

STAT 705 Chapter 22: Analysis of Covariance

Adapted from Timothy Hanson

Department of Statistics, University of South Carolina

Stat 705: Data Analysis II

Add a continuous predictor to an ANOVA model = ANCOVA.

- Mix continuous and discrete predictors.
- Useful for testing treatment effects in presence of continuous predictor(s) that may explain much variability.
- Continuous predictors, or concomitant variables, may be *supplemental or uncontrolled or controlled* (e.g. drug dose in *mg*).
- Concomitant variable should be unaffected by treatments; i.e. they should be “independent.” They are often measured before study takes place.
- Book examples (p. 919): pre-study attitude, age, SES, aptitude, *baseline outcomes* (e.g. seizure rate).

Simplest ANCOVA model

One treatment and one covariate that enters the model linearly. We have $i = 1, \dots, r$ treatment levels and $j = 1, \dots, n_i$ observations within level i . The model is

$$Y_{ij} = \mu + \tau_i + \gamma x_{ij} + \epsilon_{ij}.$$

Set $\tau_r = 0$ for identifiability.

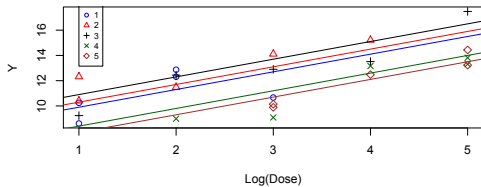
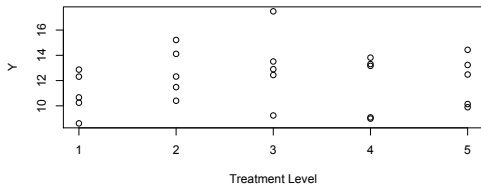
This gives r parallel regression lines, one for each treatment level (see next page). This implies, e.g., the expected mean difference between group i_1 and group i_2 , both having covariate value x , is

$$\mu + \tau_{i_1} + \gamma x - (\mu + \tau_{i_2} + \gamma x) = \tau_{i_1} - \tau_{i_2}.$$

We can obtain this effect from `estimate` or `lsmestimate`.

Your textbook recommends using the centered predictor $x_{ij}^* = x_{ij} - \bar{x}_{.}$ instead of x_{ij} ; Tim finds this to be unnecessary.

Simplest ANCOVA model, $r=5$, $n=5$



Linear model formulation

Say $r = 3$ and $n_1 = n_3 = 2$ and $n_2 = 1$. The full model is

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & x_{11} \\ 1 & 1 & 0 & x_{12} \\ 1 & 0 & 1 & x_{21} \\ 1 & 0 & 0 & x_{31} \\ 1 & 0 & 0 & x_{32} \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \gamma \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix},$$

or $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$.

Here, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, $SSE(F) = \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2$,
 $df_{E(F)} = 5 - 4 = 1$, and $MSE(F) = SSE(F)/df_{E(F)}$.

Test of $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$

The reduced model is

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{31} \\ Y_{32} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{21} \\ 1 & x_{31} \\ 1 & x_{32} \end{bmatrix} \begin{bmatrix} \mu \\ \gamma \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix},$$

or $\mathbf{Y} = \mathbf{X}_R \boldsymbol{\beta}_R + \boldsymbol{\epsilon}_R$.

Here, $\hat{\boldsymbol{\beta}}_R = (\mathbf{X}'_R \mathbf{X}_R)^{-1} \mathbf{X}'_R \mathbf{Y}$, $SSE(R) = \|\mathbf{Y} - \mathbf{X}_R \hat{\boldsymbol{\beta}}_R\|^2$, $df_{E(R)} = 5 - 2 = 3$, and $MSE(R) = SSE(R)/df_{E(R)}$. Define

$$F^* = \frac{\{SSE(R) - SSE(F)\} / \{df_{E(R)} - df_{E(F)}\}}{MSE(F)}.$$

Then if $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ is true,

$$F^* \sim F(df_{E(R)} - df_{E(F)}, df_{E(F)}).$$

Use Type III test for the treatment.

