modeling_path_efficiency

2025-04-25

Required Packages

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(lme4)
```

```
## Loading required package: Matrix
```

```
library(lmerTest)  # for p-values
```

```
##
## Attaching package: 'lmerTest'
```

```
## The following object is masked from 'package:lme4':
##
##     lmer
```

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

```
library(lattice)
library(emmeans)
library(tidyr)
```

```
##
## Attaching package: 'tidyr'

## The following objects are masked from 'package:Matrix':
##
##     expand, pack, unpack
```

Load the data

```
df <- read.csv("~/coding_projects/mouse_data_analysis/data/combined_mouse_data.csv")
head(df)
```

```
##    Id Experiment Sex   Drug Genotype Trial Outcome
## 1 101          1   M Saline       WT     1   0.014
## 2 102          1   M Saline       WT     1   0.053
## 3 103          1   M Saline       WT     1   0.154
## 4 104          1   M Saline       WT     1   0.065
## 5 105          1   M Saline       WT     1   0.038
## 6 106          1   M Saline       WT     1   0.237
```

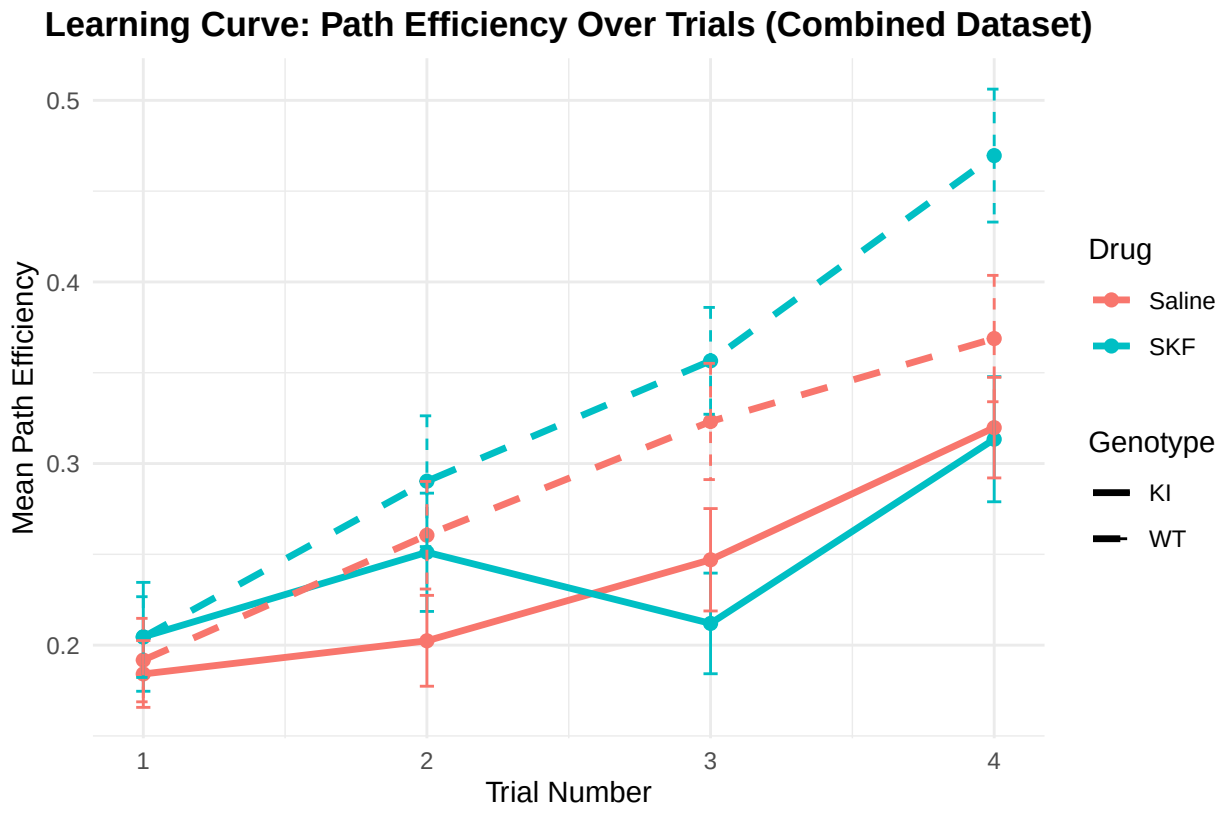
Make a plot of path efficiency accross trials

