STAT 2131
Orly Olbum

Homework 2

20

1. #1.22 from textbook

Y = hardness of items
X = elapsed time since termination
4 treatment times studied, 16 units

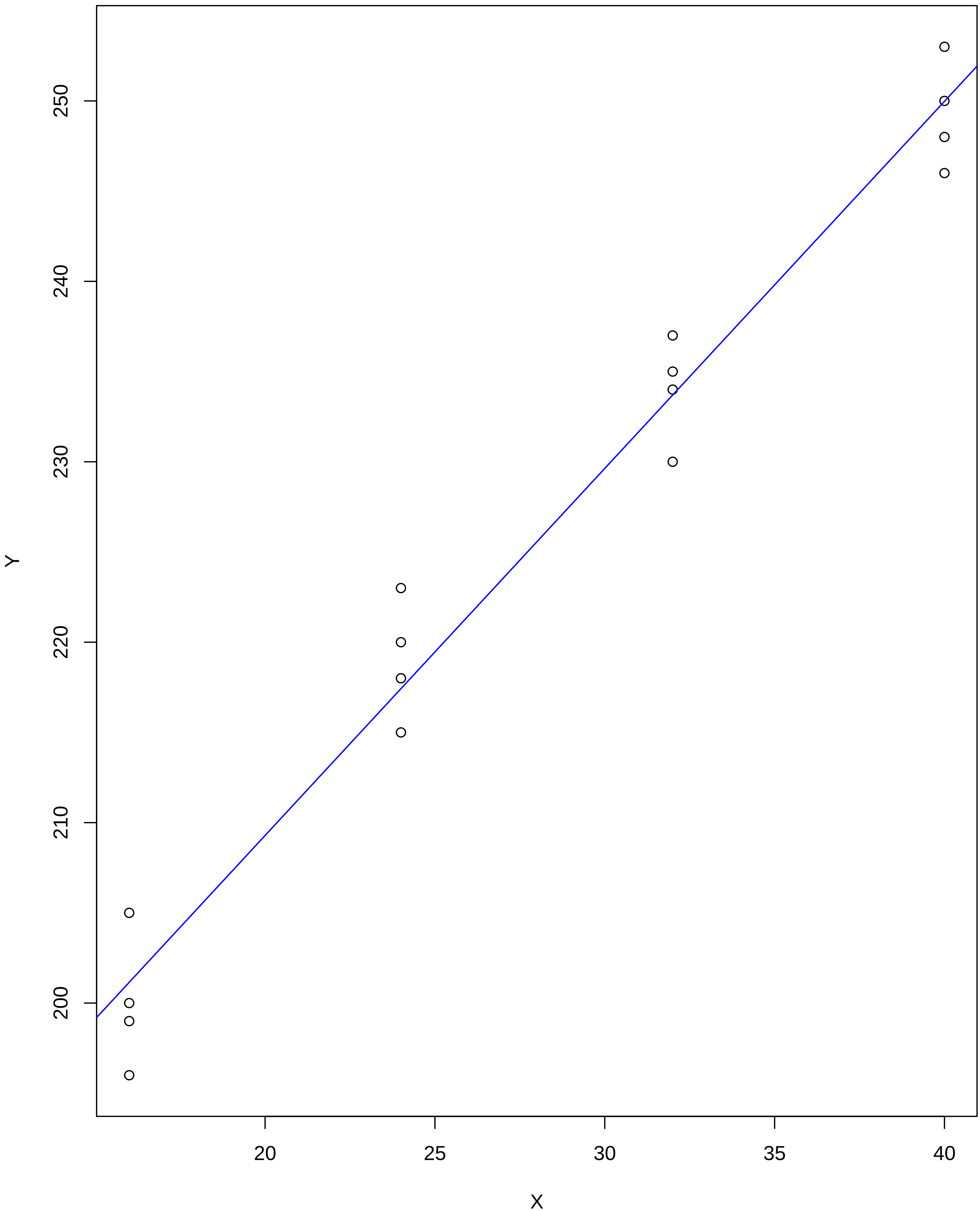
ⓐ obtain estimated regression function and plot data → is it a good fit?

