

9.5: Simple analysis of 2x2 repeated measures design

Normally in a chapter about factorial designs we would introduce you to Factorial ANOVAs, which are totally a thing. We will introduce you to them soon. But, before we do that, we are going to show you how to analyze a 2x2 repeated measures ANOVA design with paired-samples t-tests. This is probably something you won't do very often. However, it turns out the answers you get from this method are the same ones you would get from an ANOVA.

Admittedly, if you found the explanation of ANOVA complicated, it will just appear even more complicated for factorial designs. So, our purpose here is to delay the complication, and show you with t-tests what it is that the Factorial ANOVA is doing. More important, when you do the analysis with t-tests, you have to be very careful to make all of the comparisons in the right way. As a result, you will get some experience learning how to know what it is you want to know from factorial designs. Once you know what you want to know, you can use the ANOVA to find out the answers, and then you will also know what answers to look for after you run the ANOVA. Isn't new knowledge fun!

The first thing we need to do is define main effects and interactions. Whenever you conduct a Factorial design, you will also have the opportunity to analyze main effects and interactions. However, the number of main effects and interactions you get to analyse depends on the number of IVs in the design.

Main effects

Formally, main effects are the mean differences for a single Independent variable. There is always one main effect for each IV. A 2x2 design has 2 IVs, so there are two main effects. In our example, there is one main effect for distraction, and one main effect for reward. We will often ask if the main effect of some IV is significant. This refers to a statistical question: Were the differences between the means for that IV likely or unlikely to be caused by chance (sampling error).

If you had a 2x2x2 design, you would measure three main effects, one for each IV. If you had a 3x3x3 design, you would still only have 3 IVs, so you would have three main effects.

Interaction

We find that the interaction concept is one of the most confusing concepts for factorial designs. Formally, we might say an interaction occurs whenever the effect of one IV has an influence on the size of the effect for another IV. That's probably not very helpful. In more concrete terms, using our example, we found that the reward IV had an effect on the size of the distraction effect. The distraction effect was larger when there was no-reward, and it was smaller when there was a reward. So, there was an interaction.

We might also say an interaction occurs when the difference between the differences are different! Yikes. Let's explain. There was a difference in spot-the-difference performance between the distraction and no-distraction condition, this is called the distraction effect (it is a difference measure). The reward manipulation changed the size of the distraction effect, that means there was difference in the size of the distraction effect. The distraction effect is itself a measure of differences. So, we did find that the difference (in the distraction effect) between the differences (the two measures of the distraction effect between the reward conditions) were different. When you start to write down explanations of what interactions are, you find out why they come across as complicated. We'll leave our definition of interaction like this for now. Don't worry, we'll go through lots of examples to help firm up this concept for you.

The number of interactions in the design also depend on the number of IVs. For a 2x2 design there is only 1 interaction. The interaction between IV1 and IV2. This occurs when the effect of say IV2 (whether there is a difference between the levels of IV2) changes across the levels of IV1. We could write this in reverse, and ask if the effect of IV1 (whether there is a difference between the levels of IV1) changes across the levels of IV2. However, just because we can write this two ways, does not mean there are two interactions. We'll see in a bit, that no matter how do the calculation to see if the difference scores—measure of effect for one IV—change across the levels of the other IV, we always get the same answer. That is why there is only one interaction for a 2x2. Similarly, there is only one interaction for a 3x3, because there again we only have two IVs (each with three levels). Only when we get up to designs with more than 2 IVs, do we find more possible interactions. A design with three IVs, has four interactions. If the IVs are labelled A, B, and C, then we have three 2-way interactions (AB, AC, and BC), and one three-way interaction (ABC). We hold off on this stuff for much later.

Looking at the data

It is most helpful to see some data in order to understand how we will analyze it. Let's imagine we ran our fake attention study. We will have five people in the study, and they will participate in all conditions, so it will be a fully repeated-measures design. The data could look like this:

	No Reward		Reward	
	No Distraction	Distraction	No Distraction	Distraction
subject	A	B	C	D
1	10	5	12	9
2	8	4	13	8
3	11	3	14	10
4	9	4	11	11
5	10	2	13	12

Note: Number of differences spotted for each subject in each condition.

Main effect of Distraction

The main effect of distraction compares the overall means for all scores in the no-distraction and distraction conditions, collapsing over the reward conditions.

The yellow columns show the no-distraction scores for each subject. The blue columns show the distraction scores for each subject. The overall means for for each subject, for the two distraction conditions are shown to the right. For example, subject 1 had a 10 and 12 in the no-distraction condition, so their mean is 11.

