Lecture 27 Monday April 3

Announcements

>) Project 2 is posted - due April 19. You are responsible for material covered in lecture particular to the project.

>) See Lec2.r on the class web site.

>) Exam 2 will be returned Wednesday

>) NO CLASS THIS FRIDAY APRIL 7
a makeup office hour will be scheduled.

TOPICS Two-way ANOVA

>) The three F tests - generic description

>) Flowchart of analysis - A lot depends on test of interaction

>) Ex. Corn Yield in R

Compute SSA, SSB, SSAB, SSE, SSTR for corn yield data.✓ Done Friday 31 March
Put up the ANOVA table with formulas, including E(MS).

valid under both Ho & Ha

ANOVA table

Source of Variation	Sum of Squares	df	Mean Square	Expected Mean Square
Factor A	SSA	$a-1$	$\frac{\text{SSA}}{a-1}$	$\sigma^2 + nb\frac{\sum \alpha_i^2}{a-1}$
Factor B	SSB	$b-1$	$\frac{\text{SSB}}{b-1}$	$\sigma^2 + na\frac{\sum \beta_j^2}{b-1}$
AB interaction	SSAB	$(a-1)(b-1)$	$\frac{\text{SSAB}}{(a-1)(b-1)}$	$\sigma^2 + n\frac{\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$
Error	SSE	$ab(n-1)$	$\frac{SSE}{ab(n-1)}$	σ^2
Total	SSTO	$nab-1$		

There are three F tests: main effects of A, main effect of B, interaction effect AB

Each F statistic will be of the same form, $F_{obs} = \frac{MS(\text{Factor})}{MSE}$

---

① Consider the F test for main effect of Factor A.

Recall, $\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot}$, $i = 1, \ldots, a$

We write $H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$

(which is the same as $H_0: \mu_{1\cdot} = \mu_{2\cdot} = \cdots = \mu_{a\cdot}$)

H_a: At least one $\alpha_i \neq 0$

$F_{obs} = \frac{MSA}{MSE}$

FACT: Under H_0, $F_{obs} \sim F_{a-1,(n-1)ab}$ ← null distribution

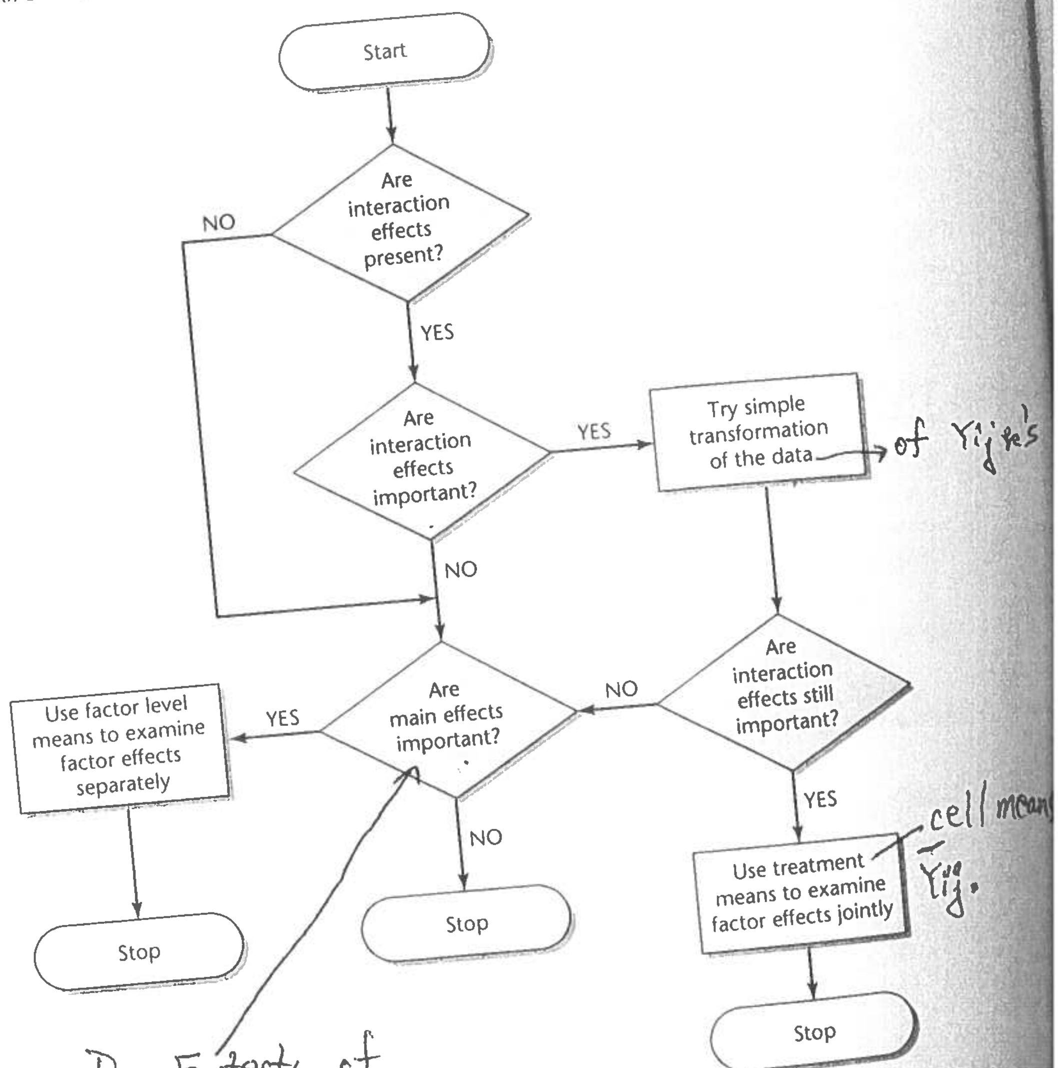
From the expected mean squares, we see that if H_a is true, then F_{obs} tends to be bigger than typical for the null distribution.

