

STAT 608 HW 1

Jack Cunningham (jgavc@tamu.edu)

9/6/24

1)

```
playbill <- read.csv("playbill.csv")

playbill_slr <- lm(data = playbill, formula = CurrentWeek ~ LastWeek)
slr_data <- summary(playbill_slr)
slr_data
```

Call:

```
lm(formula = CurrentWeek ~ LastWeek, data = playbill)
```

Residuals:

Min	1Q	Median	3Q	Max
-36926	-7525	-2581	7782	35443

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.805e+03	9.929e+03	0.685	0.503
LastWeek	9.821e-01	1.443e-02	68.071	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18010 on 16 degrees of freedom

Multiple R-squared: 0.9966, Adjusted R-squared: 0.9963

F-statistic: 4634 on 1 and 16 DF, p-value: < 2.2e-16

a)

```

B_1 <- slr_data$coefficients[2,"Estimate"]
B_1_se <- slr_data$coefficients[2,"Std. Error"]
n = length(playbill[, "CurrentWeek"])
lower <- B_1 - B_1_se*qt(1-0.05/2, df = n - 2)
upper <- B_1 + B_1_se*qt(1-0.05/2, df = n - 2)
c(lower,upper)

```

```
[1] 0.9514971 1.0126658
```

The 95% confidence interval for β_1 is (0.951, 1.013) The suggested value of 1 for β_1 is reasonable because it is within our confidence interval.

b)

We are testing the hypothesis: $H_0 : \beta_0 = 10000, H_1 : \beta_0 \neq 10000$. To do this we have the T statistic:

$T = \frac{\hat{\beta}_0 - 10000}{SE(\hat{\beta}_0)}$ and its absolute value needs to be greater than $t_{n-2}(\alpha/2)$ to reject the null hypothesis.

```

t_quantile_rejection <- qt(1 - 0.05/2, df = n - 2)
B_0_hat <- slr_data$coefficients[1,"Estimate"]
B_0_hat_se <- slr_data$coefficients[1,"Std. Error"]
T_statistic_B_0 <- (B_0_hat - 10000)/B_0_hat_se
T_statistic_B_0

```

```
[1] -0.3217858
```

In this case our T statistic equals to -0.322 whose absolute value is far lower than 2.12.

```

lower_B_0 <- B_0_hat - B_0_hat_se*qt(1-0.05/2, df = n - 2)
upper_B_0 <- B_0_hat + B_0_hat_se*qt(1 - 0.05/2, df = n - 2)
c(lower_B_0, upper_B_0)

```

```
[1] -14244.33 27854.10
```

In fact the 95% confidence interval for β_0 is extremely wide, it is $(-1.4244327 \times 10^4, 2.7854099 \times 10^4)$. Therefore I wouldn't use the currently fitted model to predict current week sales when previous week sales are closer to zero. This is mainly due to the fact that our data only covers a range of previous week sales values from \$200,000 to \$1,200,000.

c)

To construct a 95% prediction interval for $x^* = 400000$ we use the below:

$$\hat{\beta}_0 + \hat{\beta}_1 x^* \pm t(\alpha/2, n-2) S \sqrt{\frac{1}{n} + 1 + \frac{(x^* - \bar{x})^2}{SXX}}$$

We can do this manually or use the predict function.

```
prediction <- B_0_hat + B_1*400000
S <- sqrt(1/(n-2)*sum((playbill[, "CurrentWeek"]-B_0_hat - B_1*playbill[, "LastWeek"])^2))
SXX <- sum((playbill[, "LastWeek"] - mean(playbill[, "LastWeek"]))^2)

upper <- prediction + qt(1 - 0.05/2, df = n - 2)*S*sqrt(1/n + 1 + (400000 - mean(playbill[,
lower <- prediction - qt(1 - 0.05/2, df = n - 2)*S*sqrt(1/n + 1 + (400000 - mean(playbill[,

c(lower, upper)
```

```
[1] 359832.8 439442.2
```

```
current_predicted_value <- predict(playbill_slr,
                                   newdata = data.frame(LastWeek = c(400000)),
                                   interval = "prediction",
                                   confidence = 0.95)

current_predicted_value
```

```
      fit      lwr      upr
1 399637.5 359832.8 439442.2
```

