Homework 2

Louis Bensard

February 22, 2018

Problem 1

The method used to sample the points is the Inversion Sampling Method. I used it because I know that the cdf of the exponential density is easily invertible, as shown below:

If $X \sim exp(2)$, then $F_X(x) = 1 - e^{-x/2} \Rightarrow x = 1 - e^{-F_X^{-1}(x)/2}$. Thus,

$$F_X^{-1}(x) = -2ln(1-x), \quad 0 < x < 1$$

Therefore, if we sample 1000 observations from the $uniform[0,1]$ distribution, such that $U_i \sim unif[0,1] \quad (iid)$, then $F_X^{-1}(U_i) \sim f_X(x) = exp(2)$.

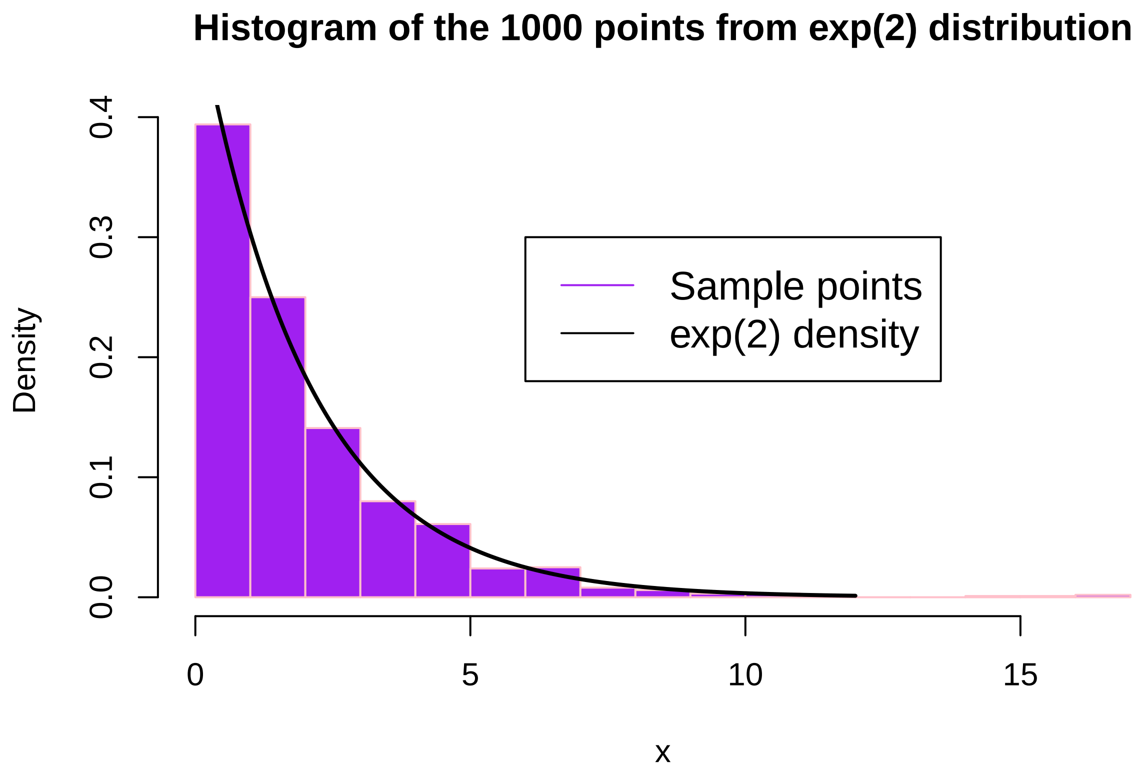
```r
n=1000

u_vect = runif(n)
F_inv_vect = -2*log(1-u_vect)

x = seq(0,12, length=n)
rexp_vect = rexp(n, rate=1/2)

hist(F_inv_vect, freq=F, col = "purple", border = "pink", breaks = 20,
    main = paste("Histogram of the 1000 points from exp(2) distribution"),
    xlab="x")

lines(x, (1/2)*exp(-x/2), lwd=2)
legend(6,0.3,c("Sample points", "exp(2) density"), col=c("purple", "black"),
      lty=1, cex= 1.25)
```



Problem 2

(i)

PROVE the least squares estimate of β is given by:

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

We want to find the $\hat{\beta}$ that minimizes $SSE = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta} x_i)^2$. Now let's set the derivative of SSE with respect to $\hat{\beta}$ equal to 0, and solve for $\hat{\beta}$:

$$\frac{\partial SSE}{\partial \hat{\beta}} = \sum_{i=1}^{n} (-2x_i)(y_i - \hat{\beta} x_i) = 0$$

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

Moreover, $\frac{\partial^2 SSE}{\partial \hat{\beta}^2} = \sum_{i=1}^{n} 2x_i^2 > 0$. Therefore $\hat{\beta}$ is a minimum.

(ii)

PROVE under the above assumptions that…

(a)

$E(\hat{\beta}|X) = \beta$

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

(b)

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

$$\begin{aligned} Var(\hat{\beta}|X) &= [(x^T x)^{-1} x^T][(x^T x)^{-1} x^T]^T Var(Y|X) \\ &= (x^T x)^{-1} x^T x[(x^T x)^{-1}]^T Var(Y|X) = (x^T x)^{-1} Var(Y|X) \\ &= (x^T x)^{-1} Var(\beta X + \epsilon|X) = (x^T x)^{-1} Var(\epsilon|X) \\ &= (x^T x)^{-1} \sigma_{err}^2 \quad \text{(by assumption)} \\ &= \frac{\sigma_{err}^2}{\sum_{i=1}^n x_i^2} \end{aligned}$$

(c)

$\hat{\beta}|X$ *is normal.*

$\hat{\beta}|X = (x^T x)^{-1} x^T Y|X$, therefore $\hat{\beta}|X$ is a linear combination of $Y|X = \beta x + \epsilon|X$. But we know by assumption that ϵ is normally distributed, therefore $\beta x + \epsilon|X = Y|X$ is normally distributed and thus $\hat{\beta}|X$ is normal.

