Homework 1 Law of Total Expectation

```
>rm(list=ls())
> a<-read.table("D:/R/data/cps09mar.txt",head=T)
> X<-a$earnings/(a$hours*a$week)
> mean(X)
[1] 23.90266
> index_f<-which(a$female==1)
> index_m<-which(a$female==0)
> X_f<-X[index_f]
> X_m<-X[index_m]
> mean_Xf<-mean(X_f)*mean(a$female)+mean(X_m)*(1-mean(a$female))
> mean_Xf
[1] 23.90266
```

Homework 1 Law of Total Variance

```
> var(X)
[1] 428.9483
> V_e<-(mean(X_f)-mean(X))^2*mean(a$female)+(mean(X_m)-mean(X))^2*(1-mean(a$female))
> E_v<-var(X_f)*mean(a$female)+var(X_m)*(1-mean(a$female))
> V_e+E_v
[1] 428.9554
```

- $var()$ 函数计算的为无偏方差, 而原公式中的 $Var()$ 指总体方差, 因此结果不一致. 下面计算总体方差.

```
> nrow(a[which(a$female==1),])
[1] 21602
> nrow(a[which(a$female==0),])
[1] 29140
> E_v<-var(X_f)*21601/21602*mean(a$female)+var(X_m)*29139/29140*(1-mean(a$female))
> V_e+E_v
[1] 428.9399
> nrow(a)
[1] 50742
> var(X)*50741/50742
[1] 428.9399
```

Homework 2

无截距项的一元线性回归模型的数学形式为:

$$y = \beta_1 x + \varepsilon; \tag{1}$$

通常假定:

$$\begin{cases} E(\varepsilon) = 0 \\ Var(\varepsilon) = \sigma^2 \end{cases} ;$$

对 (1) 式两端求条件期望, 得到回归方程:

$$E(y|x) = \beta_1 x.$$

如果获得 n 组样本观测值 $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, 样本模型:

$$y_i = \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \cdots, n; \tag{2}$$

满足 $Gauss-Markov$ 条件:

$$\begin{cases} E(\varepsilon_i) = 0, Var(\varepsilon_i) = \sigma^2, \quad i = 1, 2, \cdots, n \\ Cov(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j \end{cases} ;$$

对 (2) 两端分别求期望和方差, 得

$$E(y_i) = \beta_1 x_i, \quad Var(y_i) = \sigma^2, \quad i = 1, 2, \cdots, n.$$

$E(y_i) = \beta_1 x_i$ 从平均意义上表达了变量 y 与 x 的统计规律性。
用 $\hat{\beta}_1$ 表示 β_1 的估计值，获得 y 关于 x 的一元线性经验回归方程

$$\hat{y} = \hat{\beta}_1 x.$$

求最小二乘估计即求参数 β_1 的估计值使离差平方和达到极小, 即

$$\begin{aligned} Q(\hat{\beta}_1) &= \sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i)^2 \\ &= \min_{\beta_1} \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2 \end{aligned}$$

$$\frac{dQ}{d\beta_1}|_{\beta_1 = \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i) x_i = 0;$$

可以得到残差的性质:

$$\sum_{i=1}^{n} x_i e_i = 0;$$

整理得正规方程:

$$(\sum_{i=1}^{n} x_i^2) \hat{\beta}_1 = \sum_{i=1}^{n} x_i y_i;$$

则

$$\hat{\beta}_1 = \frac{\sum\limits_{i=1}^{n} x_i y_i}{\sum\limits_{i=1}^{n} x_i^2}.$$