```
# Step 1: Average per mouse across experiment sessions
mouse_summary <- df %>%
  group_by(Id, Trial, Genotype, Drug) %>%
  summarize(Mean_Outcome = mean(Outcome, na.rm = TRUE), .groups = "drop")

# Step 2: Compute group mean and standard error
plot_data <- mouse_summary %>%
  group_by(Trial, Genotype, Drug) %>%
  summarize(
    Mean_Path_Efficiency = mean(Mean_Outcome),
    SE = sd(Mean_Outcome) / sqrt(n()),  # Standard Error
    .groups = "drop"
  )

# Step 3: Plotting
ggplot(plot_data, aes(
  x = Trial,
  y = Mean_Path_Efficiency,
  color = Drug,
  group = interaction(Drug, Genotype),
  linetype = Genotype
)) +
  geom_line(linewidth = 1.2) +
  geom_point(size = 2) +
  geom_errorbar(
    aes(ymin = Mean_Path_Efficiency - SE, ymax = Mean_Path_Efficiency + SE),
```

```
    width = 0.05
  ) +
  labs(
    title = "Learning Curve: Path Efficiency Over Trials (Combined Dataset)",
    x = "Trial Number",
    y = "Mean Path Efficiency",
    caption = "Error bars represent ±1 standard error of the mean (SEM) across mice."
  ) +
  scale_linetype_manual(values = c("WT" = "dashed", "KI" = "solid")) +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    plot.caption = element_text(hjust = 0.5)   # ← centers the caption
  )
```




```
plot_data
```

```
## # A tibble: 16 x 5
##    Trial Genotype Drug   Mean_Path_Efficiency     SE
##    <int> <chr>    <chr>                 <dbl>  <dbl>
##  1     1 KI       SKF                   0.205 0.0300
##  2     1 KI       Saline                0.184 0.0184
##  3     1 WT       SKF                   0.204 0.0222
##  4     1 WT       Saline                0.192 0.0230
##  5     2 KI       SKF                   0.251 0.0326
```

```
##  6     2 KI       Saline               0.202 0.0250
##  7     2 WT       SKF                  0.290 0.0360
##  8     2 WT       Saline               0.261 0.0296
##  9     3 KI       SKF                  0.212 0.0277
## 10     3 KI       Saline               0.247 0.0282
## 11     3 WT       SKF                  0.357 0.0294
## 12     3 WT       Saline               0.323 0.0320
## 13     4 KI       SKF                  0.313 0.0344
## 14     4 KI       Saline               0.320 0.0277
## 15     4 WT       SKF                  0.470 0.0366
## 16     4 WT       Saline               0.369 0.0348
```

```r
# Step 1: Compute trial-by-trial slopes
slope_table <- plot_data %>%
  arrange(Genotype, Drug, Trial) %>%
  group_by(Genotype, Drug) %>%
  mutate(Slope = Mean_Path_Efficiency - lag(Mean_Path_Efficiency)) %>%
  filter(!is.na(Slope)) %>%
  mutate(Comparison = paste0("Trial ", Trial - 1, " to ", Trial)) %>%
  select(Genotype, Drug, Comparison, Slope)

# Step 2: Compute average slope per group
avg_slope <- slope_table %>%
  group_by(Genotype, Drug) %>%
  summarise(Average_Slope = mean(Slope), .groups = "drop")

# Step 3: Merge and pivot for final table
final_table <- left_join(avg_slope, slope_table, by = c("Genotype", "Drug")) %>%
  pivot_wider(
    names_from = Comparison,
    values_from = Slope
  )

