Two-Way ANOVA

Applied Regression and Other Multivariable Methods
Sections 19-1 - 19-7

Two Factor Analysis of Variance

- In one-way, trts often different levels of one factor
- What if you're interested in combos of two factors?
  - Temperature and Humidity
  - Diet and Exercise Regime
- Could treat each combination as trt and do ANOVA
  - Example : KKMN 19-1 (Four trts each with 5 obs)
  - Looking at $Y =$ cortical sterone level (i.e., stress)
  - Trts are (C)ontrol, (L)evorphanol, (E)pinephrine, (B)oth
  - Consider this a FIXED factor experiment
  - Can use contrasts to study effects
    $\overline{Y}_L - \overline{Y}_C$ measures L effect (no E)
    $\overline{Y}_B - \overline{Y}_E$ measures L effect (with E)
    Is there an L effect?
    $H_0 : \mu_L - \mu_C = 0$ or $H_0 : \mu_B - \mu_E = 0$
    $H_0 : .5(\mu_B + \mu_L) - .5(\mu_C + \mu_E) = 0$
    Is the L effect different for different levels of E?
    $H_0 : \mu_L - \mu_C = \mu_B - \mu_E$

Example

An animal experiment is designed to investigate whether levorphanol reduces stress as reflected in the cortical sterone level. The four treatments each contained five animals.

Treatment	Responses				
Control	1.90	1.80	1.54	4.10	1.89
Levorphanol	0.82	3.36	1.64	1.74	1.21
Epinephrine	5.33	4.84	5.26	4.92	6.07
Both	3.08	1.42	4.54	1.25	2.57

Three contrasts of interest are

Comparison	C	L	E	B
L Effect	-1	1	-1	1
E Effect	-1	-1	1	1
Interaction	1	-1	-1	1

These contrasts are orthogonal so the SST will be broken up into separate sums of squares that add up to SST.

```
data example1;
input trt y @@;
cards;
1 1.90 1 1.80 1 1.54 1 4.10 1 1.89
2 0.82 2 3.36 2 1.64 2 1.74 2 1.21
3 5.33 3 4.84 3 5.26 3 4.92 3 6.07
4 3.08 4 1.42 4 4.54 4 1.25 4 2.57
;

proc glm;
 class trt;
 model y=trt;
 contrast 'L' trt -1 1 -1 1;
 contrast 'E' trt -1 -1 1 1;
 contrast 'interaction' trt 1 -1 -1 1;
run;
quit;
```



```
The GLM Procedure
                                       Sum of
Source              DF        Squares     Mean Square    F Value    Pr > F
Model                3    37.57844000     12.52614667      12.30    0.0002
Error               16    16.29784000      1.01861500
Corrected Total     19    53.87628000

R-Square     Coeff Var      Root MSE        y Mean
0.697495      34.05076      1.009265      2.964000

Source              DF      Type I SS     Mean Square    F Value    Pr > F
trt                  3    37.57844000     12.52614667      12.30    0.0002

Source              DF    Type III SS     Mean Square    F Value    Pr > F
trt                  3    37.57844000     12.52614667      12.30    0.0002

Contrast            DF    Contrast SS     Mean Square    F Value    Pr > F
L                    1    12.83202000     12.83202000      12.60    0.0027
E                    1    18.58592000     18.58592000      18.25    0.0006
interaction          1     6.16050000      6.16050000       6.05    0.0257
```

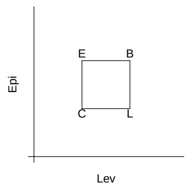
Because these are orthogonal contrasts, $12.83202 + 18.58592 + 6.1605 = 37.57844$.

At $\alpha = .05$ all three contrasts are significantly different from zero. This suggests that there is an E and L effect as well as an interaction between E and L. An interaction means that the E effect is different at different levels of L (or vice versa).

