (2) pass the fitted model into the powerTransform function. No plots required.

- I first did the same thing for Assets that I did for sales, and again found that the log transform worked. CI for λ_{max} $(-0.203, 0.115)$ w/ $\hat{\lambda}_{max} = -0.043$
- Fitting a regression model of Assets vs log(sales), passing through the fitted model into powerTransform we get $\hat{\lambda} = -0.01658166$. A 95% CI for lambda is given by $(-0.1658, 0.1357)$. Thus, the log transformation seems appropriate (b/c our CI contains 0).

see stackOH! 4.0.9 pg 42 for CI

ask how to get CI for λ from powerTransform function in R ↓ use summary(powerTransform)

is the process normalizing the residuals? and the above is normalizing y independent of x? both work b/c when normalizing residuals, we get the fit we want, & when normalizing y independent of x, x & y are plot of the same location scale family so we get a linear fit? (Is there a way to see if we should normalize just one or the other, or both simultaneously)

(d) Call the model w/ both variables transformed model 2. Create diagnostic plots for this model & discuss any weaknesses of this model.

- After the log-log transformation, sales & assets seem to be linearly related.
- The diagnostic plots are not perfect, but have definitely improved over the original model w/ the untransformed variables. The residuals vs fitted plot is less concentrated & seem to have constant variance, the Normal Q-Q plot shows our residuals might have lighter tails than if they were normally distributed, but they are closer to normal than they were before we transformed the variables.
- When looking at the plot of sales vs cooks distance, we still seem to have 2 influential pts (16, 40)
- Looking at the plot of fitted values vs $(\text{standardized residuals})^{1/2}$ we seem to have constant error variance.
- Overall model 2 seems to be an improvement over model 1 wrt meeting the assumptions of a regression model.

(e) Compare models 1 & model 2. Which is preferable?

As discussed in (d) model 2 better meets the assumptions of a linear regression model. Thus, I prefer model 2.

8.) (f) Using the model $\log(\text{Assets}) = \beta_0 + \beta_1(\log(\text{sales}))$, interpret the slope in the context of this problem.

- Model 2 gives us $\hat{\beta}_1 = 0.5870$. In the context of this model, that means

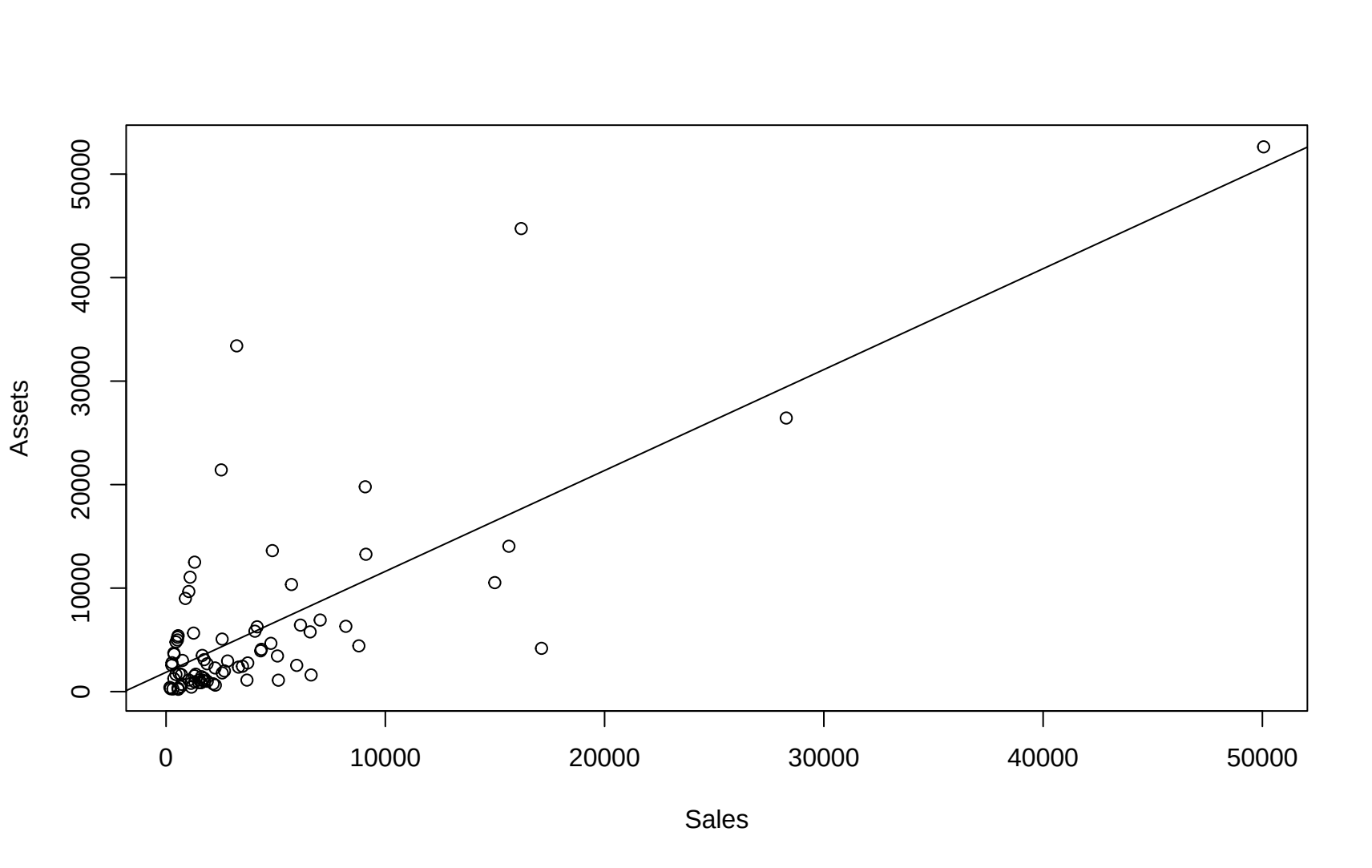
For a 1 percentage point increase in the sales of a company on the Forbes 500 list, we expect, on average, to see a corresponding 0.5870 percentage point increase in the assets of a company in the Forbs 500 list

