Module_C

Xu Shi

Sep 13, 2020

```
## here() starts at /Users/wakim/Library/CloudStorage/OneDrive-O
```

Example 2:

- A study is carried out in order to evaluate the relationship between birth weight (measured in g/100) and estriol levels (measured as mg/24h based on urine samples). Data were collected on a total of n=31 mothers-to-be and their newborns. A straight line regression model was used for analysis.
- Read in the data file (estriol1.sas7bdat), which is stored as a permanent SAS set.

```
library(sas7bdat)
data = read.sas7bdat(file="estriol1.sas7bdat")
attach(data)
```

Obtain a listing of the variables. Check if there is any missing data.

- Let's install and load a useful package called psych (or Hmisc)

```
library(psych)
```

Obtain a listing of the variables. Check if there is any missing data.

- The describe() function is handy

```
psych::describe(data)
```

```
##          vars  n  mean   sd median trimmed   mad min max
## subj        1 31 16.00 9.09     16   16.00 11.86   1  31
## estriol     2 31 17.23 4.75     16   17.28  2.97   7  27
## bwt         3 31 32.00 4.74     32   31.84  4.45  24  43
##          range skew kurtosis   se
## subj        30 0.00    -1.32 1.63
## estriol     20 0.07    -0.30 0.85
## bwt         19 0.21    -0.63 0.85
```

Fit an SLR model

```
m = lm(bwt~estriol); summary(m)
```

```
##
## Call:
## lm(formula = bwt ~ estriol)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.1200 -2.0381 -0.0381  3.3537  6.8800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  21.5234     2.6204   8.214 4.68e-09 ***
## estriol       0.6082     0.1468   4.143 0.000271 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.821 on 29 degrees of freedom
## Multiple R-squared:  0.3718, Adjusted R-squared:  0.3501
## F-statistic: 17.16 on 1 and 29 DF,  p-value: 0.0002712
```

By hand calculate parameter estiamtes

```r
X=estriol; Y=bwt
Xbar = mean(X); Ybar = mean(Y)
n = length(Y)
( SSXY = sum((Y-Ybar)*(X-Xbar)) )
```

```
## [1] 412
```

```r
cov(X,Y)*(n-1)  # compare to SSXY
```

```
## [1] 412
```

```r
( SSX =  sum((X-Xbar)^2) )
```

```
## [1] 677.4194
```

```r
var(X)*(n-1)  # compare to SSX
```

```
## [1] 677.4194
```

By hand calculate parameter estiamtes (cont'd)

```r
(beta1hat = SSXY/SSX)                    #beta1
```

```
## [1] 0.6081905
```

```r
as.numeric(coef(m)["estriol"])           #beta1 from lm
```

```
## [1] 0.6081905
```

```r
(beta0hat = Ybar-beta1hat*Xbar)          #beta0
```

```
## [1] 21.52343
```

```r
as.numeric(coef(m)["(Intercept)"]) #beta0 from lm
```

```
## [1] 21.52343
```

Obtain the F statistic and corresponding p value

- Print ANOVA table

```
anova(m)
```

```
## Analysis of Variance Table
##
## Response: bwt
##           Df Sum Sq Mean Sq F value    Pr(>F)
## estriol    1 250.57 250.574  17.162 0.0002712 ***
## Residuals 29 423.43  14.601
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By hand calculate the F statistic and corresponding p value

```r
Yhat = m$fitted.values
( SSE = sum((Y-Yhat)^2) )
```

```
## [1] 423.4255
```

```r
( SSR = sum((Yhat-Ybar)^2))
```

```
## [1] 250.5745
```

```r
## Compare to ANOVA table
anova(m)
```

```
## Analysis of Variance Table
##
## Response: bwt
##           Df Sum Sq Mean Sq F value    Pr(>F)
## estriol    1 250.57 250.574  17.162 0.0002712 ***
## Residuals 29 423.43  14.601
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

By hand calculate the F statistic and corresponding p value (cont'd)

```r
MSE = SSE/(n-2)
MSR = SSR/1
( Fstat = MSR/MSE )          ## F statistic
```

```
## [1] 17.1616
```

```r
anova(m)$F[1]                ## F statistic from ANOVA table
```

```
## [1] 17.1616
```

```r
1-pf(q=Fstat,df1=1,df2=n-2) ## P-value
```

```
## [1] 0.0002712299
```

```r
anova(m)$Pr[1]               ## P-value from ANOVA table
```

```
## [1] 0.0002712299
```

Estimate the marginal and conditional variance of birth weight.

- Marginal variance of bwt
- Variance of birth weight conditional on estriol level

```
anova(m)  #var(bwt); sd(bwt)^2; sum((Y-Ybar)^2)/(n-1); MSE
```

```
## Analysis of Variance Table
##
## Response: bwt
##           Df Sum Sq Mean Sq F value    Pr(>F)
## estriol    1 250.57 250.574  17.162 0.0002712 ***
## Residuals 29 423.43  14.601
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Does birth weight depend on estriol level? Carry out an appropriate t test

```r
sigma_squared = MSE
var_beta1hat = sigma_squared/SSX
se_beta1hat = sqrt(var_beta1hat)
t = beta1hat/se_beta1hat    ## t statistic
2*(1 - pt(q=t,df=n-2))      ## p-value
```

```
## [1] 0.0002712299
```

```r
summary(m)$coef["estriol",] ## compare to lm result
```

```
##     Estimate   Std. Error      t value     Pr(>|t|)
## 0.6081904762 0.1468117168 4.1426562497 0.0002712299
```

Use an F test to determine whether estriol levels are associated with birth weight

```
( Fstat = MSR/MSE )         ## F statistic
```

```
## [1] 17.1616
```

```
1-pf(q=Fstat,df1=1,df2=n-2) ## P-value
```

```
## [1] 0.0002712299
```

What percentage of the variation in birth weight is explained by estriol level?

```r
(R_squared = SSR / (SSR+SSE))## R-squared
```

```
## [1] 0.3717722
```

```r
summary(m)$r.squared          ## compare to R-squared from lm
```

```
## [1] 0.3717722
```