# View the final table
print(final_table)
```

```
## # A tibble: 4 x 6
##   Genotype Drug   Average_Slope `Trial 1 to 2` `Trial 2 to 3` `Trial 3 to 4`
##   <chr>    <chr>          <dbl>          <dbl>          <dbl>          <dbl>
## 1 KI       SKF           0.0363         0.0466        -0.0392         0.101
## 2 KI       Saline        0.0452         0.0183         0.0446         0.0728
## 3 WT       SKF           0.0884         0.0858         0.0663         0.113
## 4 WT       Saline        0.0590         0.0688         0.0626         0.0457
```

Based upon the graph, all the mice seem to be learning as their path efficiency is greater in trial 4 than in trial 1. WT also seems to have the greatest rate of improvement, especially when treated with SKF. In contrast, KI mice improve more gradually, and their learning appears less sensitive to drug treatment.

Convert data types to factors

```r
df$Id <- factor(df$Id)
df$Experiment <- factor(df$Experiment)
df$Sex <- factor(df$Sex)
df$Drug <- factor(df$Drug)
df$Genotype <- factor(df$Genotype)
```

Fit the LMM

(baseline: Drug: Saline, Genotype: WT, Sex: F, Trial: 1)

```r
lmm <- lmer(
  log(Outcome) ~ Drug*Genotype*Trial + Sex*Trial + (1 | Id/Experiment),
  data = df,
  REML = FALSE
)
summary(lmm)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula:
## log(Outcome) ~ Drug * Genotype * Trial + Sex * Trial + (1 | Id/Experiment)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##   3631.4   3698.2  -1802.7   3605.4     1252
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.2468 -0.6303 -0.0370  0.7069  2.2984
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  Experiment:Id (Intercept) 0.05039  0.2245
##  Id            (Intercept) 0.03438  0.1854
##  Residual                  0.94895  0.9741
## Number of obs: 1265, groups:  Experiment:Id, 159; Id, 61
##
## Fixed effects:
##                             Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)              -2.424e+00  1.538e-01  8.533e+02 -15.764  < 2e-16 ***
## DrugSKF                   2.542e-01  1.912e-01  1.106e+03   1.330   0.1838
## GenotypeWT                4.260e-04  1.993e-01  9.290e+02   0.002   0.9983
## Trial                     2.420e-01  5.321e-02  1.106e+03   4.548 6.01e-06 ***
## SexM                     -5.684e-02  1.505e-01  5.141e+02  -0.378   0.7058
## DrugSKF:GenotypeWT       -1.290e-01  2.683e-01  1.106e+03  -0.481   0.6307
## DrugSKF:Trial            -1.212e-01  6.977e-02  1.106e+03  -1.737   0.0827 .
## GenotypeWT:Trial          4.855e-02  6.945e-02  1.107e+03   0.699   0.4846
## Trial:SexM               -3.267e-02  5.022e-02  1.107e+03  -0.651   0.5155
## DrugSKF:GenotypeWT:Trial  1.315e-01  9.798e-02  1.107e+03   1.343   0.1797
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##             (Intr) DrgSKF GntyWT Trial  SexM   DrSKF:GWT DSKF:T GnWT:T Trl:SM
## DrugSKF     -0.621
## GenotypeWT  -0.659  0.479
## Trial       -0.865  0.598  0.577
## SexM        -0.389 -0.001  0.010  0.313
## DrgSKF:GnWT  0.442 -0.712 -0.676 -0.426  0.002
## DrugSKF:Trl  0.567 -0.913 -0.438 -0.655  0.000  0.650
## GntypWT:Trl  0.573 -0.458 -0.871 -0.662 -0.008  0.646     0.502
## Trial:SexM   0.324  0.000 -0.008 -0.375 -0.834 -0.001     0.000  0.010
## DrSKF:GWT:T -0.403  0.650  0.617  0.467 -0.001 -0.913    -0.712 -0.709  0.000
```

