Same-different and A-not A tests with **sensR**

Christine Borgen Linander

DTU Compute
Section for Statistics
Technical University of Denmark
chjo@dtu.dk

August 20th 2015

DTU Compute
Department of Applied Mathematics and Computer Science

A huge thank to a former colleague of mine Rune H B Christensen.

Outline

Same-Different and the Degree-of-Difference tests

2 products — 2 confusable stimuli:

A Chocolate bar (standard)

B Chocolate bar with less saturated fat

Setting:

- One pair of samples evaluated at each trial
- Question: Are the samples the *same* or *different*?

Stimuli:

- Same stimuli pairs: AA and BB
- Different stimuli pairs: AB and BA

Same-Different test:

Same	Different
☐	☐

Degree-of-Difference test:

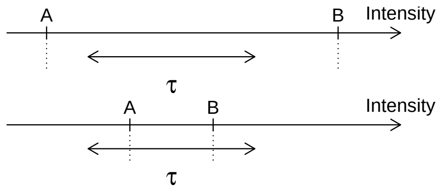
Same	2	3	4	Different
☐	☐	☐	☐	☐

Characteristics of the DOD test

- An unspecified test (like Triangle, Duo-Trio, Tetrad)
- Only 2 samples compared at each trial
- Easily understood test (by consumers) (O'Mahony and Rousseau, 2002)
- No prior knowledge of products required (unlike A-not A)
- Response bias (like A-not A)

Giving answers — τ criteria and the decision rule

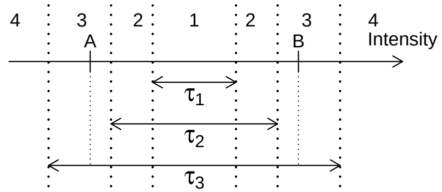
Same-Different:



$|B - A| > \tau \quad \rightarrow$ "different"

$|B - A| < \tau \quad \rightarrow$ "same"

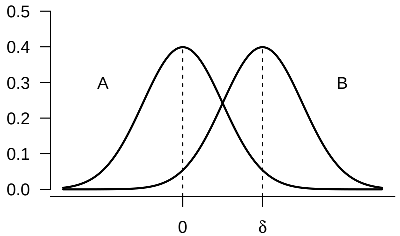
Degree of difference:



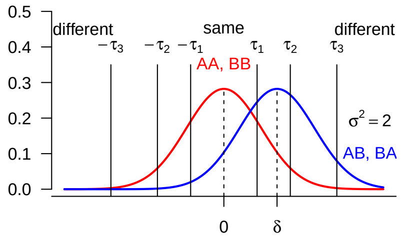
Rating scale: 1 ☐ 2 ☐ 3 ☐ 4 ☐

Thurstonian model for the DOD test

Thurstonian distributions:



Difference distributions



Probability of answer in the jth category:

$$P(\text{"}j\text{"}|\text{Same-pair}) = f_s(\boldsymbol{\tau})$$
$$P(\text{"}j\text{"}|\text{Different-pair}) = f_d(\boldsymbol{\tau}, \delta)$$

Maximum likelihood estimation of parameters:

$$\text{likelihood} \sim f_s(\boldsymbol{\tau}) + f_d(\boldsymbol{\tau}, \delta)$$

Same-different example

Examples in R — difference and similarity assessments.

Sample	Response		
	"Same"	"Different"	Total
Same	8	5	13
Different	4	9	13

The A-not A protocol

Situation:

- 2 products: A and B ("not A")
- Assessors are familiarized with A samples (and sometimes B samples as well)
- Assessors are served one sample — either A or B
- Question: Is the sample an A or a *not* A sample?

