

12.3: Practice with RM ANOVA Summary Table

Let's use a real scenario to practice with the Repeated Measures ANOVA Summary Table.

Scenario

The data are taken from a recent study conducted by Behmer and Crump (a co-author of this chapter), at Brooklyn College (Behmer & Crump, 2017).

Behmer and Crump (2017) were interested in how people perform sequences of actions. One question is whether people learn individual parts of actions, or the whole larger pattern of a sequence of actions. We looked at these issues in a computer keyboard typing task. One of our questions was whether we would replicate some well known findings about how people type words and letters.

From prior work we knew that people type words way faster than random letters, but if you made the random letters a little bit more English-like, then people who can read English type those letter strings a little bit faster, but not as slow as random string.

In the study, 38 participants sat in front of a computer and typed five-letter strings one at a time. Sometimes the five letters made a word (Normal condition: TRUCK), sometimes they were completely random (Random condition: JWYFG), and sometimes they followed patterns like you find in English but were not actual words (Bigram condition: QUEND). What makes this repeated measures is that each participant received each condition; some trials was a word, some trials were a random string of letters, and some trials were a string of letters that looked like a word in English but was not a word. The order for each trial were randomly assigned. We measured every single keystroke that participants made, and we'll look at the reaction times (how long it took for participants to start typing the first letter in the string, in milliseconds).

✓ Example 12.3.1

Answer the following questions to understand the variables and groups that we are working with.

1. Who is the sample?
2. Who might be the population?
3. What is the IV (groups being compared)?
4. What is the DV (quantitative variable being measured)?

Solution

1. The sample is 38 participants.
2. Maybe anyone who types on a keyboard? English-speaker typists? There's not much info in the scenario to determine a specific population.
3. The IV is something like "word status" with the three levels being Normal (English word), Random (letter string), and Bigram (English-like letter string).
4. Reaction time (how long it took for participants to start typing the first letter) in milliseconds.

Step 1: State the Hypotheses

Based on the means from Table 12.3.1, we can see the means look different. What could be a directional research hypothesis?

Table 12.3.1- Descriptive Statistics for Reaction Time by Word Conditions

	N:	Mean:	SD:
Normal (English word)	38	779.00	20.40
Bigram (English-like non-word)	38	869.00	24.60
Random (non-word)	38	1037.00	29.30

? Exercise 12.3.1

Determine the research hypothesis in words and symbols. You can fill in the following underlined spot with the symbols for greater than (>), less than (<), or equal signs. Just remember, at least one pair of means must be predicted to be different from each other.

Symbols:

- \bar{X}_N _____ \bar{X}_B
- \bar{X}_N _____ \bar{X}_R
- \bar{X}_B _____ \bar{X}_R

Answer

Based on the means, I might predict that the Normal condition will react fastest and will be significantly faster than the Bigram condition and the Random condition. I also might hypothesis that the Bigram condition would have a significantly shorter reaction time as the Random condition, as well.

Symbols:

- $\bar{X}_N < \bar{X}_B$
- $\bar{X}_N < \bar{X}_R$
- $\bar{X}_B < \bar{X}_R$

Notice that we are predicting that the Normal words will have a smaller reaction time, meaning that they will respond faster and the time to respond will be shorter.

What about the null hypothesis? What might that look like?

? Exercise 12.3.2

State the null hypothesis in words and symbols. .

Answer

The reaction time will be similar for the Normal condition, the Bigram condition, and the Random condition.

$$\bar{X}_N = \bar{X}_B = \bar{X}_R$$

Step 2: Find the Critical Values

Using the sample size information included in Table 1, you can now find the critical values from the Critical Values of F Table found on this page in the chapter that first discussed ANOVAs, or find a list of critical value tables at the end of this textbook on the Common Critical Value Tables page.

As shown on the bottom of the critical values page, the two Degrees of Freedom that you'll use is still from the numerator (Between Groups) and the denominator (Within Group or Error), but the denominator's df is calculated slightly differently.

✓ Example 12.3.2

What is the critical value for this scenario?

Solution

The df for the numerator is still $k-1$; $3-1 = 2$.

The df for the denominator is $(k-1)*(P-1)$, which means that we need to figure out $P-1$ first. Since P stands for the number of participants, that would be $38-1 = 37$.

$$(k-1) * (P-1) = (3-1) * (38-1) = 2 * 37 = 74$$

The critical value of F for 2 and 74 in the 0.05 row is 3.15.

Step 3: Compute the Test Statistic

Using the Sum of Squares provide in the following ANOVA Summary Table (Table 12.3.2) and the information from the scenario about the sample size and number of conditions, fill in the ANOVA Summary Table to determine the calculated F-value.

Table 12.3.2- RM ANOVA Summary Table with Some Sums of Squares

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Between	1,424,914.00			
Participants	2,452,611.90			
Error				
Total	4,101,175.30			

✓ Example 12.3.3

Complete the ANOVA Summary Table in Table 12.3.2 to determine the calculated F-score.

