Lecture 29 Monday April 10

- HW06 is due Wednesday April 12
- Project 2 is due Wednesday April 19

TOPICS

Pooling sums of squares

Example of transformable (removable) interaction

Section 19.10 Pooling Sums of Squares in Two-Factor Analysis of Variance

If interaction effect is statistically insignificant with a P-value well over $.20$, we can drop the interaction term. The advantages of this approach:

- We gain degree(s) of freedom for error.
- The model and followup analysis are logically consistent.

The model equation for the additive model (model w/out interaction):

$E(Y_{ijk}) = \mu_{..} + \alpha_i + \beta_j$ ~~$+ \epsilon_{ij}$~~ no error!

Analysis Notes:

1 The sums of squares for A and B are the same as before. because of balance

2 The new error sum of squares is the sum of the error and interaction sums of squares from the interaction model.

3 Degrees of freedom for error is
$ab(n-1) + (a-1)(b-1) = abn - a - b + 1$

df(Error) df(AB)

from the original, interaction model

Interaction Model v. Additive Model: Decomposition of Observations, SS, df

Interaction

$$Y_{ijk} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad (1)$$

Obs: $$Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot} = (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}) + (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot}) + (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot}) + (Y_{ijk} - \bar{Y}_{ij\cdot}) \quad (2)$$

SS: $$SSTO = SSA + SSB + SSAB + SSE \quad (3)$$

df: $$nab - 1 = a-1 + b-1 + (a-1)(b-1) + (abn - ab)$$

Additive

$$Y_{ijk} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + \varepsilon_{ijk}$$

Obs $$Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot} = (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}) + (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot}) + (Y_{ijk} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})$$

SS $$SSTO = SSA + SSB + SSE_{Additive}$$

df $$nab - 1 = a-1 + b-1 + nab - a - b + 1$$

where $SSE_{Additive} = SSAB + SSE$ from (2) above.

Ex Corn Yield

See p. 135 for ANOVA table for interaction model.

Fit the additive model. ((*Additive* model means the model without interaction.)

```
> m2 <- aov(yield ~ fertilizer + manure, data=yield.df)
> summary(m2)
            Df Sum Sq Mean Sq F value Pr(>F)
fertilizer   1  17.67  17.672   6.332 0.0222 *
manure       1  19.21  19.208   6.883 0.0178 *
Residuals * 17  47.44   2.791
```

Do the tests for **main effects** of fertilizer and manure, exactly as in the interaction model. The mean squares for fertilizer and manure are exactly the same as before, but the F statistics and P-values are slightly different, because there is a different mean square for residuals, and more d.f. for the denominator of the F statistics.

* · df(Error) = 16 + 1

Interaction model → df(Error) + df(Interaction) p. 135

SS(Error) = SS_I(Error) + SS_I(Interaction) = 44.4 + 3.04 = 47.44

Interaction model fit for Corn Yield data:

```
> m1 <- aov(yield ~ manure*fertilizer,data=yield.df)
> anova(m1)
Analysis of Variance Table

Response: yield
                  Df Sum Sq Mean Sq F value  Pr(>F)
manure             1 19.208  19.208  6.9218 0.01816 *
fertilizer         1 17.672  17.672  6.3683 0.02258 *
manure:fertilizer  1  3.042   3.042  1.0962 0.31066
Residuals         16 44.400   2.775
```

Three F tests are given in the table.

Each F statistic (labelled “F value”) is a ratio of mean squares:

$$F_{\text{obs}} = \frac{\text{MS(Factor)}}{\text{MS(Residuals)}}$$

Example Corn yield

Follow-up analysis

Let's find the 95% CI for effect of fertilizer (High - Low).

The point estimate is: $\bar{Y}_{\text{high}} - \bar{Y}_{\text{low}} = 1.88$

Estimate the model parameter σ^2 by the MS(Residuals) $= 2.7907$.

There are 10 observations in each fertilizer level, so the SE of the effect estimate is $\sqrt{2.7907(1/10 + 1/10)} = .7471$.

Resulting 95% confidence interval is $(.30, 3.46)$ (work not shown). This is based on a t distribution with 17 d.f., after pooling sums of squares. The CI does not include zero, so we conclude the main effect of fertilizer is significant at level $\alpha = .05$.

