

8.5: Real Data

Let's look at some real data from a published experiment that uses a repeated measures design. This is the same example that you will be using in the lab for repeated measures ANOVA. The data happen to be taken from a recent study conducted by Lawrence Behmer and myself, at Brooklyn College (Behmer and Crump 2017).

We 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 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 5 letter strings one at a time. Sometimes the 5 letter made a word (Normal condition, TRUCK), sometimes they were completely random (Random Condition, JWYFG), and sometimes they followed patterns like you find in English (Bigram Condition, QUEND), but were not actual words. So, the independent variable for the typing material had three levels. We measured every single keystroke that participants made. This gave us a few different dependent measures. Let's take a look at the reaction times. This is how long it took for participants to start typing the first letter in the string.

```
library(data.table)
library(ggplot2)
suppressPackageStartupMessages(library(dplyr))
expl_data <- fread(
  'https://raw.githubusercontent.com/CrumpLab/statistics/master/data/expl_BehmerCrumpAPP.csv')
expl_data$Block<-as.factor(expl_data$Block)
levels(expl_data$Block) <- c("Visible keyboard", "Covered Keyboard")
## get subject mean RTs
subject_means <- expl_data %>%
  filter(Order==1, Correct==1, PureRTs<5000) %>%
  dplyr::group_by(Subject, Block, Stimulus) %>%
  dplyr::summarise(mean_rt = mean(PureRTs), .groups='drop_last')
subject_means$Subject<-as.factor(subject_means$Subject)
subject_means$Block<-as.factor(subject_means$Block)
subject_means$Stimulus<-as.factor(subject_means$Stimulus)
## get condition mean RTs
plot_means <- subject_means %>%
  arrange(match(Stimulus,c("Normal", "Bigrams", "Random")) %>%
  dplyr::group_by(Block, Stimulus) %>%
  dplyr::summarise(means = mean(mean_rt),
    SEs = sd(mean_rt)/sqrt(length(mean_rt)), .groups='drop_last')
## plot the condition means
# re-order stimulus factor for plotting
plot_means$Stimulus <- factor(plot_means$Stimulus,
  levels = c("Normal", "Bigrams", "Random"))
ggplot(plot_means, aes(x=Stimulus, y=means, group=Block, color=Block))+
  geom_point()+
  geom_line()+
  geom_errorbar(aes(ymin=means-SEs, ymax=means+SEs), width=.2)+
  theme_classic()+
  ylab("Mean Reaction Time (ms)")+
  xlab("Typing Material")
```

run

restart

restart & run all

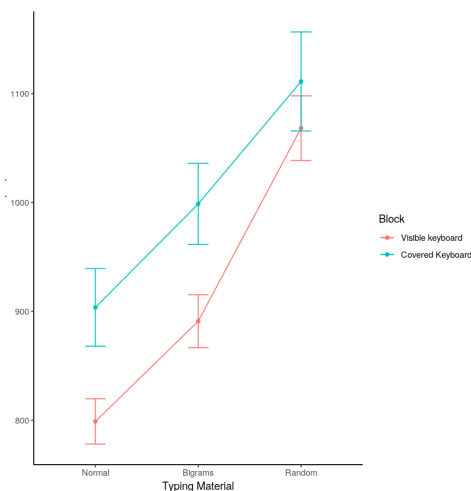


Figure \(\PageIndex{1}\): Results from Behmer & Crump (2017).

OK, I made a figure showing the mean reaction times for the different typing material conditions. You will notice that there are two sets of lines. That's because there was another manipulation I didn't tell you about. In one block of trials participants got to look at the

keyboard while they typed, but in the other condition we covered up the keyboard so people had to type without looking. Finally, the error bars are standard error of the means.

Note

Note, the use of error bars for repeated-measures designs is not very straightforward. In fact the standard error of the means that we have added here are not very meaningful for judging whether the differences between the means are likely not due to chance. They would be if this was a between-subjects design. We will update this textbook with a longer discussion of this issue, for now we will just live with these error bars.

For the purpose of this example, we will say, it sure looks like the previous finding replicated. For example, people started typing Normal words faster than Bigram strings (English-like), and they started typing random letters the most slowly of all. Just like prior research had found.

Let's focus only on the block of trials where participants were allowed to look at the keyboard while they typed, that's the red line, for the "visible keyboard" block. We can see the means look different. Let's next ask, what is the likelihood that chance (random sampling error) could have produced these mean differences. To do that we run a repeated-measures ANOVA in R. Here is the ANOVA table.

```
library(data.table)
library(ggplot2)
library(xtable)
suppressPackageStartupMessages(library(dplyr))
expl_data <- fread(
  "https://raw.githubusercontent.com/Crumplab/statistics/master/data/expl_BehmerCrumplabAPP.csv"
)
expl_data$Block <- as.factor(expl_data$Block)
levels(expl_data$Block) <- c("Visible keyboard", "Covered Keyboard")
## get subject mean RTs
subject_means <- expl_data %>%
  filter(Order==1, Correct==1, PureRTs<5000) %>%
  dplyr::group_by(Subject, Block, Stimulus) %>%
  dplyr::summarise(mean_rt = mean(PureRTs), .groups='drop_last')
subject_means$Subject <- as.factor(subject_means$Subject)
subject_means$Block <- as.factor(subject_means$Block)
subject_means$Stimulus <- as.factor(subject_means$Stimulus)

visible_means <- subject_means %>%
  filter(Block=="Visible keyboard")
s_out <- summary(aov(mean_rt~Stimulus + Error (Subject/Stimulus),
  visible_means))
knitr::kable(xtable(s_out))
```

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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	37	2452611.9	66286.808	NA	F)" style="vertical-align:middle;" class="lt-stats-7937">NA
Stimulus	2	1424914.0	712457.010	235.7342	F)" style="vertical-align:middle;" class="lt-stats-7937">0

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals1	74	223649.4	3022.289	NA	F)" style="vertical-align:middle;" class="lt-stats-7937">NA

Alright, we might report the results like this. There was a significant main effect of Stimulus type, $F(2, 74) = 235.73$, $MSE = 3022.289$, $p < 0.001$.

Notice a couple things. First, this is a huge F -value. It's 253! Notice also that the p -value is listed as 0. That doesn't mean there is zero chance of getting an F -value this big under the null. This is a rounding error. The true p -value is 0.000000000000000... The zeros keep going for a while. This means there is only a vanishingly small probability that these differences could have been produced by sampling error. So, we reject the idea that the differences between our means could be explained by chance. Instead, we are pretty confident, based on this evidence and previous work showing the same thing, that our experimental manipulation caused the difference. In other words, people really do type normal words faster than random letters, and they type English-like strings somewhere in the middle in terms of speed.

This page titled 8.5: Real Data 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.