We are interested in the main effect of distraction. This is the difference between the AC column (average of subject scores in the no-distraction condition) and the BD column (average of the subject scores in the distraction condition). These differences for each subject are shown in the last green column. The overall means, averaging over subjects are in the bottom green row.

All Conditions							
	No Reward		Reward		Distraction Means		Distraction Effect
	No Distraction	Distraction	No Distraction	Distraction	No Distraction	Distraction	Difference
subject	A	B	C	D	AC	BD	AC.minus.BD
1	10	5	12	9	11	7	4
2	8	4	13	8	10.5	6	4.5
3	11	3	14	10	12.5	6.5	6
4	9	4	11	11	10	7.5	2.5
5	10	2	13	12	11.5	7	4.5
Means					11.1	6.8	4.3

Just looking at the means, we can see there was a main effect of Distraction, the mean for the no-distraction condition was 11.1, and the mean for the distraction condition was 6.8. The size of the main effect was 4.3 (the difference between 11.1 and 6.8).

Now, what if we wanted to know if this main effect of distraction (the difference of 4.3) could have been caused by chance, or sampling error. You could do two things. You could run a paired samples t -test between the mean no-distraction scores for each subject (column AC) and the mean distraction scores for each subject (column BD). Or, you could run a one-sample t -test on the difference scores column, testing against a mean difference of 0. Either way you will get the same answer.

Here's the paired samples version:

```
A <- c(10,8,11,9,10) #nD_nR
B <- c(5,4,3,4,2) #D_nR
C <- c(12,13,14,11,13) #nD_R
```

```
D <- c(9,8,10,11,12) #D_R
AC<- (A+C)/2
BD<- (B+D)/2
t.test(AC,BD, paired=TRUE,var.equal=TRUE)
```

[run](#)[restart](#)[restart & run all](#)

Paired t-test

```
data: AC and BD
t = 7.6615, df = 4, p-value = 0.00156
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.741724 5.858276
sample estimates:
mean of the differences
          4.3
```

Here's the one sample version:

```
A <- c(10,8,11,9,10) #nD_nR
B <- c(5,4,3,4,2) #D_nR
C <- c(12,13,14,11,13) #nD_R
D <- c(9,8,10,11,12) #D_R
AC<- (A+C)/2
BD<- (B+D)/2
t.test(AC-BD, mu=0)
```

[run](#)[restart](#)[restart & run all](#)

One Sample t-test

```
data: AC - BD
t = 7.6615, df = 4, p-value = 0.00156
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 2.741724 5.858276
sample estimates:
mean of x
          4.3
```

If we were to write-up our results for the main effect of distraction we could say something like this:

The main effect of distraction was significant, $t(4) = 7.66$, $p = 0.001$. The mean number of differences spotted was higher in the no-distraction condition ($M = 11.1$) than the distraction condition ($M = 6.8$).

Main effect of Reward

The main effect of reward compares the overall means for all scores in the no-reward and reward conditions, collapsing over the reward conditions.

The yellow columns show the no-reward scores for each subject. The blue columns show the reward scores for each subject.

The overall means for each subject, for the two reward conditions are shown to the right. For example, subject 1 had a 10 and 5 in the no-reward condition, so their mean is 7.5.

We are interested in the main effect of reward. This is the difference between the AB column (average of subject scores in the no-reward condition) and the CD column (average of the subject scores in the reward condition). These differences for each subject are shown in the last green column. The overall means, averaging over subjects are in the bottom green row.

All Conditions							
	No Reward		Reward		Reward Means		Reward Effect
	No Distraction	Distraction	No Distraction	Distraction	No Reward	Reward	Difference
subject	A	B	C	D	AB	CD	CD.minus.AB
1	10	5	12	9	7.5	10.5	3
2	8	4	13	8	6	10.5	4.5
3	11	3	14	10	7	12	5
4	9	4	11	11	6.5	11	4.5
5	10	2	13	12	6	12.5	6.5
Means					6.6	11.3	4.7

Just looking at the means, we can see there was a main effect of reward. The mean number of differences spotted was 11.3 in the reward condition, and 6.6 in the no-reward condition. So, the size of the main effect of reward was 4.7.

