STA221

Neil Montgomery

Last edited: 2017-03-20 16:19

the analysis

We have new sums of squares decompositions. The details can get brutal - so we'll stay symbolic.

We'll call the factor variables `A` and `B`, with I and J levels respectively.

In the additive case we get, assuming a common sample size of n for each combination of factor levels (strongly encouraged in practice!):

$$SS_{Total} = SS_A + SS_B + SS_{Error}$$

the analysis

We have new sums of squares decompositions. The details can get brutal - so we'll stay symbolic.

We'll call the factor variables `A` and `B`, with I and J levels respectively.

In the additive case we get, assuming a common sample size of n for each combination of factor levels (strongly encouraged in practice!):

$$SS_{Total} = SS_A + SS_B + SS_{Error}$$

Everything has χ^2 distributions. The degrees of freedom add up (N is the grand sample size):

$$N - 1 = (I - 1) + (J - 1) + (N - I - J + 1)$$

the analysis

We have new sums of squares decompositions. The details can get brutal - so we'll stay symbolic.

We'll call the factor variables `A` and `B`, with I and J levels respectively.

In the additive case we get, assuming a common sample size of n for each combination of factor levels (strongly encouraged in practice!):

$$SS_{Total} = SS_A + SS_B + SS_{Error}$$

Everything has χ^2 distributions. The degrees of freedom add up (N is the grand sample size):

$$N - 1 = (I - 1) + (J - 1) + (N - I - J + 1)$$

When the experiment is *balanced* (strongly encouraged!) with n in each combination of levels, the error degrees of freedom simplifies to $IJ(n - 1)$.

fire example - no interaction (?!)

