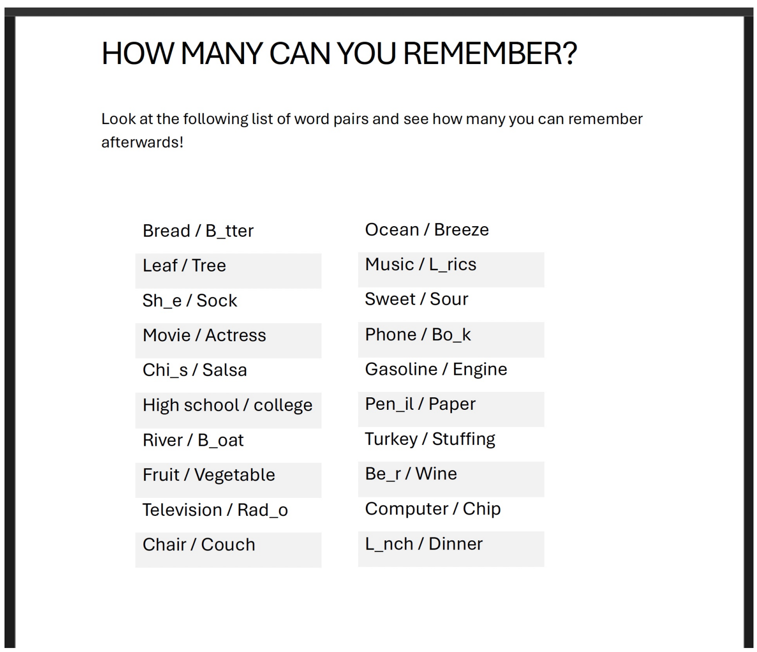
STAT332 FINAL PROJECT

Peter, Ogorogile, Wenhao

2025-04-08



HOW MANY CAN YOU REMEMBER?

Look at the following list of word pairs and see how many you can remember afterwards!

Bread / B_tter	Ocean / Breeze
Leaf / Tree	Music / L_rics
Sh_e / Sock	Sweet / Sour
Movie / Actress	Phone / Bo_k
Chi_s / Salsa	Gasoline / Engine
High school / college	Pen_il / Paper
River / B_oat	Turkey / Stuffing
Fruit / Vegetable	Be_r / Wine
Television / Rad_o	Computer / Chip
Chair / Couch	L_nch / Dinner

