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)
```

Load the data

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

```
##    Id Expiriment 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
```

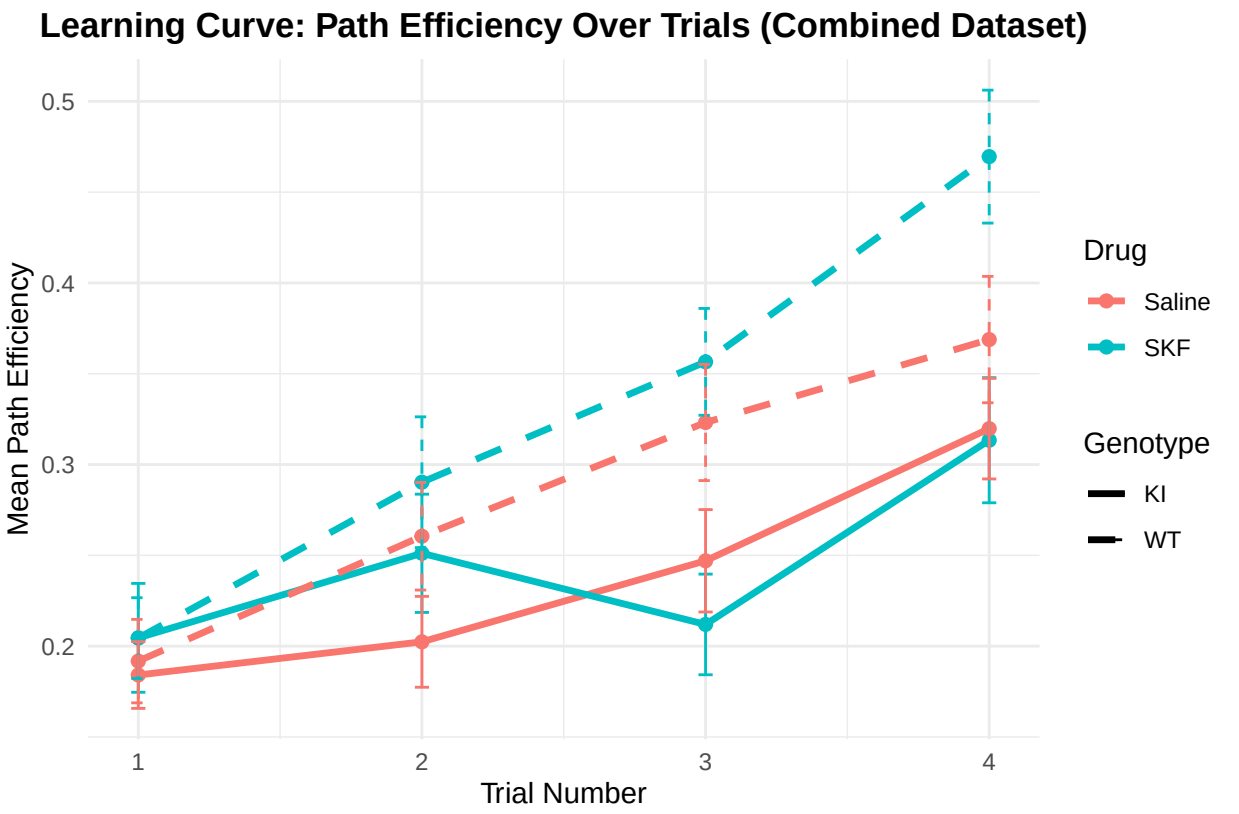
Make a plot of path efficiency accross trials

```r
# 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
  )
```



Error bars represent ±1 standard error of the mean (SEM) across mice.

