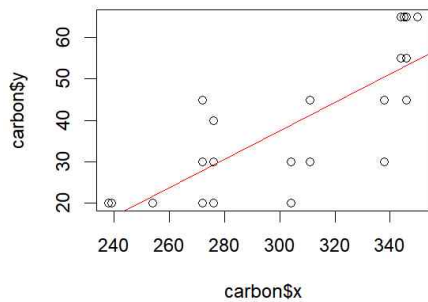


〈기초통계실습2 기말고사 대체과제〉

#1.

(1) 산점도를 그려라.



(2) 추정회귀직선을 구하라.

$$y = -65.7197811 + 0.3440518 \cdot x$$

```
> fit<-lm(data=carbon, y~x)
> coef(fit) # y = -65.7197811 + 0.3440518*x
(Intercept)          x
-65.7197811    0.3440518
```

(3) 위의 자료가 단위면적당 탄소 알갱이의 수가 많을수록 열전도율이 커진다는 주장을 뒷받침해 줄 수 있을지 검정하라. (유의수준 $\alpha=0.05$)

```
> tobs=qt(0.025, df=fit$df.residual, lower.tail = F);tobs
[1] 2.085963
> Sxx=sum((carbon$x-mean(carbon$x))^2);Sxx
[1] 31195.09
> t<-(fit$coef[2]-0)/(sigma(fit)/sqrt(Sxx));t
      x
5.742951
> t>tobs
      x
TRUE
# t 검정통계량이 기준치 보다 크기 때문에 주장을 뒷받침 할 수 있다.
```

(4) $x=300$ 일 때, 평균 열전도율에 대한 95% 신뢰구간을 구하라.

```
> predict(fit, interval = "confidence",method="confidence",newdata =  
data.frame(x=300),level=0.95)  
      fit      lwr      upr  
1 37.49576 32.75449 42.23703
```

(5) 총제곱합의 분해를 제시하고, 선형회귀모형의 적합성에 대하여 평가하라.

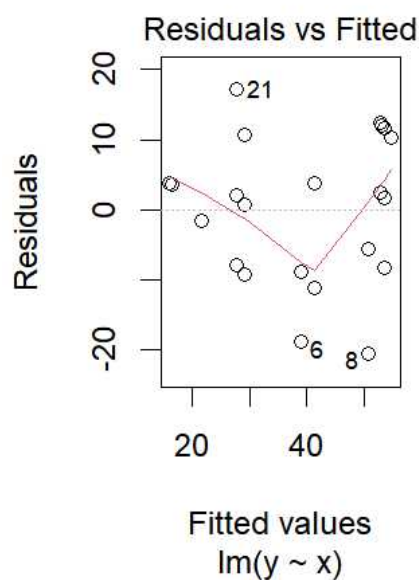
```
> SSE=sum(resid(fit)^2)  
> SST=sum((carbon$y-mean(carbon$y))^2)  
> SSR=SST-SSE  
> rbind(c(SST=SST,SSE=SSE,SSR=SSR))  
      SST      SSE      SSR  
[1,] 5931.818 2239.204 3692.614  
> R_squared=SSR/SST;R_squared  
[1] 0.6225097
```

반응변수 변동의 62.25%가 회귀모형으로 설명된다고 볼 수 있다.

(6) 잔차와 예측값의 산점도를 그리고 선형회귀모형의 가정 중에서 위배되는지를 검토하라.

```
plot(resid(fit)~fitted(fit))
```

그림에서 특별한 형태 없이 잘 산개해 있다고 볼 수 있다.



#2.국제 표준협회에서는 네 가지 직물에 대한 인화성을 검사하였다. 다음 자료는 각 직물로 만든 옷에서 잘라낸 조각에 불을 붙여 인화되는 시간을 초 단위로 기록한 것이다.

```
> textile<-read.csv("textile.csv")
>
> dat=data.frame(y=as.numeric(as.matrix(textile)),textile=rep(1:4,each=5))
> m=aov(y~textile, data=dat)
> summary(m)
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|-----------|
| textile | 1 | 63.36 | 63.36 | 11.03 | 0.0038 ** |
| Residuals | 18 | 103.40 | 5.74 | | |

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#P-value가 0.0038로 0.05보다 작기 때문에 귀무가설을 기각하고 대립가설을 채택할 수 있다.

#즉 인화성이 textile마다 서로 다르다고 볼 수 있다.

#3.

(1) 유의수준 $\alpha=0.05$ 로 "헌혈한 사람의 혈액형은 같은 비율이다" 라는 가설을 검정하라.

