1. If F is a F-variate follows $F(a,b)$. then verify:

(i) $X = \dfrac{aF/b}{1+aF/b} \sim Beta(a/2, b/2)$

(ii) $E(B) = \dfrac{a}{a+b}$

pf: (i) $X = \dfrac{aF/b}{1+aF/b}$, $x \in (0,1)$ $\Leftrightarrow F = \dfrac{b}{a} \cdot \dfrac{x}{1-x}$, $\dfrac{\partial F}{\partial x} = \dfrac{b}{a} \cdot \dfrac{1}{(1-x)^2}$

pdf of F: $f(y) = B(\frac{a}{2}, \frac{b}{2})^{-1} (\frac{a}{b})^{\frac{a}{2}} \cdot y^{\frac{a}{2}-1} (1+\frac{a}{b}y)^{-\frac{a+b}{2}}$, $B(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

$\Rightarrow$ pdf of X: $g(x) = B(\frac{a}{2}, \frac{b}{2})^{-1} (\frac{a}{b})^{\frac{a}{2}} \cdot (\frac{b}{a} \cdot \frac{x}{1-x})^{\frac{a}{2}-1} (1+\frac{a}{b} \cdot \frac{b}{a} \frac{x}{1-x})^{-\frac{a+b}{2}} \cdot \frac{b}{a} \frac{1}{(1-x)^2}$

$= B(\frac{a}{2}, \frac{b}{2})^{-1} (\frac{x}{1-x})^{\frac{a}{2}-1} \cdot (\frac{1}{1-x})^{-\frac{a}{2}-\frac{b}{2}+2}$

$= B(\frac{a}{2}, \frac{b}{2})^{-1} x^{\frac{a}{2}-1} (1-x)^{\frac{b}{2}-1}$, $x \in (0,1)$

$\Rightarrow X \sim Beta(\frac{a}{2}, \frac{b}{2})$ $\square$

(ii) $E(X) = \int_0^1 x g(x)\, dx$

$= \int_0^1 B(\frac{a}{2}, \frac{b}{2})^{-1} x^{(\frac{a}{2}+1)-1} (1-x)^{\frac{b}{2}-1} dx$

$= \dfrac{B(\frac{a}{2}+1, \frac{b}{2})}{B(\frac{a}{2}, \frac{b}{2})} \cdot \int_0^1 B(\frac{a}{2}+1, \frac{b}{2})^{-1} x^{(\frac{a}{2}+1)-1} (1-x)^{\frac{b}{2}-1} dx$

$= B(\frac{a}{2}+1, \frac{b}{2}) / B(\frac{a}{2}, \frac{b}{2}) \cdot 1$

$= \dfrac{\Gamma(\frac{a}{2}+1)\Gamma(\frac{a}{2}+\frac{b}{2})}{\Gamma(\frac{a}{2})\Gamma(\frac{a}{2}+\frac{b}{2}+1)}$

$= \frac{a}{2} \cdot (\frac{a}{2}+\frac{b}{2})^{-1}$

$= \dfrac{a}{a+b}$ $\square$

2.

table 1

```
data1 <- data.frame(matrix(c(1,2,3,4,5,6,15,37,52,59,83,92), nrow = 6, ncol = 2))
```

```
mod1 <- lm(X2 ~ 1 + X1, data1)
mod2 <- lm(X2 ~ 0 + X1, data1)
mod3 <- lm(log(X2) ~ 1 + log(X1), data1)
```

