Factorial Designs

Marek J. Drużdżel

Politechnika Białostocka
Wydział Informatyki

m.druzdzel@pb.edu.pl
http://www.wi.pb.edu.pl/~druzdzel/

Overview

- **Experimental designs (review)**
- **Factorial designs**
- **Concluding remarks**

The necessity of empirical work

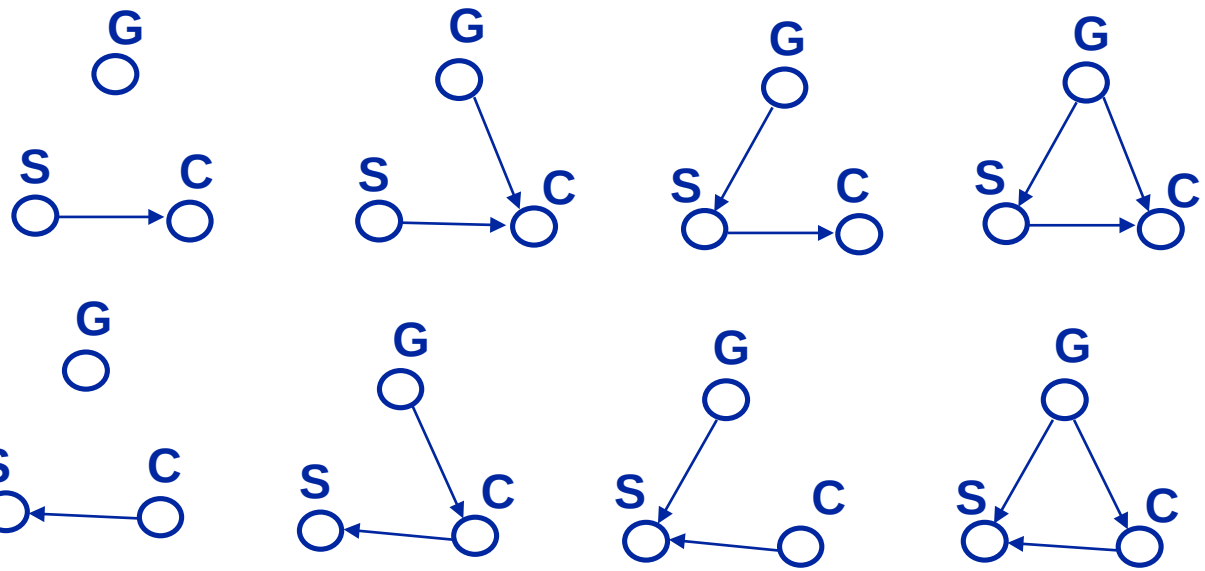
Smoking and lung cancer are correlated.
Can we reduce the incidence of lung cancer by reducing smoking?
In other words: Is smoking a cause of lung cancer?

Each of the following causal structures is compatible with the observed correlation:

G = genetic factors

S = smoking

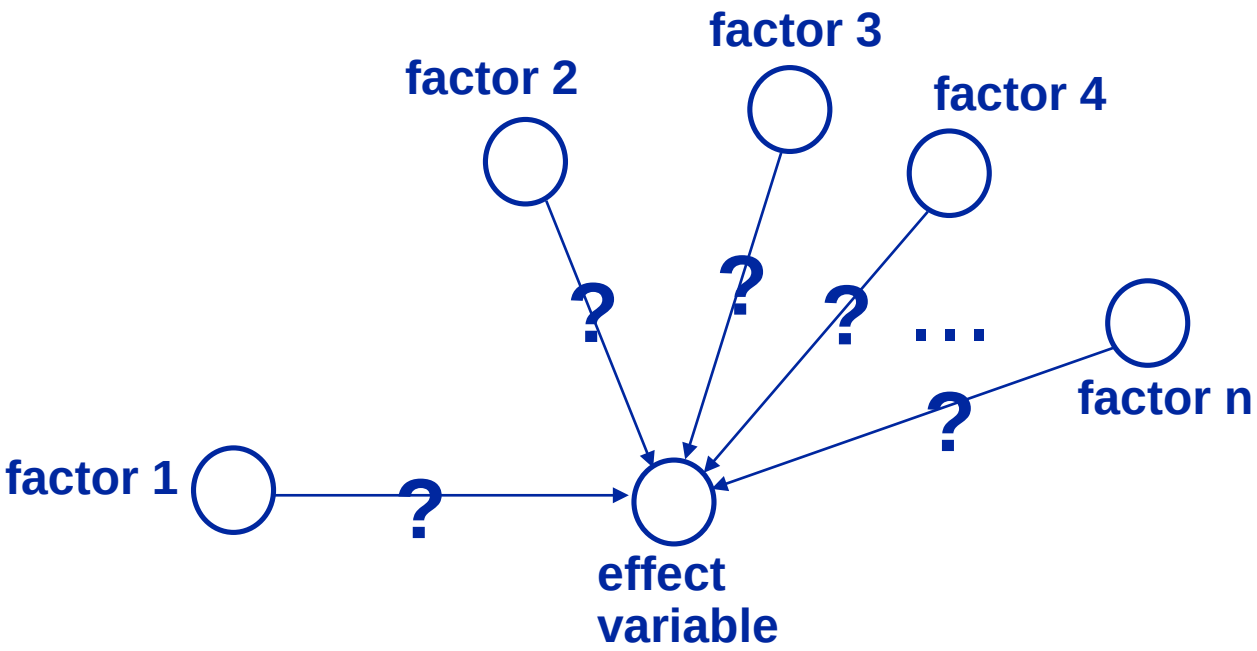
C = lung cancer



Single factor designs

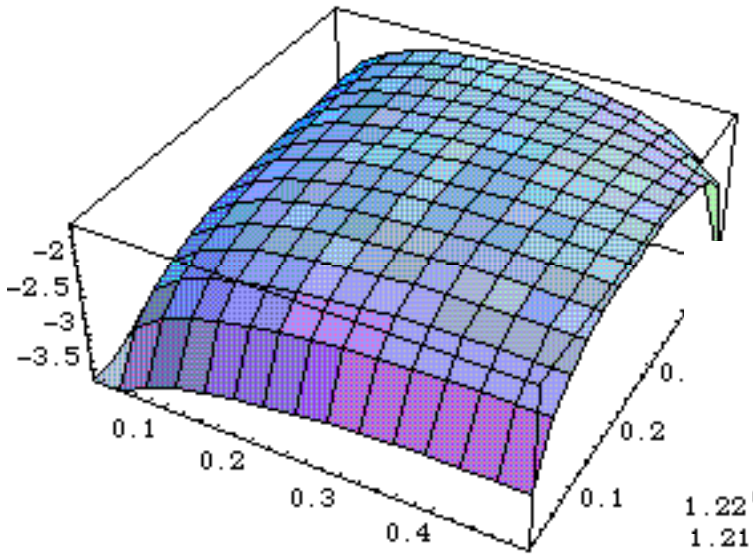
Factorial designs

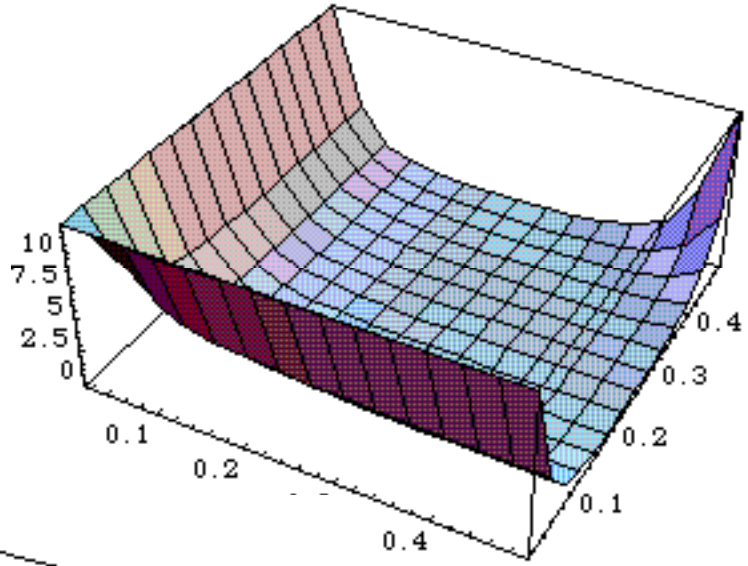
Often we would like to know the effect of various independent variables on the dependent variable, including interaction among them.

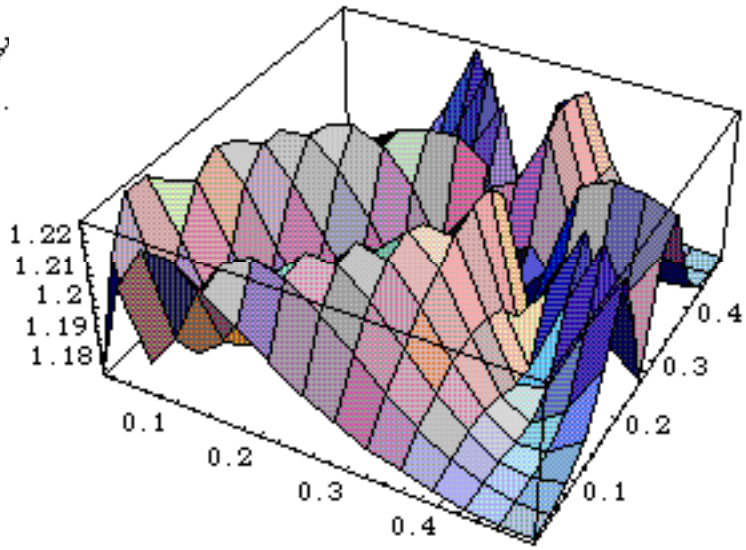


Interaction effects

Two variables can interact with one another producing the value of a third variable in a variety of ways.







Factorial designs

In general, we would have to multiply the number of values of each factor by the number of factors and have that many groups.

There are two problems with that:

sample size (the number of subjects in each group).

statistical decision making: What happens if we perform several one-way instead of one multi-way? There will usually be significant effects observed just by accident (i.e., if you perform 3x3x2=18 t-tests, then at p=0.05 level you will discover some that support H_0 by chance).

Factorial designs

The answer to that: an ingenious idea (due to Fisher) that allows to practically multiply the sample size.

We look at all of the results in a creative way and decompose them into row and column effects (a nice metaphor for two factor designs). This allows us to use all the data for each of the factors.

Factorial designs

		Factor 2: low	Factor 2: high
Factor 1	low	m	n
	high	q	r

Factorial designs

A basic equation:

Group_Mean = Grand_Mean + Row_Effects + Column_Effects + Interaction_Effects

Grand_Mean - mean of all cell means
Row_Effect - mean of the row minus the Grand_Mean
Column_Effect - mean of the column minus the Grand_Mean

Individual differences are seen as errors.

This assumes a very special type of function, it almost is like regressing a line to a table.

Watch out because in this equation the Grand_Mean is not the population mean, so the absolute size of the effect is not meaningful in general.

Analysis of variance (ANOVA)

ANOVA is a statistical method of comparing several population means.

Basic vocabulary:

factor - a categorical variable that distinguishes the population to be compared.

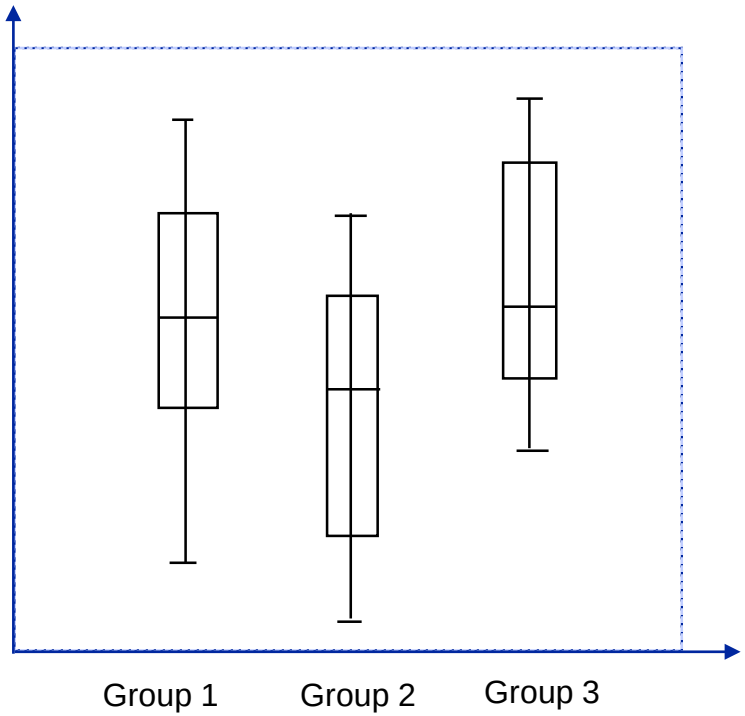
groups - factors in one-way ANOVA.

levels - different values of a factor.

Analysis of variance (ANOVA)

- The means in various groups will be (of course) different. The question is whether they are really different in the population or whether they just came out different by chance. In other words, are the differences "significant"? Null hypothesis, H_0, is "there are no differences in means," H_1 is “there are differences in means.”
- We compare the variation among the means of several groups with the variation within groups.
- The power of ANOVA depends on the effect sizes (difference in means between groups), group sizes, and within-group variances.

Analysis of variance (ANOVA)

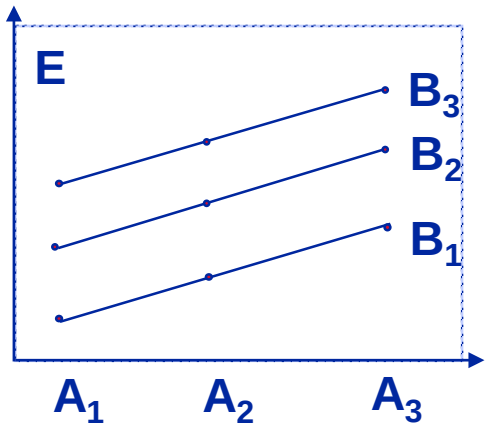


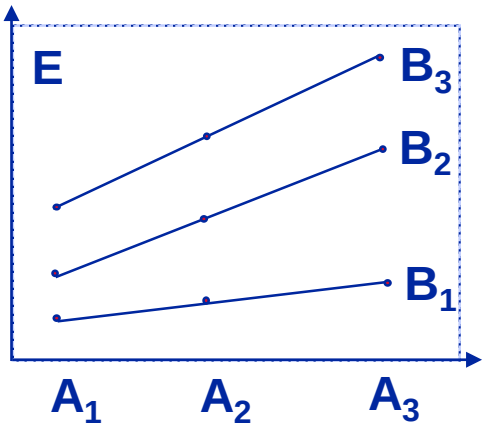
More than one-way ANOVA

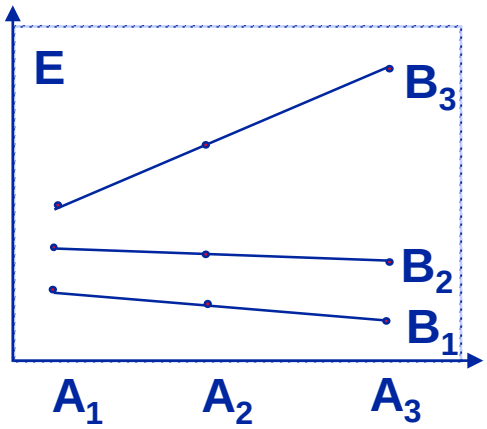
2x3 ANOVA means that there are two factors, the first has 2 levels and the second has three levels.

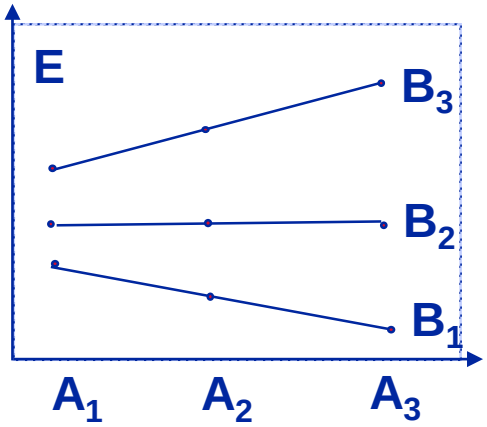
Interaction effects

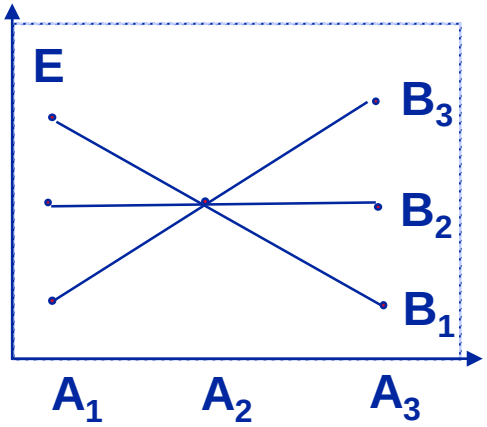
Multiple-way ANOVA simplifies the class of possible interactions. A simple graphical method for illustrating interaction, used a lot in social sciences is the following:









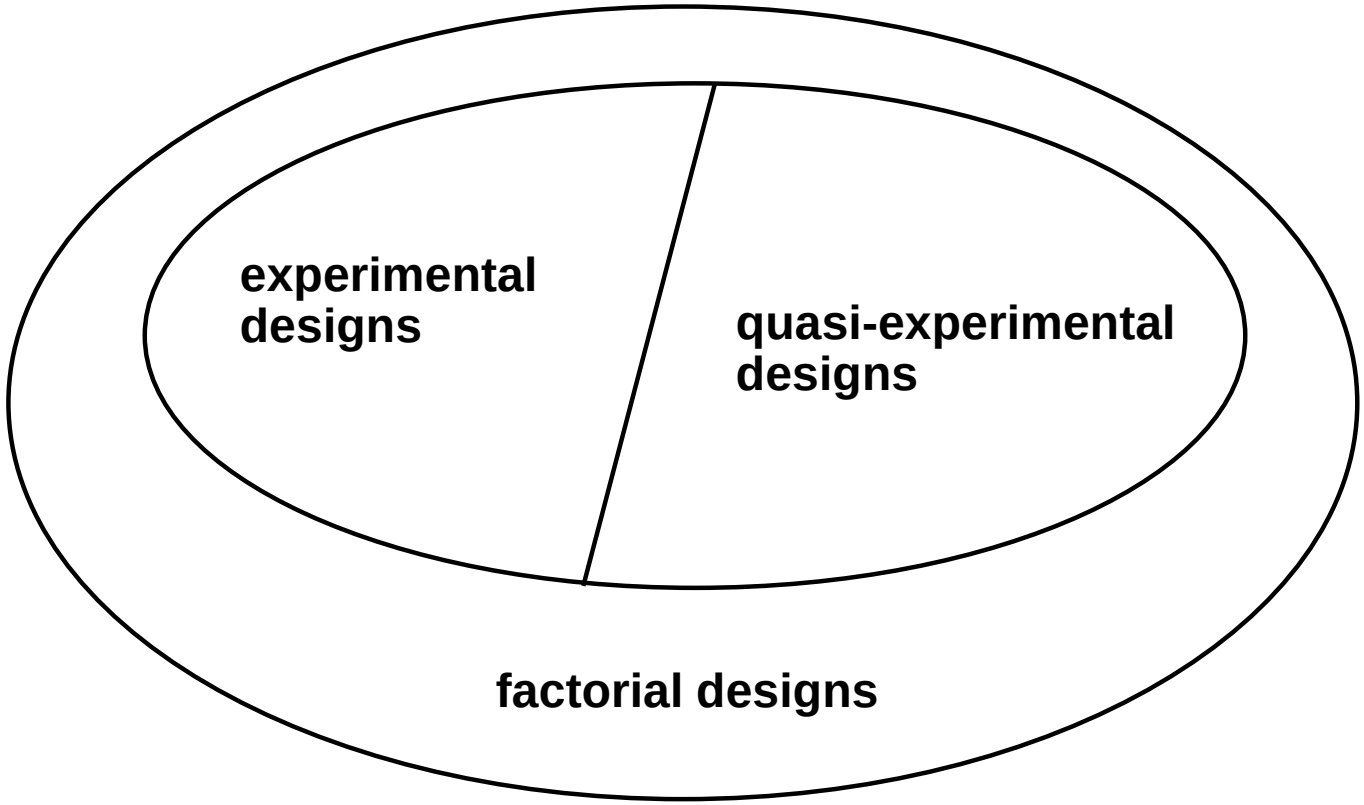


Can you identify these: monotonic interactions, no interaction effect (just summation of effects), synergistic effect, crossed-line interactions?

What will we observe in each case if we study one factor only?

Experimental designs roadmap

How different experimental and quasi-experimental designs are related



Concluding Remarks

- Factorial designs allow for studying interaction effects.
- ANOVA allows for a smart management of resources in terms of the number of subjects in each group.