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

```
df$Id <- factor(df$Id)
df$Expiriment <- factor(df$Expiriment)
```

Fit the LMM

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

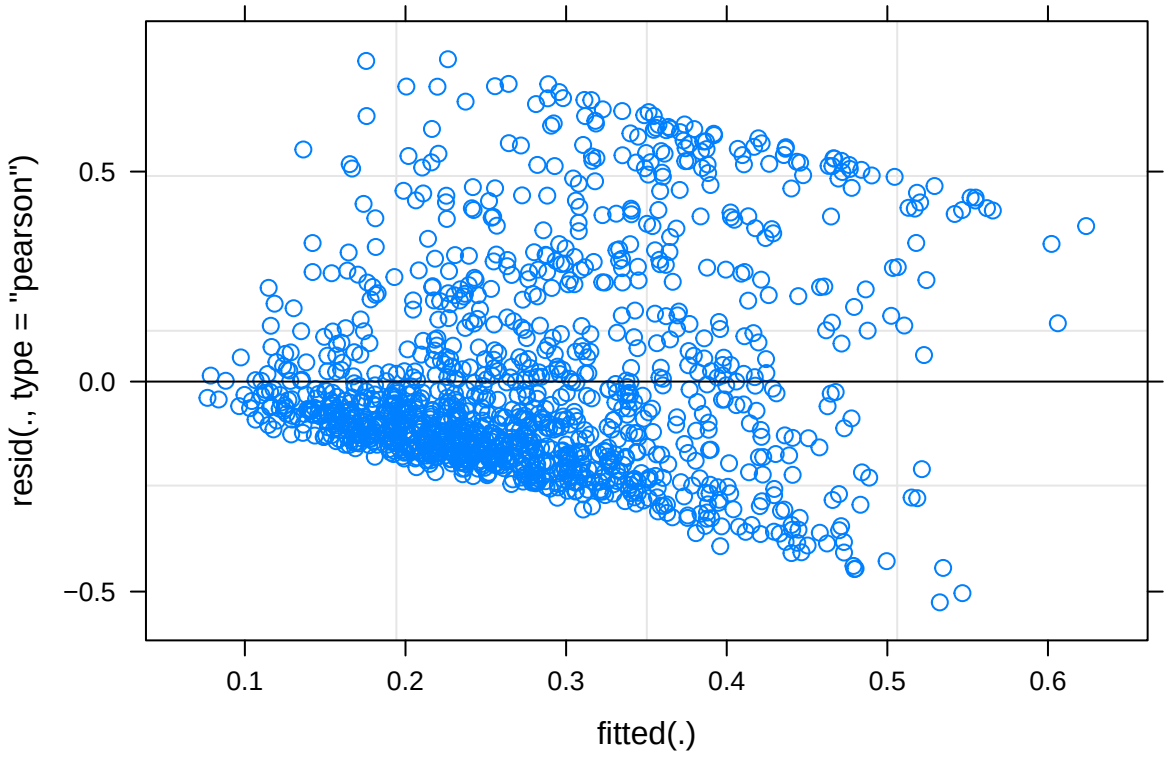
```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: Outcome ~ Drug * Genotype * Trial + Sex * Trial + (1 | Id/Expiriment)
##    Data: df
##
##      AIC      BIC   logLik deviance df.resid
##    241.5    308.4   -107.8    215.5     1252
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.0754 -0.6390 -0.3207  0.3830  3.0304
##
## Random effects:
##  Groups        Name        Variance Std.Dev.
##  Expiriment:Id (Intercept) 0.005450 0.07382
##  Id            (Intercept) 0.001765 0.04201
##  Residual                  0.064155 0.25329
## Number of obs: 1265, groups:  Expiriment:Id, 159; Id, 61
##
## Fixed effects:
##                             Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)                1.304e-01  4.011e-02  8.512e+02   3.251 0.001195 **
## DrugSKF                    6.243e-02  4.970e-02  1.105e+03   1.256 0.209322
## GenotypeWT                 1.571e-02  5.195e-02  9.275e+02   0.302 0.762413
## Trial                      5.025e-02  1.384e-02  1.105e+03   3.632 0.000294 ***
## SexM                      -9.368e-03  3.932e-02  5.122e+02  -0.238 0.811791
## DrugSKF:GenotypeWT        -8.699e-02  6.976e-02  