

# Stats\_ex\_08

Dennis Perrett

12/16/2021

**Exercise 23.** Hedeker and Gibbons (2006) report on a study designed to determine the efficacy of a school-based smoking prevention curriculum in conjunction with a television- based prevention program. The study was based on a  $(2 \times 2)$ -factorial design with the factors prevention curriculum (yes = 1, no = 0) and TV prevention program (yes = 1, no = 0). Students were randomly assigned to one of the four groups at the school level. Students' knowledge of tobacco-related health risks was assessed by administering the tobacco and health knowledge scale (THKS) before and after the intervention.

- a. Set up the general model equation of an analysis of covariance (i.e., the adjusted follow-up values) for this study, using the labels  $y_{i1}$  and  $y_{i2}$  for individual  $i$ 's THKS scores before and after the intervention. In addition, let  $c_i = 1$  if person  $i$  completes the intervention with prevention curriculum ( $c_i = 0$  otherwise), and let  $tv_i = 1$  if intervention is completed with TV prevention program ( $tv_i = 0$  otherwise).

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 c_i + \beta_3 tv_i + \beta_4 c_i tv_i + \epsilon$$

- b. Interpret the parameters included in the model equation.

- $\beta_0$  represents the mean follow-up THKS score for subject not exposed to either treatment.
- $\beta_1$  is the effect of the  $y_i$  THKS scores.
- $\beta_2$  represents the mean difference from the baseline mean for subjects exposed to the  $c_i$  treatment.
- $\beta_3$  represents the mean difference from the baseline mean for subjects exposed to the  $tv_i$  treatment.
- $\beta_4$  is there effect of the interaction term  $c_i$  and  $tv_i$ , that is, what happens if someone has attended both treatments.

- c. Read the data contained in the file television.txt into R. (Note: Skip the data description at the beginning of the file in lines 1-42 by specifying the argument skip = 42 in the function read.table()).

```
dat <- read.table("../television.txt",header=T,skip=42)
riesby <- read.table("../reisby.txt",header=T)
```

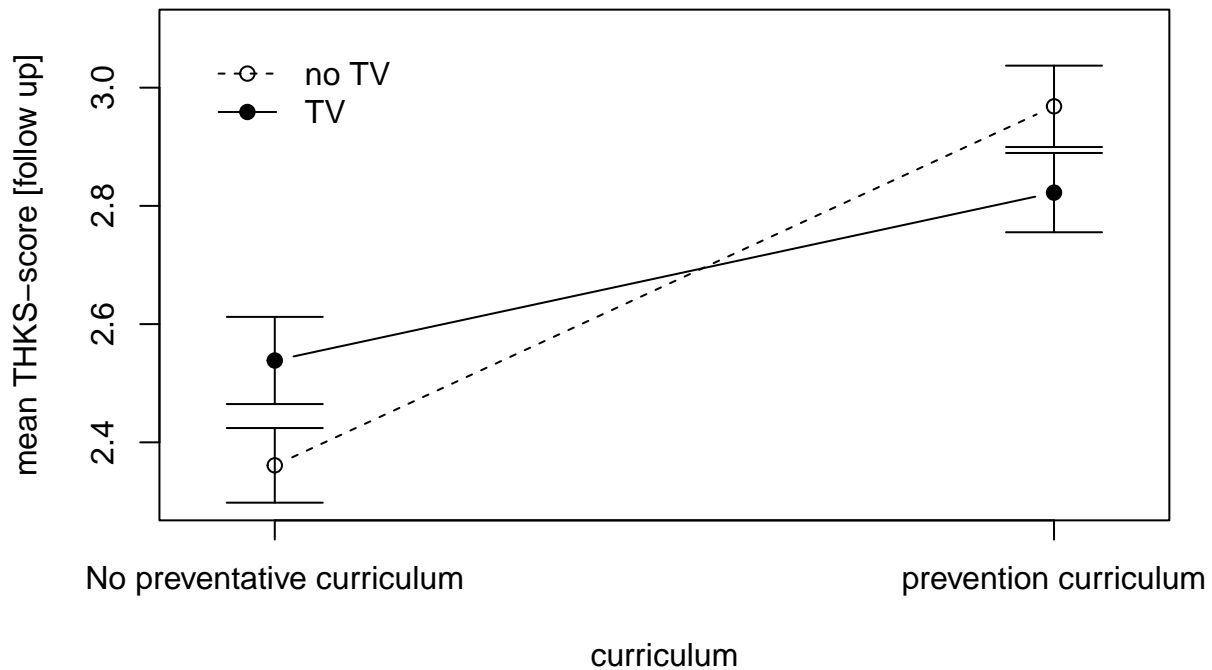
- d. Calculate means and standard deviations for baseline, follow-up, and change scores for each of the four groups separately. How many students are in each group? (Note: R help pages provide an example of the calculation of mean ozone and temperature values separately by month in the airquality data set included in R. Data may be accessed via data(airquality) and the calculation of month-specific means proceeds by aggregate(cbind(Ozone, Temp) ~ Month, data = airquality, mean) The number of cases may be determined using function length()).

```
SD <- aggregate(cbind(pre, post,pre-post) ~ curri + tv, data = dat, sd)
M <- aggregate(cbind(pre, post,pre-post) ~ curri + tv, data = dat, mean)
N <- aggregate(cbind(pre, post, pre-post) ~ tv+curri, data = dat, length)
SE <- M
SE[,3:5] <- SD[,3:5]/sqrt(N[,3:5])
```

- e. Plot the follow-up means as a function of the presence of the curriculum separately for the two TV conditions. What effects are you supposing? (Note: You may use function interaction.plot.)

