Longitudinal data

Anaranya Basu

November 2023

Problem Statement: Consider the Orthodont dataset in the "nlme" package or R. Consider age and sex as independent variable

- Find the WLS estiamtor with W=I
- Test for significance of age and sex
- Repeat first and second for different choices of W and comment.

Solution : We know that Weighted Least Square method is robust for estimating coefficients.Where W= any p.d known matrix
$\hat{\beta} = argmin(Y - X\beta)'W(Y - X\beta)$
i.e. $\hat{\beta_w} = (X'WX)^{-1}X'WY$
$Var(\hat{\beta}_w) = (X'WX)^{-1}X'WVW'X(XW'X')^{-1}$
for W=I
$\hat{\beta}_1 = (X'X)^{-1}X'Y$

Now, when W=1

```
    Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 17.70671    1.11221  15.920  < 2e-16 ***
age          0.66019    0.09776   6.753 8.25e-10 ***
sexFemale   -2.32102    0.44489  -5.217 9.20e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.272 on 105 degrees of freedom
Multiple R-squared:  0.4095,Adjusted R-squared:  0.3983
F-statistic: 36.41 on 2 and 105 DF,  p-value: 9.726e-13
```

The estimated intercept($\hat{\beta}_0$) is 17.70671, estimated coefficient for age($\hat{\beta}_{age}$) is is 0.66019, estimated coefficient for sex($\hat{\beta}_{sex}$) is -2.32102.
The entire model should be represented as

dist=17.70671+0.66019 age -2.32102 sex

Test for significance of Age and Sex

Here the null hypothesis is all the β's are zero against the alternative any of one is non zero.

Wald's test statistics = $\frac{\theta}{s.e(\theta)} \asymp N(0,1)$

Test Statistics for $\hat{\beta}_0$= 16.00105, $\hat{\beta}_{age}$= 6.851961, $\hat{\beta}_{sex}$= -5.35514

Observed P values in R are as below-

```
                      [,1]
(Intercept) 1.000000e+00
age         1.000000e+00
sexFemale   4.265686e-08
```

Comment : For the both age and sex p values is less than 0.05 so we reject the null hypothesis at the 5% level of significance and conclude that age and sex has a significance effect on distance.

Now we will repeat the entire procedure for $W = S_1$ where S_1 is -

```
  [1]  491  295  334  112  655  133  935  619  981  471  732  992  793
 [14]  426  959  239  998  519  785  642  515  302  915  648  946  795
 [27]  488  504  536  882  678  420  777  113  169  635  229  337  186
 [40]  177  339  310  385  838  944  455  442  434  812  165  267  750
 [53]  578  271  212  210  708  902  507 1000  235  724  522  387  957
 [66]  482  232  819  663  634  645  216  686  567  668  460  702  628
 [79]  197  925  535  799  398  849  479  817  188  147  541  745  452
 [92]  830  934  887  389  461  884  299  151  580  273  749  828  797
[105]  109  717  277  771
```

The summary as we get in R is attached as -

```
   Coefficients:
             Estimate Std. Error t value Pr(>|t|)
X(Intercept) 17.32774    1.00432  17.253  < 2e-16 ***
Xage          0.66185    0.08895   7.441 2.87e-11 ***
XsexFemale   -1.93040    0.41641  -4.636 1.03e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 50.43 on 105 degrees of freedom
Multiple R-squared:  0.9922,Adjusted R-squared:  0.992
F-statistic:  4459 on 3 and 105 DF,  p-value: < 2.2e-16
```

Comment: WLS estimators as we get $\hat{\beta}_0 = 17.32774, \hat{\beta}_{age} = 0.66185, \hat{\beta}_{sex} = -1.93040$

Observe that distance has a significant dependence on sex and age in both the cases. For choosing the different weight W=I,S_1 we are getting different WLS estimators but there is no changes in significant dependence.

Appendix

```
    library(nlme)
library(MASS)
Orthodont
dist= Orthodont$distance
age=Orthodont$age
sex=as.factor(Orthodont$Sex)
ols=lm(dist~age+sex)
summary(ols)
ols
X=model.matrix(dist~age+sex)
t=t(X)%*%X
t1=X%*%t(X)
solve(t)
r=resid(ols)
V=r%*%t(r)
diag(V)
v=diag(diag(V))
var_beta=solve(t)%*%t(X)%*%v%*%X%*%solve(t)
var_beta
c=ols$coefficients
c=as.matrix(c)
std_err=sqrt(diag(var_beta))
test_beta=c/std_err
test_beta=as.matrix(test_beta)
beta_0=test_beta[1,1]
beta_1=test_beta[2,1]
beta_2=test_beta[3,1]
pnorm(test_beta,0,1)

pos_weights=sample(100:1000,108,replace=F)
wls_mod=lm(dist~X-1,data=Orthodont,weights = pos_weights)
summary(wls_mod)
```