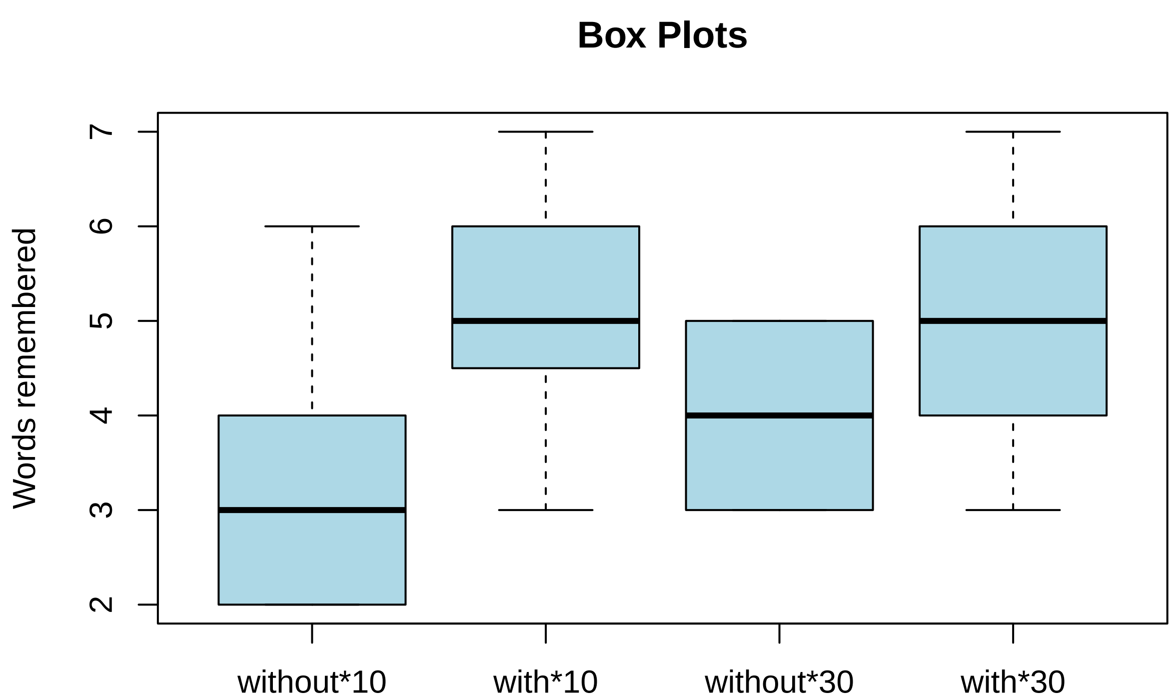
Figure 1: survey sheet

- **Aim**: of the experiment is to determine whether making the brain do some work will improve information retention/ remembering and thus learning
- **How?:** we presented a list of word pairs to individuals. Some of the word pairs had some letters omitted from them. These 2 different word pairs were evenly spread across the document. We gave each individual 40 seconds to study the words, and then waited for either 10 seconds or 30 seconds before we asked them to write down all the words they remembered. Then we noted down the number of words pairs with holes that they remembered and word pairs without holes that they remembered
- we treated this as an experiment with 2 factors:
  - type of word pairs: this factor had 2 levels, which were word pairs with holes and without holes
  - waiting time after viewing the list: this factor had 2 levels, which were 10 seconds and 30 seconds
- we are interested in testing the hypothesis that making the brain work through providing words with holes will result in more words remembered than if you provide words without holes

Exploratory analysis:

```
data <-read.csv("STAT 332 Project Data.csv", header = TRUE)
data <- read.table("STAT 332 Project Data.csv", header = TRUE, sep = "", stringsAsFactors = FALSE)
group10 <- data[data$Wait_Time ==10,]
group30 <-data[data$Wait_Time==30,]
wo10 <- group10$Words_wo_Holes
w10 <- group10$Words_Holes
wo30 <-group30$Words_wo_Holes
w30 <- group30$Words_Holes


boxplot(
  list(wo10, w10, wo30, w30),
  main = "Box Plots",
  ylab = "Words remembered",
  col = "lightblue",
  names = c("without*10", "with*10", "without*30", "with*30")
)
```



- The above boxplots seem to indicate that the amount of words with holes remembered is in general greater than the amount of words without holes remembered, but the difference is more significant after 10s seconds than at 30s. We will use contrasts later to find the significance levels of these differences.

Modelling

- we treat each individual as a block, presenting both types of word pairs to one individual in the hope of reducing variance. We present the lists at the same time (in the same document), because we found that if we presented one after the other, the list presented first interferes with remembering words in the second list
- each treatment has 15 replications, so this is a balanced experiment
- the model is:
  - The additive model with treatments and blocks is given by:

$$\text{words\_remembered}_{ij} = \mu + \tau_i + b_j + \epsilon_{ij}$$

where:

- μ is the grand mean
- τ_i is the effect of the i^{th} treatment $(i = 1, \ldots, 4)$ with $\sum_{i=1}^{4} \tau_i = 0$
- b_j is the effect of the j^{th} block $(j = 1, \ldots, 30)$ with $\sum_{j=1}^{30} b_j = 0$
- $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ is the random error term
- The experiment implements a 2×2 single factorial design with treatments defined as:

$$\text{Treatment} = \begin{cases} 1, & \text{if } \texttt{list == "w/o holes"} \text{ and } \texttt{wait\_time == 10} \\ 2, & \text{if } \texttt{list == "w/o holes"} \text{ and } \texttt{wait\_time == 30} \\ 3, & \text{if } \texttt{list == "with holes"} \text{ and } \texttt{wait\_time == 10} \\ 4, & \text{if } \texttt{list == "with holes"} \text{ and } \texttt{wait\_time == 30} \end{cases}$$