- The plot suggests an interaction effect of curri and tv. This is suggested by the cross-over by the lines.

```
plot(post ~ curri, M[M$tv == 1,], type="b", pch=19, ylim=c(2.3, 3.1), axes = F, xlim=c(-.1,1.1), xlab="curri")
points(post ~ curri, M[M$tv==0,], type="b", lty=2)
axis(2)
axis(1, at= 0:1, labels=c("No preventative curriculum","prevention curriculum"))
legend("topleft", legend=c("no TV", "TV"), pch=c(1,19), lty=c(2,1), bty="n", inset=0.05)
arrows(c(0,1,0,1), M$post - SE$post, y1=M$post + SE$post, code=3, angle=90)
box()
```



f. Perform an analysis of the follow-up scores (ANOVA), change scores, adjusted follow-up scores (ANCOVA), and adjusted change scores.

- \* Follow-up score analysis: Here we have no linear trends. We want to see if any of the `post` results are significant.
- \* The interaction coefficient `curri:tv` is significant at the 10% level, with a p-value of 0.018. We can conclude that there is an interaction effect between curriculum and TV.

*# Follow up scores*

```
m1 <- lm(post ~ curri*tv, dat)
summary(m1)
```

```
##
## Call:
## lm(formula = post ~ curri * tv, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9684 -0.9684  0.0316  1.0316  4.6390
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.36105    0.06646  35.524 < 2e-16 ***
## curri         0.60738    0.09649   6.294 3.98e-10 ***
## tv           0.17742    0.09427   1.882  0.060 .
## curri:tv     -0.32338    0.13652  -2.369  0.018 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.364 on 1596 degrees of freedom
## Multiple R-squared:  0.02944,    Adjusted R-squared:  0.02761
## F-statistic: 16.14 on 3 and 1596 DF,  p-value: 2.464e-10

* For the change score analysis, we want to see if the groupings have made any difference to the size of the change.
* Here we see that the interaction term is significant p = 0.0411.
```

#### *# Change score analysis*

```
m2 <- lm(post ~ offset(pre) + curri*tv, dat)
summary(m2)
```

```
##
## Call:
## lm(formula = post ~ offset(pre) + curri * tv, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.4519 -0.9184  0.0816  1.0816  5.1567
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.20903    0.07573   2.760  0.00585 **
## curri        0.70939    0.10995   6.452 1.46e-10 ***
## tv          0.24290    0.10742   2.261  0.02389 *
## curri:tv    -0.31798    0.15556  -2.044  0.04111 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.554 on 1596 degrees of freedom
## Multiple R-squared:  0.405,    Adjusted R-squared:  0.4039
## F-statistic: 362.1 on 3 and 1596 DF,  p-value: < 2.2e-16
```

\* Adjusted follow-Up-score analysis (ANCOVA). As above, but we include a coefficient for the pre-score.  
 \* Here we again conclude that the interaction term is significant. (p=0.0136). m

```
m3 <- lm(post ~ pre + curri*tv, dat)
summary(m3)
```

```
##
## Call:
## lm(formula = post ~ pre + curri * tv, data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6025 -0.9522  0.0136  0.8456  4.1645
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.66126    0.08436  19.693 < 2e-16 ***
## pre          0.32518    0.02585  12.578 < 2e-16 ***
## curri        0.64055    0.09210   6.955 5.14e-12 ***
## tv          0.19871    0.08996   2.209  0.0273 *
## curri:tv    -0.32162    0.13025  -2.469  0.0136 *
```

