

# Stat 324: Lecture 24

## Testing equality of regression parameters in two groups

Moo K. Chung  
mchung@stat.wisc.edu

May 3, 2005

1. As an application of multiple regression, we show how to perform a test on equality of regression parameters in two groups. Suppose there are  $m = 43$  autistic and  $n = 32$  normal subjects. The data can be found at <http://www.stat.wisc.edu/~mchung/teaching/data/autism/>. The fractional anisotropy (FA) measure  $y_i$  is regressed over age  $x_i$  and plotted in Figure 1 for each group. Let

$$Y_i = \beta_{a0} + \beta_{a1}x_i + \epsilon_i$$

be the linear model for the autism group and

$$Y_i = \beta_{c0} + \beta_{c1}x_i + \epsilon_i$$

be the linear model for the control group.

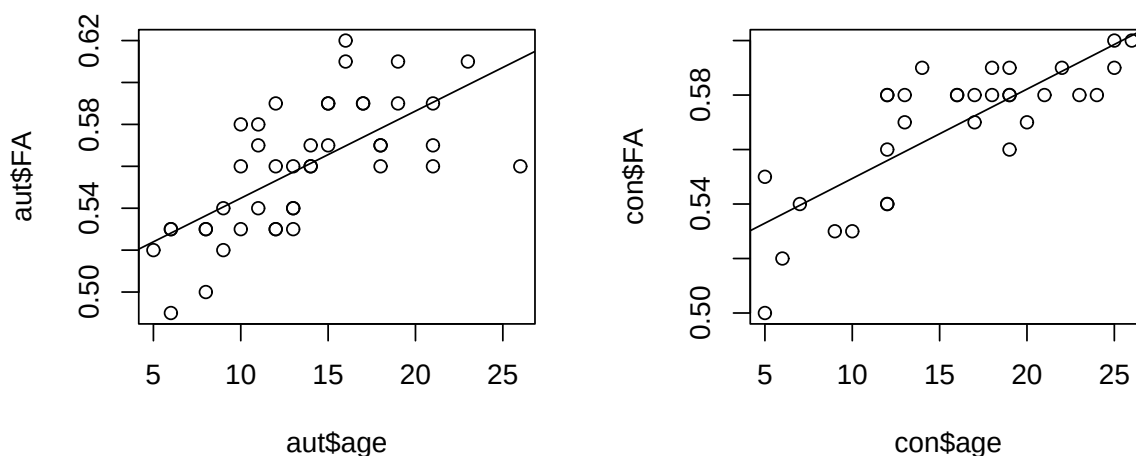


Figure 1: Regression lines for each group.

```
> summary(lm(aut$FA~aut$age))
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.5031720  0.0104110  48.331  < 2e-16 ***
aut$age      0.0041600  0.0007205   5.773  9.08e-07 ***

> summary(lm(con$FA~con$age))
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.5165714  0.0076978  67.107  < 2e-16 ***
con$age      0.0032801  0.0004563   7.189  5.33e-08 ***
```

Then we test if  $H_0 : \beta_{a1} = \beta_{c1}$  based on the two sample  $t$  test with unequal variance. We present a different method using the multiple regression framework.

2. We combine both the autistic subjects data with the normal control data in such a way that  $x_i, i = 1, \dots, m$  are ages for the autistic subjects and  $x_i, i = m+1, m+n$  are ages for the normal controls. We also combine FA measures  $y_i$  together. Then we can combine two independent linear models into a single big linear model:

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + \epsilon_i, i = 1, \dots, m+n$$

where  $z_i$  is a dummy variable taking values 0 if the  $i$ -th subject is autistic and 1 otherwise. For autistic subjects,

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \dots, m$$

For normal subjects,

$$Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x_i + \epsilon_i, i = 1, \dots, n$$

Hence we have  $\beta_{a1} = \beta_1$  and  $\beta_{c1} = \beta_1 + \beta_3$  and testing if the regression slopes of the both groups are equal is equivalent to testing the following:

$$H_0 : \beta_3 = 0 \text{ vs. } H_1 : \beta_3 \neq 0.$$

The null hypothesis corresponds to the reduced model

$$H_0 : Y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i$$

and the alternate hypothesis corresponds to the full model

$$H_1 : Y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + \epsilon_i, \beta_3 \neq 0.$$

The fit of model is measured by the sum of squared errors (SSE). Let  $SSE_0$  and  $SSE_1$  be the SSEs for the null and the alternate model respectively. Then

$$SSE_0 = \sum_{i=1}^{m+n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i)^2,$$

where  $\hat{\beta}_i$  are the least squares estimates of  $\beta_i$ .

$$SSE_1 = \sum_{i=1}^{m+n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i - \hat{\beta}_2 z_i - \hat{\beta}_3 x_i z_i)^2,$$

where  $\hat{\beta}_i$  are the least squares estimates of  $\beta_i$ . Then under  $H_0$ , the test statistic is

$$F = \frac{(SSE_0 - SSE_1)/1}{SSE_0/(m+n-1-3)} \sim F_{1, m+n-1-3}$$

```
> summary(lm(FA~age+group))
Residual standard error: 0.02007 on 72 degrees of freedom Multiple
R-Squared: 0.5126, Adjusted R-squared: 0.4991 F-statistic: 37.86
on 2 and 72 DF, p-value: 5.805e-12
```

```
> summary(lm(FA~age+group+age*group))
Residual standard error: 0.02007 on 71 degrees of freedom Multiple
R-Squared: 0.5195, Adjusted R-squared: 0.4992 F-statistic: 25.59
on 3 and 71 DF, p-value: 2.485e-11
```

```
> (0.5195-0.5126)/0.5195*(43+32-1-3)
[1] 0.9430221
> qf(0.95, 1, 43+32-1-3)
[1] 3.97581
```

In our example, the resulting  $F$  value is 0.943. Hence we can not reject  $H_0$  at level  $\alpha = 0.05$  (95% significance) and conclude that there is no difference in regression slope between the two groups.

**Assigned problems.** Exercise 13.36., Exercise 13.40. HW 7 due last day of class. Solve the assigned problems in Lecture 23-24 and Exercise 13.25 from Lecture 22. (generate a similar computer output using R).