```
confint(lmm, method = "Wald")
```

```
##                                 2.5 %      97.5 %
## .sig01                             NA          NA
## .sig02                             NA          NA
## .sigma                             NA          NA
## (Intercept)              -2.72569776 -2.12286201
## DrugSKF                  -0.12042029  0.62888454
## GenotypeWT               -0.39010905  0.39096113
## Trial                     0.13772205  0.34631078
## SexM                     -0.35179997  0.23812499
## DrugSKF:GenotypeWT       -0.65489194  0.39684446
## DrugSKF:Trial            -0.25794339  0.01555627
## GenotypeWT:Trial         -0.08756330  0.18466318
## Trial:SexM               -0.13109617  0.06575546
## DrugSKF:GenotypeWT:Trial -0.06049485  0.32358397
```

Log transform as ratio is not normal

Post Hoc Analysis

```
# Get estimated slopes of Trial within each Drug × Genotype group
em_trends <- emtrends(lmm, ~ Drug * Genotype, var = "Trial")
summary(em_trends)
```

```
##  Drug   Genotype Trial.trend     SE   df lower.CL upper.CL
##  Saline KI             0.226 0.0498 1113  0.12806    0.323
##  SKF    KI             0.104 0.0498 1113  0.00684    0.202
##  Saline WT             0.274 0.0494 1116  0.17733    0.371
##  SKF    WT             0.285 0.0489 1113  0.18860    0.381
##
## Results are averaged over the levels of: Sex
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
```

```r
# Contrast the learning slopes between SKF KI and SKF WT
contrast(em_trends, method = list("SKF KI - SKF WT" = c(0, -1, 0, 1)))
```

```
##  contrast        estimate     SE   df t.ratio p.value
##  SKF KI - SKF WT     0.18 0.0693 1113   2.597  0.0095
##
## Results are averaged over the levels of: Sex
## Degrees-of-freedom method: kenward-roger
```

Clear significance, good!

```r
contrast(em_trends, interaction = "pairwise")
```

```
##  Drug_pairwise Genotype_pairwise estimate     SE   df t.ratio p.value
##  Saline - SKF  KI - WT              0.132 0.0983 1114   1.338  0.1811
##
## Results are averaged over the levels of: Sex
## Degrees-of-freedom method: kenward-roger
```

```r
contrast(em_trends, method = list("Interaction (SKF effect diff by Genotype)" = c(-1, 1, 1, -1)))
```

```
##  contrast                                  estimate     SE   df t.ratio p.value
##  Interaction (SKF effect diff by Genotype)   -0.132 0.0983 1114  -1.338  0.1811
##
## Results are averaged over the levels of: Sex
## Degrees-of-freedom method: kenward-roger
```

Now let us check sex difference.

```r
# Get EMMs for Sex
em_sex <- emmeans(lmm, ~ Sex)
```

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