CI for $\tau_2 - \tau_1$, for example

$$\tau_2 - \tau_1 = [0 \quad -1 \quad 1 \quad 0] \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \gamma \end{bmatrix} = \mathbf{c}'\boldsymbol{\beta}.$$

The point estimate is $\hat{\tau}_2 - \hat{\tau}_1 = \mathbf{c}'\hat{\boldsymbol{\beta}} = \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

$$s^2\{\hat{\tau}_1 - \hat{\tau}_2\} = \text{MSE}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}.$$

$$\frac{\hat{\tau}_2 - \hat{\tau}_1}{s\{\hat{\tau}_2 - \hat{\tau}_1\}} \sim t(df_E).$$

So the 95% CI is

$$\hat{\tau}_2 - \hat{\tau}_1 \pm t(0.975; df_E)\sqrt{\text{MSE}\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}.$$

We can obtain pairwise contrasts via `estimate`, `lsmeans`, or `lsmestimate` commands. Note that the latter two commands evaluate the means μ_i at the overall mean $\bar{x}_{..} = \frac{1}{n_T} \sum_{i=1}^r \sum_{j=1}^{n_i} x_{ij}$. The continuous part of the model cancels for pairwise differences.

Estimating a mean response

If you want to estimate mean response given x and i , you need to estimate $\mu + \tau_i + \gamma x$.

This is a simple linear combination of model parameters, and easily obtained via estimate. For example, if $r = 3$, estimate "group 2, x=5" `intercept 1 group 0 1 0 gamma 5;`

LSMEANS evaluates treatment means at the grand mean $\bar{x}_{..}$.

22.3 Cracker sales

- CRD where $n_T = 15$ stores were randomly assigned one of three “promotion” treatment levels:
 - ① $i = 1$ sampling of product by customers in store and regular shelf space,
 - ② $i = 2$ additional shelf space,
 - ③ $i = 3$ special display shelves at ends of aisle in addition to regular shelf space.
- Y_{ij} is number of cases sold during the promotional period.
- x_{ij} is number of cases sold during the previous (non-promotional) period.
- Model fit in SAS is $Y_{ij} = \mu + \tau_i + \gamma x_{ij} + \epsilon_{ij}$ where $\tau_3 = 0$.

Cracker sales in SAS

```
data cracker;
input treatment cases preceding @@;
datalines;
  1 38 21 1 39 26 1 36 22 1 45 28 1 33 19
  2 43 34 2 38 26 2 38 29 2 27 18 2 34 25
  3 24 23 3 32 29 3 31 30 3 21 16 3 28 29
;

* Figure, p. 927;
proc sgscatter;
plot cases*preceding / group=treatment reg;
run;

* Type III p-value for treatment is <0.0001 ;
* Analyses similar to those on pp. 930-932 ;
* Note all analyses conditional on preceding;
proc glimmix;
class treatment;
model cases=treatment preceding / solution;
lsestimate treatment "mu1-mu2" 1 -1 0,
              "mu1-mu3" 1 0 -1,
              "mu2-mu3" 0 1 -1 / adjust=t cl alpha=0.05;
* gives expected cases for each treatment at average value of preceding;
lsmeans treatment/adjust=tukey diff cl;
run;
```

Checking for non-constant slopes

The assumption of parallel slopes should be checked, via plots and/or Type III tests. Again, say $r = 3$. A model that allows for slopes to change with treatment is

$$Y_{ij} = [\mu + \tau_j] + [\gamma + \beta_j]x_{ij} + \epsilon_{ij},$$

where $\tau_r = \beta_r = 0$. This is fit by SAS using zero-one dummy variables

$$Y_{ij} = \mu + \tau_1 I\{i = 1\} + \tau_2 I\{i = 2\} + \gamma x_{ij} + \beta_1 x_{ij} I\{i = 1\} + \beta_2 x_{ij} I\{i = 2\} + \epsilon_{ij}.$$

Checking for non-constant slopes

We have $r = 3$ different lines. Let $\mu(i, x)$ be the mean for level i and covariate x .

$$\mu(1, x) = (\mu + \tau_1) + (\gamma + \beta_1)x$$

$$\mu(2, x) = (\mu + \tau_2) + (\gamma + \beta_2)x$$

$$\mu(3, x) = \mu + \gamma x$$

This is fit via `class treatment; model response=treatment covar treatment*covar; β_1 and β_2 provide offsets to the slope of level $i = 3$. A test for equal slopes is simply $H_0 : \beta_1 = \beta_2 = 0$, available as a Type III nested linear hypothesis test.`

```
* test for parallel slopes ;  
* pp. 932-933, p-value=0.40;  
proc glimmix; * can also use proc glm;  
  class treatment;  
  model cases=treatment|preceding;  
run;
```

