

1 Factorial Designs

In the designs discussed so far, completely randomized one-way ANOVA and Randomized Block Design, included only one factor variable of interest.

In a factorial design we will now discuss how more than one factor can be included in the model, and how we study the interaction between such factors.

Assume the presence of one factor A at a levels and a factor B at b levels, then there are ab different factor combinations, or treatments.

	Level	Factor B at b levels			
		1	2	...	b
Factor A at a levels	1				
	2				
	$\vdots$				
	a				

Each combination of a row and column lead to a different treatment. That is each combination of factors lead to a different treatment.

Example 1

Researchers at a trauma center wished to develop a program to help brain-damaged trauma victims regain an acceptable level of independence. An experiment involving 72 subjects with the same degree of brain damage was conducted. The objective was to compare different combinations of psychiatric treatment and physical therapy. Each subject was assigned to one combinations of four types of psychiatric treatment and six physical therapy programs. There were three subjects in each combination. The response variable is the number of months elapsing between initiation of therapy and time at which the patient was able to function independently.

response variable = month from therapy begin to independence

factor A = psychiatric treatment (1,2,3,4), $a = 4$

factor B = physical therapy (I, II, III, IV, V, VI), $b = 6$

There are $24 = ab = 4 \times 6$ possible treatments, that is factor combinations, since 3 patients were treated with each combination this gives a total sample size of $3 \times 24 = 72$.

The data:

Physical Therapy	Psychiatric Treatment			
	1	2	3	4
I	11.0	9.4	12.5	13.2
	9.6	9.6	11.5	13.2
	10.8	9.6	10.5	13.5
II	10.5	10.8	10.5	15.0
	11.5	10.5	11.8	14.6
	12.0	10.5	11.5	14.0
III	12.0	11.5	11.8	12.8
	11.5	11.5	11.8	13.7
	11.8	12.3	12.3	13.1
IV	11.5	9.4	13.7	14.0
	11.8	9.1	13.5	15.0
	10.5	10.8	12.5	14.0
V	11.0	11.2	14.4	13.0
	11.2	11.8	14.2	14.2
	10.0	10.2	13.5	13.7
VI	11.2	10.8	11.5	11.8
	10.8	11.5	10.2	12.8
	11.8	10.2	11.5	12.0

Data handed out in class as data sheet.

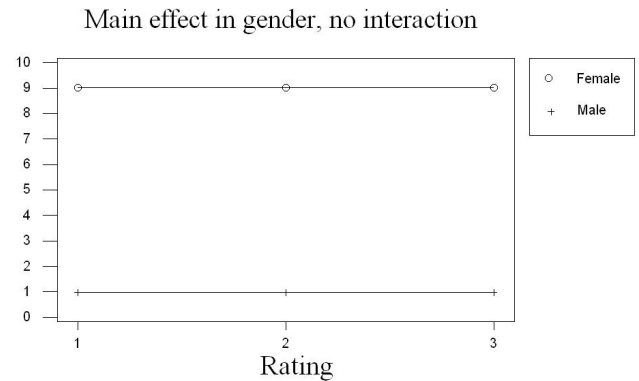
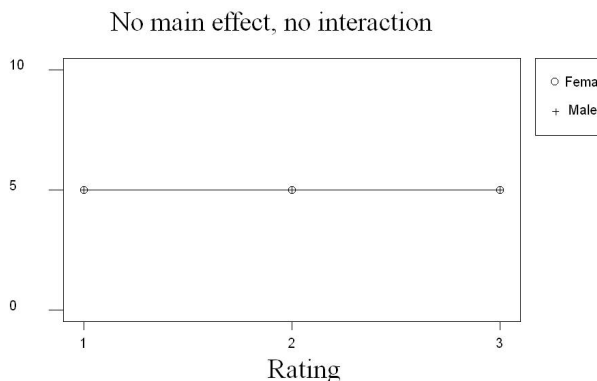
The questions to be answered through this study include:

- if certain physical or psychiatric therapies, or a certain combination of these is more successful than the others, i.e. results in a shorter time until independence of the patient.