Two-Way ANOVA

- Break up trts into the two factors
  - Factor 1: Levorphanol (Present/Absent)
  - Factor 2: Epinephrine (Present/Absent)
- Also known a two-factor factorial
- Investigates all combinations of two factors
- Here, single replicate of exp involves $2 \times 2 = 4$ trials
- Design often illustrated as table or graph
- Rows/columns represent levels of two factors

	Epinephrine	
Levorphanol	Absent	Present
Absent	XXXXX	XXXXX
Present	XXXXX	XXXXX



Statistical Model

- Consider r levels of factor 1 and c levels of factor 2
- Statistical model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijk} \quad \begin{cases} i = 1, 2, \ldots, r \\ j = 1, 2, \ldots, c \\ k = 1, 2, \ldots, n \end{cases}$$

μ - grand mean

α_i - ith level effect of factor 1 (row i)

β_j - jth level effect of factor 2 (col j)

γ_{ij} - interaction effect in cell ij

$E_{ijk} \sim \mathrm{N}(0, \sigma^2)$

- Everything again considered a deviation from grand mean so the parameter restrictions are

$$\sum_i \alpha_i = 0 \quad \sum_j \beta_j = 0 \quad \sum_i \gamma_{ij} = 0 \quad \sum_j \gamma_{ij} = 0$$

Analysis of Variance Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Row	SSR	$r-1$	MSR	MSR/MSE
Column	SSC	$c-1$	MSC	MSC/MSE
Interaction	SSRC	$(r-1)(c-1)$	MSRC	MSRC/MSE
Error	SSE	$rc(n-1)$	MSE	
Total	TSS	$rcn-1$		

$$\text{TSS} = \sum\sum Y_{ijk}^2 - G^2/rcn = \sum\sum\sum(\overline{Y}_{ijk} - \overline{Y}_{...})^2$$

$$\text{SSR} = \tfrac{1}{cn}\sum_i R_i^2 - G^2/rcn = cn\sum(\overline{Y}_{i..} - \overline{Y}_{...})^2$$

$$\text{SSC} = \tfrac{1}{rn}\sum_j C_j^2 - G^2/rcn = rn\sum(\overline{Y}_{.j.} - \overline{Y}_{...})^2$$

$$\text{SST} = \tfrac{1}{n}\sum\sum RC_{ij}^2 - G^2/rcn$$

$$\text{SSRC} = \text{SST - SSR - SSC} = n\sum\sum(\overline{Y}_{ij.} - \overline{Y}_{i..} - \overline{Y}_{.j.} + \overline{Y}_{...})^2$$

$\text{df}_\text{E} > 0$ only if $n > 1$. When $n = 1$, cannot separate interaction from error (confounded). Recall typical RCBD uses $n = 1$. Assuming no interaction (additivity) allows us to estimate error and test for treatment differences.

Example : Comparing Factors/Levels

- The mean at each combination is

	Epinephrine	
Levorphanol	Absent	Present
Absent	2.246	5.284
Present	1.754	2.572

- Can look at one factor by averaging out other
- Only meaningful if no interaction
- Comparing Levorphanol
  - Average out epinephrine

$$\frac{(1.754 + 2.572)}{2} - \frac{(2.246 + 5.284)}{2} = -1.602$$

- Comparing Epinephrine
  - Average out levorphanol

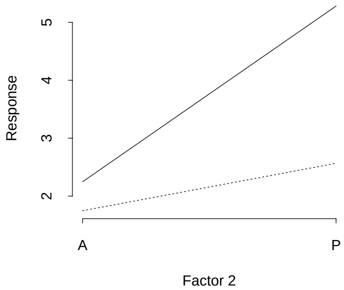
$$\frac{(5.284 + 2.572)}{2} - \frac{(2.246 + 1.754)}{2} = 1.928$$

Example : Comparing Factors/Levels

- Interaction
  - Difference in the effect of L when E present vs absent

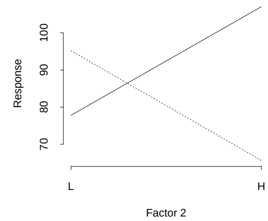
$$(2.572 - 5.284) - (1.754 - 2.246)$$
$$(-2.712) - (-0.492) = -2.220$$

  - L effect much larger (negatively) when E present
- Interaction Plot
  - Can view changes in response difference
  - Are lines parallel or not?

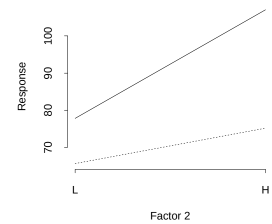


Two Factor Experiment

- Completely opposite behavior (no Factor 2 effect?)