- though there are four treatments, we are effectively randomizing over 2 classes of treatments
- we start of with ANOVA

```r
##
data2 = cbind(as.numeric(data$Words_wo_Holes),data$Participant_ID,data$Wait_Time )
t1 = rep("wo", nrow(data2))
data2 = cbind(data2, t1)
colnames(data2) = c("words_remembered", "block", "wait_time", "list")
data3 = cbind(as.numeric(data$Words_Holes), data$Participant_ID, data$Wait_Time)
t2 = rep("w", nrow(data3))
data3 = cbind(data3, t2)
colnames(data3) = c("words_remembered", "block", "wait_time", "list")

data_all = data.frame(rbind(data2, data3))
anva = aov(words_remembered~list+ wait_time+block, data = data_all)


# Creating the treatment column based on conditions
data_all <- data_all %>%
  mutate(treatment = case_when(
    list == "wo" & wait_time == 10 ~ 1,
    list == "wo" & wait_time == 30 ~ 2,
```

```r
    list == "w" & wait_time == 10 ~ 3,
    list == "w" & wait_time == 30 ~ 4,
    TRUE ~ NA_real_  # Default case if no condition is met (optional)
  ))
data_all$treatment = as.factor(data_all$treatment) # change treatment to factor variable
data_all$block = as.factor(data_all$block)
data_all$wait_time = as.factor(data_all$wait_time)
# View the updated dataframe
anva3 = aov(words_remembered~treatment+block, data = data_all)
model = lm(words_remembered~treatment+block, data = data_all)
#summary(model)

anova_table <- anova(model)

# Extract p-values and terms
p_table <- data.frame(
  Term = rownames(anova_table),
  `P-value` = signif(anova_table[["Pr(>F)"]], 4)
)

# View the table
#print(p_table)
#mean(as.numeric(data_all$words_remembered))
summary(anva3)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## treatment    3  34.73  11.578   5.966 0.00281 **
## block       28  13.67   0.488   0.252 0.99976
## Residuals   28  54.33   1.940
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- from ANOVA we see that the pvalue is 0.00281 associated with treatment, which means that there is strong evidence that the treatments are significantly different from each other. This is a green light for us to go ahead with analysis
- We also see that there is **not** enough evidence that the blocks, which are the individuals participating in the experiment, are significantly different from each other.
- analysis for the degrees shows that blocking used up 28 degrees of freedom, contrary to the expected value of 29. The residuals also have 28 degrees of freedom, contrary to 30 that we expected. We will address this under potential issues.

Contrasts

```r
library(emmeans)
```

```r
emm<-emmeans(anva3, ~ treatment)
pairs(emm, adjust = "tukey") [c(2,5)]
```

```
## Note: adjust = "tukey" was changed to "sidak"
## because "tukey" is only appropriate for one set of pairwise comparisons
```

```
##  contrast                 estimate    SE df t.ratio p.value
##  treatment1 - treatment3   -1.933 0.509 28  -3.801  0.0014
##  treatment2 - treatment4   -0.867 0.509 28  -1.704  0.1891
##
## Results are averaged over the levels of: block
## P value adjustment: sidak method for 2 tests
```

- treatment1 - treatment3 is the difference in words remembered between the 2 lists after 10 seconds. The above shows this difference is very significant, in line with what we observed from the box plots. The same difference is however not very significant after 30 seconds as indicated by the p-value.

Potential issues

- we lost 2 degrees of freedom but we can't track where we lost them. A guess is that we lost them because the treatments where grouped by time, (30 seconds and 10 seconds), so though we have 4 treatments, we are essentially randomizing over two groups of treatments to individuals.

Conclusion

- from this experiment, we see that there is evidence that making the brain work will increase ones chances of remembering and strengthen memory.
- Making the brain work is the basis of active learning, so the results of this experiment show why active learning is superior to passive learning