숙제1

통계학전공 3학년 안효준 (5463312)

Problem 1.

모형 $y_{ij}=\alpha+\beta x_{ij}+\epsilon_{ij}$에 대한 총평균, 처리평균, 처리제곱합, 잔차제곱합의 식

$\overline{y}_{..}=\dfrac{y_{11}+y_{12}+\cdots+y_{an}}{N}$ (단, $N=\sum_{i=1}^{a}n_i$), $\overline{y}_{i.}=\dfrac{y_{i1}+y_{i2}+\cdots+y_{in}}{n_i}$

$SS_{treat}=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(\overline{y}_{i.}-\overline{y}_{..})^2$, $SSE=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{i.})^2$

a) 모든 y_{ij}에 상수 c를 더할 경우

총평균: $\dfrac{(y_{11}+c)+(y_{12}+c)+\cdots+(y_{an}+c)}{N}=\dfrac{y_{11}+y_{12}+\cdots+y_{an}}{N}+c=\boxed{\overline{y}_{..}+c}$

처리평균: $\dfrac{(y_{i1}+c)+(y_{i2}+c)+\cdots+(y_{in}+c)}{n_i}=\dfrac{y_{i1}+y_{i2}+\cdots+y_{in}}{n_i}+c=\boxed{\overline{y}_{i.}+c}$

처리제곱합: $\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left\{(\overline{y}_{i.}+c)-(\overline{y}_{..}+c)\right\}^2=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(\overline{y}_{i.}-\overline{y}_{..})^2=\boxed{SS_{treat}}$

잔차제곱합: $\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left\{(y_{ij}+c)-(\overline{y}_{i.}+c)\right\}^2=\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{i.})^2=\boxed{SSE}$

$\therefore F_0=\dfrac{SS_{treat}/(a-1)}{SSE/(N-a)}=\boxed{\dfrac{MS_{treat}}{MSE}}$ (즉, F_0값은 변하지 않는다.)

b) 모든 y_{ij}에 상수 d를 곱할 경우

총평균: $\dfrac{(y_{11}\times d)+(y_{12}\times d)+\cdots+(y_{an}\times d)}{N}=\dfrac{y_{11}+y_{12}+\cdots+y_{an}}{N}\times d=\boxed{\overline{y}_{..}\times d}$

처리평균: $\dfrac{(y_{i1}\times d)+(y_{i2}\times d)+\cdots+(y_{in}\times d)}{n_i}=\dfrac{y_{i1}+y_{i2}+\cdots+y_{in}}{n_i}\times d=\boxed{\overline{y}_{i.}\times d}$

처리제곱합: $\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left\{(\overline{y}_{i.}\times d)-(\overline{y}_{..}\times d)\right\}^2=\sum_{i=1}^{a}\sum_{j=1}^{n_i}d^2(\overline{y}_{i.}-\overline{y}_{..})^2=\boxed{d^2SS_{treat}}$

잔차제곱합: $\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left\{(y_{ij}\times d)-(\overline{y}_{i.}\times d)\right\}^2=\sum_{i=1}^{a}\sum_{j=1}^{n_i}d^2(y_{ij}-\overline{y}_{i.})^2=\boxed{d^2SSE}$

$\therefore F_0=\dfrac{d^2SS_{treat}/(a-1)}{d^2SSE/(N-a)}=\dfrac{SS_{treat}/(a-1)}{SSE/(N-a)}=\boxed{\dfrac{MS_{treat}}{MSE}}$ (즉, F_0값은 변하지 않는다.)

Problem 2.

a) 랜덤화의 원리, b) 반복의 원리, c) 블록화의 원리

Problem 3.

반응변수: 전구 수명, 설명변수: 두 회사 제품

$\overline{y}_{1\cdot} = \dfrac{6.1+7.1+\cdots+8.2}{6} = \boxed{7.28}$, $\overline{y}_{2\cdot} = \dfrac{9.1+8.2+\cdots+7.9}{6} = \boxed{8.03}$, $\overline{y}_{\cdot\cdot} = \dfrac{7.28+8.03}{2} = \boxed{7.66}$

$$SS_{treat} = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot})^2 = 6\{(7.28-7.66)^2 + (8.03-7.66)^2\} = \boxed{1.687}$$

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij} - \overline{y}_{i\cdot})^2 = (6.1-7.28)^2 + \cdots + (7.9-8.03)^2 = \boxed{5.862}$$

요인	제곱합	자유도	평균제곱	F값	유의확률
처리	1.687	1	1.687	2.879	0.121
잔차	5.862	10	0.586		
계	7.549	11			

H_0 : 두 제품 간 평균 수명에 차이가 없다. H_0 : 두 제품 간 평균 수명에 차이가 있다.