Removable interaction

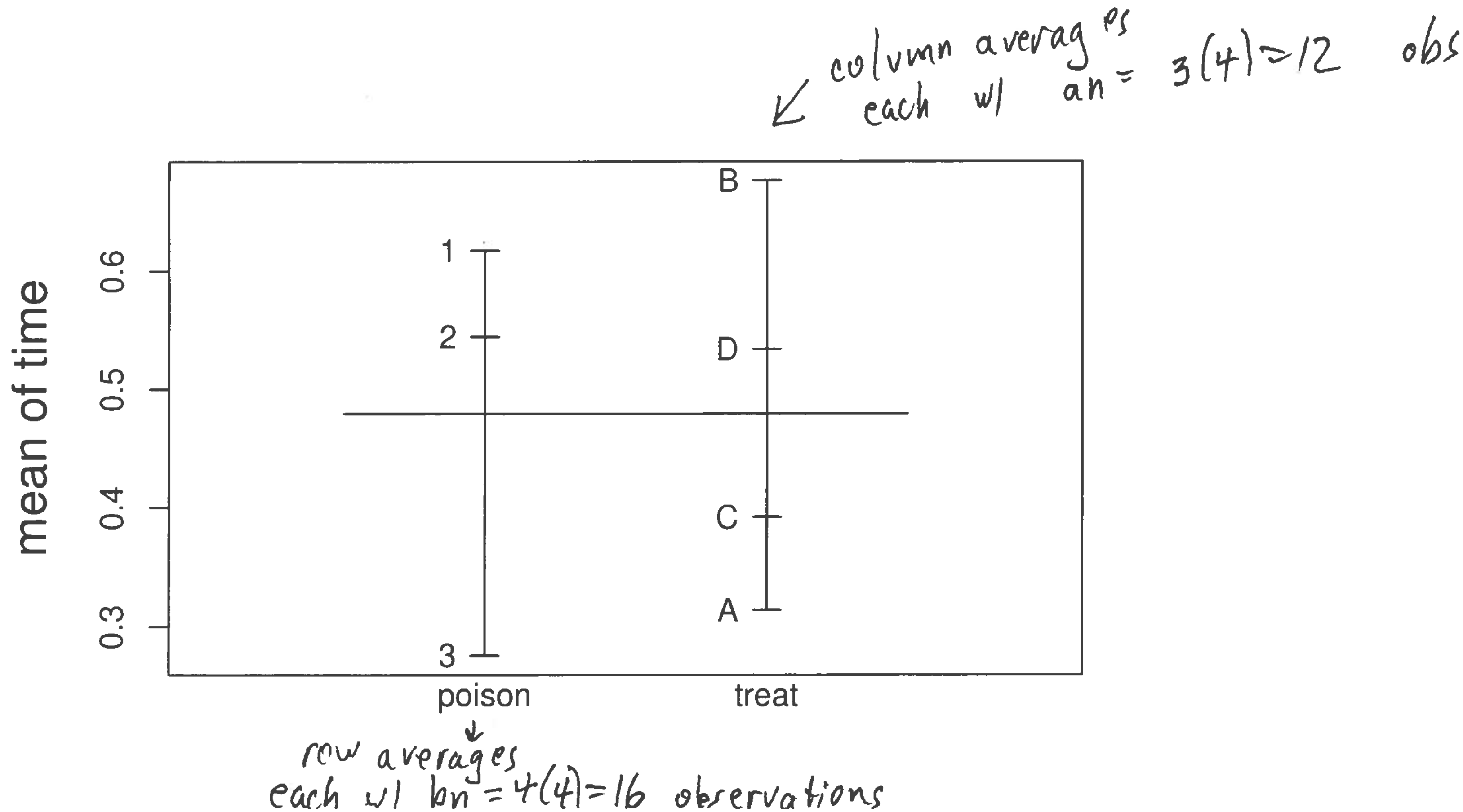
Sometimes the Y variable can be transformed so as to "remove the interaction." So long as this does not cause unequal variance or other model violations, it is a good thing.