Problem 3

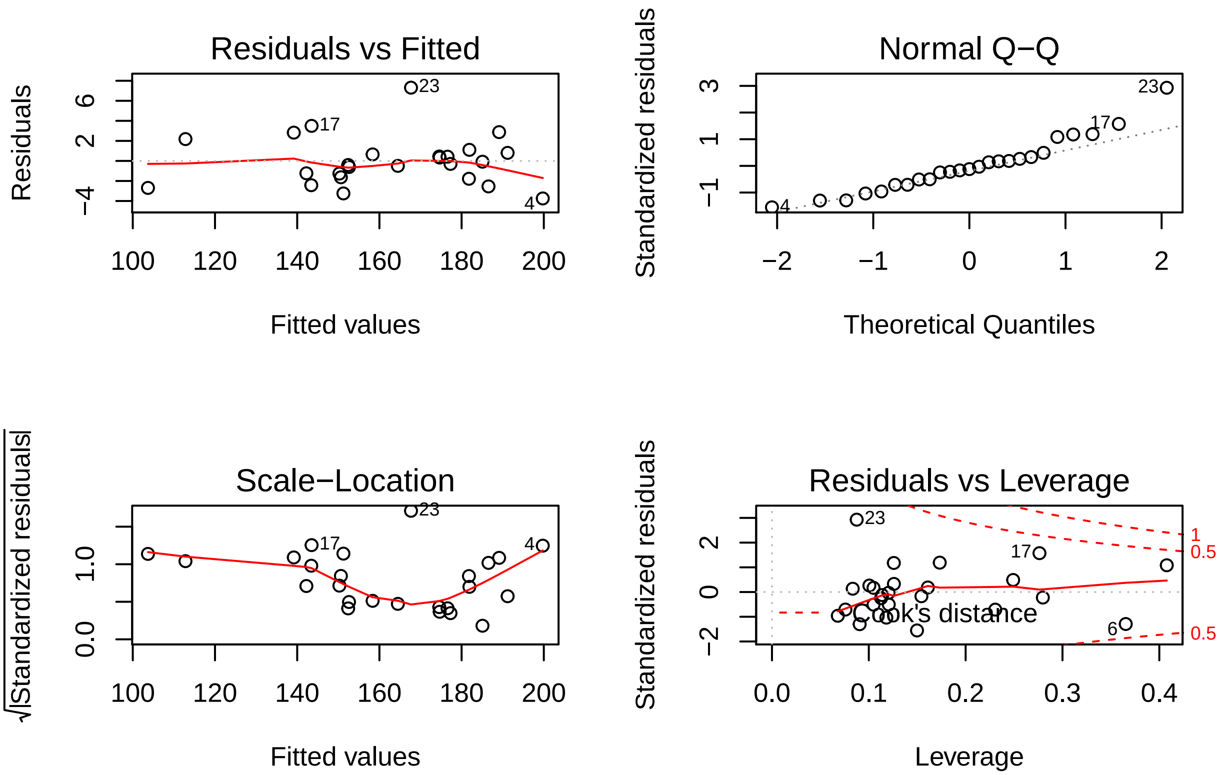
(a)

```
data = read.csv("C:/Users/Louis/Documents/UPMC/M1/Spring 2018/MATH 535/Homework 2/hw2p3.csv")

attach(data)

model1 = lm(FINAL~EXAM1+EXAM2+EXAM3)

par(mfrow=c(2,2))
plot(model1)
```



- The Residual vs Fitted values looks good, even though a small pattern can be observed, therefore the assumption$E(\hat{\beta}|X) = \beta$ seem valid.
- The Scale-Location graph does not look good, there is a clear pattern observed, therefore the constant variance assumption is invalid.
- The Normal Q-Q plot does not look good as the points are clearly not aligned. Therefore the normality assumption is invalid.
- The assumption of independence is considered always met.

(b)

Since the summary of the model gives us the p-value for $\beta_{EXAM1} = 0$ by default, then we need to compute it ourselves. We simply need to compute another t^* adapted to the new problem and then compute the new p-value. We know that for a null hypothesis: $H_0 : \beta_{EXAM1} = c$, then:

$$t^* = \frac{\hat{\beta}_{EXAM1} - c}{Std.Error(\hat{\beta}_{EXAM1})}$$

The following code computes this new p-value:

```
#just getting the values we need in the summary table
new_t_star = (coef(summary(model1))[2,1] - 0.2)/coef(summary(model1))[2,2]

new_p_value = 2*pt(abs(new_t_star), length(FINAL)-3-1, lower.tail = FALSE)

print(new_p_value)
```

```
## [1] 0.2129166
```

($p - value > 0.05$, then we do not reject the null hypothesis).

(c)

```r
predict(model1,newdata=data.frame(EXAM1=70,EXAM2=70,EXAM3=70),
        interval="confidence")
```

```
##       fit      lwr      upr
## 1 140.277 138.7054 141.8485
```

Therefore, a 95% confidence interval for the average of all individuals who score 70's on all three previous exams is [138.7054, 141.8485].

(d)

```r
predict(model1,newdata=data.frame(EXAM1=70,EXAM2=70,EXAM3=70),
        interval="prediction")
```

```
##       fit      lwr      upr
## 1 140.277 134.6191 145.9348
```

Therefore, a 95% prediction interval for an individual who scores 70 on their first three exams is [134.6191, 145.9348].