mod1

```
r11 <- 1 - sum(mod1$residuals^2)/sum((data1$X2 - mean(data1$X2))^2)
r12 <- sum((mod1$fitted.values - mean(data1$X2))^2)/sum((data1$X2 - mean(data1$X2))^2)
r13 <- sum((mod1$fitted.values - mean(mod1$fitted.values))^2)/sum((data1$X2 - mean(data1$X2))^2)
r14 <- 1 - sum((mod1$residuals - mean(mod1$residuals))^2)/sum((data1$X2 - mean(data1$X2))^2)
r15 <- (cor(data1$X1, data1$X2))^2
r16 <- (cor(data1$X1, mod1$fitted.values))^2
r17 <- 1 - sum(mod1$residuals^2)/sum(data1$X2^2)
r18 <- sum(mod1$fitted.values^2)/sum(data1$X2^2)
```

```
rmse1 <- (sum(mod1$residuals^2)/6)^.5
mae1 <- sum(abs(mod1$residuals)/6)
mse1 <- sum(mod1$residuals^2)/(6-2)
```

mod2

```
r21 <- 1 - sum(mod2$residuals^2)/sum((data1$X2 - mean(data1$X2))^2)
r22 <- sum((mod2$fitted.values - mean(data1$X2))^2)/sum((data1$X2 - mean(data1$X2))^2)
r23 <- sum((mod2$fitted.values - mean(mod2$fitted.values))^2)/sum((data1$X2 - mean(data1$X2))^2)
r24 <- 1 - sum((mod2$residuals - mean(mod2$residuals))^2)/sum((data1$X2 - mean(data1$X2))^2)
r25 <- (cor(data1$X1, data1$X2))^2
r26 <- (cor(data1$X1, mod2$fitted.values))^2
r27 <- 1 - sum(mod2$residuals^2)/sum(data1$X2^2)
r28 <- sum(mod2$fitted.values^2)/sum(data1$X2^2)
```

```
rmse2 <- (sum(mod2$residuals^2)/6)^.5
mae2 <- sum(abs(mod2$residuals)/6)
mse2 <- sum(mod2$residuals^2)/(6-1)
```

mod3

```
r31 <- 1 - sum((exp(mod3$fitted.values) - data1$X2)^2)/sum((data1$X2 - mean(data1$X2))^2)
r32 <- sum((exp(mod3$fitted.values) - mean(data1$X2))^2)/sum((data1$X2 - mean(data1$X2))^2)
r33 <- sum((exp(mod3$fitted.values) - mean(exp(mod3$fitted.values)))^2)/sum((data1$X2 - mean(data1$X2))
r34 <- 1 - sum(((exp(mod3$fitted.values) - data1$X2)
                - mean((exp(mod3$fitted.values) - data1$X2)))^2)/sum((data1$X2 - mean(data1$X2))^2)
r35 <- (cor(log(data1$X1), log(data1$X2)))^2
r36 <- (cor(data1$X2, exp(mod3$fitted.values)))^2
r37 <- 1 - sum((exp(mod3$fitted.values) - data1$X2)^2)/sum(data1$X2^2)
r38 <- sum(exp(mod3$fitted.values)^2)/sum(data1$X2^2)
```

```
rmse3 <- (sum((exp(mod3$fitted.values) - data1$X2)^2)/6)^.5
mae3 <- sum(abs(exp(mod3$fitted.values) - data1$X2))/6
```

```r
mse3 <- sum((exp(mod3$fitted.values) - data1$X2)^2)/(6-2)
```

