

```
#####
# simple linear regression
#####

lungca<-matrix(scan('A:/reg.txt'),byrow=T,ncol=2)
lungca<-read.table('A:/reg.txt',header=T)
attach(lungca)

x<-CIG
y<-LUNG

## beta1
beta1hat<-sum( (x-mean(x))*(y-mean(y)) ) / sum( (x-mean(x))^2 )
## beta0
mean(y)-beta1hat*mean(x)

## lm
> fit<-lm(LUNG~CIG)
> summary(fit)

Call:
lm(formula = LUNG ~ CIG)

Residuals:
    Min     1Q   Median     3Q      Max
-6.943 -1.656  0.382  1.614  7.561

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.4717     2.1407   3.023  0.00425 **
CIG            0.5291     0.0839   6.306 1.44e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.066 on 42 degrees of freedom
Multiple R-Squared: 0.4864,    Adjusted R-squared: 0.4741
F-statistic: 39.77 on 1 and 42 DF,  p-value: 1.439e-07

## residuals
fit$res

## SE for beta1hat
SE1<-sqrt( sum((fit$res)^2)/( (length(x)-2)*sum((x-mean(x))^2) ) )

## decision rule
> help(qt)
> qt(.975,df=42)
[1] 2.018082

detach()

#####
# multiple linear regression
#####

infect<-read.table('A:/multreg.txt',header=T)
attach(infect)

plot(stay,risk)
plot(age,risk)
plot(surv,risk)

fit1<-lm(risk~stay+age+surv)
summary(fit1)
Call:
lm(formula = risk ~ stay + age + surv)
```

```
Residuals:
      Min       1Q   Median       3Q      Max
-2.2248 -0.7125  0.2134  0.5602  2.1870
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.95982     3.34407  -0.586  0.56331
stay         0.22458     0.10241   2.193  0.03824 *
age          0.06113     0.06329   0.966  0.34379
surv         0.06479     0.01893   3.423  0.00223 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.074 on 24 degrees of freedom
Multiple R-Squared: 0.5743, Adjusted R-squared: 0.5211
F-statistic: 10.79 on 3 and 24 DF, p-value: 0.0001111
```

```
anova(fit1)
Response: risk
      Df Sum Sq Mean Sq F value    Pr(>F)
stay   1 22.0103 22.0103 19.0782 0.0002073 ***
age    1  1.8264  1.8264  1.5831 0.2204190
surv   1 13.5189 13.5189 11.7180 0.0022271 **
Residuals 24 27.6886  1.1537
(The SS's from R are sequential.)
```

```
## R^2=SSR/SST
(22.0103+1.8264+13.5189)/(22.0103+1.8264+13.5189+27.6886)
## R^2 adjusted: 1-(SSE/(n-k-1))/(SST/(n-1))
1-(27.6886/(nrow(infect)-4))/
  ((22.0103+1.8264+13.5189+27.6886)/(nrow(infect)-1))
```

```
## overall F=MSR/MSE
((22.0103+1.8264+13.5189)/3)/(27.6886/24)=10.79
> 1-pf(10.79,df1=3,df2=24)
[1] 0.0001113451 ## p-value
```

```
## testing for beta_surv=0:
## 1. t test, t=3.423, p-value=0.00223, significant
## 2. partial F test using anova output: (13.5189/1)/(27.6886/24)=11.72,
##    p-value=1-pf(11.72,df1=1,df2=24)=0.00223
## (t and partial F are equivalent in this case, 3.423^2 = 11.72)
## 3. partial F test by fitting Full and Reduced models
```

```
fit2<-lm(risk~stay+age)
anova(fit2)
Response: risk
      Df Sum Sq Mean Sq F value    Pr(>F)
stay   1 22.010  22.010 13.3533 0.001197 **
age    1  1.826   1.826  1.1081 0.302574
Residuals 25 41.208  1.648
```

```
SSR(surv|stay,age)=41.208-27.6886=13.519
partial F*=(13.519/1)/(27.6886/24)=11.72
```

```
#####
# model diagnostics
#####
```

```
fit<-lm(risk~stay+log(surv))
plot(predict(fit),resid(fit))
qqnorm(resid(fit))
```