- Basic model is $Y_{ij} = \mu + \tau_i + \gamma x_{ij} + \epsilon_{ij}$.
 - Response mean is a linear function of x given i .
 - $i = 1, \dots, r$ levels of a single treatment modeled.
 - $\tau_{i_1} - \tau_{i_2}$ gives mean treatment differences for a given level of x .
 - Similar, but *simpler* than a RCB with x chopped up into categories like age group. Just treat age as continuous.
 - Increased *efficiency* if age really is linear.
- Nonlinear mean, e.g. $Y_{ij} = \mu + \tau_i + \gamma_1 x_{ij} + \gamma_2 x_{ij}^2 + \epsilon_{ij}$.
 - Mean response is parallel *curves* in x , one for each i .
 - Might be necessary if e_{ij} vs \hat{Y}_{ij} shows a parabolic (or otherwise nonlinear) shape.
 - $\tau_{i_1} - \tau_{i_2}$ again gives mean treatment differences for a given level of x .

Generalizations

- More factors, e.g. $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma x_{ijk} + \epsilon_{ijk}$.
 - Here $i = 1, \dots, a$ levels of A, $j = 1, \dots, b$ levels of B, and $k = 1, \dots, n_{ij}$ replicates in $A = i$ and $B = j$.
 - If this fits, we should see approximately parallel curves in scatterplot stratified by (i, j) .
 - If $H_0 : (\alpha\beta)_{ij} = 0$ then analysis simplifies; we can look at differences in main effects. Pairwise differences, e.g. $\beta_3 - \beta_1$ do not change with either i or x .
- More concomitant variables, e.g.
 $Y_{ijk} = \mu + \tau_i + \gamma_1 x_{i1k} + \gamma_2 x_{i2k} + \epsilon_{ijk}$ where x_{ijk} is variable j on k th subject with treatment i .
 - Mean response is parallel *surfaces* in (x_1, x_2) .
 - Here we are assuming parallel *planes*, one for each level of i .
For example when $r = 3$ there are three parallel mean planes:

$$\mu(1, x_1, x_2) = \mu + \tau_1 + \gamma_1 x_1 + \gamma_2 x_2$$

$$\mu(2, x_1, x_2) = \mu + \tau_2 + \gamma_1 x_1 + \gamma_2 x_2$$

$$\mu(3, x_1, x_2) = \mu + \gamma_1 x_1 + \gamma_2 x_2$$

22.4 Horticulture example

- Factor A is flower variety: $i = 1$ LP, $i = 2$ WB.
- Factor B is moisture level: $j = 1$ low, $j = 2$ high.
- $n_T = 24$ plots total; $n_{ij} = 6$ replications of each pairing (i, j) .
- Y_{ijk} is yield of salable flowers.
- x_{ijk} is plot size; expect $\gamma > 0$.
- Model is $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma x_{ijk} + \epsilon_{ijk}$.
- CRD with factorial treatment structure.

Horticulture in SAS

```
data flowers;
  input Yield Plotsize Variety Moisture @@;
  if variety=1 and moisture=2 then group='LP, High';
  else if variety=2 and moisture=1 then group='WB, Low';
  else if variety=2 and moisture=2 then group='WB, High';
  else group='LP, Low';
datalines;
  98 15 1 1 60 4 1 1 77 7 1 1 80 9 1 1 95 14 1 1 64 5 1 1
  55 4 2 1 60 5 2 1 75 8 2 1 65 7 2 1 87 13 2 1 78 11 2 1
  71 10 1 2 80 12 1 2 86 14 1 2 82 13 1 2 46 2 1 2 55 3 1 2
  76 11 2 2 68 10 2 2 43 2 2 2 47 3 2 2 62 7 2 2 70 9 2 2
;

* parallel lines indicates no interaction between plotsize and treatments;
proc sgscatter;
  plot yield*plotsize / group=group loess;
run;

* p-value for treatment interaction is 0.13 so drop it;
proc glm;
  class variety moisture;
  model yield=plotsize variety|moisture/ss2; *Type II SS;
  lsmeans variety*moisture/out=outls; *Save LSMEANS for interaction plot (see code online);
run;

* change to alpha=0.025 below to do Bonferroni on p. 937;
proc glimmix;
  class variety moisture;
  model yield=plotsize variety moisture;
  lsmestimate variety "alpha1-alpha2" 1 -1 / cl alpha=0.05;
  lsmestimate moisture "beta1-beta2" 1 -1 / cl alpha=0.05;
run;
```