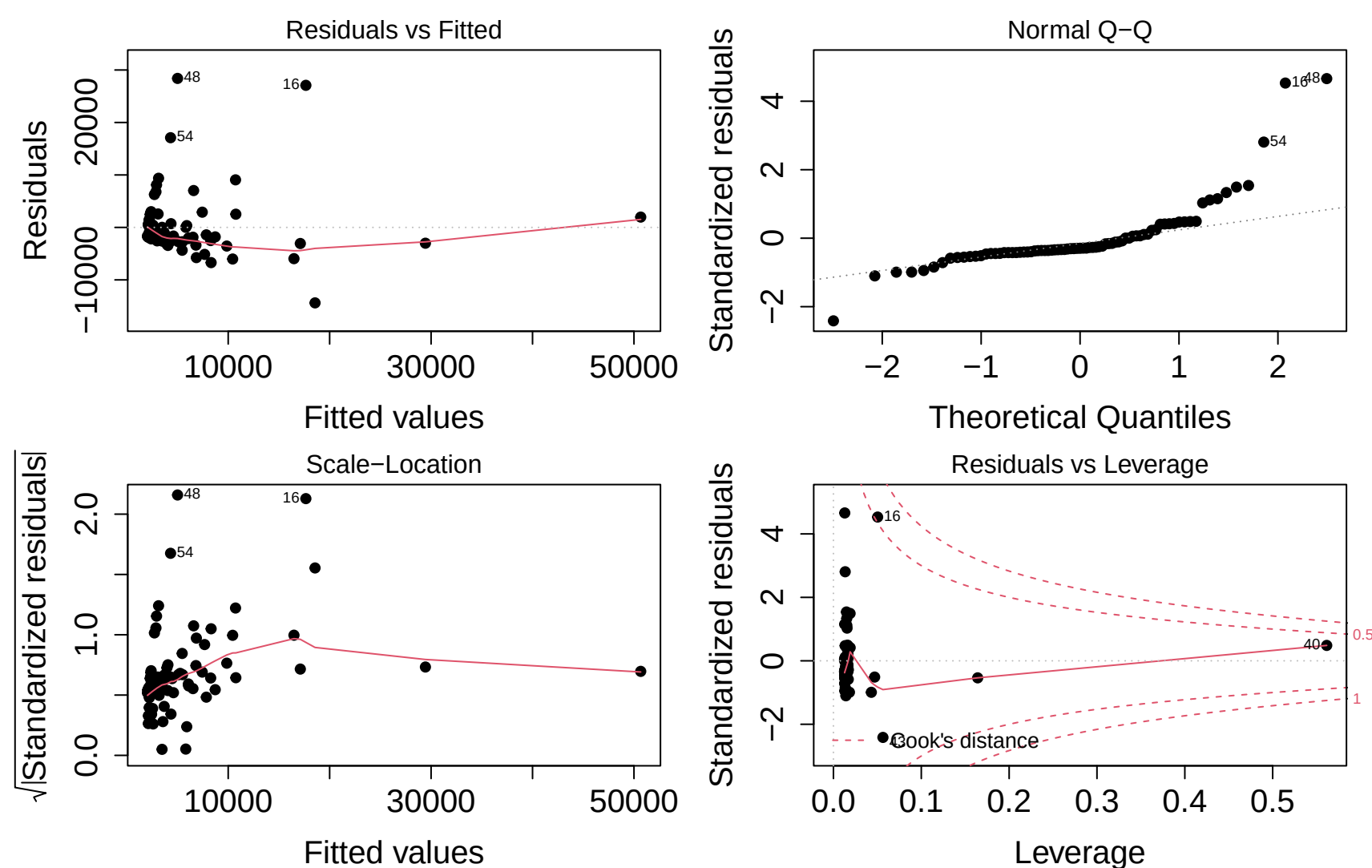
(g) Again, using the model: $\log(\text{Assets}) = \beta_0 + \beta_1 \log(\text{sales})$, find a 95% CI for the average assets of a company w/ 6,571 million in sales, as HP did. Interpret your confidence interval in context.

