MA 575: HW2

Benjamin Draves

September 19, 2017

Exercise 2.2

First, we'll read in the data and take a look at the data frame containing the housing data.

```
#read in data
dat = read.table("~/Desktop/Fall 2017/MA 575/book_data/indicators.txt", header = TRUE)

#take a peak
head(dat)
```

```
##   MetroArea PriceChange LoanPaymentsOverdue
## 1   Atlanta         1.2                4.55
## 2    Boston        -3.4                3.31
## 3   Chicago        -0.9                2.99
## 4    Dallas         0.8                4.26
## 5    Denver        -0.7                3.56
## 6   Detroit        -9.7                4.71
```

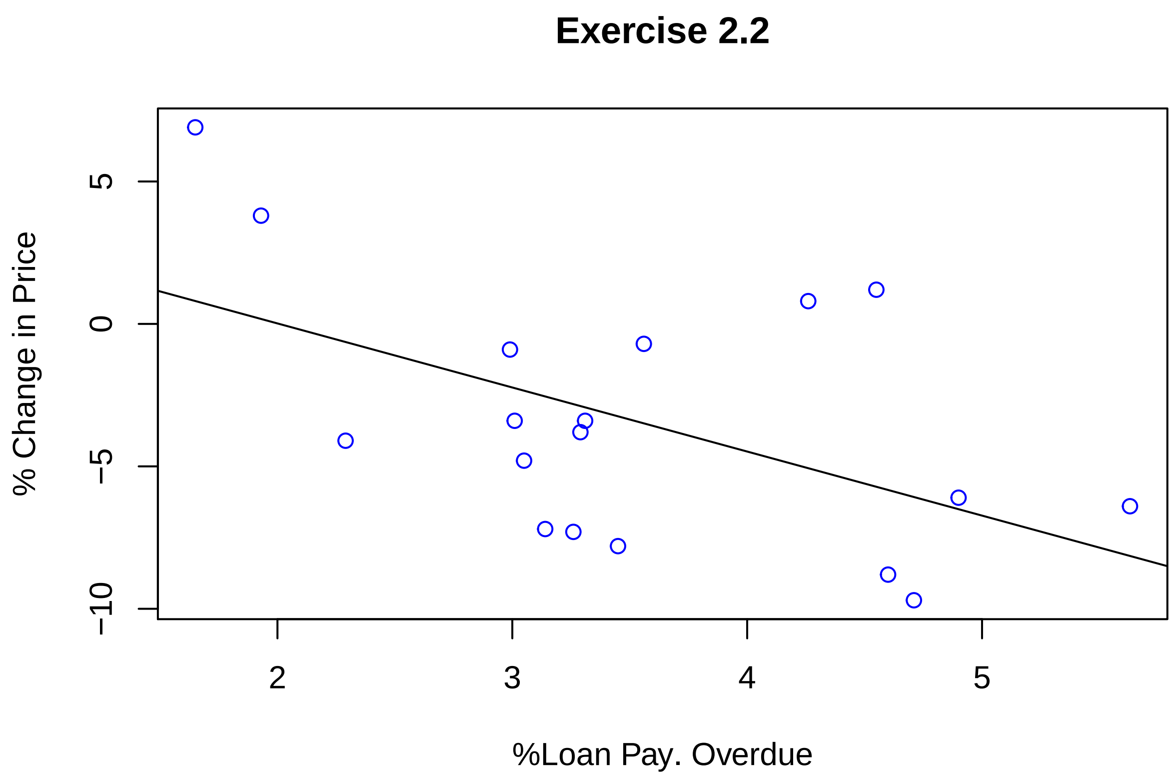
Next, we'll actually build the model. From here, we'll visualize the relationship and the simple regression fit.

```
#attach the dataframe
attach(dat)

#build the model
model = lm(PriceChange ~ LoanPaymentsOverdue)

#plot the relationship
plot(LoanPaymentsOverdue, PriceChange, main = "Exercise 2.2", xlab = "%Loan Pay. Overdue", ylab = "% Cha

#add the regression line
abline(model)
```



Now we'll take a look at the fit statistics.

```
summary(model)
```

```
##
## Call:
## lm(formula = PriceChange ~ LoanPaymentsOverdue)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6541 -3.3419 -0.6944  2.5288  6.9163
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           4.5145     3.3240   1.358   0.1933
## LoanPaymentsOverdue  -2.2485     0.9033  -2.489   0.0242 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.954 on 16 degrees of freedom
## Multiple R-squared:  0.2792, Adjusted R-squared:  0.2341
## F-statistic: 6.196 on 1 and 16 DF,  p-value: 0.02419
```

(a)