Is a difference of this size likely or unlikely due to chance? We could conduct a paired-samples t -test on the AB vs. CD means, or a one-sample t -test on the difference scores. They both give the same answer:

Here's the paired samples version:

```
A <- c(10,8,11,9,10) #nD_nR
B <- c(5,4,3,4,2) #D_nR
C <- c(12,13,14,11,13) #nD_R
D <- c(9,8,10,11,12) #D_R
AB<- (A+B)/2
CD<- (C+D)/2
t.test(CD,AB, paired=TRUE,var.equal=TRUE)
```

run

restart

restart & run all

Paired t-test

data: CD and AB

$t = 8.3742$, $df = 4$, $p\text{-value} = 0.001112$

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

3.141724 6.258276

sample estimates:

mean of the differences

4.7

Here's the one sample version:

```
A <- c(10,8,11,9,10) #nD_nR
B <- c(5,4,3,4,2) #D_nR
C <- c(12,13,14,11,13) #nD_R
D <- c(9,8,10,11,12) #D_R
AB<- (A+B)/2
CD<- (C+D)/2
t.test(CD-AB, mu=0)
```

run

restart

restart & run all

One Sample t-test

```
data: CD - AB
t = 8.3742, df = 4, p-value = 0.001112
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 3.141724 6.258276
sample estimates:
mean of x
 4.7
```

If we were to write-up our results for the main effect of reward we could say something like this:

The main effect of reward was significant, $t(4) = 8.37$, $p = 0.001$. The mean number of differences spotted was higher in the reward condition ($M = 11.3$) than the no-reward condition ($M = 6.6$).

Interaction between Distraction and Reward

Now we are ready to look at the interaction. Remember, the whole point of this fake study was what? Can you remember?

Here's a reminder. We wanted to know if giving rewards versus not would change the size of the distraction effect.

Notice, neither the main effect of distraction, or the main effect of reward, which we just went through the process of computing, answers this question.

In order to answer the question we need to do two things. First, compute distraction effect for each subject when they were in the no-reward condition. Second, compute the distraction effect for each subject when they were in the reward condition.

Then, we can compare the two distraction effects and see if they are different. The comparison between the two distraction effects is what we call the interaction effect. Remember, this is a difference between two difference scores. We first get the difference scores for the distraction effects in the no-reward and reward conditions. Then we find the difference scores between the two distraction effects. This difference of differences is the interaction effect (green column in the table)

All Conditions

subject	No Reward		Reward		Distraction Effects		Interaction Effect
	No Distraction	Distraction	No Distraction	Distraction	No Reward	Reward	Difference
	A	B	C	D	A-B	C-D	(A-B)-(C-D)
1	10	5	12	9	5	3	2
2	8	4	13	8	4	5	-1
3	11	3	14	10	8	4	4
4	9	4	11	11	5	0	5
5	10	2	13	12	8	1	7
Means					6	2.6	3.4

The mean distraction effects in the no-reward (6) and reward (2.6) conditions were different. This difference is the interaction effect. The size of the interaction effect was 3.4.

How can we test whether the interaction effect was likely or unlikely due to chance? We could run another paired-sample t -test between the two distraction effect measures for each subject, or a one sample t -test on the green column (representing the difference between the differences). Both of these t -tests will give the same results:

Here's the paired samples version:

```
A <- c(10,8,11,9,10) #nD_nR
B <- c(5,4,3,4,2) #D_nR
C <- c(12,13,14,11,13) #nD_R
D <- c(9,8,10,11,12) #D_R
A_B <- A-B
C_D <- C-D
t.test(A_B,C_D, paired=TRUE,var.equal=TRUE)
```

run restart restart & run all

Paired t-test

```
data: A_B and C_D
t = 2.493, df = 4, p-value = 0.06727
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.3865663 7.1865663
sample estimates:
mean of the differences
3.4
```

Here's the one sample version:

```
A <- c(10,8,11,9,10) #nD_nR
B <- c(5,4,3,4,2) #D_nR
C <- c(12,13,14,11,13) #nD_R
D <- c(9,8,10,11,12) #D_R
```

```
A_B <- A-B
C_D <- C-D
t.test(A_B-C_D, mu=0)
```

run restart restart & run all

One Sample t-test

```
data:  A_B - C_D
t = 2.493, df = 4, p-value = 0.06727
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.3865663  7.1865663
sample estimates:
mean of x
      3.4
```

Oh look, the interaction was not significant. At least, if we had set our alpha criterion to 0.05, it would not have met that criteria. We could write up the results like this. The two-way interaction between between distraction and reward was not significant, $t(4) = 2.493$, $p = 0.067$.

Often times when a result is “not significant” according to the alpha criteria, the pattern among the means is not described further. One reason for this practice is that the researcher is treating the means as if they are not different (because there was an above alpha probability that the observed idfferences were due to chance). If they are not different, then there is no pattern to report.

There are differences in opinion among reasonable and expert statisticians on what should or should not be reported. Let’s say we wanted to report the observed mean differences, we would write something like this:

The two-way interaction between between distraction and reward was not significant, $t(4) = 2.493$, $p = 0.067$. The mean distraction effect in the no-reward condition was 6 and the mean distraction effect in the reward condition was 2.6.