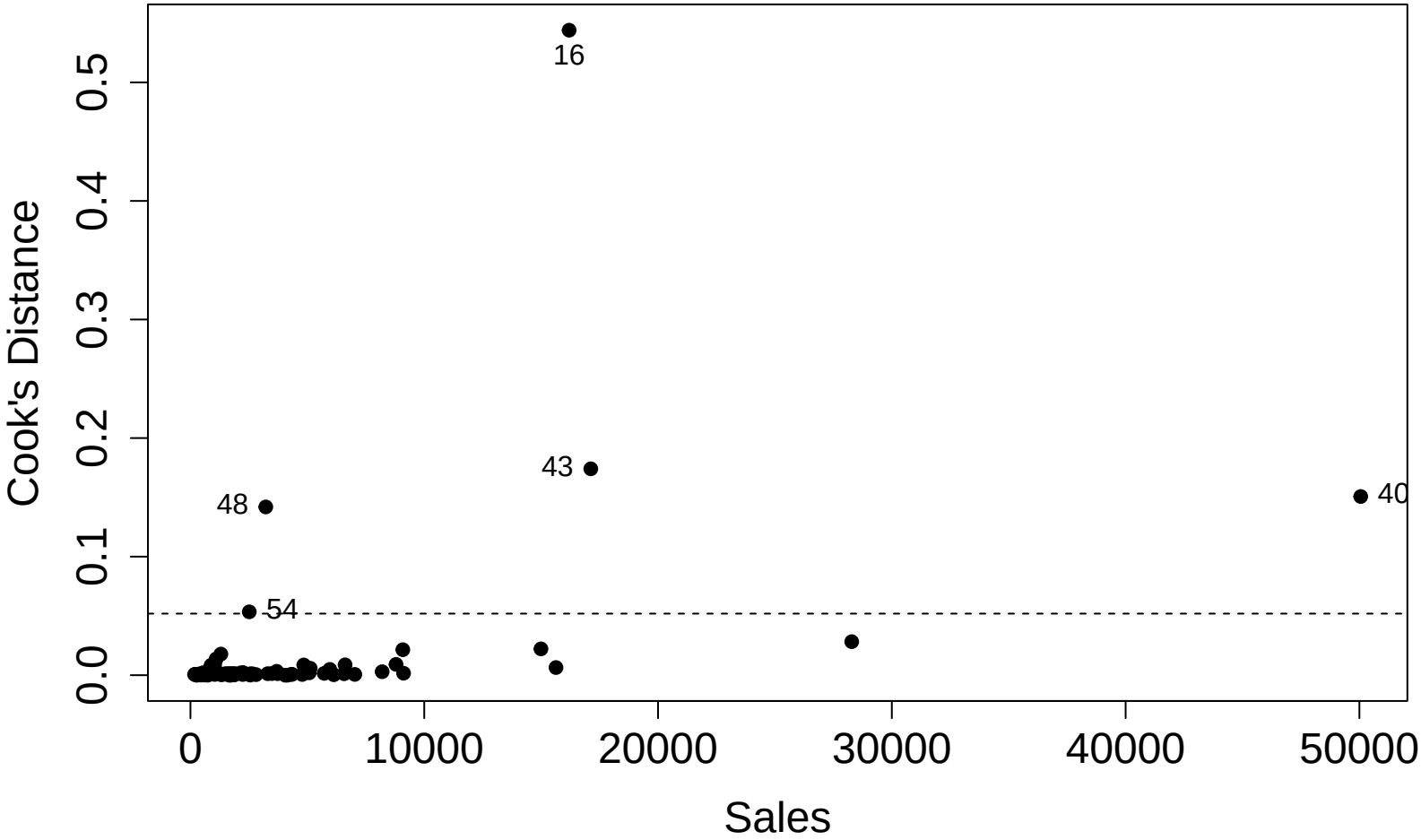
Q: Do we just back transform the mean, or do we back transform the whole pt. est. ± $t^* \cdot s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{SXX}}$?
See Chp 2 pg 65, Chp 3 pg 80.