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Bath         1  715.3   715.3  23.954 1.31e-05
## Launderings  1  202.1   202.1   6.769   0.0125
## Residuals   45 1343.8    29.9
```

sums of squares - a few details

The total sum of squares SS_{Total} is (as always) $(N-1)$ times the sample varianace of the response variable:

$$\sum_{i,j,k} \left(y_{ijk} - \overline{\overline{y}}\right)^2$$

The treatment sums of squares will be:

$$SS_A = nI \sum_{i} \left(\overline{y}_{i\cdot\cdot} - \overline{\overline{y}}\right)^2$$

$$SS_B = nJ \sum_{j} \left(\overline{y}_{\cdot j\cdot} - \overline{\overline{y}}\right)^2$$

where the dots in the subscript mean “averaged over this index.”

sums of squares - a few details

The total sum of squares SS_{Total} is (as always) $(N-1)$ times the sample varianace of the response variable:

$$\sum_{i,j,k} \left(y_{ijk} - \overline{\overline{y}}\right)^2$$

The treatment sums of squares will be:

$$SS_A = nI \sum_i \left(\overline{y}_{i\cdot\cdot} - \overline{\overline{y}}\right)^2$$

$$SS_B = nJ \sum_j \left(\overline{y}_{\cdot j \cdot} - \overline{\overline{y}}\right)^2$$

where the dots in the subscript mean "averaged over this index."

The key thing to note is that each treatment sum of squares is computed using the average over all levels of the other treatment...

sums of squares - a few details

The total sum of squares SS_{Total} is (as always) $(N-1)$ times the sample varianace of the response variable:

$$\sum_{i,j,k} (y_{ijk} - \overline{\overline{y}})^2$$

The treatment sums of squares will be:

$$SS_A = nI \sum_i (\overline{y}_{i\cdot\cdot} - \overline{\overline{y}})^2$$

$$SS_B = nJ \sum_j \left(\overline{y}_{\cdot j\cdot} - \overline{\overline{y}}\right)^2$$

where the dots in the subscript mean "averaged over this index."

The key thing to note is that each treatment sum of squares is computed using the average over all levels of the other treatment...

...which only makes sense when there is no interaction.

error sum of squares

For the sake of completeness:

$$SS_{Error} = \sum_{i,j,k} \left(y_{ijk} - \overline{y}_{ij\cdot} \right)^2$$

Note that $y_{ijk} - \overline{y}_{ij\cdot}$ is also called a "residual".

model assumptions

Mostly the same as with one treatment factor, with the same verification techniques.

1. Observations are independent (this is assumed - only knowledge of the experiment itself is of any help in satisfying this assumption.)

model assumptions

Mostly the same as with one treatment factor, with the same verification techniques.

1. Observations are independent (this is assumed - only knowledge of the experiment itself is of any help in satisfying this assumption.)
2. Equal variance for all combinations of levels of treatment factors. (Plots, or Levene's test when n is not too small.) (Fatal if violated.)

model assumptions

Mostly the same as with one treatment factor, with the same verification techniques.

1. Observations are independent (this is assumed - only knowledge of the experiment itself is of any help in satisfying this assumption.)
2. Equal variance for all combinations of levels of treatment factors. (Plots, or Levene's test when n is not too small.) (Fatal if violated.)
3. Normal error (normal quantile plot of residuals) (Violation OK if sample size is large enough.)

model assumptions

Mostly the same as with one treatment factor, with the same verification techniques.

1. Observations are independent (this is assumed - only knowledge of the experiment itself is of any help in satisfying this assumption.)
2. Equal variance for all combinations of levels of treatment factors. (Plots, or Levene's test when n is not too small.) (Fatal if violated.)
3. Normal error (normal quantile plot of residuals) (Violation OK if sample size is large enough.)

model assumptions

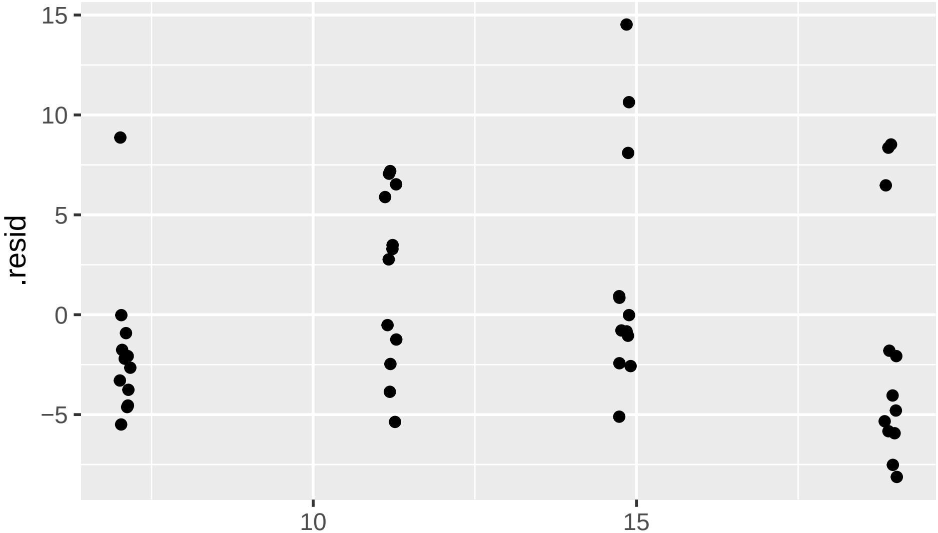
Mostly the same as with one treatment factor, with the same verification techniques.

1. Observations are independent (this is assumed - only knowledge of the experiment itself is of any help in satisfying this assumption.)
2. Equal variance for all combinations of levels of treatment factors. (Plots, or Levene's test when n is not too small.) (Fatal if violated.)
3. Normal error (normal quantile plot of residuals) (Violation OK if sample size is large enough.)

When $n = 1$, the lack of interaction is also an *assumption*.

fire retardant model assumptions - equal variance

Plot of residuals versus “fitted values” (in this case, just the group averages):

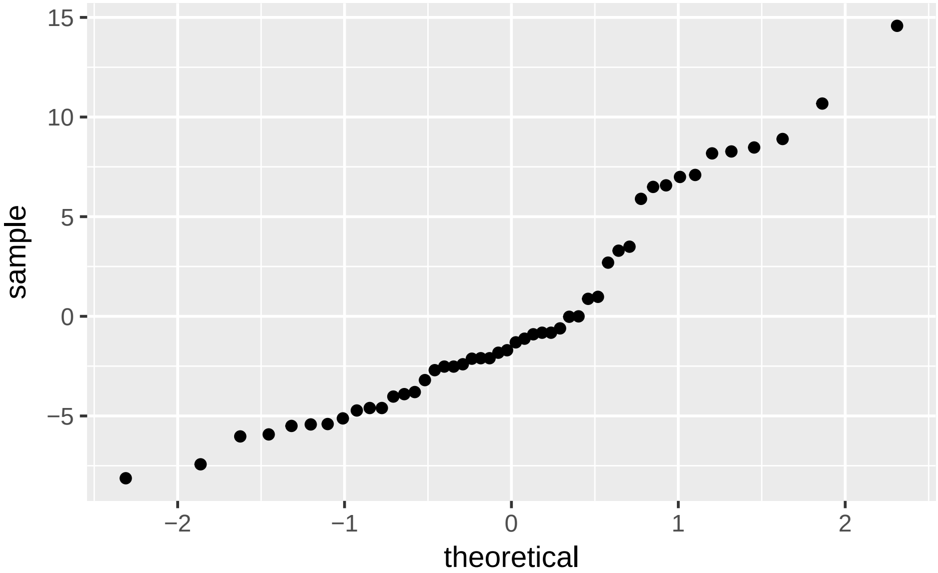


fire retardant model assumptions - equal variance

Since $n = 12$ Levene's test also works:

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3  0.7138  0.549
##       44
```

fire retardant model assumptions - normality



the general model and analysis (with interaction)