mod4

```r
data2 <- data.frame(matrix(c(6,7,8,9,10,11,12,13,3882,1266,733,450,410,305,185,112)
                           , nrow = 8, ncol = 2))
```

```r
data2$X2 <- data2$X2/7343
```

```r
mod4 <- lm(log(X2) ~ log(X1), data2)
```

```r
r41 <- 1 - sum((exp(mod4$fitted.values) - data2$X2)^2)/sum((data2$X2 - mean(data2$X2))^2)
r42 <- sum((exp(mod4$fitted.values) - mean(data2$X2))^2)/sum((data2$X2 - mean(data2$X2))^2)
r43 <- sum((exp(mod4$fitted.values) - mean(exp(mod4$fitted.values)))^2)/sum((data2$X2 - mean(data2$X2))
r44 <- 1 - sum(((exp(mod4$fitted.values) - data2$X2)
                - mean((exp(mod4$fitted.values) - data2$X2)))^2)/sum((data2$X2 - mean(data2$X2))^2)
r45 <- (cor(log(data2$X1), log(data2$X2)))^2
r46 <- (cor(data2$X2, exp(mod4$fitted.values)))^2
r47 <- 1 - sum((exp(mod4$fitted.values) - data2$X2)^2)/sum(data2$X2^2)
r48 <- sum(exp(mod4$fitted.values)^2)/sum(data2$X2^2)
```

```r
rmse4 <- (sum((exp(mod4$fitted.values) - data2$X2)^2)/8)^.5
mae4 <- sum(abs(exp(mod4$fitted.values) - data2$X2))/8
mse4 <- sum((exp(mod4$fitted.values) - data2$X2)^2)/(8-2)
```

result

```r
table1 <- data.frame(row.names = c("b0", "b1","Rsq1","Rsq2","Rsq3"
                                   ,"Rsq4","Rsq5","Rsq6","Rsq7","Rsq8"
                                   ,"RMSE", "MAE", "MSE"))
```

```r
table1$mod1 <- t(t(c(mod1$coefficients, r11, r12,r13,r14,r15,r16,r17,r18,rmse1,mae1,mse1)))
table1$mod2 <- t(t(c(NA, mod2$coefficients, r21, r22,r23,r24,r25,r26,r27,r28,rmse2,mae2,mse2)))
table1$mod3 <- t(t(c(exp(mod3$coefficients[1]), mod3$coefficients[2],
                     r31, r32,r33,r34,r35,r36,r37,r38,rmse3,mae3,mse3)))
table1$mod4 <- t(t(c(exp(mod4$coefficients[1]), mod4$coefficients[2],
                     r41, r42,r43,r44,r45,r46,r47,r48,rmse4,mae4,mse4)))
```

```r
print(table1)
```

```
##            mod1       mod2       mod3          mod4
## b0    3.3333333         NA 16.3756622 594.619866683
## b1   15.1428571 15.9120879  0.9900216  -4.082594933
## Rsq1  0.9808189  0.9776853  0.9777150   0.901851109
## Rsq2  0.9808189  1.0836003  1.0983583   0.585771192
## Rsq3  0.9808189  1.0829977  1.0983013   0.582510144
## Rsq4  0.9808189  0.9782880  0.9777719   0.905112158
## Rsq5  0.9808189  0.9808189  0.9816110   0.966777175
## Rsq6  1.0000000  1.0000000  0.9810787   0.949777500
## Rsq7  0.9966075  0.9960532  0.9960585   0.939182129
## Rsq8  0.9966075  0.9960532  1.0231547   0.687877041
## RMSE  3.6165405  3.9007842  3.8981900   0.049984275
## MAE   3.5238095  3.6520147  3.6334210   0.028314020
## MSE  19.6190476 18.2593407 22.7938279   0.003331237
```

3. Commentary to Reading Material 2

In paper *Another Cautionary Note About R^2: Its Use in Weighted Leaset-Squares Regression Analysis*, the author discussed why R^2 is a problematic measure of goodness of fitness under senarios of weighted least square estimation. Besides the theoriotical explanation, he furtherly introduced a new measure, pseudo WLS R^2, and conducted a numerical experiment to demonstrate how R^2 can fail representing goodness of fitness in WLS situations.

For the ordinary least square estimate of linear regression models, we always report R^2, the percentage of variation in outcome explained by the predictors, as a measure of goodness of fit. However, OLS are valid only when we assume the error (noise in the outcome) terms follows normal distribution with mean zero and same variance independently. When the error terms have varince of different quantity or have dependency on each other, we need to adjust the estimation by assigning different weights to observations. To do so, we only have to make linear transformation for both the outcome and design matrix based on our knowledge about the distribution of error terms, and conduct ordinary least square estimation on the transformed dataset. The process is defined as "Weighted Least Square" estimation.