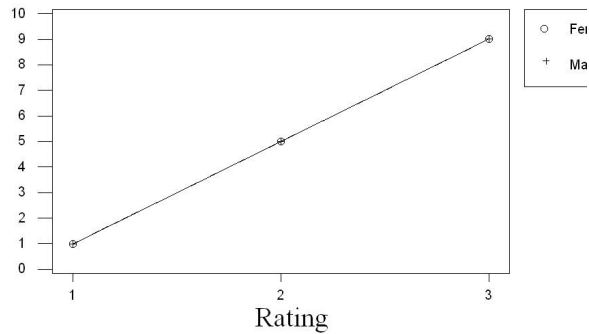
We will see how to use an ANOVA table to make decisions about such questions. Followed by multiple comparisons if effects for the different factors can be proven.

A new concept that is studied within factorial design is interaction between two factors.

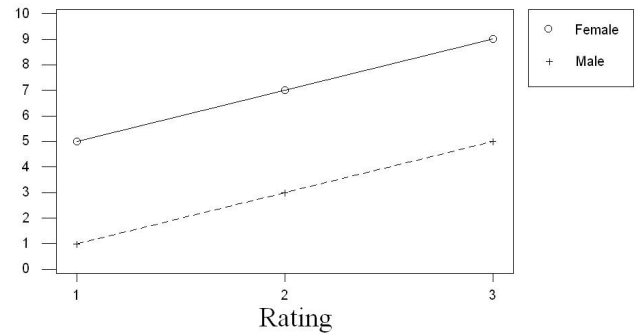
To illustrate this concept we will assume a 2 factorial design with factor A = gender at 2 levels and factor B = rating at 3 levels.



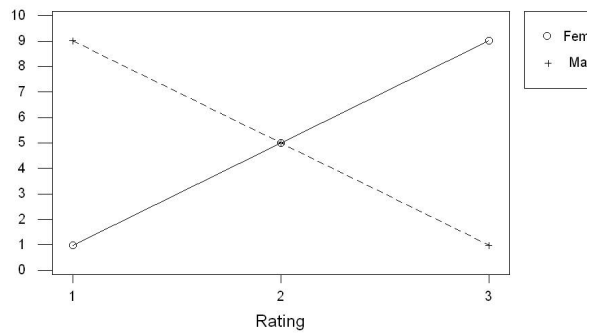
main effect in rating, no main effect in gender, no interaction



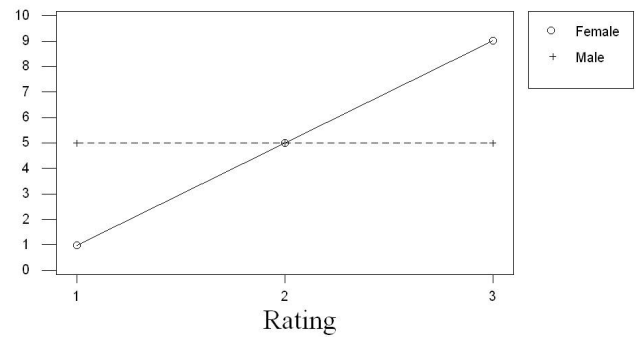
main effect in rating and gender, no interaction



no main effect in gender and rating, interaction

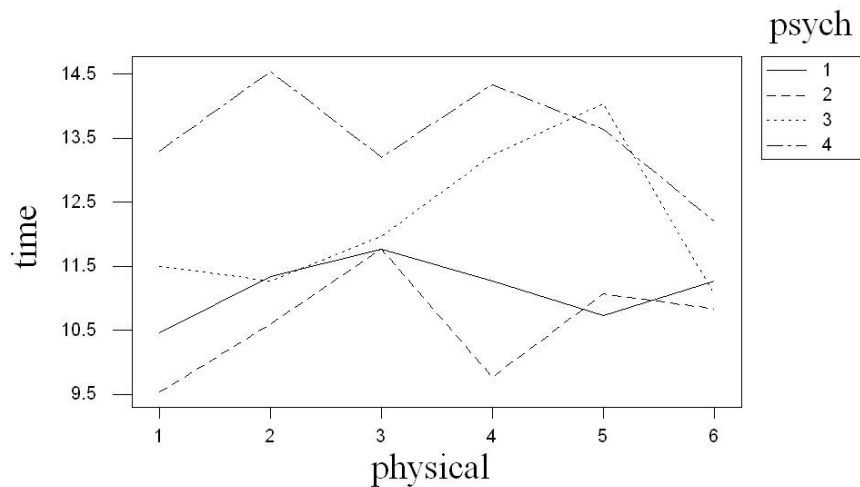


main effect in rating, no main effect in gender, interaction



Example 2

The equivalent graph for the the example is: (this is the minitab graph, SPSS would not let me save as jpeg file):