The following example is to illustrate an interaction removable by transformation.

Example: Poisons The data is $T =$ survival time (in ten-hour time units) of groups of four animals randomly allocated to three poisons and four treatments. The experiment was part of an investigation to combat the effects of certain toxic agents. [4]

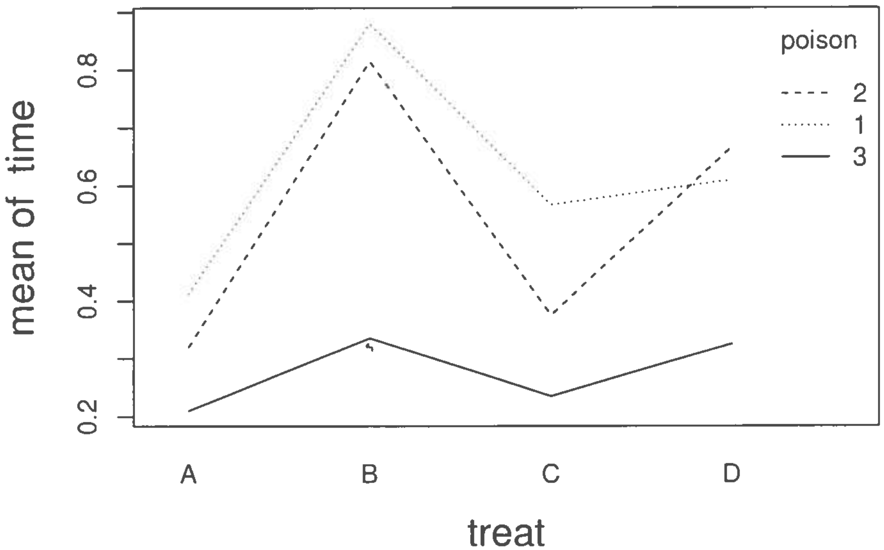
```
> library("boot")
> data(poisons)
> plot.design(poisons)
```

Plot of factor level means:



```
>  interaction.plot(treat,poison,response=time)
```

Shows the cell means graphically



← There's some hope to remove interaction here because
1) The crossing of lines is minor
2) Lines (all 3) move consistently in same direction.

```
> toxic.aov ← aov(time ~ poison*treat, data = poison)
> anova(toxic.aov)
Analysis of Variance Table

Response: time
             Df  Sum Sq Mean Sq F value    Pr(>F)
poison        2 1.03301 0.51651 23.2217 3.331e-07 ***
treat         3 0.92121 0.30707 13.8056 3.777e-06 ***
poison:treat  6 0.25014 0.04169  1.8743    0.1123
Residuals    36 0.80072 0.02224
```

→ poison:treat ← we can't accept Ho because P < .20

Example: Poisons Data were transformed to $Y = 1/T =$ rate of dying because the transformed data satisfied model assumptions better than the original data.

Questions: Is there a difference in mean rate of dying among the three poisons? Is there a difference in mean rate of dying among the four treatments?

And first, we have to answer the question: Does effect of treatment depend upon the poison; that is, is there an interaction effect?

This is a 3×4 factorial design with four observations per cell (four replications). Both of the factors, poisons and treatments, are of equal interest.

Let our response variable be $Y = 1/T$ = rate of dying
this as number of animals who die in a ten-hr period if they died at the given time. For example, the first animal in the dataset had a survival time of $.31 = .31 \times 10$hrs $= 3.1$hrs. The reciprocal $1/.31 = 3.2258$ is the expected number of animals who would die in ten hours if they die at that rate.