→ using R software (lm(Y~X, data)) we find the following estimated regression line: $\hat{Y} = 168.60 + 2.03438\hat{X}$

with an adjusted R^2 of .9712, the regression line is a good fit for our data. See attached/next page for a plot of the data and the regression line. However, it may be better plotted as a time series

ⓑ When $X=40$, $\hat{Y} = 168.60 + 2.03438(40) = 249.9752$

ⓒ Find change in hardness when X increases by 1 hour → this is equal to $\hat{\beta_1}$, or 2.03438



200
210
220
230
240
250
Y
20
25
30
35
40
X

20

2. #2.16 from text book

ⓐ Find 98% CI for mean Y when $X=30$, interpret.

with $\alpha=.02$ $z_\alpha = 2.06$

$s = 9.0507$, $n=16$ $\hat{Y}(30) = 229.6312$

CI: $\hat{Y}(30) \pm t_{\alpha/2,14}\left(\frac{s}{\sqrt{n}}\right) \to df = 16-2$ → predictors $\Rightarrow 2.624$

$= (227.4569, 231.8056)$

Our interval means that with 98% confidence, the real value of Y for $X=30$ falls in this range.

ⓑ 98% PI for mean Y when $X=30$

$\hat{Y}(30) \pm t_{\alpha/2,14} \cdot var(prediction)$; $var(pred) = MSE + s^2$ of $\hat{Y}(30)$

$\Rightarrow$ 229.6312 ± 8.7618

$\Rightarrow$ $(220.8695, 238.3931)$

ⓒ 98% PI for 10 new test items, each with $X=30$.

$\hat{Y}(30) \pm t_{\alpha/2,14} \cdot var(prediction-mean)$

$var(prediction-mean) = MSE/10 + s^2$ of $\hat{Y}(30)$

$\Rightarrow$ 229.6312 ± 3.4542

$\Rightarrow$ $(226.1771, 233.0855)$

ⓓ The interval in ⓒ is narrower than ⓑ → and it should be! with a prediction MEAN interval and all values at the same X, the variance is lower which narrows our interval. ⓐ

ⓔ Boundary values: $\hat{Y}(30) \pm Boundary(\hat{Y}(30))$

$Boundary^2 = 2F \to F$ has $.98$, $df=2,14$, we get 5.24

So, $\Rightarrow$ $\hat{Y}(30) \pm (\sqrt{2 \cdot 5.24})(.8285) \Rightarrow 229.6312 \pm 2.6821$

$\Rightarrow (226.9491, 232.3133)$. $.8285 \to s(\hat{Y}(30))$. It is wider, and it should be because it is for a model and not just one point

20

3. unknown σ^2

$y = \beta_0 + x_i\beta_1 + \epsilon_i$ for $i=1,\dots,n$ subjects

$\epsilon_i \overset{iid}{\sim} N(0,1)$

For $\lambda \geq 0$, estimate β_1 with $\tilde{b}_1$ to minimize

$$PSS(\beta_0,\beta_1) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 + \lambda\beta_1^2$$

so that

$$(\tilde{b}_0, \tilde{b}_1) = \operatorname{argmin}_{\beta_0,\beta_1 \in \mathbb{R}} PSS_\lambda(\beta_0,\beta_1).$$

ⓐ Find $\tilde{b}_0, \tilde{b}_1$ for: $\lambda = 0$, $\lambda = \infty$

For $\lambda = 0$, the term disappears and we want only to minimize summed square of error term

so $(\tilde{b}_0, \tilde{b}_1)$ will be β_0, β_1

$E\{\tilde{b}_0, \tilde{b}_1\} = \beta_0, \beta_1$ (OLS estimators)

For $\lambda = \infty$, the estimate for β_1 $(\tilde{b}_1)$ shrinks.

$\lim_{\lambda\to\infty} E\{\tilde{b}_1\} = 0$

Minimize $PSS = \sum(y_i - \beta_0 - \beta_1 x_i)^2 + \lambda\beta_1^2$

$\rightarrow$ $\partial PSS / \partial\beta_0 = \sum -2(y_i - \beta_0 - \beta_1 x_i) = 0$

$\rightarrow$ $\partial PSS / \partial\beta_1 = \sum 2(y_i - \beta_0 - \beta_1 x_i)(-x_i) + 2\lambda\beta_1 = 0$

Through what we've seen in lecture, these come down to:

$\Rightarrow$ $\tilde{b}_0 = \bar{y} - \hat{\beta}_1\bar{x}$

and $\tilde{b}_1 = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2 + \lambda}$