The graph illustrates that we should expect to find

- an effect due to the psychiatric treatment (the lines differ in their mean time).
- an effect due to the physical treatment (the mean times for the different physical therapies are different).

- interaction to be present between the factors (the lines are not parallel).

The model for a two factorial ANOVA:

x_{ijl} = the l^{th} measurement for level i of factor A and level j of factor B.

then

$$x_{ijl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijl}$$

where:

μ = overall mean

α_i = main effect of level i of factor A (difference with μ)

β_j = main effect of level j of factor B (difference with μ)

$(\alpha\beta)_{ij}$ = interaction effect of level i of factor A and level j of factor B

e_{ijl} = error in the l^{th} measurement for level i of factor A and level j of factor B.

In order to generalize this model to 3 or more factors you would add one main effect term for each factor and interaction terms for each combination of factors. It is also possible to include higher degree interaction terms, i.e. interaction of three or more factors.

To analyze such a model and test for main and interaction effects an ANOVA table is again utilized. The ANOVA table will include each factor and each interaction as a source of variation in the data. Each measured through their specific sum of squares.

2-way ANOVA table				
Source	Sum of Squares	df	Mean Square	F
Factor A	SS(A)	a-1	MS(A)	MS(A)/MSE
Factor B	SS(B)	b-1	MS(B)	MS(B)/MSE
Interaction Factor A * Factor B	SS(AB)	(a-1)(b-1)	MS(AB)	MS(AB)/MSE
Error	SSE	n-a-b-(a-1)(b-1)+1	MSE	
Total	TotalSS	n-1		

Three different tests shall be considered in this situation

1. In a first step we should ask if we have sufficient evidence that the mean response is not the same for all **treatments** (i.e. factor level combinations). Because otherwise no further analysis is needed, the factors have no influence on the mean of the response variable.

For testing the null hypothesis that all treatment means are equal versus at least one is different, we will use an F-statistic based on the Mean Squares for Treatment (MST)

$$MST = \frac{SS(A) + SS(B) + SS(AB)}{(a-1) + (b-1) + (a-1)(b-1)}$$

and the Mean Square for Error (MSE).

For the example above this would mean, that we ask if all therapy combinations of psychiatric and physical therapy result in the same mean recovery time until independence, or the therapy combinations are having all the same effect on the mean time until independence.

This test is also called model utility test, because through it we test if the model is meaning or useful in describing the relationship between the response and the factor variables.

For the example on the impact of the different physical and psychiatric therapies, this test would be testing if the effect on the time until independence of all therapy combinations is the same.

If we can not reject this null hypothesis, it would mean that the choice of therapy is irrelevant when investigating the time until independence. No further questions need to be asked. None of the factors show an influence on the response.

2. Once we have established that NOT all treatment means are equal, we continue to ask if the two factors show an interaction effect by using an F-statistic based on $SS(AB)$ and SSE and their degrees of freedom.

For our example this means, asking if there are certain psychiatric/physical therapy combinations that work particularly well or bad, beyond what we can explain through the sum of the effects caused by each factor separately.

3. In addition (independent from the result of the test for interaction) we can ask for main effects of the two factors. (This is different from the text book). By comparing the mean squares for each effect with the mean squares for error (through an F statistic), we are able to decide if the factor in question influences the mean of the response variable of interest, e.g. does the choice of physical therapy have a significant influence on the mean time until independence, after accounting for the different psychiatric treatments?

Instead of outlining all three categories of tests we will only present the test for a main effect of factor A

The 2 way ANOVA F-Test for an effect of factor A

1. The Hypotheses are

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_a = 0 \text{ versus} \\ H_a : \text{at least one of the values differs from the others}$$

2. Assumption: The response variable follows a normal distribution within each treatment (that is each factor combination) and that the standard deviations for all factor combinations are equal.