The value \$450,000 is not a reasonable value to expect for the current week given a previous week of \$400,000. This is because it falls outside of our 95% prediction interval at $x^* = 400000$: $(3.5983275 \times 10^5, 4.394422 \times 10^5)$.

d)

The model that the promoters of Broadway plays are using can be represented mathematically as $\hat{Y} = \beta_0 + \beta_1 x$ where $\beta_0 = 0$ and $\beta_1 = 1$. Thus to see if this is an appropriate rule we can test to see if the predictions generated by this simplified model are within the prediction interval for various values of last week data. Since we only have values of x from \$200,000 to \$1,200,000 in our dataset, I use evenly spaced values of x in that range and see if the predictions are within the prediction intervals generated from our simple linear regression model.

```

x_to_test <- seq(200000, 1200000, 10000)
conf_intervals_for_x <- predict(playbill_slr,
                                newdata = data.frame>LastWeek = x_to_test),
                                interval = "prediction",
                                confidence = 0.95)
x_in_range <- x_to_test < conf_intervals_for_x[, "upr"] & x_to_test > conf_intervals_for_x[, "lwr"]

x_percent_in_range <- sum(x_in_range)/length(x_in_range)
x_percent_in_range

```

[1] 1

From this test we can see that the promoters prediction rule of using last week's gross box office result to predict the current week's box office result is a good rule of thumb (for last week gross box office amounts between \$200,000 and \$1,200,000). However it is worth noting that the promoters should expect the actual value of current week sales to not exactly match their prediction and rather randomly deviate around it.

2)

$$Var(Y_i | X_i = x_i) = Var(\beta_0 + \beta_1 x_i + e_i) = Var(e_i).$$

We assume that Y is related to X through the linear regression model $Y_i = \beta_0 + \beta_1 x_i + e_i$.

We assume that x_i are known "known fixed constants".

3)

To create a linear regression model we need a method to estimate the parameters. The least squares criterion estimates the parameters by finding the values that minimize the sum of squared differences between our predicted values of the response variable and the actual values of the response variable in the data.

4)

i)

Since there is no intercept the design matrix is:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{bmatrix}$$

ii)

$$y = X\beta + e_i$$

$$e_i^2 = (y - X\beta)^T(y - X\beta)$$

$$e_i^2 = y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta$$

$$e_i^2 = y^T y - 2y^T X\beta + \beta^T X^T X\beta$$

We minimize e_i^2 . So we take the derivative of this with respect to β .

$$\frac{dRSS(\beta)}{d\beta} = -2y^T X + (X^T X + (X^T X)^T)\beta$$

We set this equal to zero:

$$-2y^T X + (2X^T X)\beta = 0$$

Recalling that X is a vector of length n consisting of all x values we have $y^T X = \sum_{i=1}^n y_i x_i$ and $X^T X = \sum_{i=1}^n x_i^2$. We have:

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

b)

i)

$$E(\hat{\beta}|X) = E\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}\right) = \frac{\sum_{i=1}^n E(x_i y_i)}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i E(y_i)}{\sum_{i=1}^n x_i^2}$$

Using that $E(y_i) = \beta x_i$:

$$E(\hat{\beta}|X) = \frac{\beta \sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} = \beta$$

ii)

$$Var(\hat{\beta}|X) = Var\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}\right) = \frac{\sum_{i=1}^n x_i^2 Var(y_i)}{(\sum_{i=1}^n x_i^2)^2}$$

Since $Var(y_i) = \sigma^2$:

$$Var(\hat{\beta}|X) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$$

iii)

We have $E(\hat{\beta}|X) = \beta$ and $Var(\hat{\beta}|X) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$.

Since $Y = X\beta + e$. We can say that $Y \sim N_n(X\beta, \sigma^2 I_n)$.

