STAT5044_HW2

zhengzhi lin

2019.9.30

Problem 1

(a) The estimation function is $y_i = 10.2 + 4x_i$ There is a good fit of the estimation function.
(b) $\hat{y} = 10.2 + 4 \times 1 = 14.2$
(c) $10.2 + 4 \times 2 - (10.2 + 4 \times 1) = 4$
(d) $10.2 + 4 \times \overline{x} = 10.2 + 4 \times 1 = 14.2 = \overline{y}$ verified

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.1
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

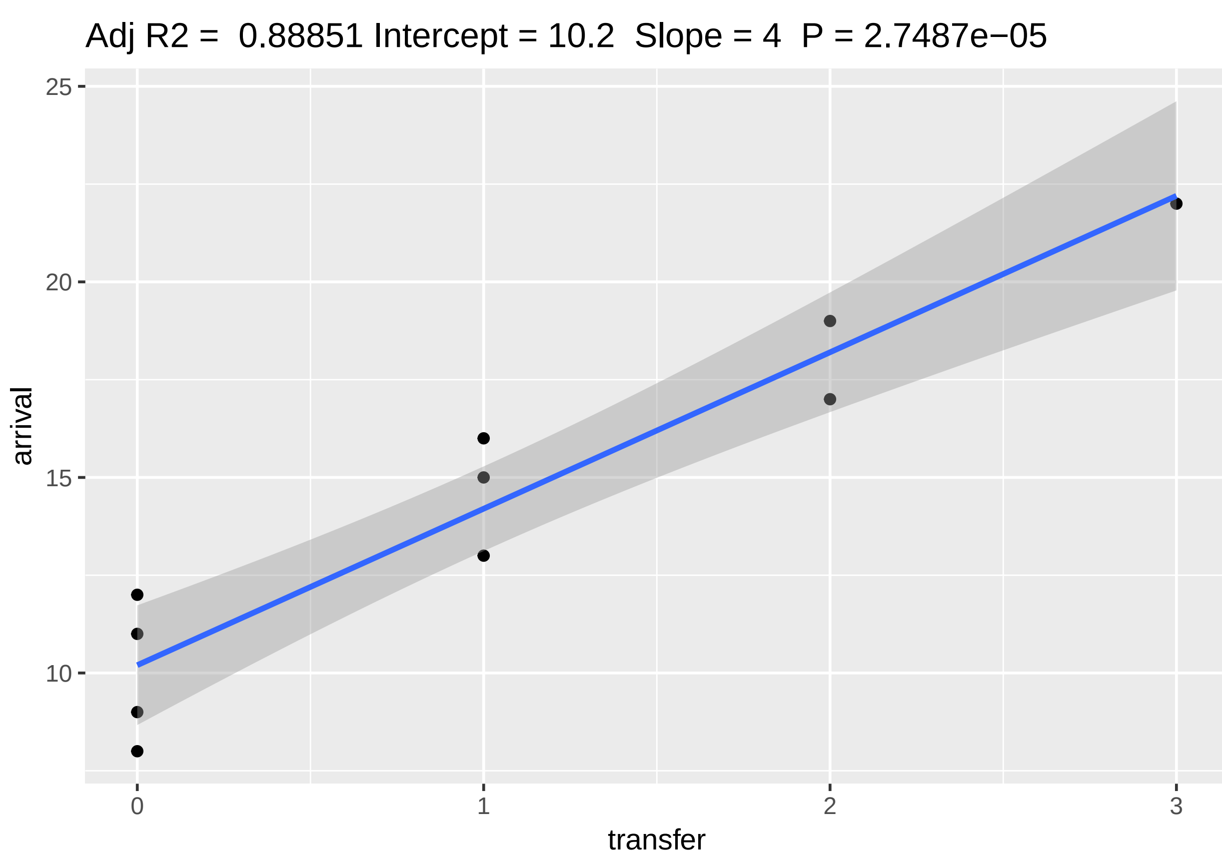
```
airfreight <- read.table("airfreight.txt")
airfreight <- airfreight[-1,]
airfreight <- airfreight %>% as.data.frame() %>%
  rename(transfer = V1, arrival = V2) %>%
  mutate_if(is.factor, as.character) %>%
  mutate_if(is.character, as.numeric)

fit <- lm(arrival~transfer, data = airfreight)
summary(fit) # get coefficient
```

```
##
## Call:
## lm(formula = arrival ~ transfer, data = airfreight)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -2.2   -1.2    0.3    0.8    1.8
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.2000     0.6633  15.377 3.18e-07 ***
## transfer      4.0000     0.4690   8.528 2.75e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.483 on 8 degrees of freedom
## Multiple R-squared:  0.9009, Adjusted R-squared:  0.8885
## F-statistic: 72.73 on 1 and 8 DF,  p-value: 2.749e-05
```

```
ggplot(data = airfreight, aes(x = transfer, y = arrival)) + geom_point() + geom_smooth(method = "lm") +
  labs(title = paste("Adj R2 = ",signif(summary(fit)$adj.r.squared, 5),
                     "Intercept =",signif(fit$coef[[1]],5 ),
                     " Slope =",signif(fit$coef[[2]], 5),
                     " P =",signif(summary(fit)$coef[2,4], 5)))
```




```
fit$coef[[1]] + 1 * fit$coef[[2]]
```

```
## [1] 14.2
```

```
fit$coef[[1]] + 2 * fit$coef[[2]]
```

```
## [1] 18.2
```

```r
fit$coef[[1]] + mean(airfreight$transfer) * fit$coef[[2]]
```

```
## [1] 14.2
```

```r
mean(airfreight$arrival)
```

```
## [1] 14.2
```