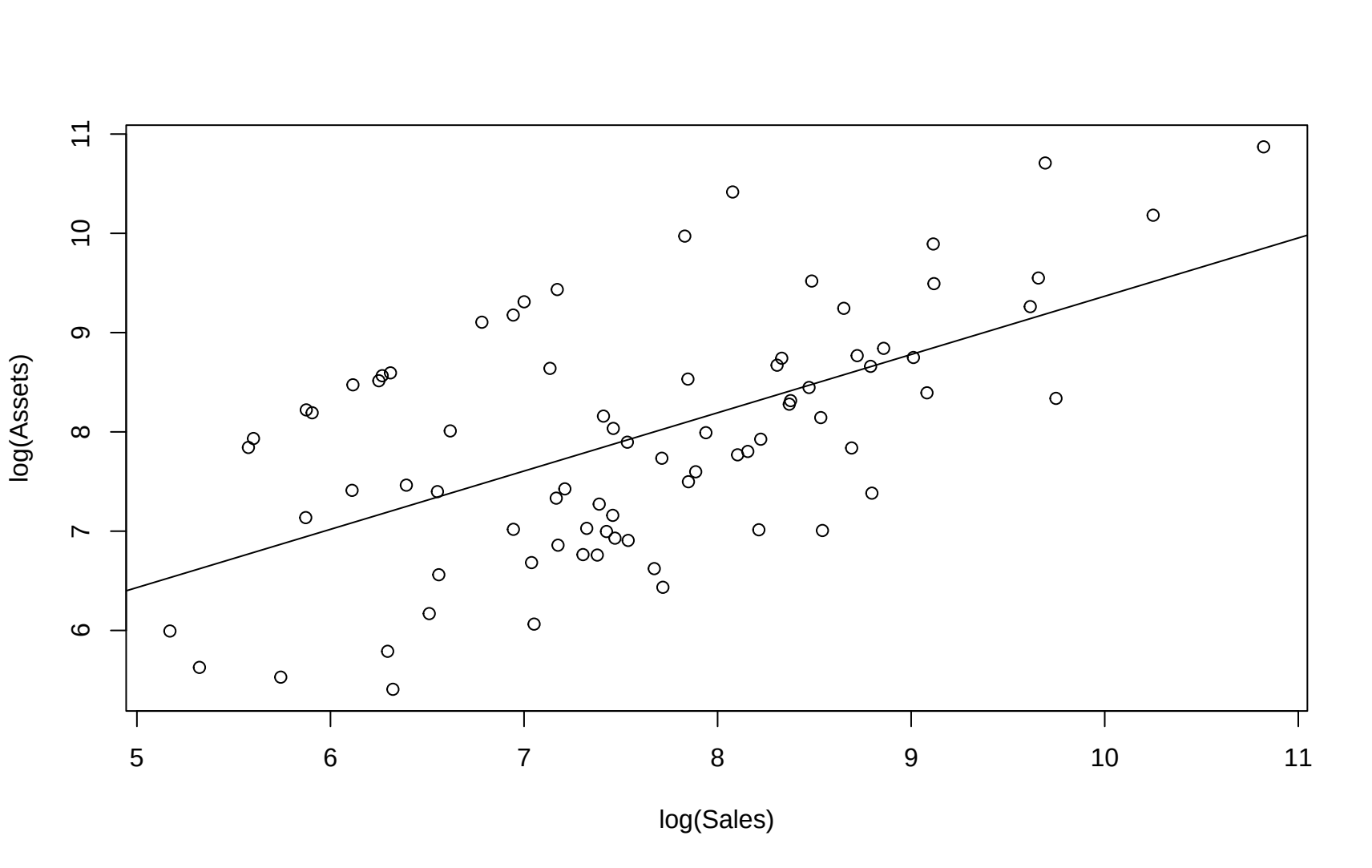
*[NOTE: we transformed the data so we must apply a back transformation.
- Ask about calculation of MSE. relation to Residual standard error in R.]

