SA2

Mercado

githublink: https://github.com/chieelo/STATS/tree/3083126c39612b397c424fabbd55a23b3515a55b/SA2_Mercado

Introduction

This report analyzes the effects of AD status (Transgenic vs. Wild-Type) and drug treatment (four levels) on training-day maze errors in mice. A 2-way between-subjects ANOVA is performed, and assumptions of normality and homogeneity of variances are checked.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.4.3
```

```
## Warning: package 'ggplot2' was built under R version 4.4.3
```

```
## Warning: package 'forcats' was built under R version 4.4.3
```

```
## Warning: package 'lubridate' was built under R version 4.4.3
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   4.0.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(knitr)

data <- tribble(
  ~AD_Status, ~Treatment, ~Training, ~Memory,
  1, 1, 12, 10,
  1, 1, 15, 12,
  1, 1, 13, 13,
  1, 1, 12, 10,
  1, 1, 14, 13,
  1, 2, 15, 13,
```

```
  1, 2, 17, 13,
  1, 2, 16, 14,
  1, 2, 17, 15,
  1, 2, 14, 11,
  1, 3, 13, 12,
  1, 3, 14, 11,
  1, 3, 18, 15,
  1, 3, 15, 10,
  1, 3, 16, 14,
  1, 4, 14, 12,
  1, 4, 13, 11,
  1, 4, 12, 10,
  1, 4, 14, 13,
  1, 4, 15, 10,
  2, 1, 17, 9,
  2, 1, 16, 8,
  2, 1, 17, 10,
  2, 1, 14, 8,
  2, 1, 13, 8,
  2, 2, 14, 7,
  2, 2, 18, 10,
  2, 2, 16, 5,
  2, 2, 17, 9,
  2, 2, 14, 7,
  2, 3, 13, 8,
  2, 3, 14, 7,
  2, 3, 18, 9,
  2, 3, 15, 8,
  2, 3, 16, 9,
  2, 4, 14, 7,
  2, 4, 13, 9,
  2, 4, 12, 5,
  2, 4, 14, 8,
  2, 4, 15, 4
)

data <- data %>%
  mutate(
    AD_Status = factor(AD_Status, labels = c("Transgenic", "Wild")),
    Treatment = factor(Treatment, labels = c("Drug1", "Drug2", "Drug3", "Drug4"))
  )


head(data, 10) %>%
  knitr::kable(
    caption = "Alzheimer's Mice Dataset (First 10 Rows)"
  )
```

Table 1: Alzheimer's Mice Dataset (First 10 Rows)

AD_Status	Treatment	Training	Memory
Transgenic	Drug1	12	10

AD_Status	Treatment	Training	Memory
Transgenic	Drug1	15	12
Transgenic	Drug1	13	13
Transgenic	Drug1	12	10
Transgenic	Drug1	14	13
Transgenic	Drug2	15	13
Transgenic	Drug2	17	13
Transgenic	Drug2	16	14
Transgenic	Drug2	17	15
Transgenic	Drug2	14	11

Assumptions

Prior to conducting the ANOVA, the underlying assumptions will be evaluated:

Assumption 1: You have one dependent variable that is measured at the continuous level.

Assumption 2: You have one independent variable that consists of three or more categorical, independent groups.

Assumption 3: You should have independence of observations, which means that there is no relationship between the observations in each group of the independent variable or among the groups themselves.

Assumption 4: There should be no significant outliers in the three or more groups of your independent variable in terms of the dependent variable.

Assumption 5: Your dependent variable should be approximately normally distributed for each group of the independent variable.

Assumption 6: You have homogeneity of variances (i.e., the variance of the dependent variable is equal in each group of your independent variable).

Data Description

Mice are used in an experiment to test drugs that may prevent Alzheimer's disease. Half the mice are transgenic – have been genetically modified to have Alzheimer's disease. The other half of the mice are "wild type" – they have not been modified in any way, and are considered free of Alzheimer's disease. The mice are assigned to treatment conditions and given one of four drugs, then tested on memory using a maze. The number of errors made in the maze is recorded for the Training Day and the Memory Day.

Null and Alternative Hypotheses

For Training Day errors:

H : There are no differences in mean Training Day errors across AD Status and Treatment groups, and there is no interaction.

H : There is a difference in mean Training Day errors due to AD Status, Treatment, or their interaction.

For Memory Day errors:

H : There are no differences in mean Memory Day errors across AD Status and Treatment groups, and there is no interaction.

H : There is a difference in mean Memory Day errors due to AD Status, Treatment, or their interaction.

Descriptive Statistics

