Albert Lockett, 3254354, k44if@unb.ca

Q1

Exercise 8 a, c on page 624.

a)

$\mathrm{E}(Y|x^* = 2000) = 1800 + 1.3 \cdot (2000) = 4400$

$\mathrm{P}(Y > 5000) = \mathrm{P}\left(Z > \frac{5000-\mathrm{E}(Y|x^*=2000)}{\sigma/\sqrt{(n)}}\right) = \mathrm{P}\left(Z > \frac{5000-4400}{350}\right) = \mathrm{P}(Z > 1.71)$

$1 - \Phi 1.71 = 1 - 0.958 = \mathbf{0.0432}$

c)

$\mathrm{E}(Y_1) = \mathrm{E}(Y|x^* = 2000) = 4400$

$\mathrm{E}(Y_2) = \mathrm{E}(Y|x^* = 2500) = 1800 + 1.3 \cdot (2500) = 5050$

$\mathrm{E}(Y_2 - Y_1) = 5050 - 4400 = 650$

$\mathrm{P}(Y_2 - Y_1 > 100) = \mathrm{P}\left(Z > \frac{100-\mathrm{E}(Y_2-Y_1)}{\sigma/\ sqrtn}\right) = \mathrm{P}\left(Z > \frac{100-650}{350}\right) = \mathrm{P}(Z > -1.11)$

$1 - \Phi(1.11) = 1 - 0.133 = \mathbf{0.866}$

Q2

Exercise 34 a, b on page 652.

a)

```
x <- c( 50,   71, 55, 50, 33, 58, 79, 26,  69, 44,  37, 70, 20,  45, 49)
y <- c(152, 1992, 48, 22,  2,  5, 35,  7, 269, 38, 171, 13, 43, 185, 25)
n <- length(x)

S_xy <- sum(x*y) - (sum(x)*sum(y)) / n
S_xx <- sum(x^2) - (sum(x))^2 / n
```

$S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = 42402$

$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 4125.6$

```
x_bar <- mean(x)
y_bar <- mean(y)
Beta_1 <- S_xy/S_xx
Beta_0 <- y_bar - Beta_1 * x_bar
```

$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} =$ **-317.54**

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} =$ **10.278**

```
SSE <- sum(y^2) - Beta_0*sum(y) - Beta_1*sum(x*y)
SST <- sum((y - y_bar)^2)
r_2 <- 1 - SSE / SST
```

$\text{SSE} = \sum y_i^2 - \hat{\beta}_0 \sum y_i = \hat{\beta}_1 \sum x_i y_i = 3,096,567.3$

$\text{SST} = \sum (y_i - \bar{y})^2 = 3,532,368.7$

$r^2 = 1 - \frac{SSE}{SST} =$ **0.1233**

b)

```
s <- SSE / (n - 2)
s_Beta1 <- s / sqrt(S_xx)

Beta_1_0 <- 0
t <- (Beta_1 - Beta_1_0) / s

alpba <- 0.05
t_val = qt(alpba/2, n-2, lower.tail = FALSE)
reject <- t >= t_val || t <= -1 * t_val
```

$s = \frac{\text{SSE}}{n-2} = 237,187.4$

$s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{xx}}} = 3,708.46$

$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = 4.314 \cdot 10^{-5}$

$t_{0.05/2,n-2} = 2.16$

Do not reject null hypothsis because $t_{0.05/2,n-2} > t > -t_{0.05/2,n-2}$

d)

sorry

e)

90% Confidence interval

$(\hat{\beta}_1 \mp t_{0.1/2,n-2})$

$(8.507, 12.05)$

Q3

Exercise 46 a, b on page 660.

a)

$n = 20$

$\sum x_i = 2817.9$
$\bar{x} = \frac{\sum x_i}{n} = 140.9$

$\sum y_i = 1574.8$
$\bar{y} = \frac{\sum y_i}{n} = 78.74$

$x^* = 125$

$S_{xx} = 18921.8295$
$s = 0.665$
$s_{\hat{y}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} = 0.1673$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x^* = 124.24$

90% Confidence interval
$\hat{y} \mp t_{0.1/2,n-1} \implies 124.24 \mp 1.729$

(123.9613, 124.5402)

```
Beta_1 <- 0.4103377
Beta_0 <- 72.958547
n <- 20

sum_y <- 1574.8
y_bar <- sum_y/n

x_star <- 125

sum_x <- 2817.9
x_bar <- sum_x / n

S_xx <- 18921.8295
s <- 0.665
s_yhat <- s * sqrt(1/n + (x_star
     - x_bar)^2/S_xx)

y_hat <- Beta_0 + Beta_1 * x_
    star

alpha <- 0.10

t_val <- qt(alpha/2, n-1, lower.
    tail=FALSE)
ci <- c(y_hat + t_val*s_yhat, y_
    hat - t_val* s_yhat)
```

b)

Prediction Interval $\hat{y} \mp t_{0.1/2,n-1}\sqrt{s^2 + s_{\hat{y}}^2}$

(123.0650, 125.4365)

```
pi <- c(
  y_hat - t_val * sqrt(s^2 + s_
      yhat^2),
  y_hat + t_val * sqrt(s^2 + s_
      yhat^2)
)
```

It is wider than the confidence interval because the confidence interval preresents the confidence interval for the regression line expected values but the prediction interval is for the actual values of the data.

Q4

Exercise 60 on page 673.

a)

$S_{xy} = \frac{\sum x_i \sum y_i}{n} = 1.872$

$\bar{y} = 0.5557143$

$\bar{x} = 0.2407143$

$r = \frac{\sum\left((x_i - \bar{x})(y_i - \bar(y))\right)}{\sqrt{\sum(x_i-\bar{x})^2 \sum(y_i - \bar{y})^2}} = 0.4406567$

$t = \frac{r\sqrt{n-2}}{\sqrt(1-r^2)} = 1.700$

$t_{0.1/2,n-2} = 1.7822$

```
x <- c(0.18, 0.20, 0.21, 0.21, 0.21,
     0.22, 0.23,
       0.23, 0.24, 0.24, 0.25, 0.28,
          0.30, 0.37)
y <- c(0.46, 0.70, 0.41, 0.45, 0.55,
     0.44, 0.24,
       0.47, 0.22, 0.80, 0.88, 0.70,
          0.72, 0.74)
n <- length(x)

S_xy <- sum(x)*sum(y)/n

x_bar <- mean(x)
y_bar <- mean(y)

r <-  sum((x - x_bar)*(y - y_bar)) /
      (sqrt(sum((x - x_bar)^2)) *
         sqrt(sum((y - y_bar)^2)))

t <- r * sqrt(n - 2) / sqrt(1 - r^2)

alpha <- 0.10

t_val <- qt(alpha/2, n-2,lower.tail=
    FALSE)
reject = t >= t_val || t <= -1 * t_
    val
```

Rejection region:
either $t \geqslant t_{\alpha/2,n-1}$ or $t \leqslant -t_{\alpha/2,n-1}$

the t value is not in the rejection region

Do not reject the null hypothesis. The data does not show that the correlation coefficient differs from 0 at the given significance level.