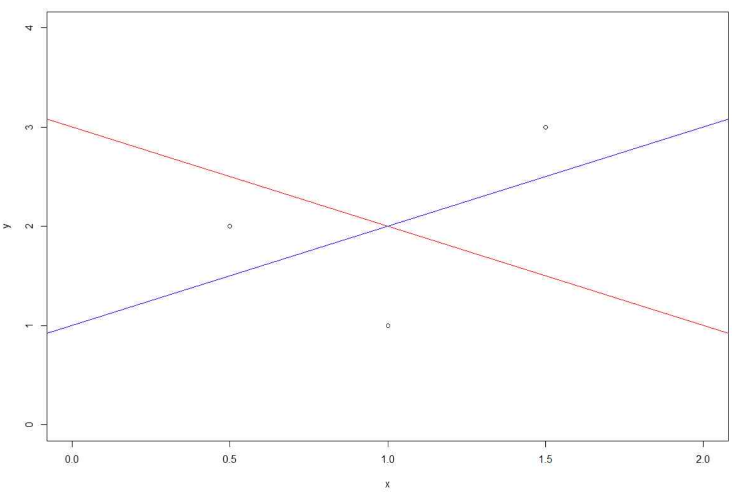
유의수준 0.05 하에서 유의확률이 0.121 이므로 귀무가설 채택

(즉, 두 제품 간 평균 수명에 차이가 없다고 할 수 있다.)

Problem 4.

a)

```
> x <- c(0.5, 1, 1.5)
> y <- c(2, 1, 3)
> plot(x, y, xlim = c(0, 2), ylim = c(0, 4))
> abline(a = 3, b = -1, col = 'red') # y = 3 - x
> abline(a = 1, b = 1, col = 'blue') # y = 1 + x
```



b) 자료를 더 잘 설명해 주고 있다고 생각하는 직선: $y = 1 + x$

이유: 산점도만 확인했을 때는 직관적으로 파란색 직선의 SSE가 더 작아보이기 때문

c) 직선 $y=3-x$의 잔차의 합: $(2-2.5)+(1-2)+(3-1.5)=-0.5-1+1.5=\boxed{0}$

직선 $y=1+x$의 잔차의 합: $(2-1.5)+(1-2)+(3-2.5)=0.5-1+0.5=\boxed{0}$

d) 직선 $y=3-x$의 SSE: $(2-2.5)^2+(1-2)^2+(3-1.5)^2=(-0.5)^2+(-1)^2+1.5^2=\boxed{3.5}$

직선 $y=1+x$의 SSE: $(2-1.5)^2+(1-2)^2+(3-2.5)^2=0.5^2+(-1)^2+0.5^2=\boxed{1.5}$

e) $\overline{x}=\dfrac{0.5+1+1.5}{3}=\boxed{1}$, $\overline{y}=\dfrac{2+1+3}{3}=\boxed{2}$, $S_{(xx)}=\sum_{i=1}^{3}(x_i-\overline{x})^2=\boxed{0.5}$

$S_{(yy)}=\sum_{i=1}^{3}(y_i-\overline{y})^2=\boxed{2}$, $S_{(xy)}=\sum_{i=1}^{3}(x_i-\overline{x})(y_i-\overline{y})=\boxed{0.5}$

$\hat{\beta}=\dfrac{S_{(xy)}}{S_{(xx)}}=\dfrac{0.5}{0.5}=\boxed{1}$, $\hat{\alpha}=\overline{y}-\hat{\beta}\overline{x}=2-1\times 1=\boxed{1}$

따라서 추정된 회귀직선은 $\hat{y}=\hat{\alpha}+\hat{\beta}x=\boxed{1+x}$

Problem 5.

$y_i=a+bx_i+\epsilon_i$, $\epsilon_i=y_i-a-bx_i$, $\sum_{i=1}^{n}\epsilon_i^2=\sum_{i=1}^{n}(y_i-a-bx_i)^2=Q(a,b)$

$\dfrac{\partial}{\partial a}Q(a,b)=-2\sum_{i=1}^{n}(y_i-a-bx_i)=0$, $n\overline{y}=na+nb\overline{x}$, $\hat{a}=\overline{y}-\hat{b}\overline{x}$

$\dfrac{\partial}{\partial b}Q(a,b)=-2\sum_{i=1}^{n}(y_i-a-bx_i)x_i=0$, $\sum_{i=1}^{n}x_iy_i=a\sum_{i=1}^{n}x_i+b\sum_{i=1}^{n}x_i^2=(\overline{y}-b\overline{x})\sum_{i=1}^{n}x_i+b\sum_{i=1}^{n}x_i^2$

$b\left(\sum_{i=1}^{n}x_i^2-\overline{x}\sum_{i=1}^{n}x_i\right)=\sum_{i=1}^{n}x_iy_i-\overline{y}\sum_{i=1}^{n}x_i$

$$\therefore\ \hat{b}=\frac{\sum_{i=1}^{n}x_iy_i-\overline{y}\sum_{i=1}^{n}x_i}{\sum_{i=1}^{n}x_i^2-\overline{x}\sum_{i=1}^{n}x_i}=\frac{\sum_{i=1}^{n}x_iy_i-n\overline{x}\,\overline{y}}{\sum_{i=1}^{n}x_i^2-n\overline{x}^2}=\frac{\sum_{i=1}^{n}x_iy_i-2n\overline{x}\,\overline{y}+n\overline{x}\,\overline{y}}{\sum_{i=1}^{n}x_i^2-2n\overline{x}^2+n\overline{x}^2}=\boxed{\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}$$