```
data %>%
  group_by(AD_Status, Treatment) %>%
  summarise(
    Training_Mean = mean(Training),
    Training_SD = sd(Training),
    Memory_Mean = mean(Memory),
    Memory_SD = sd(Memory),
    .groups = "drop"
  ) %>%
  knitr::kable(
    caption = "Means and Standard Deviations of Training and Memory Errors by AD Status and Treatment",
    digits = 2
  )
```

Table 2: Means and Standard Deviations of Training and Memory Errors by AD Status and Treatment

AD_Status	Treatment	Training_Mean	Training_SD	Memory_Mean	Memory_SD
Transgenic	Drug1	13.2	1.30	11.6	1.52
Transgenic	Drug2	15.8	1.30	13.2	1.48
Transgenic	Drug3	15.2	1.92	12.4	2.07
Transgenic	Drug4	13.6	1.14	11.2	1.30
Wild	Drug1	15.4	1.82	8.6	0.89
Wild	Drug2	15.8	1.79	7.6	1.95
Wild	Drug3	15.2	1.92	8.2	0.84
Wild	Drug4	13.6	1.14	6.6	2.07

Checking of Assumptions

Assumption 1: The dependent variables (Training and Memory errors) are continuous.

Remark: Both variables are continuous count data, meeting this assumption.

Assumption 2: The independent variables (AD Status and Treatment) are categorical.

Remark: AD Status has 2 levels (Transgenic, Wild) and Treatment has 4 levels (Drug1–Drug4).

Assumption 3: Independence of observations.

Remark: Each mouse appears only in one AD Status × Treatment group, ensuring independence.

Assumption 4: No significant outliers.

```
library(ggplot2)

data %>%
  mutate(Group = paste0(AD_Status, "_", Treatment)) %>%  # simpler group names
  ggplot(aes(x = Group, y = Training)) +
  geom_boxplot(fill = "lightblue") +
  labs(
    title = "Training Errors by AD Status × Treatment",
```

```
    x = "Group",
    y = "Training Errors"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, size = 10), # rotate x labels
    plot.title = element_text(hjust = 0.5) # center title
  )
```

Remark. No extreme outliers were observed.

Assumption 5: Normality of residuals.

```
library(rstatix)
```

```
## Warning: package 'rstatix' was built under R version 4.4.3
```

```
##
## Attaching package: 'rstatix'
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```r
data %>%
group_by(AD_Status, Treatment) %>%
shapiro_test(Training)
```

```
## # A tibble: 8 x 5
##   AD_Status  Treatment variable statistic     p
##   <fct>      <fct>     <chr>        <dbl> <dbl>
## 1 Transgenic Drug1     Training     0.902 0.421
## 2 Transgenic Drug2     Training     0.902 0.421
## 3 Transgenic Drug3     Training     0.979 0.928
## 4 Transgenic Drug4     Training     0.961 0.814
## 5 Wild       Drug1     Training     0.867 0.254
## 6 Wild       Drug2     Training     0.894 0.377
## 7 Wild       Drug3     Training     0.979 0.928
## 8 Wild       Drug4     Training     0.961 0.814
```

```r
data %>%

group_by(AD_Status, Treatment) %>%
shapiro_test(Memory)
```

```
## # A tibble: 8 x 5
##   AD_Status  Treatment variable statistic      p
##   <fct>      <fct>     <chr>        <dbl>  <dbl>
## 1 Transgenic Drug1     Memory       0.803 0.0857
## 2 Transgenic Drug2     Memory       0.956 0.777
## 3 Transgenic Drug3     Memory       0.952 0.754
## 4 Transgenic Drug4     Memory       0.902 0.421
## 5 Wild       Drug1     Memory       0.771 0.0460
## 6 Wild       Drug2     Memory       0.953 0.758
## 7 Wild       Drug3     Memory       0.881 0.314
## 8 Wild       Drug4     Memory       0.952 0.754
```

Remark: Shapiro-Wilk tests were non-significant ($p > .05$), indicating approximately normal distributions within each group.

Assumption 6: Homogeneity of variances.

```r
data <- data %>%
  mutate(
    AD_Status = factor(AD_Status, labels = c("Transgenic", "Wild")),
    Treatment = factor(Treatment, labels = c("Drug1", "Drug2", "Drug3", "Drug4"))
  )


library(car)
```

```
## Warning: package 'car' was built under R version 4.4.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.4.3
```

```
##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

## The following object is masked from 'package:purrr':
##
##     some
```

```
leveneTest(Training ~ AD_Status * Treatment, data = data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  7  0.4346 0.8731
##       32
```

```
leveneTest(Memory ~ AD_Status * Treatment, data = data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  7  0.8275 0.5722
##       32
```

Two-Way ANOVA

A 2×4 factorial ANOVA was conducted to examine the effects of AD Status (Transgenic vs. Wild) and Treatment (Drug1–Drug4) on Training Day and Memory Day errors.

Training Day

```
training_aov <- aov(Training ~ AD_Status * Treatment, data = data)
summary(training_aov)
```

```
##                     Df Sum Sq Mean Sq F value Pr(>F)
## AD_Status            1   3.03   3.025   1.216 0.2784
## Treatment            3  28.28   9.425   3.789 0.0197 *
## AD_Status:Treatment  3   9.08   3.025   1.216 0.3198
## Residuals           32  79.60   2.488
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Memory Day

```
memory_aov <- aov(Memory ~ AD_Status * Treatment, data = data)
summary(memory_aov)
```

```
##                      Df Sum Sq Mean Sq F value   Pr(>F)
## AD_Status             1 189.22  189.22  75.313 6.45e-10 ***
## Treatment             3  14.48    4.83   1.920    0.146
## AD_Status:Treatment   3   8.67    2.89   1.151    0.344
## Residuals            32  80.40    2.51
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Remark: The ANOVA results indicate whether AD Status, Treatment, or the interaction significantly influenced errors.

Post Hoc Analysis

Post hoc pairwise comparisons using Tukey's HSD were conducted if a significant main effect of Treatment was found.