We reject H_0 at level α if $F_{obs} > F_{1-\alpha, a-1, (n-1)ab}$

This was a generic description of the three F tests.

But the three F tests should be viewed in a more subtle way – the interaction F test is used for <u>an unusual purpose</u>.

FIGURE 19.11 **Strategy for Analysis of Two-Factor Studies.**



Do F tests of main effects of A, B

19.8 Analysis of Factor Effects when Factors Do Not Interact

As just noted, the analysis of factor effects usually only involves the factor level means $\mu_{i\cdot}$ and $\mu_{\cdot j}$ when the two factors do not interact, or when they interact only in an unimportant fashion.

Estimation of Factor Level Mean

Unbiased point estimators of $\mu_{i\cdot}$ and $\mu_{\cdot j}$ are:

$$\hat{\mu}_{i\cdot} = \bar{Y}_{i\cdot\cdot}$$

$$\hat{\mu}_{\cdot j} = \bar{Y}_{\cdot j\cdot}$$

Question 1 for ANOVA Analysis First question we examine in the statistical analysis is: Is there an interaction between Factor A and Factor B? That is, does the effect of Factor A depend on what is the level of Factor B?

Ex. Corn Yield

Cell Means ($\bar{Y}_{ij\bullet}$'s)

Manure		Fertilizer Low	Fertilizer High	(Simple) Individual Measures of effect
Low	n=5	11.3	13.9	13.9 – 11.3 = 2.6
High		14.0	15.1	15.1 – 14.0 = 1.1
Individual Measures of effect		14.0 – 11.3 = 2.7	15.1 – 13.9 = 1.2	

The (simple) individual measures of effect are not equal, for this data.

Question: Are the observed differences in simple effects just due to chance?

Example Corn Yield

contains a single column, the Y_{ijk} values

Set up the data set in R:

There are twenty observations in the file corn-yield.txt. Do the following to make the data frame: Run the code in Lec2.r

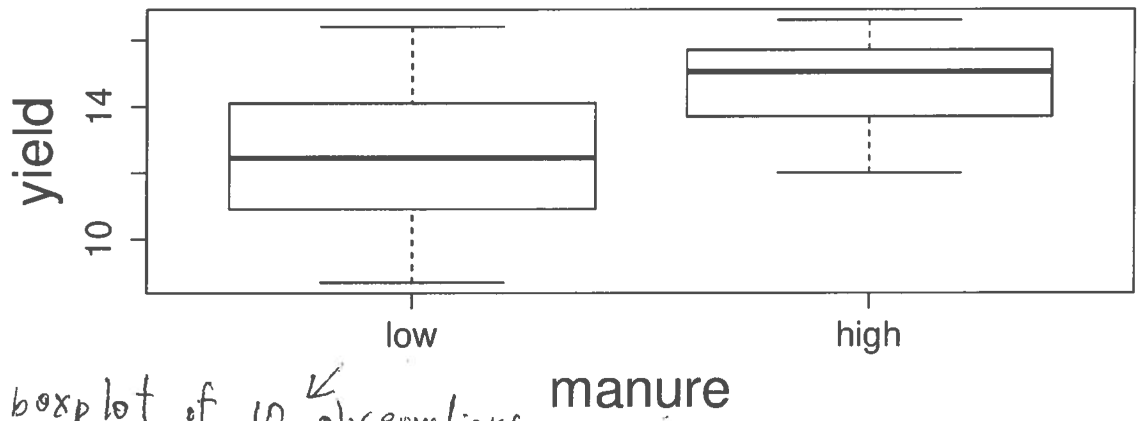
```
> yield <- read.table("corn-yield.txt",header=TRUE)
> fertilizer <- rep(c("high","low"),c(10,10))
> manure <- rep(rep(c("high","low"),c(5,5)),2)
> fertilizer <- factor(fertilizer, levels=c("low","high"))
> yield.df <- data.frame(yield=yield,
+     fertilizer=fertilizer,manure=manure)
```

```
> manure <- factor(manure, levels = c("low", "high"))
```