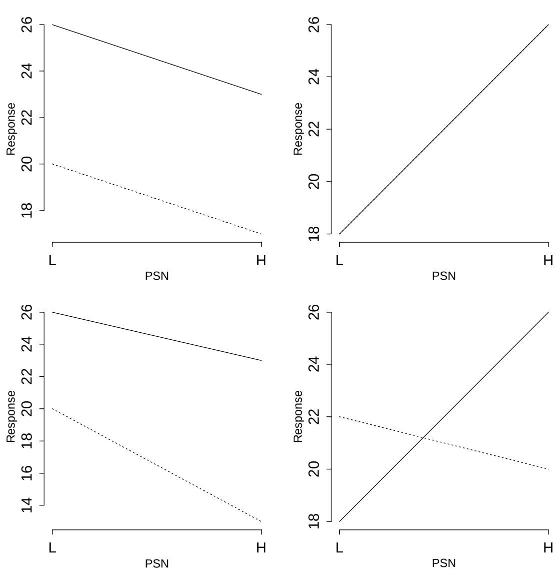


- Increase but not same amount



Interaction Plots

These plots are generated from the first table presented in TABLE 19-10 − 19-13



Interaction in Two Factor Experiment

- If interaction, how to interpret main effects?
  - Main effects average out the other factor
  - Interaction - effect depends on level of factor
  - Thus, main effects only meaningful if no interaction
  - If opposite behavior, effect might cancel out (slide 22-9)
  - May still be able to discuss main effect (but not estimate)
  - Sometimes interaction due to only a few combos
- Common to compare $\overline{Y}_{ij.}$'s (i.e., for fixed i or j).
- **Example** - From ANOVA table, the MSE=1.019 (df=16). For any two-sided t test, the critical value at $\alpha = .05$ is 2.12.

  For E absent, the estimated L effect is $1.754 - 2.246 = -0.492$. The std error is $\sqrt{1.019(1/5+1/5)} \approx 0.64$. This means $T = -0.492/.64 = -.7708$. We conclude that the L effect is NOT significantly different than zero.

  For E present, the estimated L effect is $2.572 - 5.284 = -2.712$. The std error is $\sqrt{1.019(1/5+1/5)} \approx 0.64$. This means $T = -2.712/.64 = -2.249$. We conclude that the L effect is significantly different than zero.

SAS Procedures

```
data twoway;                    /* First break trt into two factors */
 set example1;
 L=0; E=0;
 if trt=3 | trt=4 then E=1;
 if trt=2 | trt=4 then L=1;

proc glm;                   /* Now run Proc GLM */
 class E L;
 model y=E L E*L;           /* The * denotes interaction btwn E and L */
 lsmeans L*E / stderr tdiff slice=E;
run;
quit;
------------------------------------------------------------------------
The GLM Procedure
                                    Sum of
Source               DF       Squares    Mean Square  F Value  Pr > F
Model                 3   37.57844000    12.52614667    12.30  0.0002
Error                16   16.29784000     1.01861500
Corrected Total      19   53.87628000

R-Square     Coeff Var      Root MSE        y Mean
0.697495      34.05076      1.009265      2.964000

Source               DF     Type I SS    Mean Square  F Value  Pr > F
E                     1   18.58592000    18.58592000    18.25  0.0006
L                     1   12.83202000    12.83202000    12.60  0.0027
E*L                   1    6.16050000     6.16050000     6.05  0.0257

Source               DF   Type III SS    Mean Square  F Value  Pr > F
E                     1   18.58592000    18.58592000    18.25  0.0006
L                     1   12.83202000    12.83202000    12.60  0.0027
E*L                   1    6.16050000     6.16050000     6.05  0.0257
```



```
Least Squares Means

                                   Standard                    LSMEAN
E    L        y LSMEAN            Error    Pr > |t|          Number
0    0      2.24600000       0.45135684      0.0001               1
0    1      1.75400000       0.45135684      0.0013               2
1    0      5.28400000       0.45135684      <.0001               3
1    1      2.57200000       0.45135684      <.0001               4

              Least Squares Means for Effect E*L
          t for H0: LSMean(i)=LSMean(j) / Pr > |t|

                   Dependent Variable: y
i/j             1             2             3             4
   1                   0.770779      -4.75941      -0.51072
                         0.4521        0.0002        0.6165
   2     -0.77078                    -5.53019       -1.2815
           0.4521                      <.0001        0.2183
   3     4.759406      5.530185                    4.248686
           0.0002        <.0001                      0.0006
   4      0.51072      1.281499      -4.24869
           0.6165        0.2183        0.0006


                 E*L Effect Sliced by E for y

                      Sum of
E        DF          Squares      Mean Square    F Value    Pr > F
0         1         0.605160         0.605160       0.59    0.4521
1         1        18.387360        18.387360      18.05    0.0006
```