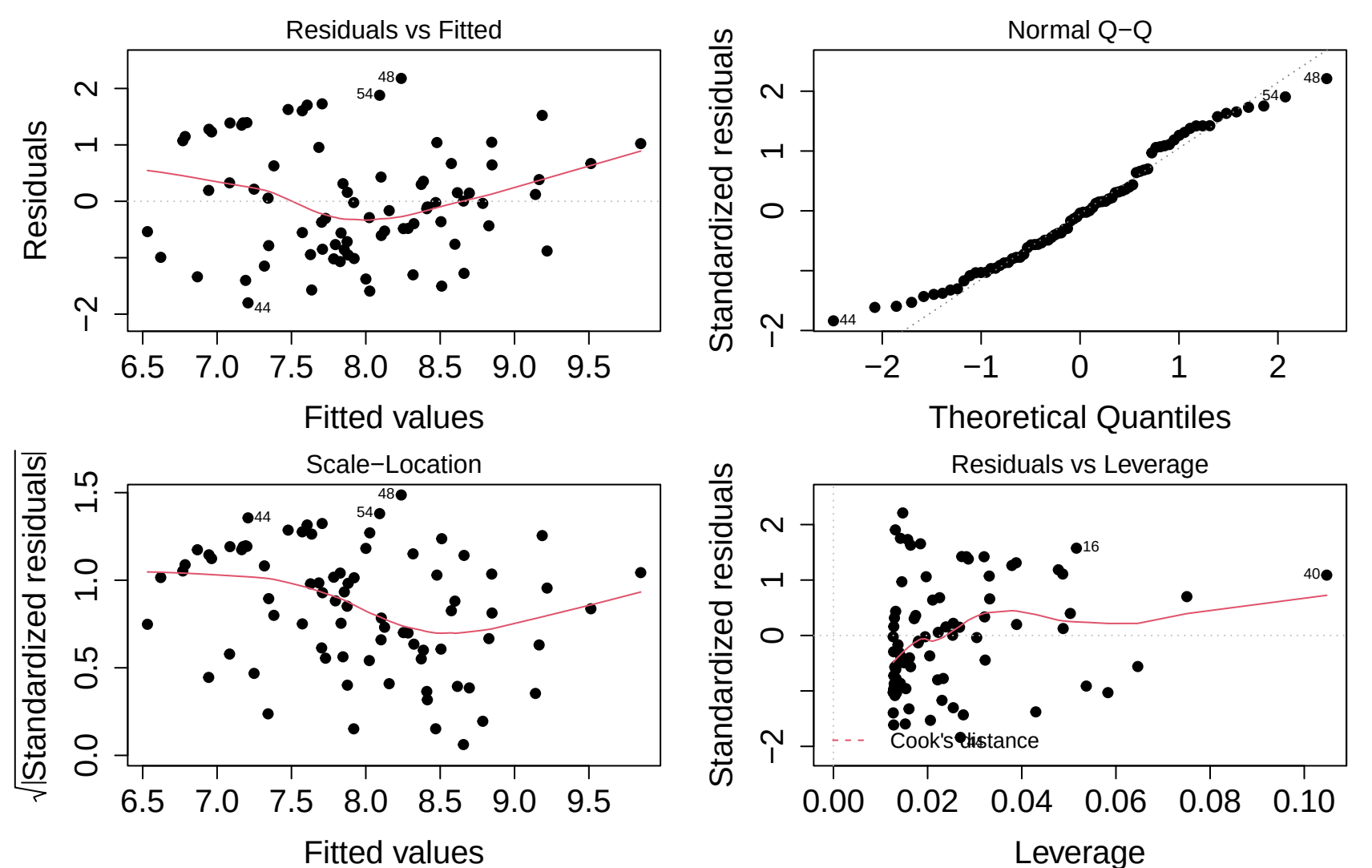
I am 95% confident that a company on the Forbes 500 list w/ 6,571 million in sales would on average have assets between 6,870 and 12,889 million

