GA17 (solutions)

Problem 1

a.

```
y<-c(1,1,5,4,2,3,6,8,9)
x<-c(60,65,70,80,80,80,90,90,100)
data<-data.frame(matrix(c(y,x), ncol=2, nrow=9))
colnames(data)=c('y','x')
lm<-lm(y~x , data=data)
summary(lm)
```

```
##
## Call:
## lm(formula = y ~ x, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4412 -0.5294 -0.3824  0.6765  2.5000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11.08824    3.56134  -3.113  0.01700 *
## x             0.19412    0.04432   4.380  0.00323 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.611 on 7 degrees of freedom
## Multiple R-squared:  0.7327, Adjusted R-squared:  0.6945
## F-statistic: 19.19 on 1 and 7 DF,  p-value: 0.003233
```

```
sum(lm$coefficients*c(1,85))
```

```
## [1] 5.411765
```

$$\hat{y} = -11.08824 + 0.19412 \cdot 85 = 5.411765$$

b.

```
confint(lm)
```

```
##                    2.5 %     97.5 %
## (Intercept) -19.50947250 -2.6669981
## x             0.08932865  0.2989066
```

```
res <- cor.test(data$x, data$y,
                    method = "pearson")
res
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$x and data$y
## t = 4.3804, df = 7, p-value = 0.003233
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
##  0.4446003 0.9691587
## sample estimates:
##       cor
## 0.8559784
```

Confidence intervals for intercept and slope, respectively:

$$[-19.50947250, -2.6669981]$$

$$[0.08932865, 0.2989066]$$

Using Pearson's correlation test we reject the hypothesis that y and x don't have a linear relationship.

Problem 2

$$\mathrm{Var}[\hat{\beta}_1] = \frac{\sigma^2}{SS_{XX}}$$

$$\mathrm{Cov}[\hat{\beta}_0, \hat{\beta}_1] = \mathrm{Cov}[\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1] = \mathrm{Cov}[\bar{y}, \hat{\beta}_1] - \bar{x}\mathrm{Var}[\hat{\beta}_1]$$

$$\mathrm{Cov}[\bar{y}, \hat{\beta}_1] = \mathrm{Cov}\left[\bar{y}, \frac{\sum_i y_i x_i - \left(\sum_i y_i\right)\left(\sum_i x_i\right)/n}{SS_{XX}}\right] = \frac{1}{nSS_{XX}} \sum_i \mathrm{Cov}\left[y_i, \sum_i y_i x_i - \left(\sum_i y_i\right)\left(\sum_i x_i\right)/n\right] =$$

$$= \frac{1}{nSS_{XX}} \sum_i \left(\mathrm{Cov}\left[y_i, \sum_i y_i x_i\right] - \mathrm{Cov}\left[y_i, \left(\sum_i y_i\right)\left(\sum_i x_i\right)/n\right]\right) =$$

$$= \frac{1}{nSS_{XX}} \sum_i \left(\sigma^2 x_i - \sigma^2 \left(\sum_i x_i\right)/n\right) = \frac{\sigma^2}{nSS_{XX}} \left(\sum_i x_i - \left(\sum_i x_i\right)\right) = 0$$

$$\Rightarrow \mathrm{Cov}[\hat{\beta}_0, \hat{\beta}_1] = -\bar{x}\mathrm{Var}[\hat{\beta}_1] = -\frac{\bar{x}\sigma^2}{SS_{XX}}.$$