Hidden Replication Feature: Main Effects

- One factor approach
  - Can estimate main effects using 3 combos (C, L, E)
  - Main effect L: $Y_L - Y_C$, Main effect E: $Y_E - Y_C$
  - Must replicate to have variance estimate
  - Cannot estimate interaction without 4th combination (B)
  - If $n = 2$ so $N = 6$, Var(effect) = $2\sigma^2/2 = \sigma^2$
- Factorial approach
  - Estimate effects using $N = 4$ observations (C, L, E, B)
  - Have variance estimate if no interaction
  - Main effect L: $.5(Y_B + Y_L) - .5(Y_E + Y_C)$
  - Var(Effect) = σ^2
  - Replication provides ability to estimate interaction

Factorial gives same accuracy with less ($N = 6$ vs $N = 4$)

Two Factor Experiments with Random Effects

- All slides before this focused on fixed effects
  - Always use MSE in denominator of F-test
  - Use MSE in linear combinations and CIs
- Not always true when random factors
  - May use interaction MS instead of MSE
- Three combinations of fixed and random factors
  - Fixed - both factors fixed
  - Random - both factors random
  - Mixed - one factor fixed and one random

Two-Factor Random Model

$$Y_{ijk} = \mu + A_i + B_j + C_{ij} + E_{ijk} \quad \begin{cases} i = 1,2,\ldots,r \\ j = 1,2,\ldots,c \\ k = 1,2,\ldots,n \end{cases}$$

$$A_i \sim \mathrm{N}(0,\sigma_R^2) \quad B_j \sim \mathrm{N}(0,\sigma_C^2) \quad C_{ij} \sim \mathrm{N}(0,\sigma_{RC}^2)$$

- $\mathrm{Var}(Y_{ijk}) = \sigma^2 + \sigma_R^2 + \sigma_C^2 + \sigma_{RC}^2$
- Expected Mean Squares similar to one-factor model

  E(MSE)= σ^2

  E(MSR)= $\sigma^2 + cn\sigma_R^2 + n\sigma_{RC}^2$

  E(MSC)= $\sigma^2 + rn\sigma_C^2 + n\sigma_{RC}^2$

  E(MSRC)=$\sigma^2 + n\sigma_{RC}^2$

- EMS determine what MS to use in denominator

  $H_0 : \sigma_R^2 = 0 \rightarrow$ MSR/MSRC

  $H_0 : \sigma_C^2 = 0 \rightarrow$ MSC/MSRC

  $H_0 : \sigma_{RC}^2 = 0 \rightarrow$ MSRC/MSE

Two-Factor Mixed Effects Model

- Assume factor 1 fixed and factor 2 random

$$Y_{ijk} = \mu + \alpha_i + B_j + C_{ij} + E_{ijk} \quad \begin{cases} i = 1,2,\ldots,r \\ j = 1,2,\ldots,c \\ k = 1,2,\ldots,n \end{cases}$$

$$\sum \alpha_i = 0 \quad B_j \sim \mathrm{N}(0,\sigma_C^2) \quad C_{ij} \sim \mathrm{N}(0,\sigma_{RC}^2)$$

- Often add restriction $\sum_i C_{ij} = 0$ for each level of j
- This is known as the **restricted** mixed model
- All but Factor 1 tested over MSE

  E(MSE)= σ^2

  E(MSR)= $\sigma^2 + cn\sum\alpha_i^2/(a-1) + n\sigma_{RC}^2$

  E(MSC)= $\sigma^2 + rn\sigma_C^2$

  E(MSRC)=$\sigma^2 + n\sigma_{RC}^2$

- If restriction not added, results similar except

  E(MSC)= $\sigma^2 + rn\sigma_C^2 + n\sigma_{RC}^2$

  Similar tests to random effects model

- For pairwise comparisons, use $\sqrt{2\mathrm{MSRC}/cn}$ (df_{AB})