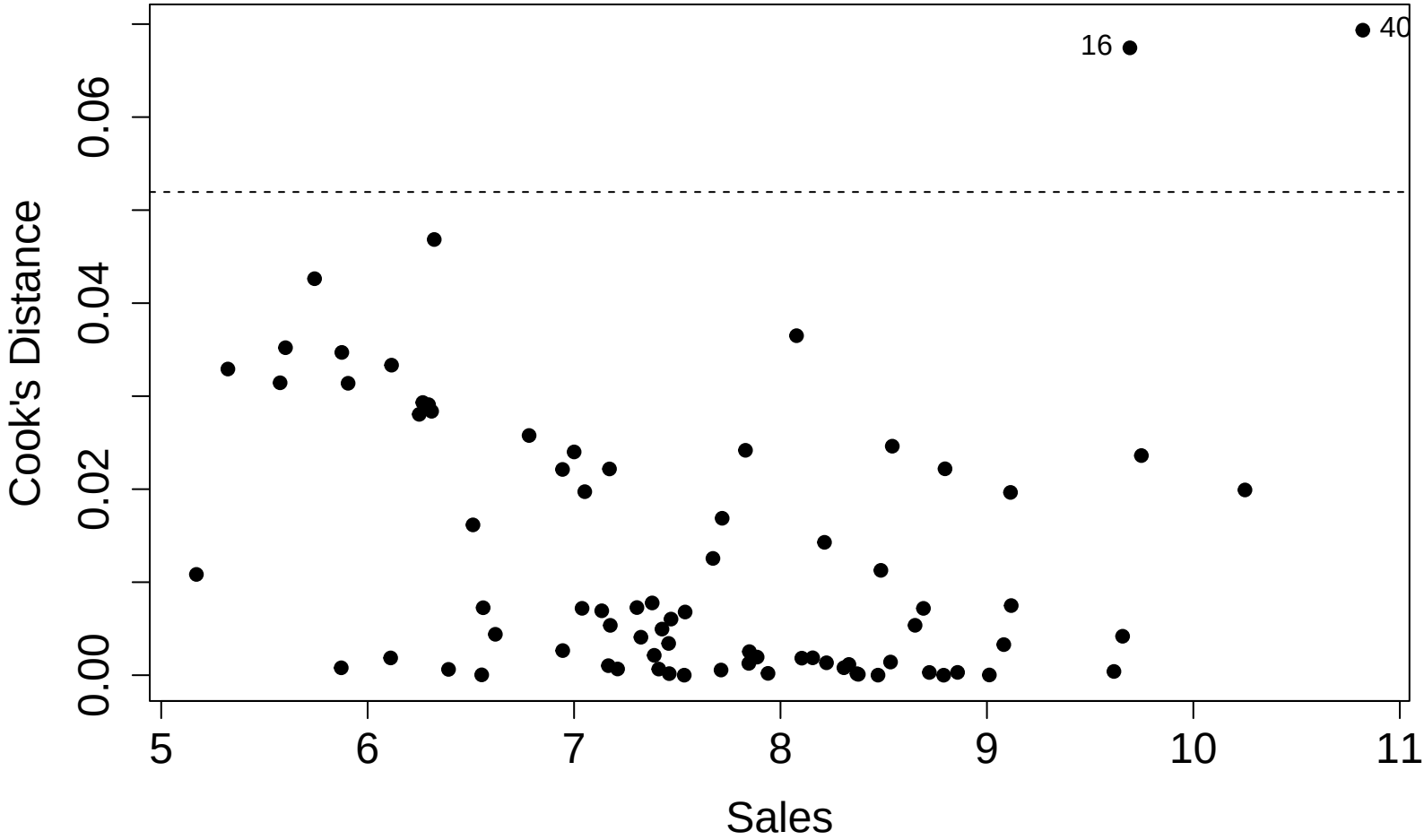

Assets
0
10000
20000
30000
40000
50000
Sales
0
10000
20000
30000
40000
50000


Residuals vs Fitted
Residuals
Fitted values
48
16
54
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Leverage
Cook's distance
40


Cook's Distance
0.0
0.1
0.2
0.3
0.4
0.5
16
43
48
40
54
0
10000
20000
30000
40000
50000
Sales


log(Assets)
log(Sales)
5
6
7
8
9
10
11


Residuals vs Fitted
Residuals
Fitted values
48
54
44
Normal Q−Q
Standardized residuals
Theoretical Quantiles
48
54
44
Scale−Location
√|Standardized residuals|
Fitted values
44
54
48
Residuals vs Leverage
Standardized residuals
Leverage
16
40
Cook's distance


Cook's Distance
0.00
0.02
0.04
0.06
16
40
5
6
7
8
9
10
11
Sales