$\tilde{b}_0$ has no λ, so is always β_1

$\tilde{b}_1$ relies on λ in the denominator, so

when $\lambda = 0$, $\tilde{b}_1 \Rightarrow \beta_1$

when $\lambda \to \infty$, $\tilde{b}_1 \Rightarrow 0$ because λ grows in the denominator

(b) The claim is true!
Variance of the new β_1 with λ should be smaller $\lambda > 0$ because λ is an additive in the denominator

$$var(\tilde{b_1}) = \frac{\Sigma(x_i-\bar{x})^2}{(\Sigma(x_i-\bar{x})^2+\lambda)^2}\sigma^2 < var(\beta_1)$$

(c) We have shown in (a) that for $\lambda > 0$, $\tilde{b_1}$ is a biased estimate for β_1 because of the penalized term, due to "Bias-Variance Tradeoff" of Ridge Regression

4 (a) Prove $\sum \hat{\epsilon_i} = \sum \hat{\epsilon_i} x_i = 0$ $\quad (\hat{\epsilon_i} = y_i - \hat{\beta_0} - \hat{\beta_1} x_i)$

20

FOC: β_0: $\frac{\partial}{\partial \beta_0} \sum (y_i - \hat{\beta_0} - \hat{\beta_1} x_i) = 0$

$\Rightarrow -2\sum (y_i - \hat{\beta_0} - \hat{\beta_1} x_i) = 0$

$\Rightarrow -2\sum \hat{\epsilon_i} = 0$, since $\hat{\epsilon_i} = y_i - \hat{\beta_0} - \hat{\beta_1} x_i$

and $\sum_{i=1}^{n} \hat{\epsilon_i} = 0$

β_1: $\frac{\partial}{\partial \beta_1} \sum (y_i - \hat{\beta_0} - \hat{\beta_1} x_i) = 0$

$\Rightarrow -2\sum x_i (y_i - \hat{\beta_0} - \hat{\beta_1} x_i) = 0$

$\Rightarrow -2\sum x_i \cdot \hat{\epsilon_i} = 0$ $\rightarrow$ x_i constants,

$\Rightarrow -2\bar{x} \sum \hat{\epsilon_i} = 0$, $\sum \hat{\epsilon_i} = 0$.

so $\sum_i^n \hat{\epsilon_i} = \sum_i^n x_i \hat{\epsilon_i} = 0$

(b) $\sum (\hat{\epsilon_i} - \hat{\bar{\epsilon}})(\hat{y_i} - \hat{\bar{y}}) = 0$

we know $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $\hat{\bar{\epsilon}} = \frac{1}{n}\sum \hat{\epsilon_i}$

$\hat{\epsilon_i} = y_i - \hat{\beta_0} - \hat{\beta_1} x_i$, $\hat{y_i} = \hat{\beta_0} - \hat{\beta_1} x_i$, $\hat{\bar{y}} = \frac{1}{n}\sum y_i$

$\hat{y_i} - \hat{\bar{y}} = \hat{\beta_0} - \hat{\beta_1} x_i - (\hat{\beta_0} + \hat{\beta_1} \cdot \frac{1}{n}\sum x_i)$

so, we write :

$= \sum (\hat{\epsilon_i} - \hat{\bar{\epsilon}})(x_i - \frac{1}{n}\sum x_i)\hat{\beta_1}$

$= \sum [\hat{\epsilon_i} x_i - \frac{1}{n}\hat{\epsilon_i}\sum x_i - \hat{\bar{\epsilon}} x_i + \hat{\bar{\epsilon}}\frac{1}{n}\sum x_i]$

$= \sum [\underset{①}{\hat{\epsilon_i} x_i} - \underset{②}{\hat{\epsilon_i}\bar{x}} - \underset{③}{\hat{\bar{\epsilon}} x_i} + \underset{④}{\hat{\bar{\epsilon}}\bar{x}}]$

① $\sum \hat{\epsilon_i} x_i = 0$ (from ⓐ)

② $\sum \hat{e_i}\bar{x} = \bar{x}\sum \hat{\epsilon_i} = 0$ from ⓐ & $E(\epsilon_i) = 0$

③ $\sum \hat{\bar{\epsilon}} x_i = \frac{1}{n}\sum \hat{\epsilon_i} x_i = 0$ (like ①, from ⓐ)

