Answers to questions (laboratory session 4)

1. Model Selection

a) Which t-test is equivalent with the F-test above? Try to confirm this using STATA function `ttail` to calculate the relevant p-value.

The t-test for cholesterol in the full model (t=-0.74, p=0.001) ,$0.74^2 \cong 0.55$.

```
. di 2*ttail(301,0.74)
.45987677
```

b) What is the relation between the F statistic in the Wald test and the t-statistic for cholesterol in the full GLM model?

The F statistic equals the (t-statistic)^2.

5. Model checking

a) What is your conclusion about the Normality assumption?

The p-value given by the `swilk` command is less than 0.05 so we cannot accept that the Normality assumption holds.

b) Check the normality assumption using the `swilk` command.

```
. swilk rstud

                 Shapiro-Wilk W test for normal data
 Variable |     Obs             W            V           z     Pr > z
 ---------+-------------------------------------------------------
    rstud |     314       0.99226        1.717       1.271   0.10181
```

Now we cannot reject the normality assumption. (p-value = 0.10181>0.05).

c) There are no observations with Cook's distance above 1, although there are several points with large residuals or leverage. What do you think about the "residual>2" criterion?

The number of points that we are testing for large residuals is so large, that the criterion of 2.0 or higher is probably very liberal (as 314 repeated tests are being conducted!) Thus, the fit is probably acceptable.