Solution

Table 12.3.3- RM ANOVA Summary Table with Formulas

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Between	1,424,914.00	$k - 1 = 3 - 1 = 2$	$\frac{SS_B}{df_B} = \frac{1424914}{2} = 712457$	$\frac{MS_B}{MS_E} = \frac{712457}{3022.29} = 235.73$
Participants	2,452,611.90	$P - 1 = 38 - 1 = 37$	leave blank	leave blank
Error	$SS_{WG} = SS_{Total} - SS_B - SS_P = 4101175.30 - 1424914.00 - 2452611.90 = 223649.40$	$df_E = (P - 1)(k - 1) = 37(2) = 74$	$\frac{SS_E}{df_E} = \frac{223649.40}{74} = 3022.29$	leave blank
Total	4,101,175.30	$N - 1 = 113$ ($N = k \times P$)	leave blank	leave blank

So the ANOVA Summary Table should end up looking like Table 12.3.4

Table 12.3.4- Completed RM ANOVA Summary Table

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Between	1,424,914.00	2	712,457.00	235.73
Participants	2,452,611.90	37		
Error	223,649.40	74	3,022.29	
Total	4,101,175.30	113		

Step 4: Make the Decision

We have the critical value (3.15) and the calculated value (235.73), so we can now make the decision just like we've been doing.

Table 12.3.5- Rejecting or Retaining the Null Hypothesis

REJECT THE NULL HYPOTHESIS		RETAIN THE NULL HYPOTHESIS	
Small p-values ($p < .05$)		Large p-values ($p > .05$)	

REJECT THE NULL HYPOTHESIS	RETAIN THE NULL HYPOTHESIS
A small p-value means a small probability that all of the means are similar. <i>Suggesting that at least one of the means is different from at least one other mean...</i>	A large p-value means a large probability that all of the means are similar.
We conclude that: At least one mean is different from one other mean. At least one group is not from the same population as the other groups.	We conclude that: The means for all of the groups are similar. All of the groups are from the same population.
The calculated F is further from zero (more extreme) than the critical F. In other words, the calculated F is bigger than the critical F. (Draw the standard normal curve and mark the calculated F and the critical F to help visualize this.)	The calculated F is closer to zero (less extreme) than the critical F. In other words, the calculated F is smaller than the critical F. (Draw the standard normal curve and mark the calculated F and the critical F to help visualize this.)
Reject the null hypothesis (which says that all of the means are similar).	Retain (or fail to reject) the null hypothesis (which says that the all of the means are similar).
Support the Research Hypothesis? MAYBE. Look at the actual means: <ul style="list-style-type: none"> Support the Research Hypothesis if the means are in the directions that were hypothesized. <ul style="list-style-type: none"> The mean of the group that you said would be bigger, really is bigger; The mean of the group that you said would be smaller really is smaller; The means of the groups that you said would be similar are actually similar. Partial support of the Research Hypothesis if some of the means are in the directions that were hypothesized, but some aren't. Do not support the Research Hypothesis if none of the means are in the direction that were hypothesized. 	Do not support the Research Hypothesis (because all of the means are similar).
Statistical sentence: $F(df) = F\text{-calc}$, $p < .05$ (fill in the df and the calculated F)	Statistical sentence: $F(df) = F\text{-calc}$, $p > .05$ (fill in the df and the calculated F)

Here's another way to show the info in Table 12.3.5

(Critical < Calculated) = Reject null = At least one mean is different from at least one other mean. = $p < .05$

(Critical > Calculated) = Retain null = All of the means are similar. = $p > .05$

? Exercise 12.3.3

Should we retain or reject the null hypothesis? Does this mean that we're saying that all of the means are similar, or that at least one mean different?

Answer

Because the calculated F-score of 235.73 is so much bigger than the critical value of 3.15, we reject the null hypothesis and say that at least one mean is different from at least one other mean.

Before we can write-up the results for this analysis, we need to determine if the mean differences are in the hypothesized direction. Behmer and Crump (2017) provided the following t-test results that tested whether pairs of means are different:

- Normal versus Bigram: $t(37) = -10.61$, $p < 0.001$
- Normal versus Random: $t(37) = -15.78$, $p < 0.001$
- Bigram versus Random: $t(37) = 13.49$, $p < 0.001$

? Exercise 12.3.4

What is one problem with using t-tests to check for multiple sets of mean differences in post-hoc analyses?

Answer

Alpha inflation; each t-test has a 5% of committing a Type I Error (rejecting the null hypothesis when there really is no difference between the sample's means in the population).

? Exercise 12.3.5

What did we learn about in the Between Groups ANOVA chapter to use instead to check for multiple sets of mean differences in post-hoc analyses?

Answer

A variety of post-hoc analyses that reduced the chance of a Type I Error in each pairwise comparison so that the total chance of a Type I Error was still 5%.

Write-up

Okay, we now have all we need to complete a conclusion that reports the results while including all of the [required components](#):

- The statistical test is preceded by the descriptive statistics (means).
- The description tells you what the research hypothesis being tested is.
- A "statistical sentence" showing the results is included.
- The results are interpreted in relation to the research hypothesis.

? Exercise 12.3.6

What could the conclusion look like for this scenario?

Answer

The research hypothesis was that the Normal condition would have the fastest reaction time ($M = 779ms$) compared to the Bigram ($M = 869ms$) and the Random ($M = 1,037ms$) conditions, and the Bigram condition would also have a faster reaction time than the Random condition. This research hypothesis was fully supported ($F(2,37)=235.73$, $p<0.05$).

That's it! Let's try one more example; this time, we'll calculate the Sum of Squares and the pairwise comparison.

This page titled [12.3: Practice with RM ANOVA Summary Table](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Michelle Oja](#).

- **8.5: Real Data** by [Matthew J. C. Crump](#) is licensed [CC BY-SA 4.0](#). Original source: <https://www.crumplab.com/statistics/>.