```r
# Contrast the levels of Sex
contrast(em_sex, method = "pairwise")
```

```
##  contrast estimate     SE   df t.ratio p.value
##  F - M       0.139 0.0853 63.3   1.624  0.1093
##
## Results are averaged over the levels of: Drug, Genotype
## Degrees-of-freedom method: kenward-roger
## Results are given on the log (not the response) scale.
```

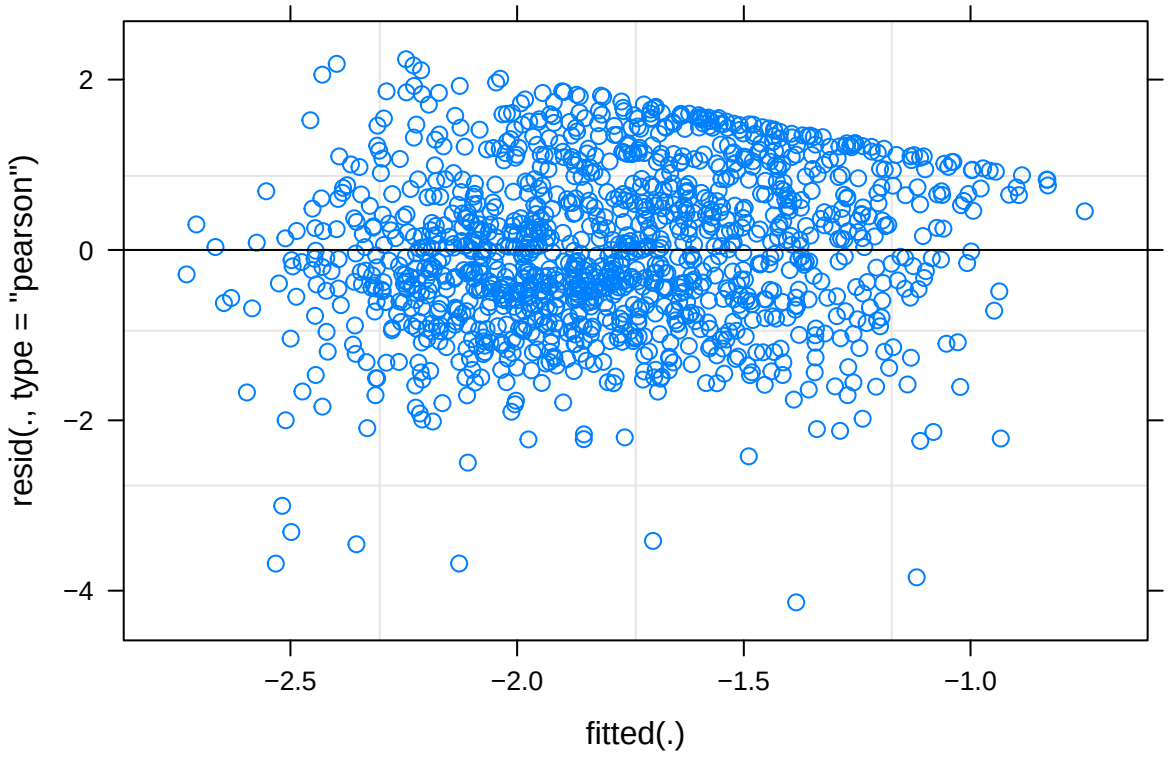
No significant difference

Model Diagnostics

We check the plot diagnostics from here