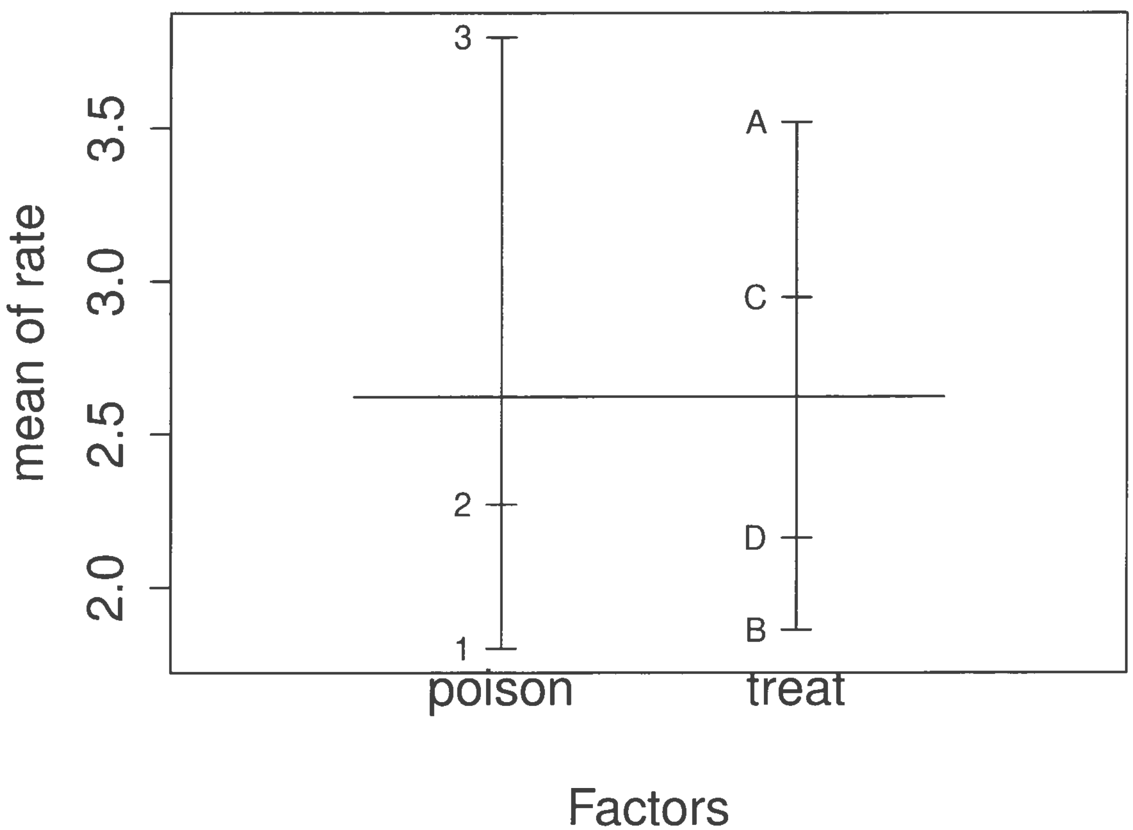
Cell means:

	A	B	C	D
I	2.49	1.16	1.86	1.69
II	3.27	1.39	2.71	1.70
III	4.80	3.03	4.26	3.09

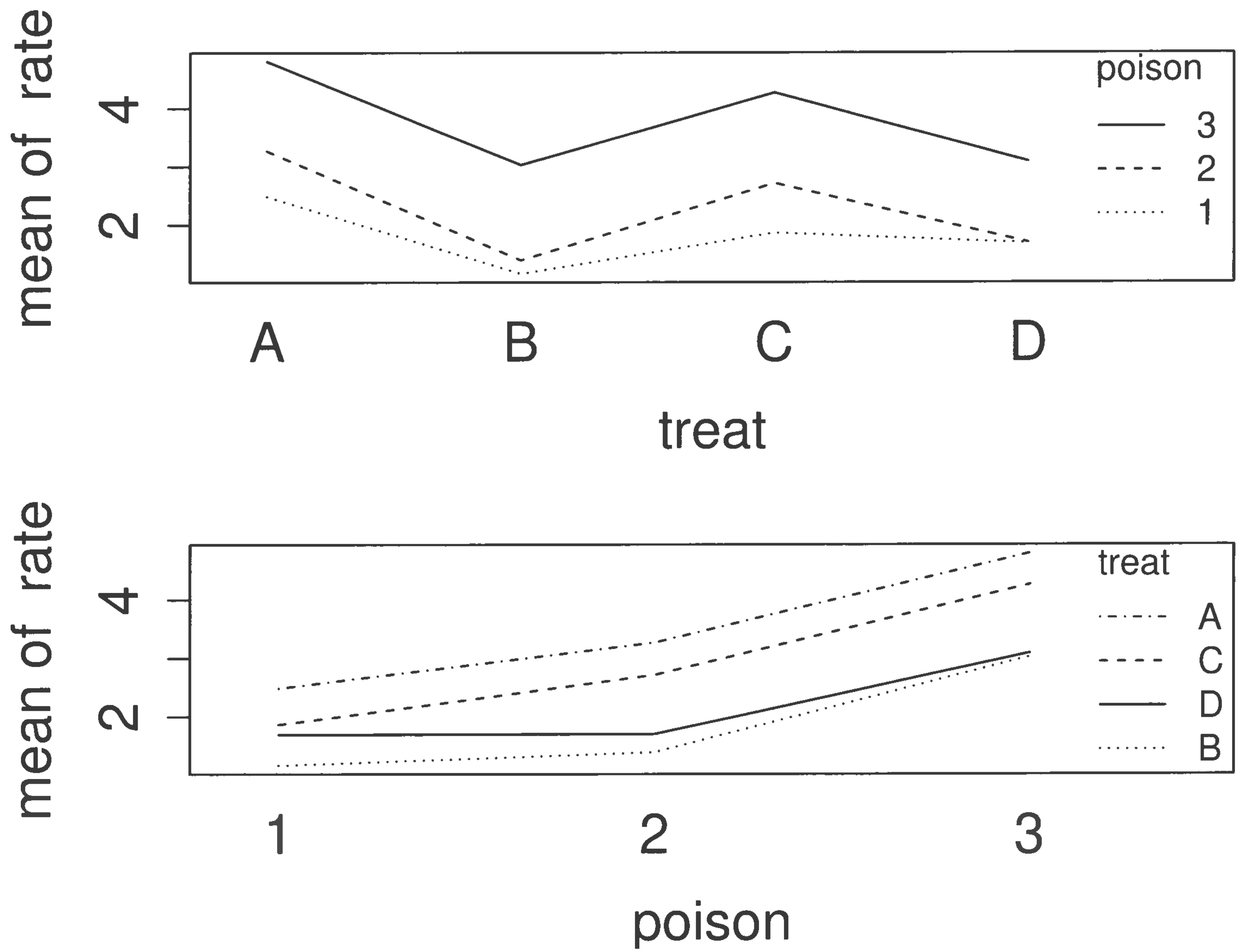
Cell SDs:

	A	B	C	D
I	0.50	0.20	0.49	0.36
II	0.82	0.55	0.42	0.70
III	0.53	0.42	0.23	0.24

Design plot, showing poison and treatment averages (main effects):



Interaction plots:



Check assumptions:

- Randomization—*we were told treatments were allocated at random, so this assumption is satisfied*.
- Normality — Normal quantile-quantile plot of studentized residuals looks more like a straight line than before transformation, so normality is better satisfied for the transformed than for the raw data. (Fill in some detail in class.)
- Constant variance—The plot of studentized residuals vs. fitted values shows no gross violation of the assumption.

And we would like to also assume no interaction. We will check this assumption with an F test.

3. Constant variance means variance should be roughly the same across all twelve experimental conditions.

Rate of death $= Y' = 1/Y = 1/\text{Time}$

First step of analysis is to check the assumption of "no interaction" We use a preliminary F test to do this.

Fit the model with interaction. Below is the resulting ANOVA table:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
poison	2	34.88	17.439	72.64	2.31e-13	***
treat	3	20.41	6.805	28.34	1.38e-09	***
poison:treat	6	1.57	0.262	1.09	0.387	
Residuals	36	8.64	0.240			

H_0 : No interaction between poison and treatment, in effect on rate
H_a : There is some interaction

$H_0 : (\alpha\beta)_{ij} = 0$ for $i = 1, \ldots, 3;\ j = 1, \ldots, 4$

H_a : Not all $(\alpha\beta)_{ij} = 0$

$F = 1.09$, null distr. $F_{6,36}$, $P = .387$, and since $P > .2$ we can accept H_0 and proceed to fit the simpler, additive model.

The reason we want to use the additive model is primarily that it is easier to analyze and to explain—the principle of parsimony.

Check df:

Sample size: