Two-way designs

In a 2-way design, 2 factors (independent variables) are studied in conjunction with the response (dependent) variable. There is thus two ways of organizing the data, as shown in a 2-way table.

	Color		
Logo	Red	Green	Blue
Logo 1	14	14	14
Logo 2	14	14	14
Logo 3	14	14	14

2-way table
(3 by 3 design to test the attractiveness of a new website)

When the dependent variable is quantitative, the data are analyzed with a two-way ANOVA procedure.

Advantages of a two-way ANOVA model

- It is **more efficient** to study 2 factors at once than separately.

  A 2-way design requires smaller sample sizes per condition than a series of one-way designs would because the samples for all levels of factor B contribute to sampling for factor A.

- Including a second factor thought to influence the response variable helps **reduce the residual** variation in a model of the data.

  In a one-way ANOVA for factor A, any effect of factor B is assigned to the residual (“error” term). In a 2-way ANOVA, both factors contribute to the fit part of the model.

- **Interactions** between factors can be investigated.

  The 2-way ANOVA breaks down the fit part of the model between each of the main components (the 2 factors) and an interaction effect. The interaction cannot be tested with a series of one-way ANOVAs.

Interaction

Two variables interact if a particular combination of variables leads to results that would not be anticipated on the basis of the main effects of those variables.

- Drinking alcohol increases the chance of throat cancer, as does smoking. However, people who both drink and smoke have an even higher chance of getting throat cancer. The combination of smoking and drinking is particularly dangerous: these risk factors interact.

An interaction implies that the effect of one variable differs depending on the level of another variable.

- The effect of smoking on the probability of getting throat cancer is greater for people who drink than for people who do not drink: the effect of smoking differs depending on whether drinkers or nondrinkers are being considered.

The two-way ANOVA model

- We record a quantitative variable in a **two-way design** with *I* levels of the first factor and *J* levels of the second factor.

- We have **independent SRSs** from each of *I x J* Normal populations. Sample sizes do not have to be identical *(although when sample sizes are equal ⇔ "balanced design").*

- All parameters are unknown. The population means may be different but **all populations have the same standard deviation σ**.

Main effects and interaction effect

- Each factor is represented by a **main effect**: this is the impact on the response (dependent variable) of varying levels of that factor, regardless of the other factor (i.e., pooling together the levels of the other factor). There are two main effects, one for each factor.

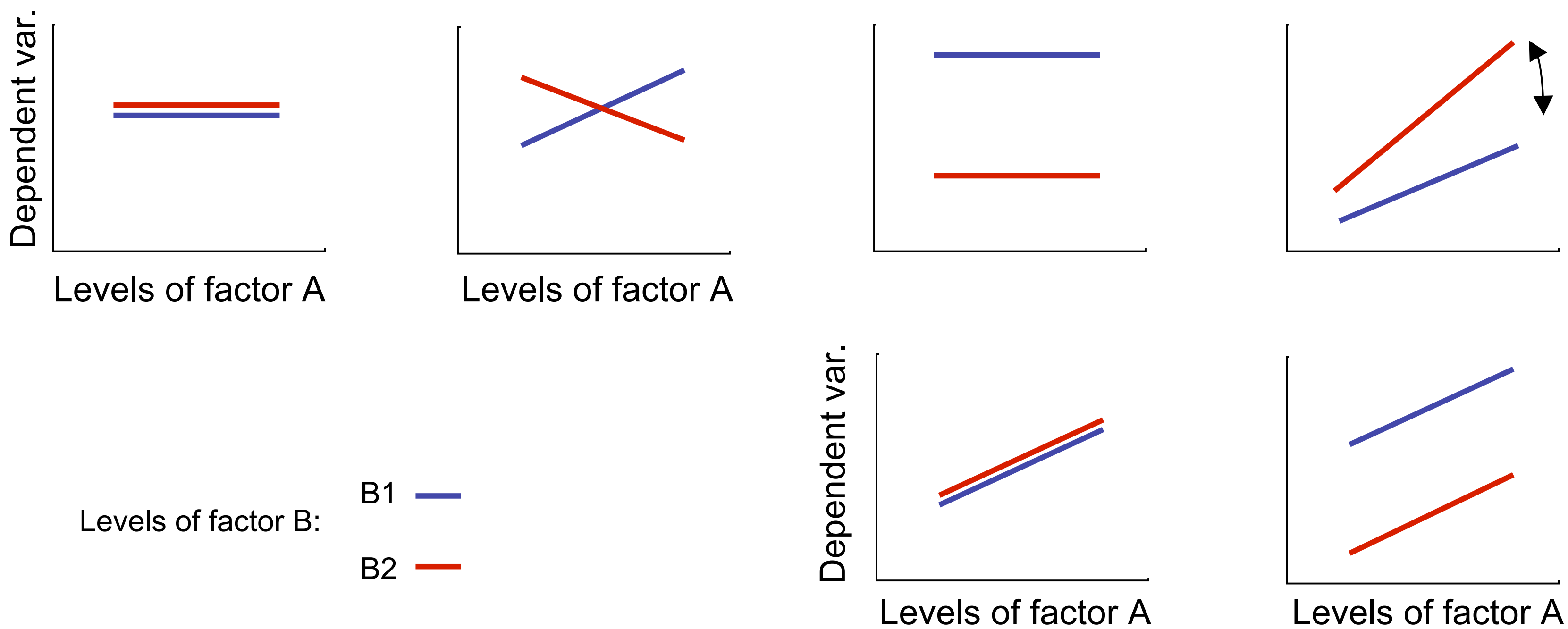
- The interaction of both factors is also studied and is described by the **interaction effect**.

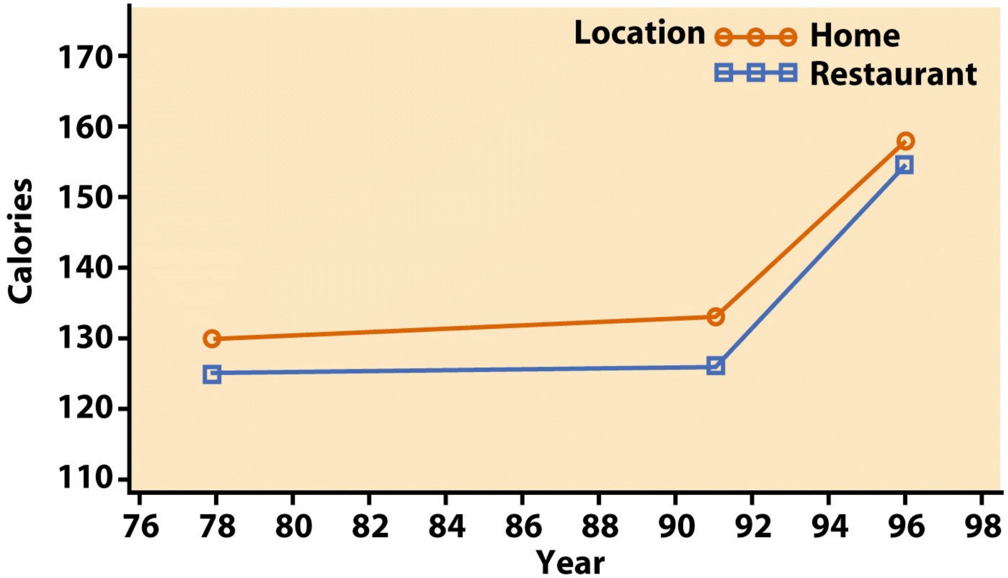
- When there is no clear interaction, the main effects are enough to describe the data. In the presence of interaction, the main effects could mask what is really going on with the data.

Major types of 2-way ANOVA outcomes

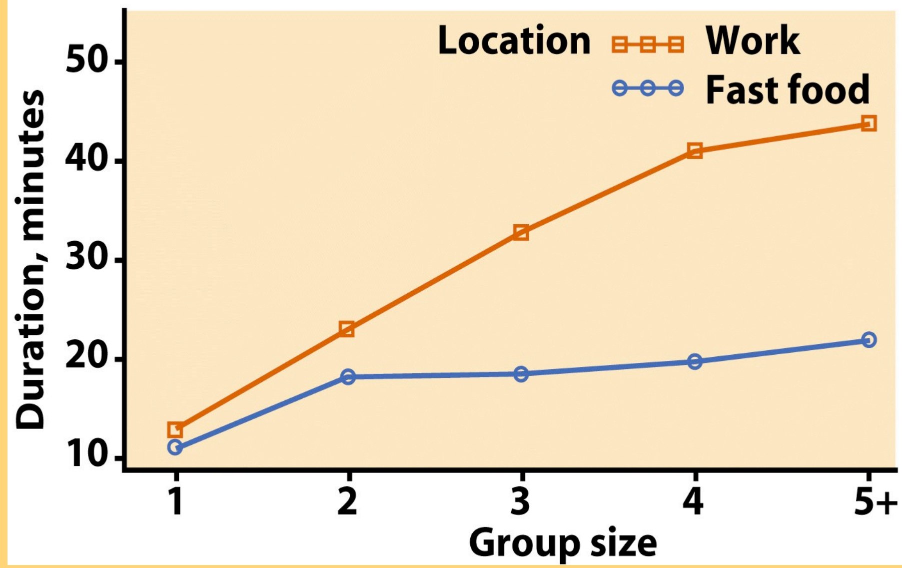
In a two-way design, statistical significance can be found for each factor, for the interaction effect, or for any combination of these.

Neither factor is significant	Neither factor is significant	Only 1 factor is significant	Both factors are significant
No interaction	Interaction effect is significant	No interaction	With or without significant interaction

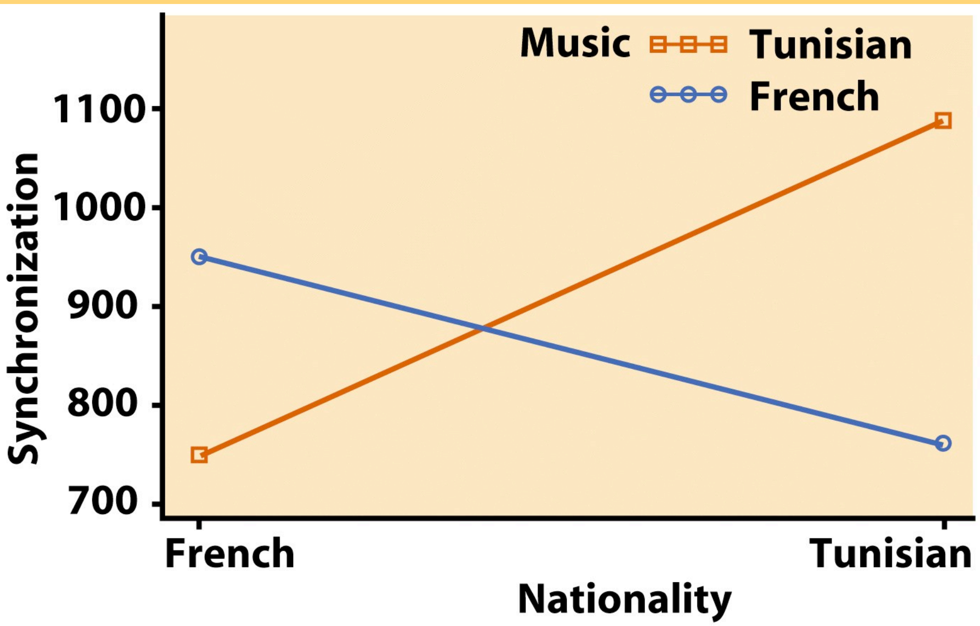




Main effects,
no interaction



Interaction effect: the main effects don’t tell the whole story.



Important interaction effect: the main effects are misleading.

Inference for two-way ANOVA

- A one-way ANOVA tests the following model of your data:

  Data ("total") = fit ("groups") + residual ("error")

  So that the sum of squares and degrees of freedom are:

  SST = SSG + SSE

  DFT = DFG + DFE

- A 2-way design breaks down the "fit" part of the model into more specific subcomponents, so that:

  SST = SSA + SSB + SSAB + SSE

  DFT = DFA + DFB + DFAB + DFE

  Where A and B are the 2 **main effects** from each of the 2 factors, and AB represents the **interaction** of factors A and B.

The two-way ANOVA table

Source of variation	DF	Sum of squares SS	Mean square MS	F	P-value
Factor A	DFA = *I* -1	SSA	SSA/DFA	MSA/MSE	for F_A
Factor B	DFB = *J* - *I*	SSB	SSB/DFB	MSB/MSE	for F_B
Interaction	DFAB = *(I-1)(J-1)*	SSAB	SSAB/DFAB	MSAB/MSE	for F_{AB}
Error	DFE = *N* - *IJ*	SSE	SSE/DFE		
Total	DFT = *N* – 1 =DFA+DFB+DFAB+DFE	SST =SSA+SSB+SSAB+SSE	SST/DFT		

- **Main effects:** P-value for factor A, P-value for factor B.
- **Interaction:** P-value for the interaction effect of A and B.
- **Error:** It represents the variability in the measurements within the groups.

MSE is an unbiased estimate of the population variance σ^2.

Nematodes and plant growth

Do nematodes affect plant growth? A botanist prepares 16 identical planting pots and adds different numbers of nematodes into the pots. Seedling growth (in mm) is recorded 2 weeks later. A one-way ANOVA can be used to analyze this data - there is only one factor: # of nematodes.

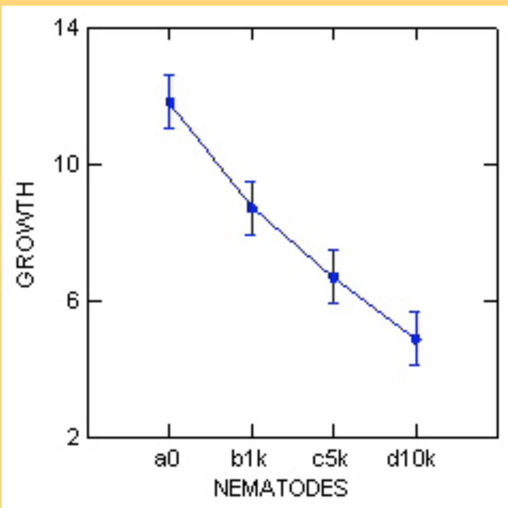
The two following slides consider two two factor problems:

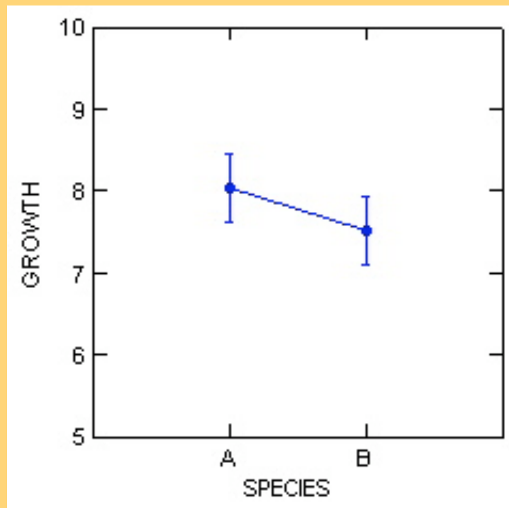
1. Suppose we also have data for a <u>second</u> plant species. We can study the effect of nematode amounts (4 levels) on seedling growth for both plant species (2 levels, species A and B).

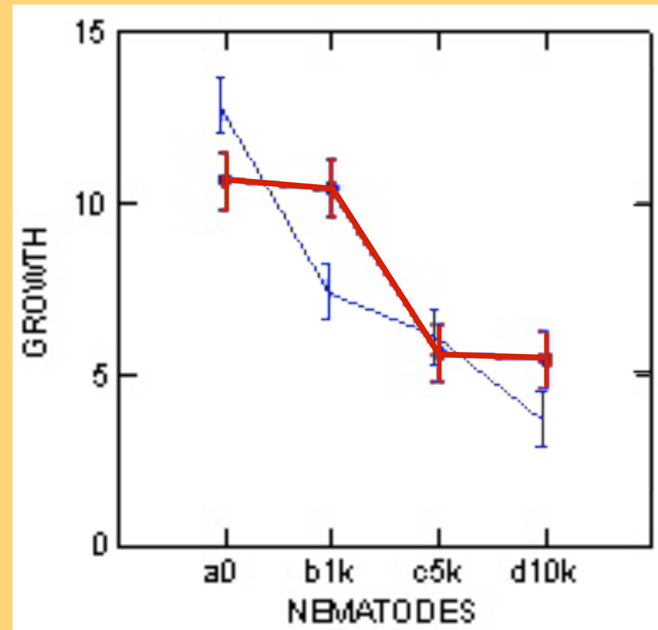
2. Suppose we have only one species A of plants, but some were grown with pesticide and some without. We can analyze seedling growth for combinations of nematodes (4 levels) and pesticide conditions ('no' and 'yes').

Nematodes level and plant species

Source	SS	df	MS	F-ratio	P
NEMATODES	254.645	3	84.882	31.002	0.000
SPECIES	2.101	1	2.101	0.767	0.390
NEMATODES*SPECIES	34.124	3	11.375	4.154	0.017
Error	65.710	24	2.738		





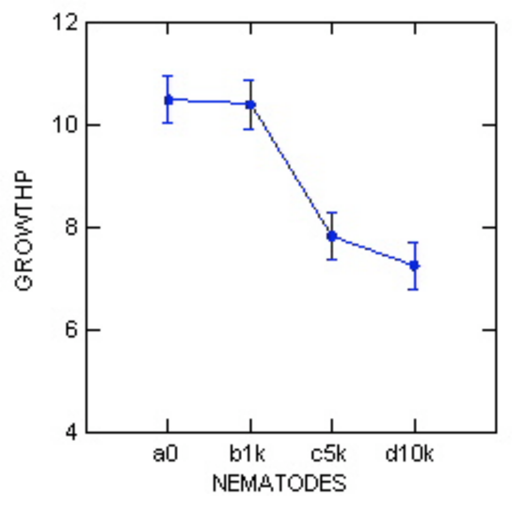


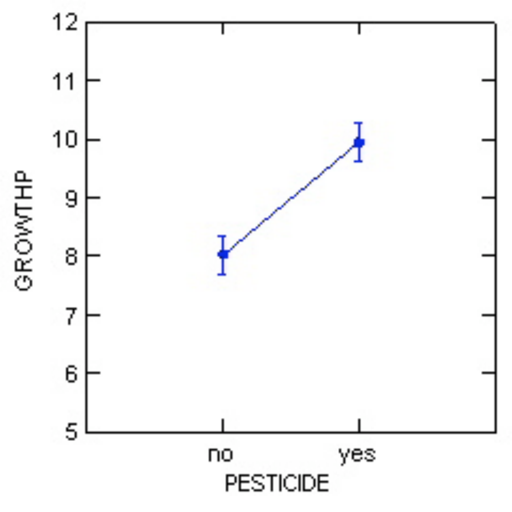
All plants suffer from the presence of nematodes (main effect $p<0.001$) but plant A and plant B do not have significantly different growth (main effect $p=0.39$).

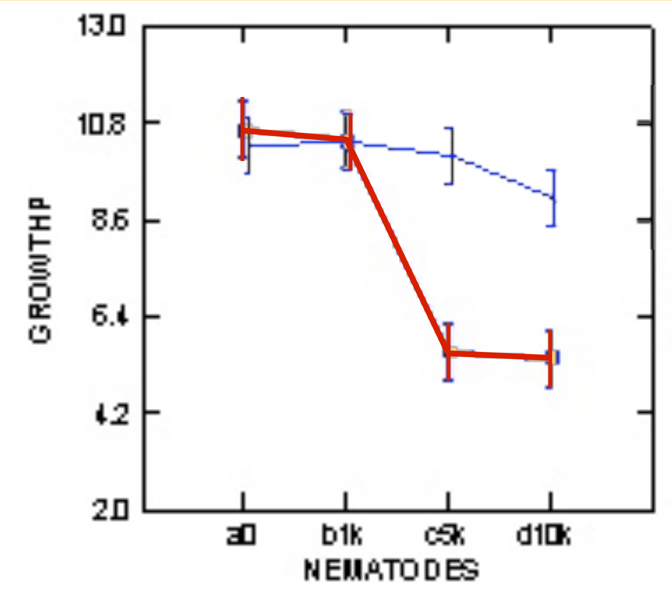
The third plot shows that the effect of nematodes is lower for plant A in red (interaction effect $p=0.017$).

Nematodes level and pesticide

Source	SS	df	MS	F-ratio	P
NEMATODES	68.343	3	22.781	13.195	0.000
PESTICIDE	29.070	1	29.070	16.837	0.000
NEMATODES*PESTICIDE	36.711	3	12.237	7.087	0.001
Error	41.438	24	1.727		







Both main effects are very significant.

The interaction is significant (p=0.001): we can see from the third plot that the detrimental effect of nematodes is much stronger in pesticide-free pots in red.