The samples are independent random samples for each factor combination.

3. Test statistic:

$$F_0 = \frac{MS(A)}{MSE}$$

$MS(A)$ are the mean squares for factor A . F_0 is based on $df_1 = (a - 1)$ and $df_2 = n - a - b - (a - 1)(b - 1) + 1$.

4. P-value: $P(F > F_0)$, where F follows an F -distribution with $df_1 = (a - 1)$ and $df_2 = n - a - b - (a - 1)(b - 1) + 1$.

5. Decision: If $P\text{-value} \leq \alpha$, then reject H_0 .
If $P\text{-value} > \alpha$, then do not reject H_0 .

6. Put into context.

Similar we can test for influence of factor B, by using $F_0 = MS(B)/MSE$ with $df_1 = (b - 1)$ and $df_2 = n - a - b - (a - 1)(b - 1) + 1$,
or for interaction, by using $F_0 = MS(AB)/MSE$ with $df_1 = (a - 1)(b - 1)$ and $df_2 = n - a - b - (a - 1)(b - 1) + 1$,
or for a treatment effect, by using $F_0 = MST/MSE$ with $df_1 = (a - 1) + (b - 1) + (a - 1)(b - 1)$ and $df_2 = n - a - b - (a - 1)(b - 1) + 1$.

Example 3

The ANOVA table for our example is (obtained from SPSS):

2-way ANOVA table for time to independence					
Source	Sum of Squares	df	Mean Square	F	P-value
Model	136.204	23	5.922	16.293	.000
psych	90.408	3	30.136	82.911	.000
physical	13.796	5	2.759	7.591	.000
psych * physical	32.001	15	2.133	5.869	.000
Error	17.447	48	.363		
Total	10283.040	71			

Interpretation:

- In the first line in the table labeled Model the information for the model utility test is given.
 H_0 : mean time until recovery is equal for all treatments, versus H_a : at least one mean is different. $\alpha = 0.05$

We will assume that the assumption of normality of the response variable is met, we will see later how we can check it.

$F_0 = 16.293$, $df_1 = 23$, $df_2 = 48$ and $P\text{-value} < 0.0001$

Since the p-value is less than 0.05, we would reject H_0 and conclude that at least one for one therapy combination the mean time until independence differs from the other treatments. It is meaningful to investigate which therapy(ies) is(are) superior to the others.

- In the second line labeled Intercept a test is reported for the hypothesis, if the overall mean of the response variable equals 0, we would reject this hypothesis, since the p-value is less than 0.05. In this context this is not a question of interest.
- The third row gives the information on testing for main effect of factor "psychiatric therapy". The null hypothesis states that the choice of psychiatric therapy has NO influence on the mean time until independence ($H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ versus H_a : at least one is different). $\alpha = 0.05$

Test statistic: $F_0 = 82.911$, $df_1 = 3$, $df_2 = 48$

P-value<0.0001

Decision/Context: At significance level 0.05 this hypothesis is rejected, and we conclude the mean time until a patient becomes independent is not the same for all psychiatric therapies, averaging over all physical therapies.

- Similar for the test if the choice of physical therapy influences the mean time until independence results are reported in line four.

In this test we would reject that the mean time until independence is independent from the choice of physical therapy, averaging over all psychiatric therapies.

- In line five the results for the test for of interaction are given:

$H_0 : (\alpha\beta)_{11} = \dots = (\alpha\beta)_{46} = 0$ versus H_a : at least one is different.

test statistic: $F_0 = 5.869, df_n = 15, df_d = 48$

p-value< 0.0001

This hypothesis is also rejected and we conclude that the two factors interact, that is different combinations of physical and psychiatric therapy result in significant different mean time until recovery than can be explained by just adding the main effects for the physical and psychiatric therapies.

This indicates that certain combinations are particularly beneficial or bad in terms of mean time until independence.

- We already predicted these results through analyzing the line graph.

Checking the Assumption

To check the assumptions for all the tests above, we should check if it is reasonable to assume that the response variable is normally distributed, and if the standard deviation is about the same for all therapies.