```
> blood<-read.csv("blood.csv")
> O<-blood[2:5]
> E<-c(0.25,0.25,0.25,0.25)*sum(O)
> Chi=sum(((O-E)^2)/E);Chi
[1] 46.41667
> Chi>qchisq(0.05, df=3, lower.tail = F)
[1] TRUE
```

#비율이 모두 같다는 귀무가설을 기각할 수 있다.

(2) "헌혈한 사람의 혈액형 O, A, B, AB 가 4 : 4 : 1 : 1 의 비율로 분포한다" 는 가설을 유의수준 $\alpha=0.05$ 로 검정하라.

```
> O<-read.csv("blood.csv")[2:5]
> E<-c(0.4,0.4,0.1,0.1)*sum(O)
> Chi=sum(((O-E)^2)/E);Chi
[1] 2.776042
> Chi>qchisq(0.05, df=3, lower.tail = F)
[1] FALSE
```

#비율이 4:4:1:1 이라는 귀무가설을 기각할 수 있다.

#4.4. 새로운 상품이 시장에 소개되고 일정기간이 지난 후, 제조업자는 그 상품이 시장에서 어떤 위치를 차지하고 있는지 알고 싶어 다음 세 개의 도시를 선택하여 각각 200명, 150명, 300명을 임의로 선택하여 새로운 상품에 대한 질문을 한 결과가 자료와 같다.

Data: goods.csv

새로운 상품의 시장에서의 위치는 도시에 따라 차이가 있는가?

방법1.

```
> goods<-read.csv("goods.csv")
> col_total<-goods[4,2:4]
> n<-sum(col_total)
> E1=goods$total[1]*col_total/n
> E2=goods$total[2]*col_total/n
> E3=goods$total[3]*col_total/n
>
Chi=sum((goods[1,2:4]-E1)^2/E1,(goods[2,2:4]-E2)^2/E2,(goods[3,2:4]-E3)^2/E3);Chi
[1] 24.60797
> Chi>qchisq(0.05, df=4, lower.tail = F)
[1] TRUE
```

방법2.

```
> chisq.test(goods[1:3,2:4])
```

Pearson's Chi-squared test

data: goods[1:3, 2:4]

X-squared = 24.608, df = 4, p-value
= 6.032e-05

```
> chisq.test(goods[1:3,2:4])$p.value<=0.05
[1] TRUE
```

#결국, 귀무가설을 기각하고 대립가설을 채택할 수 있다.

#즉, 도시에 따른 결과에 차이가 있다고 볼수 있다.

#5. 미군에 음식을 공급하는 부서에서 군인들의 출신지역과 검정 올리브를 좋아하는 것과의 관계를 조사하기 위해 435명을 임의로 추출해 다음과 같은 자료를 얻었다. Data: olive.csv

출신지역과 검정 올리브의 선호 여부와 어떤 관계가 있는가? (유의수준 $\alpha=0.05$)

```
> olive<-read.csv("olive.csv")
> olive<-data.frame(olive[,2],olive[,3]-olive[,2])
> colnames(olive)=c("likes","dislikes");rownames(olive)<-c("W","E","S","M")
>
> n<-sum(olive)
> col_total<-apply(olive, 2, sum)
> row_total<-apply(olive,1,sum)
> E_W=n*(row_total[1]/n)*(col_total/n)
> E_E=n*(row_total[2]/n)*(col_total/n)
> E_S=n*(row_total[3]/n)*(col_total/n)
> E_N=n*(row_total[4]/n)*(col_total/n)
>
>
Chi=sum((olive[1,]-E_W)^2/E_W+(olive[2,]-E_E)^2/E_E+(olive[3,]-E_S)^2/E_S+(olive[4,]-E_N)^2/E_N);Chi
[1] 4.095454
> Chi>=qchisq(0.05, df=3,lower.tail = F)
[1] FALSE

> chisq.test(olive)
```

Pearson's Chi-squared test

data: olive

X-squared = 4.0955, df = 3, p-value = 0.2513

```
> chisq.test(olive)$p.value<=0.05
```

```
[1] FALSE
```

#검정통계량의 값이 기준치보다 작고 p-value도 0.05보다 크다

#따라서 귀무가설을 기각할 수 없다.

#따라서 출신지와 올리브 선호도는 독립이라고 볼 수 있다.