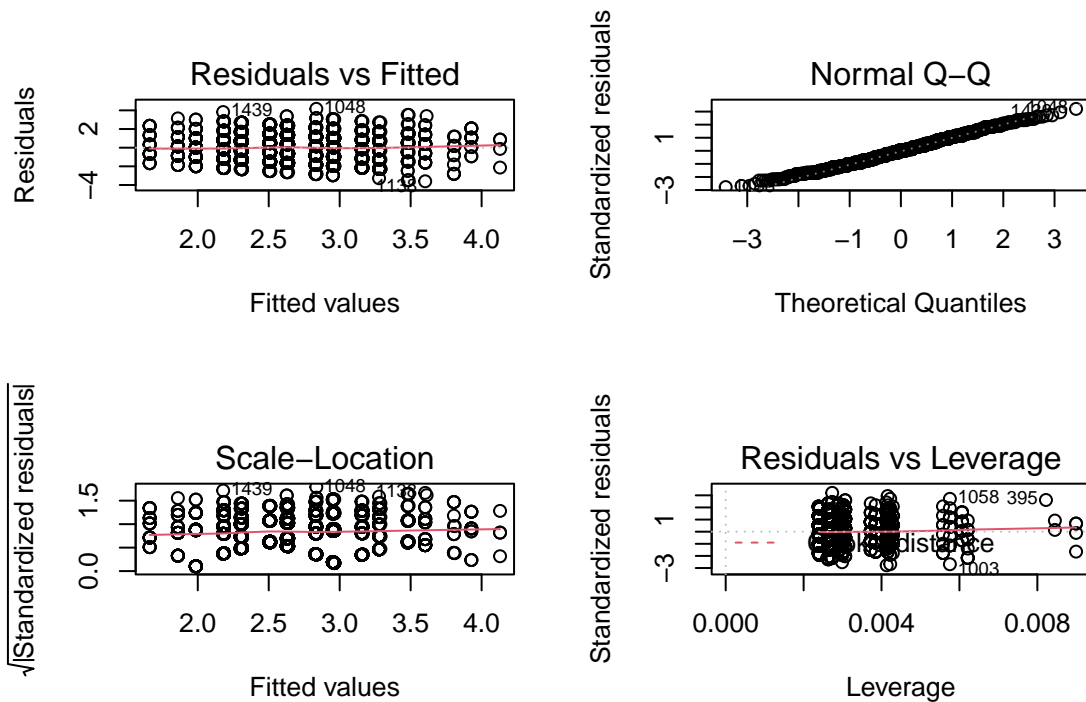
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.301 on 1595 degrees of freedom
## Multiple R-squared:  0.117, Adjusted R-squared:  0.1148
## F-statistic: 52.85 on 4 and 1595 DF, p-value: < 2.2e-16
```

g. Do the factors of the latter two models have identical effects on the change in knowledge?

- The effects of the two factors (curri, tv, & curri:tv) are identical in an ANCOVA and adjusted change score analysis.
- $post = \beta_0 + \beta_1 \times pre + \beta_2 \times c + \beta_3 \times tv + \dots$
- $(post - pre) = \beta_0 + (\beta_1 - 1) \times pre + \beta_2 \times c + \beta_3 \times tv + \dots$

h. Answer the following questions based on the parameter estimates in the ANCOVA model:

- Do the two factors interact?
    - The assumption of no interaction hypothesis between curri and tv must be rejected ( $p = 0.0136$ ). As such, Curri has a different effect depending on whether there was an addition TV prevention program, even after baseline adjustment.
  - What is the mean THKS score expected as a follow-up value for the reference group ( $c_i = 0, tv_i = 0$ ) at a baseline value of  $y_{i1} = 0$ ?
    - $\beta_0 = 1.6612561$
  - How large is the predicted change in knowledge in the remaining groups relative to the reference group?
    - Reference group (curri = 0, tv = 0) = 1.6612561
    - Ref group vs ( $c_i = 1, tv_i = 0$ ) = 0.6405502 ( $Y_d - Y_{base}$ )
    - Ref group vs ( $c_i = 0, tv_i = 1$ ) = 0.1987092 ( $Y_d - Y_{base}$ )
    - Ref group vs ( $c_i = 1, tv_i = 1$ ) = 0.5176346 ( $Y_d - Y_{base}$ )
  - Determine the follow-up value predicted for the reference group and the group with  $c_i = 1$  and  $tv_i = 1$  at a baseline value of  $y_{i1} = 2$  (median)?
    - Reference group vs group with  $c_i = 1, tv_i = 0$
- i. Use graphical procedures to check the assumptions of the ANCOVA model.
- The assumptions are:
    - Homoskedasticity (variance with respect to X remains the same): Residuals vs Fitted plot. Visually fine.
    - Linearity (??): Visually ok.
    - $Y\hat{Y}$  is normally distributed (QQ plot): Fine.
    - Outliers (Which plot?): Visually ok.



- j. Check the ANCOVA model's assumption of identical slopes in the regression of the follow-up value on the baseline value in all groups with an appropriate statistical test ( $\alpha = 0.05$ ).
- The hypothesis is that the slopes in the regression are the same in all groups. This cannot be rejected at the 5% level as the anova test returns a probability of 10.5%. As such, the simpler model without the group specific slopes is preferred.

```
# Model with group specific slopes.
m5 <- lm(post ~ pre * curri * tv, dat)
#summary(m5)
anova(m3,m5)
```

```
## Analysis of Variance Table
##
## Model 1: post ~ pre + curri * tv
## Model 2: post ~ pre * curri * tv
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1    1595 2700.2
## 2    1592 2690.0  3    10.232 2.0185 0.1094
```