Erik Parner
11 February 2016

## Standard analysis 5-1

The purpose of this note is to present a standard multiple linear regression analysis. The data concerning the lung function PEFR, height and sex were used in Exercise 5-1.

### Statistical methods

The mean PEFR of males and females were compare using the two sample t-test. The assumption of normality was accessed using QQ-plots (quantile-quantile plots) and variance homogeneity was accessed using the F-test. The mean difference in PEFR adjusted for height was estimated using a multiple linear regression analysis. The model was checked by diagnostic plots of the residuals. The assumption of the no effect measure modification of sex by height was assessed by including an interaction between sex and height in the multiple linear regression model.

### Results and conclusion

The mean PEFR of the males was 564 (95% CI: 549-579) and of the females 474 (95% CI: 459-489), resulting in an unadjusted mean difference of males compared to females of 90 (95% CI: 69-112). When adjusting for persons height the mean PEFR difference of males as compared to females was 50 (95% CI: 24-76), suggesting that almost half of the observed mean difference in PEFR between males and females can be explained by differences in height.

### Do file

```
***********************************************************************
* Standard5-1.do
* Task: A a standard multiple linear regression analysis. The data
*      concerning the lung function PEFR, height and, sex were used
*      in Exercise 5-1.
* Erik Parner: 11-2-2016.
***********************************************************************

graph drop _all

cd "D:\Teaching\BasicBiostat\Exercises"

capture log close
log using Standard5-1.log , text replace

use PEFR,clear

***********************************************************************
* Crude comparison between males and females.
***********************************************************************
qnorm PEFR if sex==1, name(p1,replace)
qnorm PEFR if sex==2, name(p2,replace)
```

```
graph combine p1 p2
* The model model seems a very good approximation.

* Comparing the means using the normal model.
sdtest PEFR , by(sex)
ttest PEFR , by(sex)


*********************************************************************
* A height adjusted comparison between males and females.
*********************************************************************
* We fist establish a linear regression model for the females.
twoway   ///
                (scatter PEFR height if sex==1, mco(blue) msy(x) ) ///
                (lfit PEFR height if sex==1, lco(blue)  )         ///
                , ytitle("PEFR") legend(off)

generate height170=height-170
regress PEFR height170 if sex==1

* Residual plots.
predict fit if e(sample), xb
predict res if e(sample), res
scatter res fit, yline(0) mco(blue) msy(x) name(p1,replace)
scatter res height, yline(0) mco(blue) msy(x) name(p2,replace)
graph combine p1 p2
drop fit res
* Both the mean and variation of the residual seem independent of the
* individual PEFR level.

* Now, we fist establish a linear regression model for the males.
twoway   ///
                (scatter PEFR height if sex==2, mco(blue) msy(x) ) ///
                (lfit PEFR height if sex==2, lco(blue)  )         ///
                , ytitle("PEFR") legend(off)

regress PEFR height170 if sex==2

* Residual plots.
predict fit if e(sample), xb
predict res if e(sample), res
scatter res fit, yline(0) mco(blue) msy(x) name(p1,replace)
scatter res height, yline(0) mco(blue) msy(x) name(p2,replace)
graph combine p1 p2
drop fit res
* Both the mean and variation of the residual seem independent of the
* individual PEFR level.

* A combined regression model for the males and females.
regress PEFR b1.sex##c.height170
```

```
* The slope can be assumed the same for males and females.

regress PEFR b1.sex c.height170


log close
```