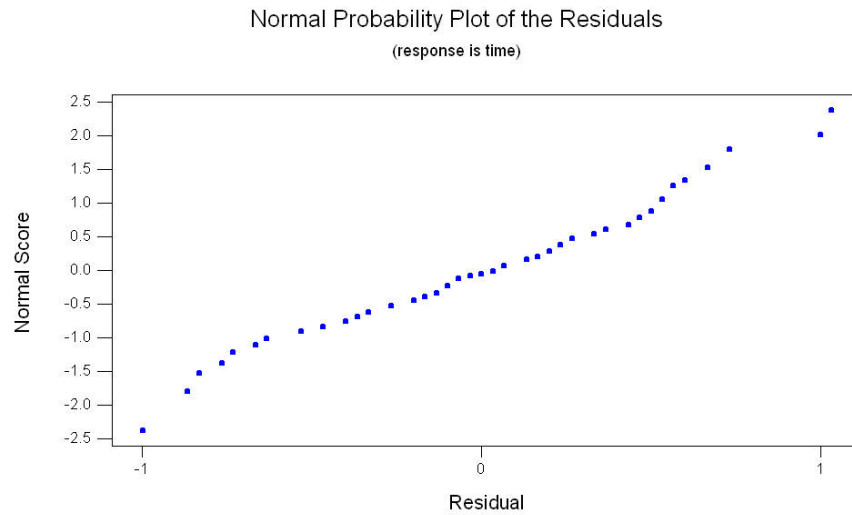
To check for normality we can use the residuals:

The residual for an individual measurement is the difference between the measured value and its estimate based on the model. For two way ANOVA this is:

$$r_{ijl} = x_{ijl} - \hat{x}_{ijl}$$

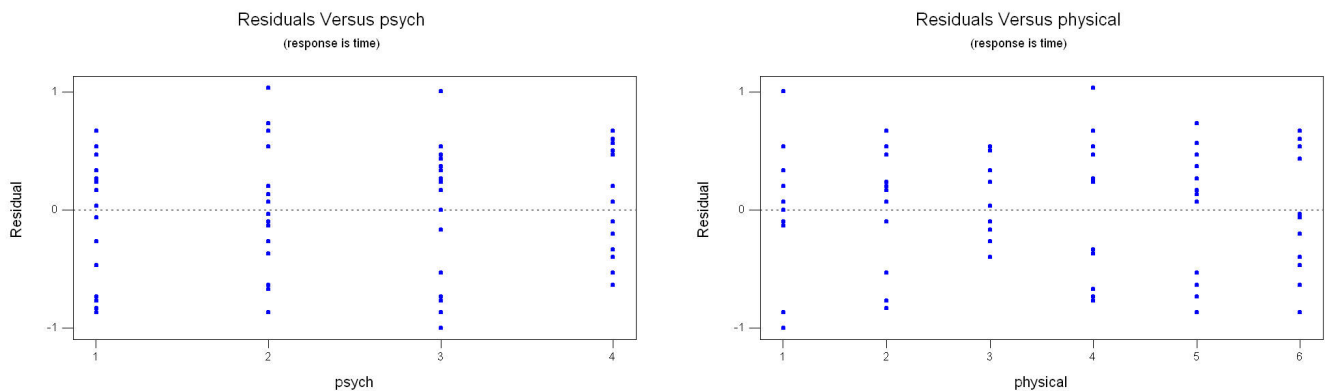
The estimate $\hat{x}_{ijl}$ is based on the estimates for the parameters in the model, $\mu, \alpha, \beta, \dots$ ($\hat{x}_{ijl} = \bar{x} + \bar{x}_{Ai} + \bar{x}_{Bj}$).

Then the residuals can be interpreted as observations on the measurement error e_{ijl} from the model, which is assumed to have a normal distribution, which we can check with a QQ-plot for the residuals.



Since the dots almost fall on a straight line it is reasonable to assume that the residuals come from a normal distribution, that is the response variable is normally distributed for each treatment (factor combination).

To check if it is reasonable to assume that the standard deviations are the same for all therapies, we can plot the residuals versus the two factors:



Since the spread in the residuals seems to be about the same for all therapies, the assumption for the 2-way ANOVA seems to be met for this data.

