Statistical Modeling in Linguistics

A systematic introduction with instructions for using R

Roland Schäfer

Dedicated to a randomly chosen individual.

Contents

Preface

Gelman & Hill (2006)

References

Gelman, Andrew & Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Acknowledgments

Abbreviations

… …	… …
… …	… …
… …	… …
… …	… …

1 Science, data, and statistics

MacDonald & Gardner (2000)

Test SY

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Test LI

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Test FI

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Test FR

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

Test FD

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

In this book, code listing are displayed as inline blocks such as the following simple code which simulates t-tests under the null hypothesis in order to demonstrate that all p-values have equal probability under the null.

```
# Set simulation parameters.
nsim  <- 1000
n     <- 100
meen  <- 0
stdev <- 1

# Data structure for results.
sims <- rep(NA, nsim)

# Simulations.
for (i in 1:nsim) {
  a <- rnorm(n, mean = meen, sd = stdev)
  b <- rnorm(n, mean = meen, sd = stdev)
  p <- t.test(a,b)$p.value
  sims[i] <- p
}
```

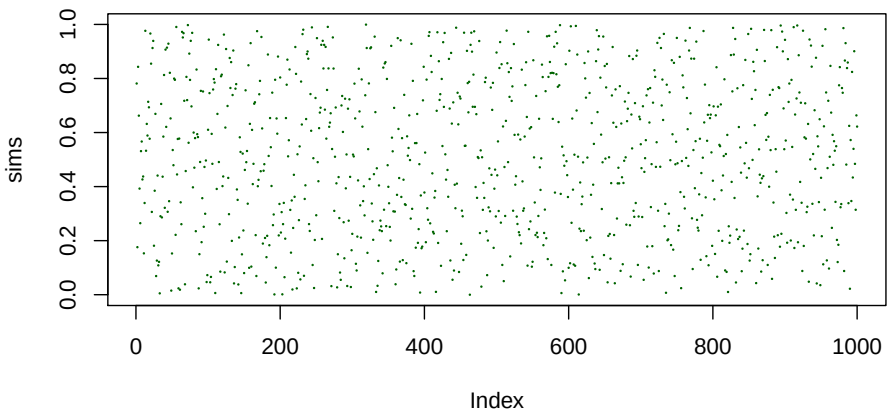

Figure 1.1: Scatterplot of p-values.

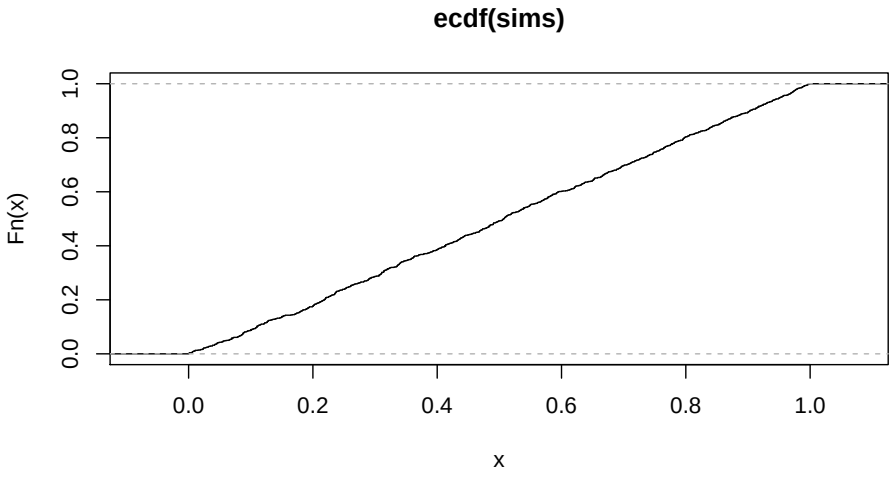

Figure 1.2: Empirical cumulative density distribution of p-values.

Figure 1.3: Histogram of p-values.

See Figure 1.2 for the cumulative density of p-values under the null in a series of 1000 t-tests. This was plotted using the following command.

```
plot(ecdf(sims))
```

2 Describing data

3 Visualising data

4 Tests

5 Models

6 Generalised models

7 Mixed models

8 Where to go from here?

References

MacDonald, Paul L. & Robert C. Gardner. 2000. Type I error rate comparisons of post hoc procedures for IxJ chi-square tables. *Educational and Psychological Measurement* 60(5). 735–754.