To build a 95% confidence interval for $\widehat{\beta}_1$, we will use the form derived in the text for the case where σ^2 is unknow. That is the $(1-\alpha)100\%$ confidence interval is given by $\widehat{\beta}_1 \pm t_{\alpha/2,n-2}\frac{S}{\sqrt{SXX}}$. R gives us $\widehat{\beta}_1 = -2.2485$ and the standard error corresponding to this estimator $se(\widehat{\beta}_1) = \frac{S}{\sqrt{SXX}} = 0.9033$. To finish building this confidence interval, we need to find $t_{\alpha/2,n-2}$. In our case there are $n = 18$ observations and we are using a

95% confindence level corresponding to $\alpha = 0.05$. Hence we seek to find $t_{\alpha/2,n-2} = t_{.05/2,18-2} = t_{.025,16}$. We can use R to find this value using

```
#find t - multiplier
qt(.975, 16)
```

```
## [1] 2.119905
```

Hence our confidence interval is given by

```
#build confidence interval
right = -2.2485 + (2.119905 * 0.9033)
left = -2.2485 - (2.119905 * 0.9033)

# Our 95% confidence interval for B_1
c(left, right)
```

```
## [1] -4.1634102 -0.3335898
```

The confidence interval covers negative values and does not include zero. This suggests that the true value of β_1 is negative. That is, this model gives us reason to believe there is a negative relationshop between percentage change in loan payment overdue and percentage change in price.

(b)

Next, we'll use our model to find the confidence interval for our regression line at $X = 4$. Using the summary above, our regression line is given by $E(Y|X = x^*) = 4.5145 - 2.2485x^*$. Hence for $x^* = 4$, we have $y^* = 4.5145 - 2.2485(4) = -4.4795$.

To construct a 95% confidence inteval for the regression line, we will use the form we derived in class given by $\hat{y}^* \pm t_{\alpha/2,n-2}S\sqrt{\frac{1}{n} + \frac{(x^*-\overline{x})^2}{SXX}}$. From part a we know $t_{\alpha/2,n-2} = 2.119905$. We will need to find the remaining terms. Consider the following

```
#find S and SXX
S = sqrt(1/(18-2)*sum((model$residuals)^2))
SXX = sum((dat$LoanPaymentsOverdue - mean(dat$LoanPaymentsOverdue))^2)

#find the part under the square root
under = 1/18 + (4 - mean(dat$LoanPaymentsOverdue))^2/SXX

#define the width variable
width = 2.119905*S*sqrt(under)
```

From here we can build our confidence interval using the formula given above.

```
#build the interval
left = -4.4795 - width
right = -4.4795 + width

c(left, right)
```

```
## [1] -6.648764 -2.310236
```

0% is likely *not* a reasonable value for the expected percentage change in housing price. We're 95% confident the true percentage change in housing price given that the percentage change in overdue loan payments is somewhere in the interval $(-6.648764, -2.310236)$. 0% is well outside this range and thus is likely not a feasible value for $E(Y|X = 4)$.

Exercise 2.3

First we'll read in the data and take a look at the dataframe.

```
#read in data
dat = read.table("~/Desktop/Fall 2017/MA 575/book_data/invoices.txt", header = TRUE)

#take a peak
head(dat)
```

```
##   Day Invoices Time
## 1   1      149  2.1
## 2   2       60  1.8
## 3   3      188  2.3
## 4   4       23  0.8
## 5   5      201  2.7
## 6   6       58  1.0
```