Known as the "yes-no" method in Signal Detection Theory (Macmillan and Creelman, 2005)

Example: the A-not A test

Example data:

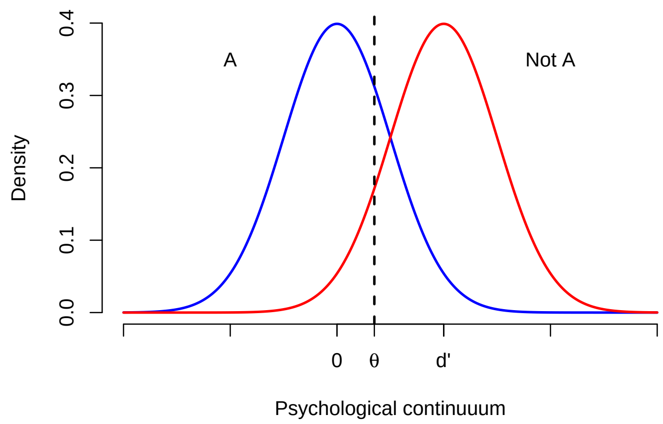
Sample	Response "A"	Response "Not-A"	Total
A	26	29	55
Not-A	14	41	55

Null hypothesis, H_0: products are similar
Alternative hypothesis, H_A: products are different

Problem: There are many tests to choose from.

What is the p-value?
Can we reject H_0?

The Thurstonian model for the A-not A test



- d': Sensory difference
- θ: Decision threshold

Estimation of d' with sensR

Estimation of d' with sensR:

```
> library(sensR)
> AnotA(26, 55, 14, 55)
Call:  AnotA(x1 = 26, n1 = 55, x2 = 14, n2 = 55)

Results for the A-Not A test:

        Estimate Std. Error     Lower    Upper    P-value
d-prime 0.591838  0.2492597 0.1032979 1.080378 0.01431559
```

Likelihood confidence intervals for A-not A tests

```
> (ana <- AnotA(26, 55, 14, 55))
Call:  AnotA(x1 = 26, n1 = 55, x2 = 14, n2 = 55)

Results for the A-Not A test:

        Estimate Std. Error     Lower    Upper    P-value
d-prime 0.591838  0.2492597 0.1032979 1.080378 0.01431559
```

Standard normal based confidence intervals: $CI_{95\%} = d' \pm 1.96\text{se}(d')$

Improved likelihood based confidence intervals:

```
> confint(ana)
              2.5 %    97.5 %
threshold -0.4007727 0.2627993
d.prime    0.1063875 1.0842269
```


Similarity testing

Aim:
Prove that products are identical ~~Prove that products are identical~~
Establish similarity within a **similarity bound** at some α-level

How:
Interchange the roles of the hypotheses:

Example:
H_0: d' is larger than 1
H_A: d' is less than 1

Huge practical challenge:
How to choose the **similarity bound**?

Similarity testing with d' for A-not A tests

Use d' for similarity testing:

Hypotheses:
H_0: d' is larger than 1
H_A: d' is less than 1

p-value $= P(Z < (d' - d'_0)/\text{se}(d') | H_0)$:

```
> ## The Wald statistic:
> statistic <- (0.592 - 1)/ 0.249
> ## Compute p-value:
> pnorm(statistic, lower.tail=TRUE)
[1] 0.05065307
```


Discrimination measures in equal-variance models

- $d' = (\mu_2 - \mu_1)/\sigma$ is the (relative) distance between normal distributions
- $\lambda = 2\Phi(-d'/2)$ is the distribution overlap $(0 < \lambda \leq 1)$
- Sensitivity, $\mathcal{S} = P(x_1 < x_2) = \Phi(d'/\sqrt{2})$ is the probability that a random sample from the low-intensity distribution has a lower intensity than a random sample from the high-intensity distribution

```
overlap 1.0 0.90 0.80 0.70 0.60 0.50 0.40 0.30 0.20 0.10 0.05 0.01
d.prime 0.0 0.25 0.51 0.77 1.05 1.35 1.68 2.07 2.56 3.29 3.92 5.15
AUC     0.5 0.60 0.69 0.78 0.85 0.91 0.95 0.98 0.99 1.00 1.00 1.00
```

The Receiver Operating Characteristic (ROC) curve

What is a ROC curve?