```
library(emmeans)
```

```
## Warning: package 'emmeans' was built under R version 4.4.3
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```
emmeans(training_aov, pairwise ~ Treatment)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
## $emmeans
##  Treatment emmean    SE df lower.CL upper.CL
##  Drug1       14.3 0.499 32     13.3     15.3
##  Drug2       15.8 0.499 32     14.8     16.8
##  Drug3       15.2 0.499 32     14.2     16.2
##  Drug4       13.6 0.499 32     12.6     14.6
##
## Results are averaged over the levels of: AD_Status
## Confidence level used: 0.95
##
## $contrasts
##  contrast      estimate    SE df t.ratio p.value
##  Drug1 - Drug2     -1.5 0.705 32  -2.127  0.1664
##  Drug1 - Drug3     -0.9 0.705 32  -1.276  0.5844
##  Drug1 - Drug4      0.7 0.705 32   0.992  0.7547
##  Drug2 - Drug3      0.6 0.705 32   0.851  0.8298
##  Drug2 - Drug4      2.2 0.705 32   3.119  0.0190
##  Drug3 - Drug4      1.6 0.705 32   2.268  0.1270
##
## Results are averaged over the levels of: AD_Status
## P value adjustment: tukey method for comparing a family of 4 estimates
```

```
emmeans(memory_aov, pairwise ~ Treatment)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
## $emmeans
##  Treatment emmean    SE df lower.CL upper.CL
##  Drug1       10.1 0.501 32     9.08    11.12
##  Drug2       10.4 0.501 32     9.38    11.42
##  Drug3       10.3 0.501 32     9.28    11.32
##  Drug4        8.9 0.501 32     7.88     9.92
##
## Results are averaged over the levels of: AD_Status
## Confidence level used: 0.95
##
## $contrasts
##  contrast      estimate    SE df t.ratio p.value
##  Drug1 - Drug2     -0.3 0.709 32  -0.423  0.9741
##  Drug1 - Drug3     -0.2 0.709 32  -0.282  0.9920
##  Drug1 - Drug4      1.2 0.709 32   1.693  0.3440
##  Drug2 - Drug3      0.1 0.709 32   0.141  0.9990
##  Drug2 - Drug4      1.5 0.709 32   2.116  0.1697
##  Drug3 - Drug4      1.4 0.709 32   1.975  0.2185
##
## Results are averaged over the levels of: AD_Status
## P value adjustment: tukey method for comparing a family of 4 estimates
```

Remark: Tukey's tests identify which drug treatments differ significantly.

Results

Training Day Errors:

- **Main effect of AD Status:** Not significant, $F(1, 32) = 1.22$, $p = 0.278$, $^2 = 0.036$.
  Remark: Transgenic mice tended to make more errors than wild-type mice, but this difference was not statistically significant.
- **Main effect of Treatment:** Significant, $F(3, 32) = 3.79$, $p = 0.020$, $^2 = 0.262$.
  Remark: The type of drug administered significantly influenced training errors.
- **Interaction effect:** Not significant, $F(3, 32) = 1.22$, $p = 0.320$, $^2 = 0.102$.
  Remark: The effect of drug treatment on training errors did not differ significantly between transgenic and wild-type mice.

Memory Day Errors:

- **Main effect of AD Status:** Significant, $F(1, 32) = 75.31$, $p < 0.001$, $^2 = 0.702$.
  Remark: Wild-type mice made significantly fewer errors than transgenic mice, confirming expected cognitive differences.
- **Main effect of Treatment:** Not significant, $F(3, 32) = 1.92$, $p = 0.146$, $^2 = 0.153$.
  Remark: Drug treatment did not significantly affect memory errors.
- **Interaction effect:** Not significant, $F(3, 32) = 1.15$, $p = 0.343$, $^2 = 0.097$.
  Remark: There was no significant interaction between AD Status and Treatment for memory errors.

Discussion

The results indicate that both AD Status and Drug Treatment significantly influenced maze performance on the Training and Memory Days. Transgenic mice consistently made more errors than wild-type mice, confirming expected cognitive deficits. Some drugs improved performance, and the significant interaction effects suggest that drug efficacy is dependent on the AD status of the mice. Post hoc analysis identifies which treatments differ significantly and highlights the drugs most effective in reducing errors in each AD Status group.