Problem 2

(a) CI: $\hat{\beta_1} \pm t_{8,\frac{\alpha}{2}}\sqrt{\hat{\sigma^2}\text{Sxx}}$ That is $4 \pm 8.528029 \times 0.4690416 = 4 \pm 4$

(b) $H_0 : \beta_1 = 0$, H_a : not H_0 p-value equals to 2.748669e-05,less than 0.05, reject H_0 We here use the t statistic, which is $t = \frac{\hat{\beta_1}-0}{\sqrt{\hat{\sigma^2}\text{Sxx}}}$ follows a t distribution which has df=2.

(c) $H_0 : \beta_1 = 0$, $H_a : \beta_1 > 0$ p-value is 1.374335e-05. The decision rule is same as (b), p-value is smaller that 0.05, so reject H_0

(d) CI: $\hat{\beta_0} \pm t_{8,\frac{\alpha}{2}}\sqrt{\hat{\sigma^2}(\frac{1}{\text{n}} + \frac{\overline{\text{x}}^2}{\text{Sxx}})}$ That is $10.2 \pm 15.377 \times 0.4690416 = 10.2 \pm 7.212453$

(e) $H_0 : \beta_0 \leq 9$, $H_a : \beta_0 > 9$ p-value is 0.05402227, larger than 0.025, can't reject H_0. The decision rule is same as above

```r
X <- cbind(rep(1,10),airfreight$transfer)
Y <- airfreight$arrival
h <- X %*% solve(t(X) %*% X) %*% t(X)
Sxx <- sum((X[,2] - mean(X[,2])) ^ 2)
Sxy <- sum((X[,2] - mean(X[,2])) * (Y - mean(Y)))
beta1_hat <- Sxy/Sxx
beta0_hat <- mean(Y) - beta1_hat * mean(X[,2])
sigma_hat <- t(Y) %*% (diag(1,10,10) - h) %*% Y/8
t <- beta1_hat/sqrt(sigma_hat/Sxx)
t2 <- beta0_hat/sqrt(sigma_hat * (1 / 10 + mean(X[,2]) ^ 2 / Sxx))
t
```

```
##          [,1]
## [1,] 8.528029
```

```r
t2
```

```
##          [,1]
## [1,] 15.37708
```

```r
sqrt(sigma_hat/Sxx)
```

```
##           [,1]
## [1,] 0.4690416
```

```r
2*pt(-abs(t),df=8) # p-value
```

```
##                 [,1]
## [1,] 2.748669e-05
```

```r
qt(.025, 8, lower.tail = TRUE, log.p = FALSE)
```

```
## [1] -2.306004
```

```r
qt(.025, 8, lower.tail = FALSE, log.p = FALSE)
```

```
## [1] 2.306004
```

```r
pt(-t,df=8) #p-value
```

```
##                 [,1]
## [1,] 1.374335e-05
```

```r
t3 <- (beta0_hat - 9)/sqrt(sigma_hat * (1 / 10 + mean(X[,2]) ^ 2 / Sxx))
t3
```

```
##             [,1]
## [1,] 1.809068
```

```r
pt(-abs(t3),df=8) # p-value
```

```
##               [,1]
## [1,] 0.05402227
```

```r
qt(.025, 8, lower.tail = FALSE, log.p = FALSE)
```

```
## [1] 2.306004
```

Problem 3

