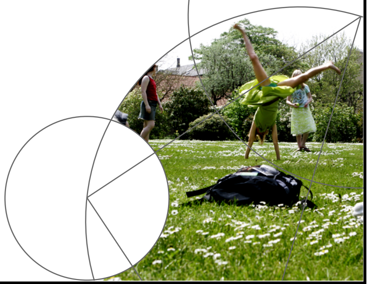
Faculty of Life Sciences

Linear models and interactions

Ib Skovgaard & Claus Ekstrøm
E-mail: ims@life.ku.dk



Program

- The (general) linear normal model
- Inference for linear models
- Product factors and interactions
- Successive tests, model reductions

The linear model

The (general) linear model is given by

$$y_i = \sum_{j=1}^{d} \beta_j \cdot x_{ij} + e_i, \; i = 1, \ldots, n,$$

where the βs are parameter, and the xs are explanatory variables — either quantitative or dummy variables.
Use `lm()`, `summary()` and `drop1()` as previously:
Example:

```
lm(blood.pres ~ gender + age)
```

Notice also that you do *not* need to define the dummy variables yourself — R does it for you.

Inference for the (general) linear model

Still the same:

estimates (LS),
test of hypotheses (F-tests),
confidence intervals,
prediction intervals,
model validation.

Degrees of freedom vary from model to model.

Example — disease in cucumber

Infection rate	Fertilizer dose		
Climate	2.0	3.5	4.0
A	51.5573	47.9937	57.9171
	51.6001	48.3387	51.3147
B	48.8981	48.2108	55.4369
	60.1747	51.0017	51.1251

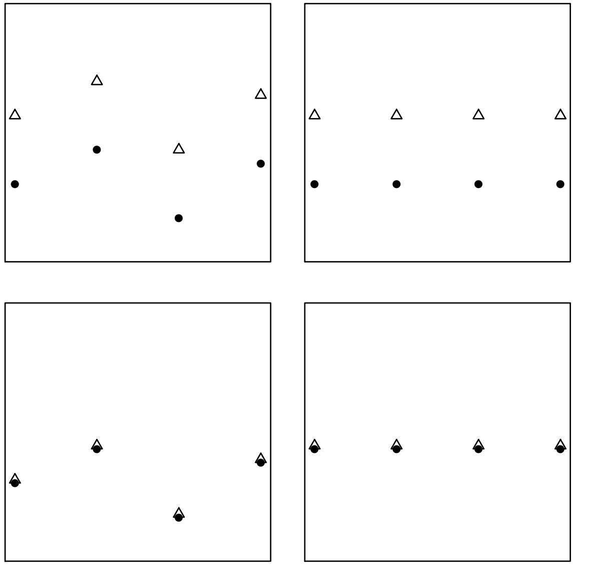
Hypothesis 1: Infection does not depend on climate:

$$H_1 : \alpha_A = \alpha_B$$

Hypothesis 2: Infection does not depend on dose:

$$H_2 : \beta_{2.0} = \beta_{3.5} = \beta_{4.0}$$

Graphical illustration of the hypotheses



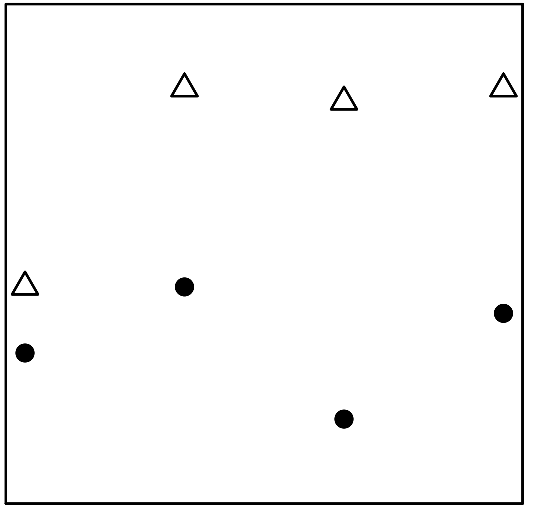
Interaction

But what if the effect of dose *depends on climate*?

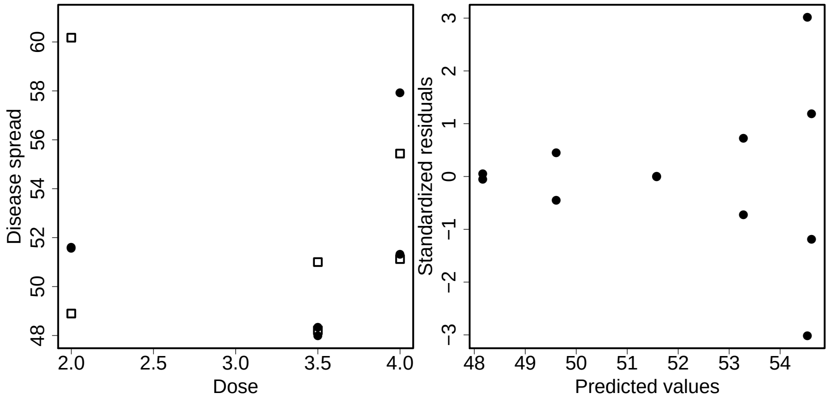
This means that the difference between dose 2.0 and dose 3.5, for example, depends on the climate.

This is in conflict with the *additive* two-way ANOVA model.
If the effect of a factor depends on another factor the two factors interact.

Graphical illustration of interaction



Cucumber experiment — interaction plot



Product factor

How can we analyze the data without assuming additivity?

Simple: Use a one-way ANOVA with 6 groups: one group for each combination of `climate` and `dose`.

The combination of two factors, is called the product factor. Here, the product factor divides the observations into $2 \times 3 = 6$ groups.

The interaction is that part of the variation between the 6 groups (from the product factor) that cannot be explained by the two factors separately (by the additive model).

In R the product factor is written with a colon, `:`, between two factors. Thus, if climate and dose are factors in R, the product factor is written

```
climate:dose
```


Testing for interaction in R

The additive two-way model is a sub-model of the model with the product-factor!

Hence, we start with the model with the product-factor, and test the hypothesis that the additive model holds (no interaction).

```
climate <- factor(climate) # if not factor already
dose <- factor(dose)       # if not factor already
model1 <- lm(infection ~ climate + dose + climatedose)
model2 <- lm(infection ~ climate + dose)
anova(model2, model1) # F-test: additivity?
```

Easier version:

```
model1 <- lm(infection <- climate*dose)
drop1(model1)         # F-test: additivity?
```


Successive tests, model reductions

Important guidelines for the analyses:

Model reductions / successive tests

If an effect in the model is insignificant it may be reasonable to remove it from the model to simplify the estimates and the description of the results.

Hierarchical principle

If a model contains an interaction between two factors, this model cannot be used to test the (main) effect of each of the factors.

Conclusions without interaction

No interaction = additivity

If interaction between two factors is clearly non-significant the effect of each factor may be reported separately, ignoring the other factor.

For example, assuming that the additive model, estimates and confidence intervals are given for

- difference between climate A and B (applies to all doses),
- differences between pairs of doses (applies to both climates).

Conclusions with interaction

Interaction = not additivity

If two factors seem to interact (significant interaction) the effect of each combination of the two factors must be estimated. It does not make sense to give estimates for any one of the factors separately.

For example, assuming significant interaction, estimates and confidence intervals are given for

- the different doses in climate A,
- the different doses in climate B.

In this way estimates are broken up according to climate.

Alternatively, break up the effect according to dose, and estimate the difference between climate A and B for each dose.

Lecture summary: main points

- Models with quantitative and qualitative variables belong to the same framework (linear models)
- Interaction between factors
  - Product factor = combination grouping
  - No interaction = additivity
  - Testing for interaction
  - Reporting results with or without interaction
- Model hierarchies and successive testing