- A visual description of the discriminative ability
- A central concept in Signal Detection Theory (A plot of False positive ratio $\sim$ True positive ratio (alt. Hit rate $\sim$ False-alarm Rate)

The real number of interest — the area under the ROC curve, AUC:

$$\mathcal{S} = AUC = \Phi(d'/\sqrt{2})$$

Examples in **R**

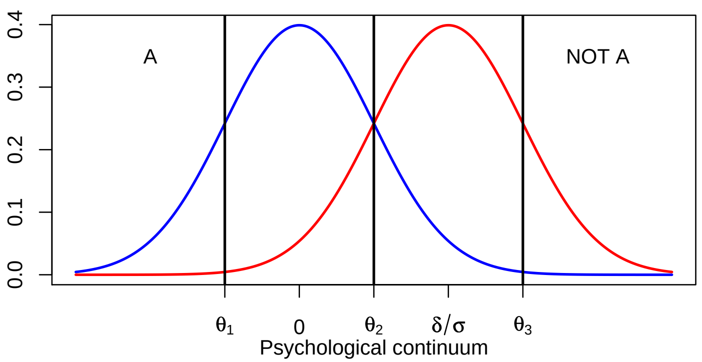
Basics of the A-not A with sureness protocol

- Answers are given on a 'sureness' scale with J categories
- The model assumes $J-1$ thresholds are adopted by the assessors
- Multinomial response: several ordered response categories
- Many parameters: Thresholds, θ_j and effect, δ

Table: Soup data

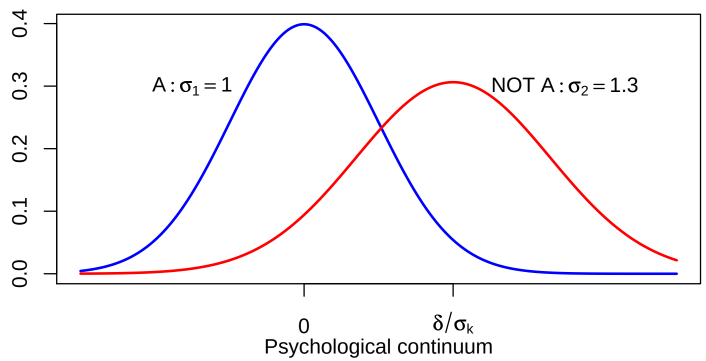
	'Reference'			'Not Reference'		
Product	Sure	Not Sure	Guess	Guess	Not Sure	Sure
Reference	134	162	66	41	122	222
Test	101	101	51	57	157	653

Thurstonian model for the A-not A with sureness protocol

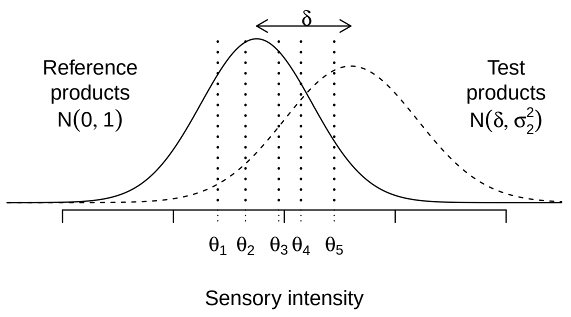


Multiple thresholds (θ) parameters.

Unequal variances model



An unequal-variance model in practice



Product	'Reference' Sure	Not Sure	Guess	'Not Reference' Guess	Not Sure	Sure
Reference	132	161	65	41	121	219
Test	96	99	50	57	156	650

Table: Discrimination of packet soup (Christensen, Cleaver and Brockhoff, 2011)

Cumulative link models for A-not A with sureness

- Christensen showed that the A-not A protocol (with and without sureness) is a version of a cumulative link model:

$$P(S_i \leq \theta_j) = \Phi\left(\frac{\theta_j - \delta(\texttt{prod}_i)}{\sigma(\texttt{prod}_i)}\right)$$

  where σ is the ratio of scales (std. dev).
- This provides (optimal) ML estimates of the parameters, standard errors etc.
- profile likelihood confidence intervals available with `confint`.

```
> fm1 <- clm(sureness ~ prod, data=my_data, link="probit")
> summary(fm1) ## print d-prime etc.
> confint(fm1) ## likelihood confidence interval for d-prime
```


Discrimination measures in unequal-variance models

- $d' = (\mu_2 - \mu_1)/\sigma$ (no change)
- Distribution overlap: use the `overlap` function.
- Sensitivity: $\mathcal{S} = \Phi(d'/\sqrt{1+\sigma_2^2})$ where σ_2 is the scale ratio of the high-intensity distribution relative to the low-intensity distribution.

All measureas are 'equivalent' in the equal-variance model, but not so in the unequal-variance model.

In the unequal-variance model d' can be a poor measure of discrimination.