If we use any mainstream statistical program to conduct WLS for our data, the R^2 provided are mostly calculated directly from the transformed outcome and predictors, whose value is usually larger than the OLS R^2 obtained directly from the unaltered datase. Since the data transformation can change the magnitude and scale, there is no point comparing OLS and WLS R^2, thus WLS R^2 having a larger value does not indicate a better fit.

To truly compare the goodness of fitness of the OLS and WLS estimation, the author introduced pseudo R^2 for WLS, which is calculated similarly to R^2 for OLS, using untransformed outcome and predictors, but replacing the estimated coefficients with that from WLS estimation. In this way, pseudo WLS R^2 can measure the fitness of WLS model with regard to the original data, therefore comparable to OLS R^2. The numerical example then illustrated that although WLS R^2 is seemingly larger than OLS R^2, the pseudo WLS R^2 is slightly smaller than OLS R^2. This example emphasizes that R^2 from WLS can mistakenly overestimate the goodness of fitness, also indicates that when comparing WLS to OLS, we shall focus on the improvement in other aspects, such as precision of estimates of coefficients etc.

4. Partial Correlation

```
data3 <- data.frame(y = c(1,2,3,5,3,1,5,0,6,3,7,4),
                    x = c(1,1,1,1,2,2,3,3,4,4,5,5),
                    z = c(12,14,16,16,18,16,12,12,10,12,10,16))
```

```
r_y1 <- cor(data3$y, data3$x)
r_y2 <- cor(data3$y, data3$z)
r_12 <- cor(data3$x, data3$z)
r_y2.1 <- (r_y2 - r_y1 * r_12)/sqrt((1 - r_y1^2) * (1 - r_12^2))
```

```
# R^2 from MLR
summary(lm(y ~ x + z, data3))$r.squared
```

```
## [1] 0.2538929
```

```
# R^2 from partial correlation
r_y1^2 + r_y2.1^2 * (1 - r_y1^2)
```

```
## [1] 0.2538929
```

5.4c/2 For generalized linear full-rank regression model, prove that R^2 and the F-statistic for testing $H: \beta_j = 0$ $(j \neq 0)$, are independent of the units in which the Y_i and the x_{ij} are measured.

pf: that is to if $Y_i^* = c_0 Y_i$; $x_{ij}^* = c_j x_{ij}$, $F^* = \dfrac{(RSS_H - RSS)/q}{RSS/(n-p)} = F$

in matrix form: $Y^* = c_0 \cdot Y$

$$X^* = X \cdot \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & c_j & \\ & & & \ddots \\ & & & & 1 \end{pmatrix} = X \cdot K$$

$H_0: \beta_j = 0$, let $X^{(-j)} = \begin{pmatrix} 1 & \cdots & x_{j-1,1} & x_{j+1,1} & \cdots \\ \vdots & & \vdots & \vdots & \\ 1 & \cdots & x_{j-1,n} & x_{j+1,n} & \cdots \end{pmatrix}$

before unit change: $RSS = Y'(I-P)Y$, $P = X(X'X)^{-1}X'$

$RSS_H = Y'(I-P_H)Y$, $P_H = X^{(-j)}(X^{(-j)'}X^{(-j)})^{-1}X^{(-j)'}$

after unit change: $RSS^* = Y^{*'}(I-P^*)Y^*$, $P^* = X \cdot K(K'X'XK)^{-1}KX'$

$= XK \cdot K^{-1}(X'X)^{-1}K' K \cdot X'$

$= X(X'X)^{-1}X = P$

$RSS_H^* = Y^*(I-P_H^*)Y^*$, $P_H^* = X^{(-j)*}(X^{(-j)*'}X^{(-j)*})X^{(-j)*)}$

since $X^{(-j)*} = \begin{pmatrix} 1 & \cdots & x_{j-1,1}, & x_{j+1,1}, & \cdots \\ \vdots & & \vdots & \vdots & \\ 1 & \cdots & x_{j-1,n}, & x_{j+1,n}, & \cdots \end{pmatrix} = X^{(-j)}$