SAS Procedures

```
proc glm;                      /* Random Effects Model */
 class E L;
 model y=E L E*L;
 random E L E*L / test;

proc glm;                     /* Mixed Effects Model */
 class E L;
 model y=E L E*L;
 random L E*L / test;
 lsmeans E / tdiff stderr;
 lsmeans E / tdiff stderr E=E*L;

proc mixed;                  /* Using Proc Mixed */
 class E L;
 model y=E;
 random L E*L;
 lsmeans E / diff;
run;
quit;
------------------------------------------------------------------
All Models using GLM give same ANOVA table

Source                    DF     Type I SS   Mean Square  F Value  Pr > F
E                          1   18.58592000   18.58592000    18.25  0.0006
L                          1   12.83202000   12.83202000    12.60  0.0027
E*L                        1    6.16050000    6.16050000     6.05  0.0257
Error                     16   16.29784000    1.01861500

The F-values and P-values, however, based on fixed model
```



```
RANDOM EFFECTS

Source                 Type III Expected Mean Square
E                      Var(Error) + 5 Var(E*L) + 10 Var(E)
L                      Var(Error) + 5 Var(E*L) + 10 Var(L)
E*L                    Var(Error) + 5 Var(E*L)


Tests of Hypotheses for Random Model Analysis of Variance

Source                    DF    Type III SS    Mean Square  F Value  Pr > F
E                          1      18.585920      18.585920     3.02  0.3326
L                          1      12.832020      12.832020     2.08  0.3857
Error: MS(E*L)             1       6.160500       6.160500

Source                    DF    Type III SS    Mean Square  F Value  Pr > F
E*L                        1       6.160500       6.160500     6.05  0.0257
Error: MS(Error)          16      16.297840       1.018615
-------------------------------------------------------------------------
MIXED EFFECTS

Source                 Type III Expected Mean Square
E                      Var(Error) + 5 Var(E*L) + Q(E)
L                      Var(Error) + 5 Var(E*L) + 10 Var(L)
E*L                    Var(Error) + 5 Var(E*L)

Tests of Hypotheses for Mixed Model Analysis of Variance

Source                    DF    Type III SS    Mean Square  F Value  Pr > F
E                          1      18.585920      18.585920     3.02  0.3326
L                          1      12.832020      12.832020     2.08  0.3857
Error: MS(E*L)             1       6.160500       6.160500

Source                    DF    Type III SS    Mean Square  F Value  Pr > F
E*L                        1       6.160500       6.160500     6.05  0.0257
Error: MS(Error)          16      16.297840       1.018615
```

Mixed Model Multiple Comparisons

- Must use MSRC in std err calculations
- Proc GLM does not do this automatically
- Must use E= option in lsmeans statement
- Proc Mixed does do this automatically

```
First analysis incorrectly uses MSE when computing t Value

Least Squares Means

                       Standard    H0:LSMEAN=0     H0:LSMean1=LSMean2
E     y LSMEAN           Error        Pr > |t|    t Value    Pr > |t|
0   2.00000000      0.31915748          <.0001      -4.27      0.0006
1   3.92800000      0.31915748          <.0001

Least Squares Means

Standard Errors and Probabilities Calculated Using the Type III
   MS for E*L as an Error Term

                       Standard    H0:LSMEAN=0     H0:LSMean1=LSMean2
E     y LSMEAN           Error        Pr > |t|    t Value    Pr > |t|
0   2.00000000      0.78488853          0.2381      -1.74      0.3326
1   3.92800000      0.78488853          0.1256
```



```
                     Iteration History
Iteration    Evaluations    -2 Res Log Like       Criterion
        0              1        67.80524284
        1              1        60.35216072      0.00000000
                   Convergence criteria met.

Covariance Parameter
      Estimates
Cov Parm     Estimate
L              0.6672    **
E*L            1.0284    ** There are the variance estimates
Residual       1.0186    **


        Type 3 Tests of Fixed Effects
              Num     Den
Effect         DF      DF    F Value    Pr > F
E               1       1       3.02    0.3326  **Test of Fixed Effect

                       Least Squares Means
                           Standard
Effect    E    Estimate       Error      DF    t Value    Pr > |t|
E         0      2.0000      0.9745       1       2.05      0.2886
E         1      3.9280      0.9745       1       4.03      0.1548


                   Differences of Least Squares Means

                                  Standard
Effect    E    _E    Estimate        Error      DF    t Value    Pr > |t|
E         0    1      -1.9280       1.1100       1      -1.74      0.3326
```