Writing it all up

We have completed an analysis of a 2x2 repeated measures design using paired-samples t -tests. Here is what a full write-up of the results could look like.

The main effect of distraction was significant, $t(4) = 7.66$, $p = 0.001$. The mean number of differences spotted was higher in the no-distraction condition ($M = 11.1$) than the distraction condition ($M = 6.8$).

The main effect of reward was significant, $t(4) = 8.37$, $p = 0.001$. The mean number of differences spotted was higher in the reward condition ($M = 11.3$) than the no-reward condition ($M = 6.6$).

The two-way interaction between between distraction and reward was not significant, $t(4) = 2.493$, $p = 0.067$. The mean distraction effect in the no-reward condition was 6 and the mean distraction effect in the reward condition was 2.6.

Interim Summary. We went through this exercise to show you how to break up the data into individual comparisons of interest. Generally speaking, a 2x2 repeated measures design would not be anlyazed with three paired-samples t -test. This is because it is more convenient to use the repeated measures ANOVA for this task. We will do this in a moment to show you that they give the same results. And, by the same results, what we will show is that the p -values for each main effect, and the interaction, are the same. The ANOVA will give us F -values rather than t values. It turns out that in this situation, the F -values are related to the t values. In fact, $t^2 = F$.

2x2 Repeated Measures ANOVA

We just showed how a 2x2 repeated measures design can be analyzed using paired-sampled t -tests. We broke up the analysis into three parts. The main effect for distraction, the main effect for reward, and the 2-way interaction between distraction and reward. We claimed the results of the paired-samples t -test analysis would mirror what we would find if we conducted the analysis using an ANOVA. Let’s show that the results are the same. Here are the results from the 2x2 repeated-measures ANOVA, using the `aov` function in R.

```
library(xtable)
```

```
A <- c(10,8,11,9,10) #nD_nR
B <- c(5,4,3,4,2) #D_nR
C <- c(12,13,14,11,13) #nD_R
D <- c(9,8,10,11,12) #D_R
Number_spotted <- c(A, B, C, D)
Distraction <- rep(rep(c("No Distraction", "Distraction"), each=5),2)
Reward <- rep(c("No Reward", "Reward"),each=10)
Subjects <- rep(1:5,4)
Distraction <- as.factor(Distraction)
Reward <- as.factor(Reward)
Subjects <- as.factor(Subjects)
rm_df <- data.frame(Subjects, Distraction, Reward, Number_spotted)
aov_summary <- summary(aov(Number_spotted~Distraction*Reward +
                          Error(Subjects/(Distraction*Reward))),
                      rm_df)
knitr::kable(xtable(aov_summary))
```

run restart restart & run all

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Distraction	1	92.45	92.450	58.698413	F)" style="vertical-align:middle;">0.0015600
Distraction:Reward	1	14.45	14.450	6.215054	F)" style="vertical-align:middle;">0.0672681
Residuals	4	3.70	0.925	NA	F)" style="vertical-align:middle;">NA
Residuals	4	6.30	1.575	NA	F)" style="vertical-align:middle;">NA
Residuals	4	9.30	2.325	NA	F)" style="vertical-align:middle;">NA
Residuals1	4	6.30	1.575	NA	F)" style="vertical-align:middle;">NA
Reward	1	110.45	110.450	70.126984	F)" style="vertical-align:middle;">0.0011122

Let's compare these results with the paired-samples t -tests.

Main effect of Distraction: Using the paired samples t -test, we found $t(4) = 7.6615$, $p = 0.00156$. Using the ANOVA we found, $F(1,4) = 58.69$, $p = 0.00156$. See, the p -values are the same, and $t^2 = 7.6615^2 = 58.69 = F$.

Main effect of Reward: Using the paired samples t -test, we found $t(4) = 8.3742$, $p = 0.001112$. Using the ANOVA we found, $F(1,4) = 70.126$, $p = 0.001112$. See, the p -values are the same, and $t^2 = 8.3742^2 = 70.12 = F$.

Interaction effect: Using the paired samples t -test, we found $t(4) = 2.493$, $p = 0.06727$. Using the ANOVA we found, $F(1,4) = 6.215$, $p = 0.06727$. See, the p -values are the same, and $t^2 = 2.493^2 = 6.215 = F$.

There you have it. The results from a 2x2 repeated measures ANOVA are the same as you would get if you used paired-samples t -tests for the main effects and interactions.

This page titled 9.5: Simple analysis of 2x2 repeated measures design is shared under a CC BY-SA 4.0 license and was authored, remixed, and/or curated by Matthew J. C. Crump via source content that was edited to the style and standards of the LibreTexts platform.