

## Answers to questions (laboratory session 3)

### 1. Single-factor analysis of variance

- a) How can we test if gender has a statistically significant impact on plasma retinol levels? How is the appropriate statistic calculated?

We can check the F-statistic for sex and the associated p-value.  
 $F = MS(\text{sex}) / MS(\text{residual}) = 533837.408 / 41943.4117 = 12.727563$

- b) How can we check now if there is a statistical significant gender effect on plasma retinol levels? What is the relation between the statistics used in the `anova` and `regress` commands?

We can check the t-statistic for sex and the associated p-value.  
 $F = t^2 \quad (3.568^2 \cong 12.73)$

- c) How can we calculate the best estimates for mean retinol level for women and men?

The best estimate for plasma retinol levels for women (reference level) will be equal to  $\hat{\beta}_0 = 587.7216$ , while the same estimate for males will be  
 $\hat{\beta}_0 + \hat{\beta}_1 = 587.7216 + 122.3759 = 710.0975$

- d) Calculate the best estimates for mean retinol level for women and men. Check the consistency of the results. (You can check the ANOVA model too, by using the following command: `oneway retplasm sex, tabulate` )

The default reference category is `sex==1` (male). The means for males are  $\hat{\beta}_0 = 710.0976$  and for females  $\hat{\beta}_0 + \hat{\beta}_1 = 710.0976 - 122.3759 = 587.7216$  consistent with the previous results.

```
. oneway retplasm sex, tab
```

Summary of Plasma retinol (ng/ml)			
Sex	Mean	Std. Dev.	Freq.
1	710.09756	305.52208	41
2	587.72161	185.43069	273
Total	603.70064	208.60239	314

  

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	533837.408	1	533837.408	12.73	0.0004
Within groups	13086344.5	312	41943.4117		
Total	13620181.9	313	43514.958		

Bartlett's test for equal variances: chi2(1) = 21.8659 Prob>chi2 = 0.000
--------------------------------------------------------------------------

- e) Try to notice the similarities between the two approaches.

The estimates produced by the two procedures are the same (in this case maximum likelihood and least squares give the same results).

## 2. Regression models for general two-way ANOVA

- a) Calculate the estimates for mean plasma retinol levels for each one of the six categories, which can be created by the combination of gender and vitamine use categories.

GENDER	VITAMINE USE		
	Often	Not Often	No use
Males	563.2184+ 166.3468+ 33.46968+ (-11.72721)= <b>751.30769</b>	563.2184+ 166.3468+ 39.49589+ (-255.6611)= <b>513.39999</b>	563.2184+ 166.3468= <b>729.5652</b>
Females	563.2184+ 33.46968= <b>596.68808</b>	563.2184+ 39.49589= <b>602.71429</b>	<b>563.2184</b>

- b) Compare the results listed above with those calculated in the previous question.

As we can see the results are exactly the same.

## 3. Regression models for the analysis of covariance

- a) Check the parallelism by creating an appropriate graph.

We can create two separate best-fitted lines for men and women as follows:

```
quietly xi: reg retplasm age i.sex i.sex*age
predict yhat
generate yhatm=yhat if sex==1
generate yhatf=yhat if sex==2
label var yhatm "males"
label var yhatf "females"
sort sex age
sc retplasm yhatm yhatf age, xlab() ylab() c(. 1 1) ms(o i i)
```