④ $\sum \hat{\bar{\epsilon}}\bar{x} = \bar{x} \cdot \frac{1}{n}\sum \hat{\epsilon_i} = 0$ from ②

All terms go to zero, so

$\sum_{i=1}^{n} (\hat{\epsilon_i} - \hat{\bar{\epsilon}})(\hat{y_i} - \hat{\bar{y}}) = 0$

$\rightarrow$

4ⓒ Define $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \epsilon_i^2$. Show $E(\hat{\sigma}^2) = \sigma^2$

we know $var(\epsilon_i) = \sigma^2$ because $cov(\epsilon_i, \epsilon_j) = \sigma^2$ for $i=j$

and we know $E(\epsilon_i) = 0$

$var(\epsilon_i) = E(\epsilon_i^2) - E(\epsilon_i)^2$, by definition of variance

$\rightarrow E(\epsilon_i^2) = var(\epsilon_i) + E(\epsilon_i)^2$ and $E(\epsilon_i)^2 = 0$,

so $E(\epsilon_i^2) = var(\epsilon_i) = \sigma^2$

In matrix form:

$\hat{\epsilon} = Y - X\hat{\beta} = (I - X(X'X)^{-1}X)Y$

$E(\hat{\sigma}^2) = \frac{1}{n-2} E[Y'(I - X(X'X)^{-1}X)Y]$

$= \frac{1}{n-2} E[\hat{\epsilon}'(I - X(X'X)^{-1}X)\epsilon]$

by trace properties, $Tr(ABC) = Tr(BCA)$

so,

$= \frac{1}{n-2} Tr[E((I - X(Y'X)^{-1}X)\epsilon\epsilon')$

$= \frac{1}{n-2} \sigma^2 Tr(\underbrace{I}_{\text{identity matrix, } Tr(I)=n \text{ by definition}} - \underbrace{X(X'X)^{-1}X}_{\text{hat matrix, } Tr(H)=2 \text{ by definition}})$

$= \frac{1}{\cancel{n-2}} \sigma^2 (\cancel{n-2}) = \sigma^2$

18

5. $Y_i \Rightarrow$ \$ by 100
$X_i \Rightarrow$ hours for i^{th} month, $i=1,\dots,24$ $(n=24)$

ⓐ Assumptions:
- relationship between X & Y is linear
- variances of residuals equal across X_i (homoscedasticity)
- observations are independent of each other
- For fixed X, Y is normally distributed

Model: $\hat{Y} = 101.57570 + 1.15806\hat{X}$

ⓑ point estimate and 95% CI for increase of 10 hours
$10 \cdot \hat{\beta}_1 = 11.5806$, $n = 24$ $t_{\alpha/2,22} = 2.0739$
CI: $11.5806 \pm t_{\alpha/2,22} \cdot \frac{s}{\sqrt{n}}$, $SE = \frac{s}{\sqrt{n}} = .04338$
$= 11.5806 \pm (2.0739)(.04338)$
$= 11.5806 \pm .08997$
$\Rightarrow (11.4906, 11.67061)$

ⓒ Manager plans no ads next month
(i) Find 95% prediction interval for sales when $X=0$
$\hat{Y}(0) = 101.57570$, $t_{\alpha/2,n-2} = 2.0739$, $S_{pred} = \sqrt{\sigma^{2(\alpha s)}\left(1+\frac{1}{24}+\frac{(X_h-\bar{X})^2}{n\sigma_x^2}\right)}$
$= 101.5757 \pm (2.0739)(1.7147)$ $\hookrightarrow = 1.7147$
$\Rightarrow (98.0196, 105.1318)$

(ii) Normality is less of an important assumption for confidence intervals because it pertains to the errors rather than the estimate itself. Data near the estimated regression line isn't such a worry. For prediction intervals however, normality is important because we are moving BEYOND known data, and our interval depends on one value, rather than the whole of our data, which somewhat limits itself by being known (and within 1 or 2 standard deviations of the mean).

(d) Does an increase in radio ads make sales \$ go up? ($\beta_1 \Rightarrow$ radio variable)

H_0: $\beta_1 = 0$
H_A: $\beta_1 > 0$
at 95%, $\alpha = 0.05$

Find $t(22, .05) = 2.0739$
t^* from R output: 26.69
p-value: <.0001

with, $|t^*| > t$, we reject the H_0 and conclude that there is evidence to suggest an increase in radio ads makes sales \$ go up