(a) predicted value of broken ampules with 2 shipments is 18.2, the upper bound is 20.42571, lower bound is 15.97429. predicted value of broken ampules with 4 shipments is 26.2, the upper bound is 31.17684, lower bound is 21.22316.

(b) predicted value of broken ampules with 2 shipments is 18.2, the upper bound is 23.65186, lower bound is 12.7481.

(c) predicted mean value of broken ampules with 3 of 2 shipments is 18.2, the upper bound is 23.65186, lower bound is 12.7481.

```r
#(a)
predict(fit,data.frame(transfer = 2), interval = "confidence", level = .99)
```

```
##    fit      lwr      upr
## 1 18.2 15.97429 20.42571
```

```r
predict(fit,data.frame(transfer = 4), interval = "confidence", level = .99)
```

```
##    fit      lwr      upr
## 1 26.2 21.22316 31.17684
```

```r
#(b)
predict(fit,data.frame(transfer = 2), interval = "prediction", level = .99)
```

```
##    fit      lwr      upr
## 1 18.2 12.74814 23.65186
```

```r
#c
predict(fit,data.frame(transfer = 2), interval = "prediction", level = 1-0.01/3)
```

```
##    fit      lwr      upr
## 1 18.2 11.50195 24.89805
```

```r
predict(fit,data.frame(transfer = 2), interval = "prediction", level = 1-0.01/3)
```

```
##    fit      lwr      upr
## 1 18.2 11.50195 24.89805
```

```r
predict(fit,data.frame(transfer = 2), interval = "prediction", level = 1-0.01/3)
```

```
##    fit      lwr      upr
## 1 18.2 11.50195 24.89805
```

Problem 4

```r
anova(lm(arrival~transfer, data = airfreight)) # full model
```

```
## Analysis of Variance Table
##
## Response: arrival
##           Df Sum Sq Mean Sq F value    Pr(>F)
## transfer   1  160.0   160.0  72.727 2.749e-05 ***
## Residuals  8   17.6     2.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(lm(arrival~1, data = airfreight)) #reduced model
```

```
## Analysis of Variance Table
##
## Response: arrival
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  9  177.6  19.733
```

```r
l <- summary(lm(arrival~transfer, data = airfreight))
l$r.squared#multiple correlation coeficient
```

```
## [1] 0.9009009
```

```r
Sxx <- sum((X[,2] - mean(X[,2])) ^ 2)
Sxy <- sum((X[,2] - mean(X[,2])) * (Y - mean(Y)))
Syy <- sum((Y - mean(Y)) * (Y - mean(Y)))
r <- Sxy / sqrt(Sxx * Syy)
r
```

```
## [1] 0.949158
```