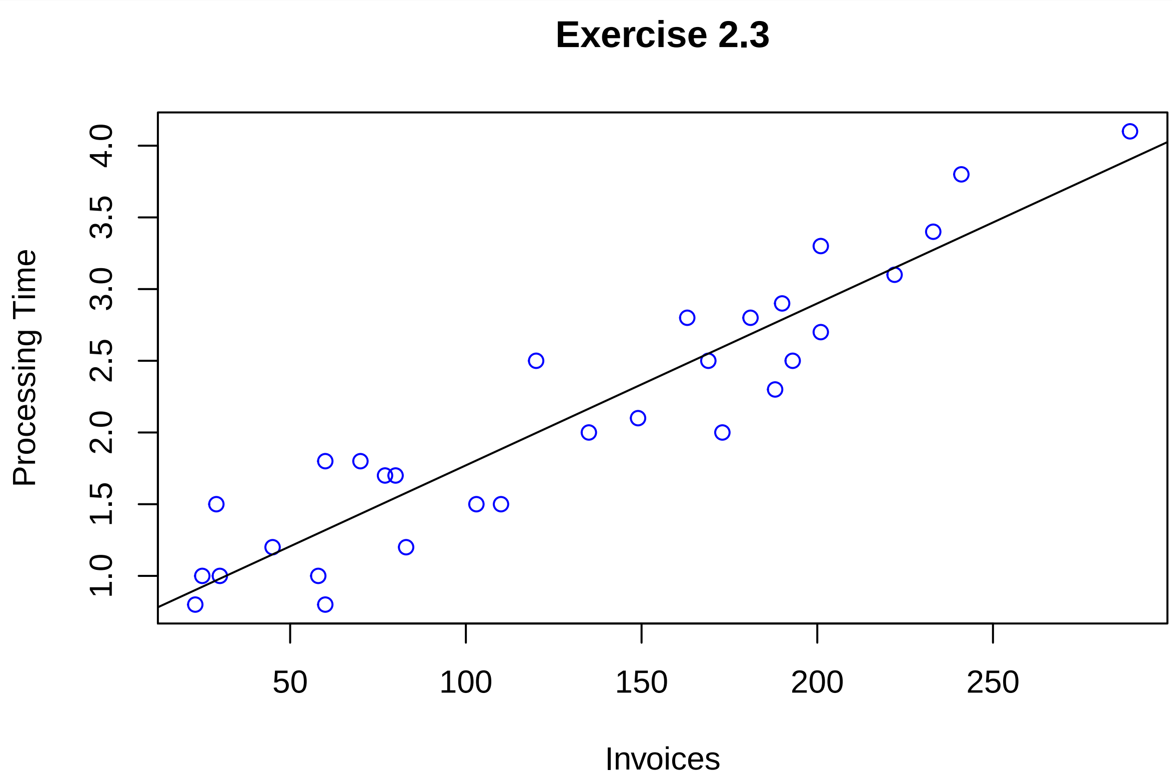
We'll plot the relationship and build the model.

```
#attach the dataframe
attach(dat)

#build the model
model = lm(Time ~ Invoices, data = dat)

#plot the relationship
plot(Invoices, Time, main = "Exercise 2.3", xlab = "Invoices", ylab = "Processing Time", pch = 1, col =

#add the regression line
abline(model)
```



Now we'll take a look at the fit statistics.

```
summary(model)
```

```
##
## Call:
## lm(formula = Time ~ Invoices, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59516 -0.27851  0.03485  0.19346  0.53083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6417099  0.1222707   5.248 1.41e-05 ***
## Invoices    0.0112916  0.0008184  13.797 5.17e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3298 on 28 degrees of freedom
## Multiple R-squared:  0.8718, Adjusted R-squared:  0.8672
## F-statistic: 190.4 on 1 and 28 DF,  p-value: 5.175e-14
```

(a)

To build the confidence for $\widehat{\beta}_0$ we will use the form where σ^2 is unknown. The $(1-\alpha)100\%$ confidence interval is given by $\widehat{\beta}_0 \pm t_{\alpha/2,n-2} S\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{SXX}}$. Here, $se(\widehat{\beta}_0) = S\sqrt{\frac{1}{n}+\frac{\overline{x}^2}{SXX}}$ and $\widehat{\beta}_0$ is given directly from R. In our case $se(\widehat{\beta}_0) = 0.1222707$ and $\widehat{\beta}_0 = 0.6417099$. We can calculate $t_{\alpha/2,n-2} = t_{.025,28}$ using the following

```
qt(.975, 28)
```

```
## [1] 2.048407
```

From here, we can construct our confidence interval for the intercept

```
width = 2.048407*0.1222707

left = 0.6417099 - width
right = 0.6417099 + width

c(left, right)
```

```
## [1] 0.3912497 0.8921701
```

(b)

Here our null hypothesis $H_0 : \beta_1 = 0.01$ and the alternative hypothesis is given by $H_A : \beta_1 \neq 0.01$. Before, we calculate our test statistic, we will identify our rejection region. We will use an $\alpha = 0.05$ level for the $T_{n-2} = T_{28}$ distribution. In part a we calculated that $t_{.025,28} = 2.048407$. So our rejection region is given by $R = (-\infty, -2.048407) \cup (2.048407, \infty)$.

Now we will calculate our test statistic given by $T = \frac{\widehat{\beta}_1 - \beta_1^{(0)}}{S/\sqrt{SXX}} = \frac{\widehat{\beta}_1 - \beta_1^{(0)}}{se(\widehat{\beta}_1)} = \frac{0.0112916 - .01}{0.0008184} = 1.578201$. Since $T \notin R$, we fail to reject H_0, that $\beta_1 = .01$. That is, we have do not have sufficent evidence to suggest that $\beta_1 \neq .01$

(c)

Here we look to find a point estimate, and corresponding prediction interval, for $E(Y|X = 130)$. Using our estimates for $\widehat{\beta}_1$ and $\widehat{\beta}_0$, we have $\hat{y}^* = 0.6417099 + 0.0112916 * 130 = 2.109618$. We will use the prediction interval of the form $\hat{y}^* \pm t_{\alpha/2,n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \overline{x})^2}{SXX}}$. We will use the built in R function *predict* to get the prediction interval for $X = 130$.

```
#Calculate prediction interval
predict(model,data.frame(Invoices = 130), interval  = "prediction")
```

```
##        fit      lwr    upr
## 1 2.109624 1.422947 2.7963
```