Get the boxplots of “high” and “low” levels, for each factor separately:

```
> plot(yield ~ fertilizer*manure, data=yield.df)
```

Shortest model notation for our initial model, the one w/ interaction —



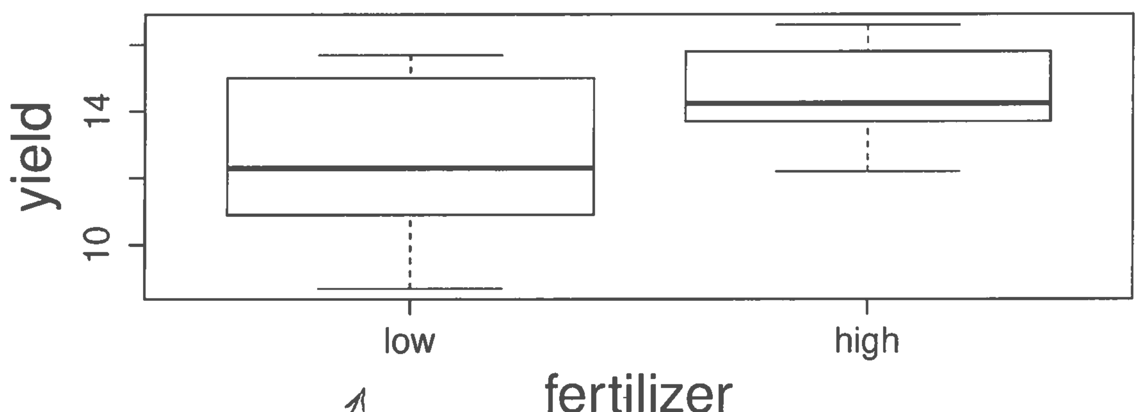
boxplot of 10 observations

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

for $i = 1, \dots, a$

$j = 1, \dots, b$

$k = 1, \dots, n$

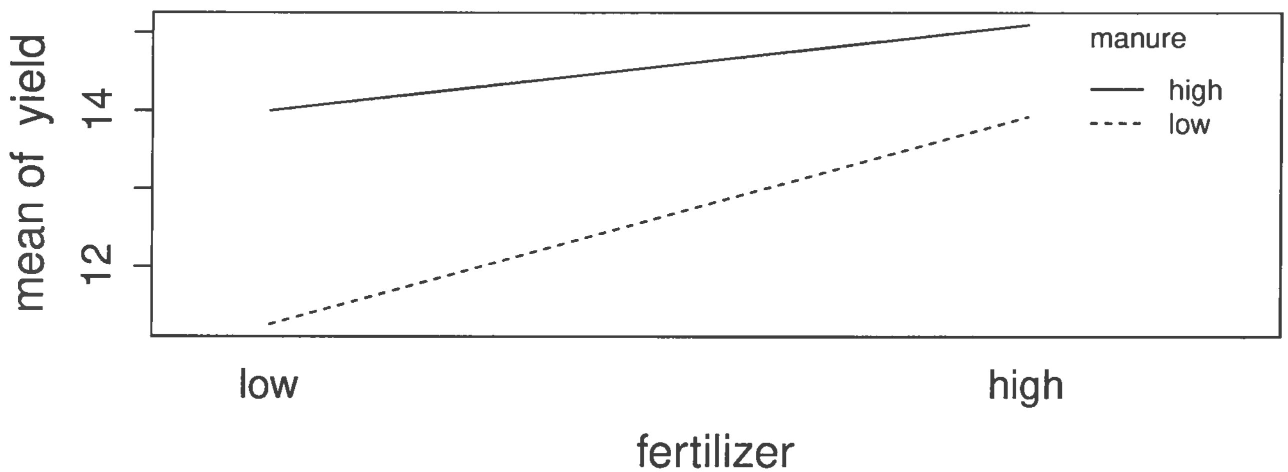


boxplot of 10 observations

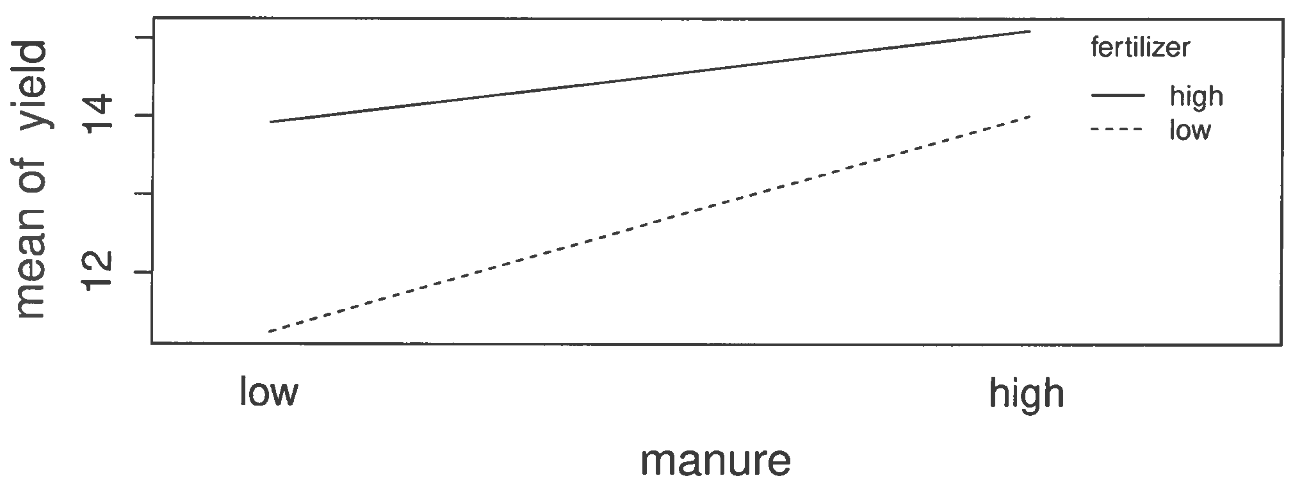
Interaction Plot We do this two possible ways; choose one.

for data

```
> interaction.plot(fertilizer,manure,response=yield)
> interaction.plot(manure,fertilizer,response=yield)
```



Looks as if interaction is present



Alternative model notation: manure + fertlizer + manure:fertilizer

Plot residuals

```
> 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
```

$= \hat{\sigma}^2 = MSE$

Three F tests are given in the table.

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

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

Test of interaction effect

H_0 : No interaction between the two factors

Or, more specifically:

H_0 : all $(\alpha\beta)_{ij} = 0$

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

The test statistic is:

$$F_{\text{obs}} = \frac{\text{MSAB}}{\text{MSE}}$$

Under H_0, the distribution of F_{obs} is $F_{(a-1)(b-1),(n-1)ab}$

Decision Rule:

WARNING - Different from generic
We want to accept Ho, but our usual test procedure is set up to reject Ho.
Think of this test as a preliminary test of ~~the~~ model assumption of additivity (which we hope holds).

Calculate $P = P\{F_{(a-1)(b-1),(n-1)ab} > F_{obs}\}$

If $P > .2$ we will accept Ho - no interaction
- analyze main effects

```
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
```

$F_{\text{obs}} = \frac{3.042}{2.775} = 1.096$; the null distribution of the F-statistic is $F_{1,16}$, $P = .3107$

(P > .2)

With this large a P-value we can "accept H_0" and conclude there is no interaction effect.

In this case, we proceed to test factor main effects. If the interaction **had** been significant, we would not have gone on to test factor main effects.

The null hypothesis means there is no interaction between fertilizer and manure in their effect on corn yield. There is no difference in the effect of fertilizer for the low and high levels of manure. There is no difference in the effect of manure for the low and high levels of fertilizer.

Note: The P-value should be at least .20 before you accept H_0. In borderline cases, consult a statistician. What if $.05 < P < .20$?

Question: Why can't you accept H_0 for $P = .06$ say?

Answer: We haven't determined $P(\text{Type II error})$.

Recall: Type I error is to reject H_0 when H_0 is true.

Type II error is to accept H_0 when H_a is true.