Using the fact that when:

$$X \sim N(\mu, \sigma^2)$$

$$AX \sim N(A\mu, A\sigma^2 A^T)$$

So since $\hat{\beta} = (X^T X)^{-1} X^T y$. We can say that:

$$\hat{\beta} \sim N((X^T X)^{-1} X^T X \beta, (X^T X)^{-1} X^T \sigma^2 I_n ((X^T X)^{-1} X^T)^T)$$

$$\hat{\beta} \sim N_n(\beta, \sigma^2 I_n (X^T X)^{-1} X^T X (X^T X)^{-1})$$

Since X is a vector of length n: $X^T X = \sum_{i=1}^n x_i^2$. So:

$$\hat{\beta} \sim N(\beta, \frac{\sigma^2}{\sum_{i=1}^n x_i^2})$$

5.

a)

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \dots \\ 1 \end{bmatrix}$$

b)

$$\hat{\beta} = (X^T X)^{-1} X^T y = (n)^{-1} \sum_{i=1}^n y_i = \bar{y}$$

6.

We know the below fact about covariance:

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

Then:

$$Cov(aX, bY) = E(aXbY) - E(aX)E(bY) = abE(XY) - abE(X)E(Y) = ab(E(XY) - E(X)E(Y)) = ab(Cov(X, Y))$$

7.

When a confidence interval for a particular x^* is created we are talking about the confidence interval of the statistic $E(Y|x^*)$, the mean value of Y at a particular x^* .

The confusion comes from thinking that the confidence interval of $E(Y|x^*)$ is the same as a prediction interval for a single observation of Y at a given x^* . The prediction interval will always be wider than the confidence interval as it also incorporates the variance of the random error e_i .

8.

Starting from $\hat{\beta} = (X^T X)^{-1} X^T y$:

$$X^T X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{n S_{XX}} \begin{bmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

$$X^T y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ \sum x_i y_i \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{n S_{XX}} \begin{bmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix} \begin{bmatrix} n\bar{y} \\ \sum x_i y_i \end{bmatrix}$$

$$\hat{\beta}_0 = \frac{1}{nSXX}(\sum x_i^2(n)(\bar{y}) - (n)(\bar{x})(\sum x_i y_i))$$

Then:

$$\hat{\beta}_0 = \frac{\bar{y}(\sum x_i^2 + n\bar{x}^2 - n\bar{x}^2) - \bar{x} \sum x_i y_i}{SXX}$$

We do this since $SXX = \sum x_i^2 - n\bar{x}^2$. So now we have:

$$\hat{\beta}_0 = \frac{\bar{y}(SXX + n\bar{x}^2) - \bar{x} \sum x_i y_i}{SXX}$$

$$\hat{\beta}_0 = \bar{y} - \frac{-n\bar{y}\bar{x}^2 + \bar{x} \sum x_i y_i}{SXX}$$

$$\hat{\beta}_0 = \bar{y} - \frac{(-n\bar{x}\bar{y} + \sum x_i y_i)\bar{x}}{SXX}$$

And we can show that $SXY = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \sum x_i(\bar{y}) - \bar{x} \sum y_i + \sum \bar{x}\bar{y} = \sum x_i y_i - 2n\bar{x}\bar{y} + n\bar{x}\bar{y} = \sum x_i y_i - n\bar{x}\bar{y}$

Then we come to the desired outcome of:

$$\hat{\beta}_0 = \bar{y} - \frac{SXY}{SXX}\bar{x}$$

9.

Starting from our estimate of $\hat{\beta}$ from earlier:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

We are looking for $Var(a^T \hat{\beta} | X)$ so:

$$Var(a^T \hat{\beta} | X) = Var(a^T (X^T X)^{-1} X^T y)$$

We then use the below fact.

$$X \sim N(\mu, \epsilon)$$

$$AX \sim N(A\mu, A\epsilon A^T)$$

Then:

$$Var(a^T \hat{\beta} | X) = a^T (X^T X)^{-1} X^T Var(y) X (X^T X)^{-1} a$$

We know $Var(y) = \sigma^2 I_n$

$$Var(a^T \beta | X) = \sigma^2 I_n a^T (X^T X)^{-1} X^T X (X^T X)^{-1} a$$

$$Var(a^T \beta | X) = \sigma^2 a^T (X^T X)^{-1} a$$