This model has the $(\tau\gamma)$ interaction term:

$$y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk} \qquad \text{general model}$$

the general model and analysis (with interaction)

This model has the $(\tau\gamma)$ interaction term:

$$y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk} \qquad \text{general model}$$

The sum of squares decomposition is now:

$$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Error}$$

the general model and analysis (with interaction)

This model has the $(\tau\gamma)$ interaction term:

$$y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk} \qquad \text{general model}$$

The sum of squares decomposition is now:

$$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Error}$$

The interaction sum of squares is:

$$n \sum_{i,j} \left(y_{ijk} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{\overline{y}} \right)^2$$

Small when additive; large when not.

degrees of freedom - balanced case

$$SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{Error}$$
$$(N-1) = (I-1) + (J-1) + (I-1)(J-1) + IJ(n-1)$$

Note: $IJ(n-1) = N - IJ$

We get (in addition):

$$\frac{SS_{AB}/(I-1)(J-1)}{SS_{Error}/IJ(n-1)} \sim F_{(I-1)(J-1),IJ(n-1)}$$

If there is evidence for interaction, do not try to interpret the “main effects”.

flame retardant with interaction

Flame retardant example, with interaction (and without):

```
##                    Df Sum Sq Mean Sq F value   Pr(>F)
## Bath                1  715.3   715.3  26.726 5.49e-06
## Laundering s        1  202.1   202.1   7.552  0.00866
## Bath:Launderings    1  166.1   166.1   6.207  0.01657
## Residuals          44 1177.7    26.8
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## Bath           1  715.3   715.3  23.954 1.31e-05
## Launderings    1  202.1   202.1   6.769   0.0125
## Residuals     45 1343.8    29.9
```

unreplicated designs

If $n = 1$, it is possible to proceed with the analysis, *if you can assume there is no interaction.*

unreplicated designs

If $n = 1$, it is possible to proceed with the analysis, *if you can assume there is no interaction.*

It isn't possible to test for interaction when $n = 1$. In principle, because there is no variation within each combination to estimate.

unreplicated designs

If $n = 1$, it is possible to proceed with the analysis, *if you can assume there is no interaction.*

It isn't possible to test for interaction when $n = 1$. In principle, because there is no variation within each combination to estimate.

Also, the ANOVA table wouldn't work!

unreplicated designs

If $n = 1$, it is possible to proceed with the analysis, *if you can assume there is no interaction.*

It isn't possible to test for interaction when $n = 1$. In principle, because there is no variation within each combination to estimate.

Also, the ANOVA table wouldn't work!

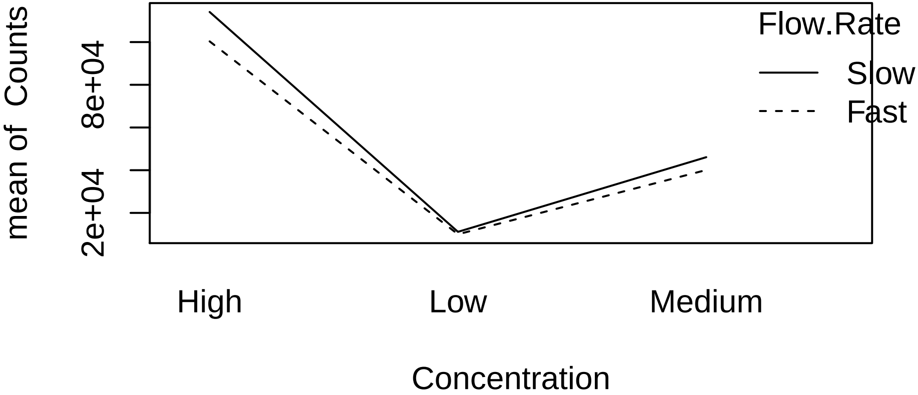
The technique is to use SS_{AB} as the error sum of squares.

another overall example

Chromatography example.

Two factors: flow rate (fast and slow); Concentration (low, med, high)

interaction plot



analysis

```
##                            Df    Sum Sq   Mean Sq  F value   Pr(>F)
## Flow.Rate                   1 3.640e+08 3.640e+08   29.645 1.35e-05
## Concentration               2 4.837e+10 2.418e+10 1969.424  < 2e-16
## Flow.Rate:Concentration     2 2.030e+08 1.015e+08    8.267  0.00186
## Residuals                  24 2.947e+08 1.228e+07
```

assumptions

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  5  1.5612 0.2089
##       24
```

assumptions