sensR part2 exercises: Similarity testing and replicated discrimination data

Introduction

Aim

The aim of this exercise is to guide students through some more advanced uses of the sensR package (compared to the initial Tutorial 3 on the sensR basics)

You can use the sensR-package as a tool for the planning and analysis of sensory discrimination and similarity experiments (see Tutorial 3 for more info). In this Tutorial 7 the focus is on the following three tasks:

1) Analysing similarity test data.
2) Planning similarity tests (power and sample size for similarity testing)
3) Analysing replicated difference test data

Method

Type in correct the commands in R or R studio using the sensR-package. For the sensR-package you will be presented with the following terms. The following table is copied from Tutorial 3, and is still relevant here:

Term	**Name**	**Explanation**
α	The level of significance	Normally set to 0.05 (5%)
1-β	The power	Dependent on YOUR choice at the planning of a study
x	Number of successes	From test
n	Number of participants / tests	Either your calculated choice or given in an already conducted test
method	Type of test	Type of test of the 6 basic protocols
δ p_C p_D	Analysis level	See the literature for M3A & B (Figure 7.1) and the lecture notes

In addition the following NEW terms will be presented and used here:

Term	**Name**	**Explanation**
d.prime0	Similarity value	The level of similarity on the d.prime-scale
d.primeA	The (assumed) true d.prime value	The assumed "alternative" situation – to be used for scenario based power and sample size computations
p_{dA}	The (assumed) true p_D-value	The assumed "alternative" situation – to be used for scenario based power and sample size computations
p_{d0}	Similarity value	The level of similarity on the p_D-scale
sd.indiv	The individual variability in d-prime values.	A value of 0 (default) corresponds to complete independence
Beta-Binomial	Standard and corrected	Models for replicated data

See this link http://finzi.psych.upenn.edu/R/library/sensR/html/00Index.html for commands, explanations and help functions in sensR.

Datasets
There are no data sets to download.

Exercise 1 (Analysing similarity test data)

You have conducted a Duo-Trio test experiment, and it yielded 17 correct answers out of 30 tests. Similarity is defined to correspond to a p_D-value less than 0.25.

1) *Have you proven that the products are similar?*
2) *What range of values for d-prime and p_D are plausible given the data you have?*
3) *Answer the same two questions with 170 correct out of 300 tests.*

Exercise 2 (Analysing similarity test data)

You have conducted a 3-AFC test experiment, and it yielded 12 correct answers out of 30 tests. Similarity is defined to correspond to a d.prime-value less than 0.5.

1) *Have you proven that the products are similar?*
2) *What range of values for d-prime and p_D are plausible given the data you have?*
3) *Answer the same two questions with 48 correct out of 120 tests.*

Exercise 3 (Planning, power)

Similarity is defined to correspond to a p_D-value less than 0.25.

1) *What is the probability that the duo-trio test with n=30 will show similarity, if the real p_D-value is at its minimum 0?*
2) *What is the probability that the duo-trio test with n=300 will show similarity, if the real p_D-value is at its minimum 0?*
3) *What is the probability that the duo-trio test with n=30 will show similarity, if the real p_D-value is 0.1?*
4) *What is the probability that the duo-trio test with n=300 will show similarity, if the real p_D-value is 0.1?*

Exercise 4 (Planning, sample size)

Similarity is defined to correspond to a d.prime-value less than 0.5. Using the usual alpha-level of 0.05, and assuming that the true d.prime is 0, what sample sizes are needed to show similarity with probability 0.90 for the following situations:

1) *Using a duo-trio protocol*
2) *Using a triangle protocol*
3) *Using a 3-AFC protocol (assuming this is possible)*
4) *Using a 2-AFC protocol (assuming this is possible)*
5) *Answer the same four questions for the scenario where the true d.prime is actually 0.25*

Exercise 5

In this exercise we do some replicated difference test data analysis on data that we simulate ourselves:

```
# Set the situation: (make your own choices)
myN=15 #Number of individuals
myK=10 #Number of replications
myd.prime=1 # Level of product difference
mysd.indiv=1 # individual heterogeneity

# Simulate and plot some data:
rs=discrimSim(myN, replicates = myK, d.prime = myd.prime, method = "triangle", sd.indiv=mysd.indiv)
barplot(rs[order(rs)],col="blue",ylim=c(0,myK), names.arg=1:myN,cex.names=0.8)
abline(h=mean(rs))

# Put the simulated data on the form needed for the betabin analysis, and have a look at the data:
X=matrix(c(rs,rep(myK,myN)),ncol=2)
X
```

```
# Do the betabin analysis:
summary(betabin(X,method="triangle"))
```

Answer the following questions for your data:

1) Does the data exhibit significant excess individual variability?
2) Are the products significantly different:
    a. By the naïve (pooled) test?
    b. By the betabin test for association?
3) Give the d-prime and its confidence interval – and compare the confidence interval with the one produced by the naïve (pooled) analysis – how much larger is it?

If time, you can try again with the same or different choices?

Ekstra (and difficult question): We have currently no function in sensR to compute power for the tests provided in the betabin function. Do you have any idea about how we could make such a function?