After the main and interaction effects were established through the 2-way ANOVA, again we have to ask where the differences can be found: A multiple comparison should be conducted.

Multiple Comparisons Dependent Variable: time Bonferroni

(I) psych	(J) psych	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	.5444	.20096	.056	-.0086	1.0975
	3	-1.0389(*)	.20096	.000	-1.5919	-.4858
	4	-2.3944(*)	.20096	.000	-2.9475	-1.8414
2	1	-.5444	.20096	.056	-1.0975	.0086
	3	-1.5833(*)	.20096	.000	-2.1364	-1.0303
	4	-2.9389(*)	.20096	.000	-3.4919	-2.3858
3	1	1.0389(*)	.20096	.000	.4858	1.5919
	2	1.5833(*)	.20096	.000	1.0303	2.1364
	4	-1.3556(*)	.20096	.000	-1.9086	-.8025
4	1	2.3944(*)	.20096	.000	1.8414	2.9475
	2	2.9389(*)	.20096	.000	2.3858	3.4919
	3	1.3556(*)	.20096	.000	.8025	1.9086

Based on observed means. * The mean difference is significant at the .05 level.

Interpretation:

2 1 3 4

The only psychiatric therapies that are not significantly different are therapy 1 and 2. The mean recovery times for therapy 1 and 2 are significantly shorter, than for therapies 3 and 4, where therapy 3 leads in average to a shorter time until independence for trauma patients with brain damage.

Multiple Comparisons Dependent Variable: time Bonferroni

(I) physical	(J)physical	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-.7333	.24613	.068	-1.4937	.0270
	3	-.9750(*)	.24613	.004	-1.7353	-.2147
	4	-.9500(*)	.24613	.005	-1.7103	-.1897
	5	-1.1667(*)	.24613	.000	-1.9270	-.4063
	6	-.1417	.24613	1.000	-.9020	.6187
2	1	.7333	.24613	.068	-.0270	1.4937
	3	-.2417	.24613	1.000	-1.0020	.5187
	4	-.2167	.24613	1.000	-.9770	.5437
	5	-.4333	.24613	1.000	-1.1937	.3270
	6	.5917	.24613	.302	-.1687	1.3520
3	1	.9750(*)	.24613	.004	.2147	1.7353
	2	.2417	.24613	1.000	-.5187	1.0020
	4	.0250	.24613	1.000	-.7353	.7853
	5	-.1917	.24613	1.000	-.9520	.5687
	6	.8333(*)	.24613	.021	.0730	1.5937
4	1	.9500(*)	.24613	.005	.1897	1.7103
	2	.2167	.24613	1.000	-.5437	.9770
	3	-.0250	.24613	1.000	-.7853	.7353
	5	-.2167	.24613	1.000	-.9770	.5437
	6	.8083(*)	.24613	.029	.0480	1.5687
5	1	1.1667(*)	.24613	.000	.4063	1.9270
	2	.4333	.24613	1.000	-.3270	1.1937
	3	.1917	.24613	1.000	-.5687	.9520
	4	.2167	.24613	1.000	-.5437	.9770
	6	1.0250(*)	.24613	.002	.2647	1.7853
6	1	.1417	.24613	1.000	-.6187	.9020
	2	-.5917	.24613	.302	-1.3520	.1687
	3	-.8333(*)	.24613	.021	-1.5937	-.0730
	4	-.8083(*)	.24613	.029	-1.5687	-.0480
	5	-1.0250(*)	.24613	.002	-1.7853	-.2647

Based on observed means. * The mean difference is significant at the .05 level.

 1 6 2 4 3 5

Therapies 1,6, and 2, and therapies 2,4,3, and 5 result in non significantly different mean time until independence for trauma patients with brain damage.

Therapies 1, 6 and 4,3,5 are significantly different (pairwise). We can conclude that the mean time until independence for trauma patients is significantly shorter for therapies 1 and 6 than for therapies 4,3, and 5. Results for therapy 2 are indecisive.

The multiple comparisons just presented only compare either physical therapies or psychiatric therapies at a time. But we found interaction effects. For that reason it would be meaningful to compare all $4 \times 6 = 24$ therapies with each other.

We might have time in the lab to illustrate how this can be done.