$\Rightarrow P_H^* = P_H$

$\Rightarrow RSS^* = c_0^2 Y'(I-P)Y = c_0^2 RSS$

$RSS_H^* = c_0^2 Y'(I-P_H)Y = c_0^2 RSS_H$

$$F^* = \frac{(RSS^* - RSS_H^*)/1}{RSS^*/(n-p)} = \frac{(RSS - RSS_H)/1}{RSS/(n-p)} = F$$

$\Rightarrow$ the F test statistic is independent of the units for Y_i's and x_{ij}'s

6. In MLR, $Y \sim x_1 + \cdots + x_k$, express the partial correlation coefficients of Y and x with the linear effects of $x_1, \ldots x_k$ removed, in terms of the test statistic for whether the coefficient of x_k in the ML model is zero.

Soln: $$r_{yx_k \cdot (x_1, \ldots x_{k-1})} = \frac{(y^{\perp}, x_k^{\perp})}{\|y^{\perp}\| \cdot \|x_k^{\perp}\|}$$

$$y^{\perp} = y - P_{C(x_1, \ldots x_{k-1})} y \triangleq (I - P_{X(k)}) y$$

$P_{X(k)}$: the orthogonal projection matrix on the space spanned by $(x_1, \cdots x_k)$

$$x_k^{\perp} = x_k - P_{X(k)} x_k = (I - P_{X(k)}) x_k$$

$$\Rightarrow r^2_{yx_k \cdot (x_1, \ldots x_{k-1})} = \frac{(y'(I - P_{X(k)}) x_k)^2}{y'(I - P_{X(k)}) y \, x_k'(I - P_{X(k)}) \cdot x_k}$$

$$F = \frac{(RSS - RSS_H)}{RSS / (n-k)} = \frac{y'(I - P_X) y - y'(I - P_{X(k)}) y}{y'(I - P_X) y \; / (n-k)}$$

$$P_X = P_{C(X(k) | x_k)} = P_{X(k)} + (I - P_{X(k)}) x_k (x_k'(I - P_{X(k)}) x_k)^{-1} \cdot x_k'(I - P_{X(k)})$$

$$y' P_X y = y' P_{X(k)} y + y'(I - P_{X(k)}) x_k (x_k'(I - P_{X(k)}) x_k)^{-1} \cdot x_k'(I - P_{X(k)}) y$$

$$y'(I - P_X) y = y'(I - P_{X(k)}) y + \frac{(y'(I - P_{X(k)}) x_k)^2}{x_k'(I - P_{X(k)}) x_k}$$

$$F = (n-k)\left[1 - \frac{y'(I - P_{X(k)}) y}{y'(I - P_X) y}\right]$$

$$= (n-k)\left[1 - \frac{y'(I - P_{X(k)}) y}{y'(I - P_{X(k)}) y + \frac{(y'(I - P_{X(k)}) x_k)^2}{x_k'(I - P_{X(k)}) x_k}}\right]$$

$$= (n-k)\left[1 - \left(1 + \frac{(y'(I - P_{X(k)}) x_k)^2}{y'(I - P_{X(k)}) y \, x_k'(I - P_{X(k)}) x_k}\right)^{-1}\right]$$

$$F = (n-k)\left[1 - (1 + r^2_{yx_k \cdot (x_1, \ldots x_{k-1})})^{-1}\right]$$

$$\Rightarrow (1 + r^2_{yx_k \cdot (x_1 \ldots x_{k-1})})^{-1} = \frac{n-k-F}{n-k}$$

$$r^2_{yx_k \cdot (x_1 \ldots x_{k-1})} = \frac{F}{n-k-F}$$