1.106e+03  -1.247 0.212672
## DrugSKF:Trial             -2.342e-02  1.814e-02  1.105e+03  -1.291 0.197032
## GenotypeWT:Trial           1.219e-02  1.806e-02  1.107e+03   0.675 0.499787
## Trial:SexM                -9.479e-03  1.306e-02  1.106e+03  -0.726 0.468038
## DrugSKF:GenotypeWT:Trial  5.001e-02  2.548e-02  1.106e+03   1.963 0.049896 *
## ---
## 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     -0.619
## GenotypeWT  -0.658  0.478
## Trial       -0.862  0.598  0.575
## SexM        -0.390 -0.001  0.010  0.311
## DrgSKF:GnWT  0.440 -0.712 -0.674 -0.426  0.002
## DrugSKF:Trl  0.565 -0.913 -0.436 -0.655  0.000  0.650
## GntypWT:Trl  0.571 -0.458 -0.868 -0.662 -0.008  0.646     0.502
## Trial:SexM   0.323  0.000 -0.008 -0.375 -0.830 -0.001     0.000  0.010
## DrSKF:GWT:T -0.402  0.650  0.615  0.467 -0.001 -0.913    -0.712 -0.709  0.000
```

Model Diagnostics

We check the plot diagnostics from here

Fitted values vs residuals

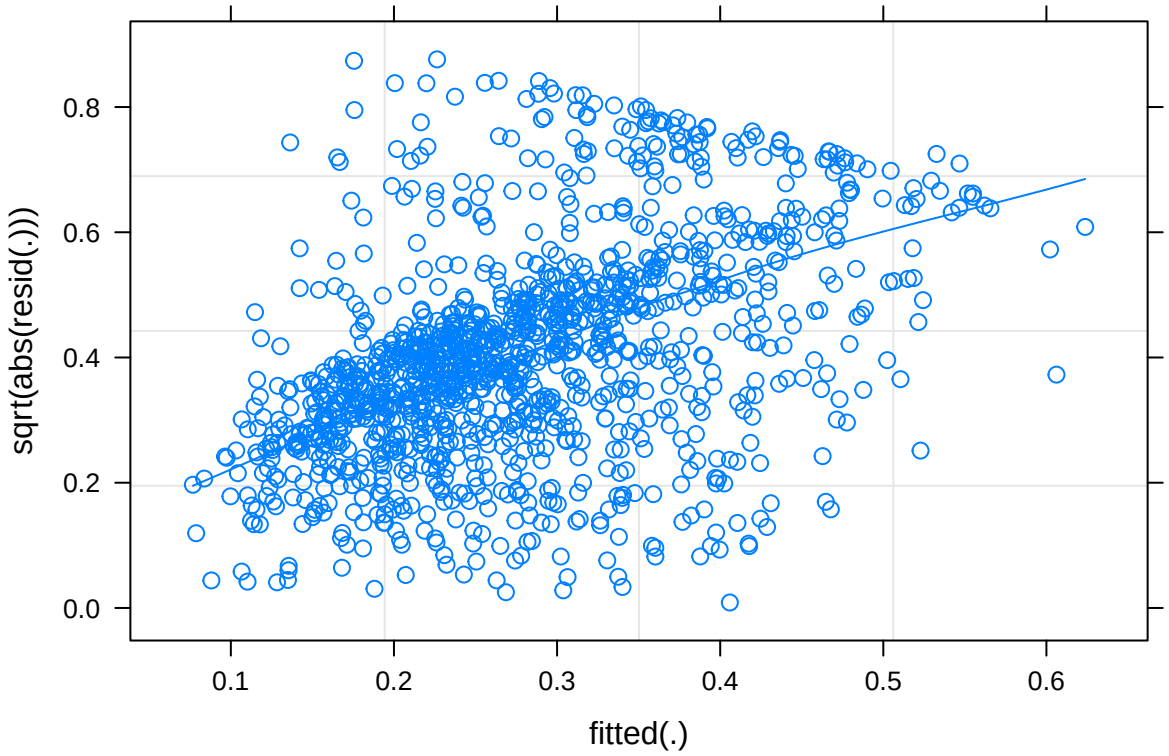
```
plot(lmm)
```



Downward trend may indicate slight nonlinearity observed

scale-location plots

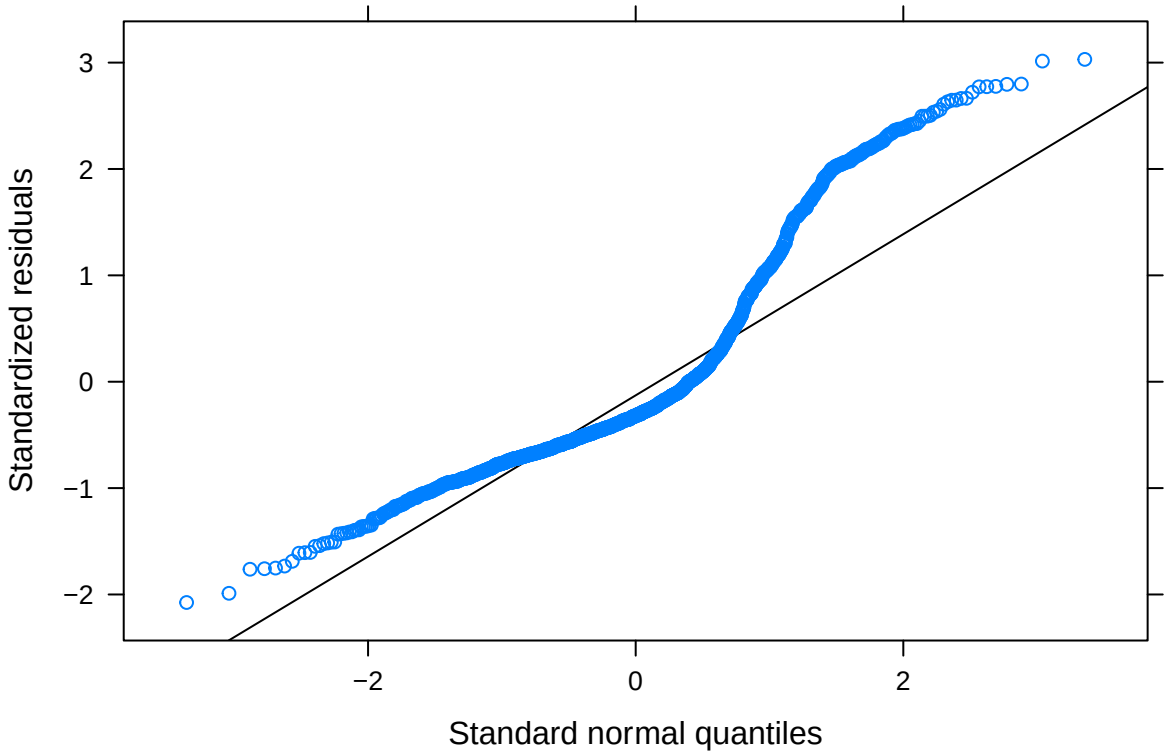
```
plot(lmm, sqrt(abs(resid(.))) ~ fitted(.), type = c("p", "smooth"))
```



Doesn't seem to be an issue with heteroscedasticity

Quantile-Quantile plot

```
qqmath(lmm)
```



Some minor deviation from the normality assumption of the residuals.