Fitted values vs residuals

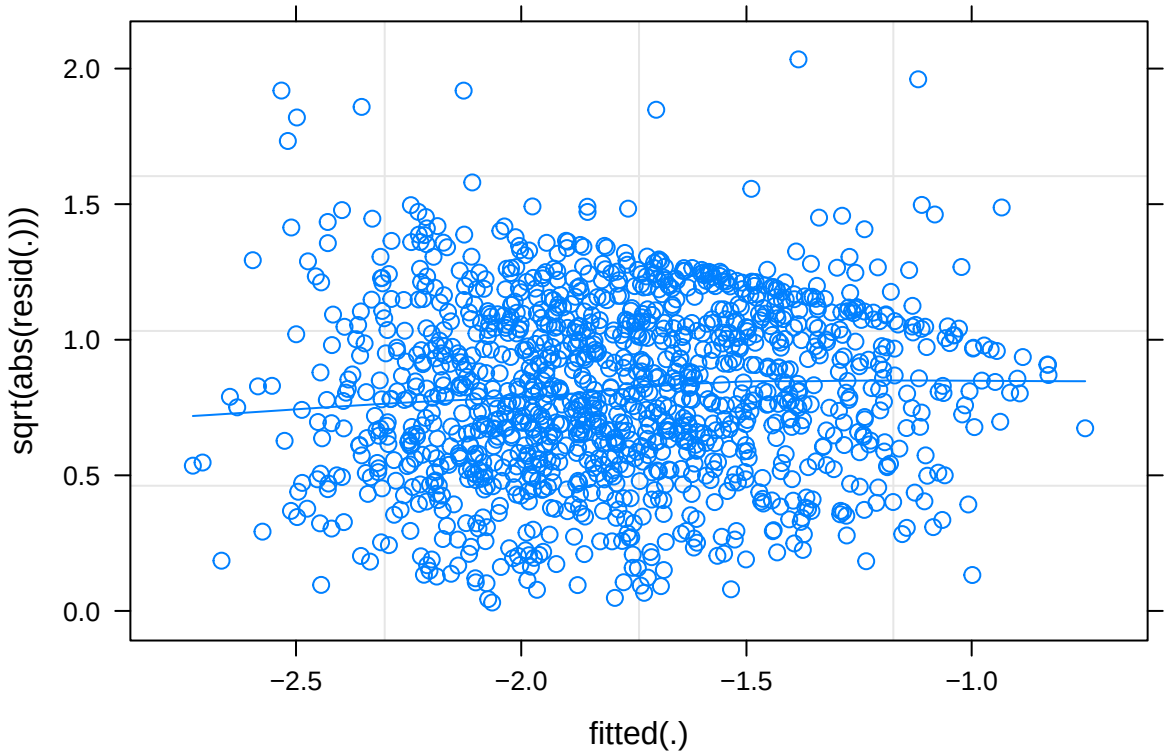
```
plot(lmm)
```



Downward trend may indicate slight nonlinearity observed

scale-location plots

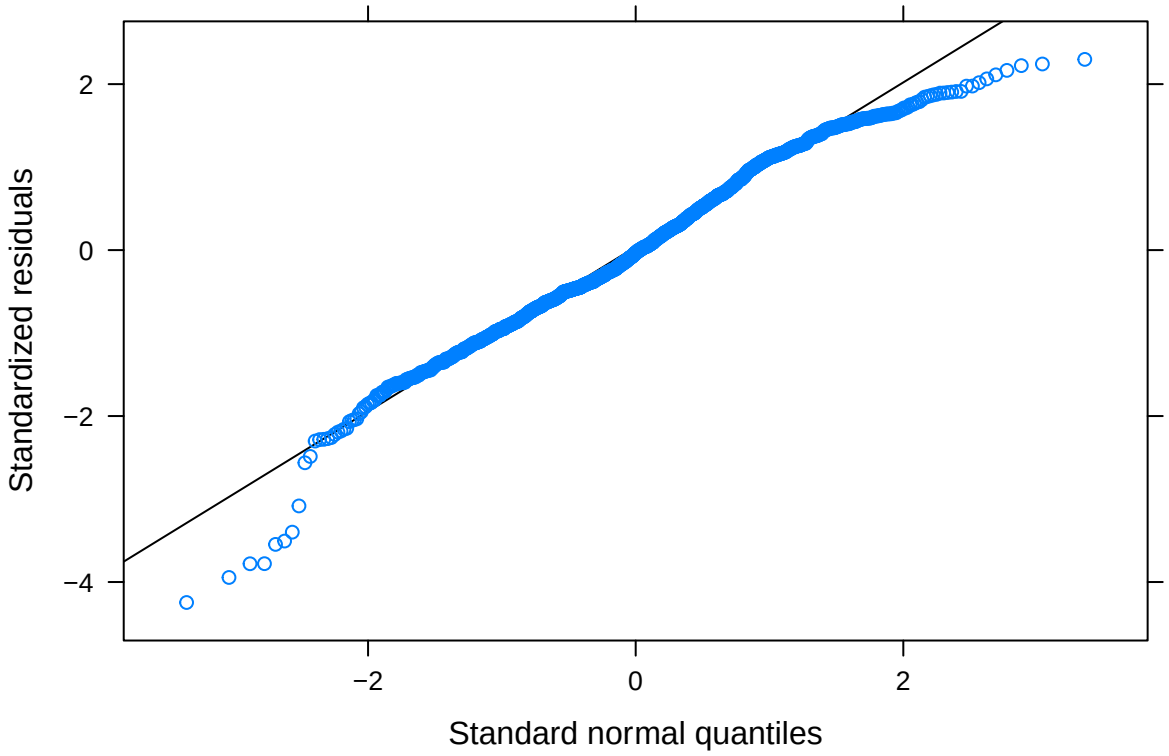
```
plot(lmm, sqrt(abs(resid(.))) ~ fitted(.), type = c("p", "smooth"))
```



Doesn't seem to be an issue with heteroscedasticity

Quantile-Quantile plot

```
qqmath(lmm)
```



Some minor deviation from the normality assumption of the residuals.