# Exercises in sensR basics
# — exercises AND solutions

Per Bruun Brockhoff

June 20, 2015
file: `sensR_part1_exercises.Rnw`

**Topics:**

- Three levels of analysis ($d'$, $p_d$ and $p_c$). Exercises {1, 2, 3, 4, 5}.
- Analysis of data from sensory discrimination experiments using the simple-binomial protocols using hypothesis/significance tests, $p$-values and confidence intervals. Exercises {1, 5}.
- Power and sample size computations for sensory discrimination tests using the simple-binomial protocols. Exercises {6, 7}.

## Exercise 1

You have conducted a Duo-Trio experiment, and it yielded 21 correct answers out of 30 tests.

1. Have you proven that the products differ?
2. What range of values for $d'$ and $p_d$ are plausible given the data you have?

**Answer to the exercise:**

To analyze the data, we use the `dicrim` function from the sensR package:

```
discrim(21, 30, method="duotrio")

##
## Estimates for the duotrio discrimination protocol with 21 correct
## answers in 30 trials. One-sided p-value and 95 % two-sided confidence
## intervals are based on the 'exact' binomial test.
##
##         Estimate Std. Error   Lower  Upper
## pc         0.700    0.08367 0.50604 0.8527
```

```
## pd         0.400    0.16733 0.01208 0.7053
## d-prime    1.715    0.49640 0.25735 2.7728
##
## Result of difference test:
## 'exact' binomial test:  p-value = 0.02139
## Alternative hypothesis: d-prime is greater than 0
```

1. Since the $p$-value is 0.021, i.e. less than 5%, we can conclude that the products are significantly different at the 5% level. The $p$-value means that the probability of observing 21 or more correct out of 30 if there is no product difference is $p = 0.021$.
2. The range of plausible values of $d'$ and $p_d$ are given by the confidence intervals for these parameters. The 95% confidence interval for $d'$ is (0.26; 2.77), and the 95% CI for $p_d$ is (0.01; 0.71).

# Exercise 2

If you have a $d'$ of 1.0 with a 3-AFC test, what is $p_d$ and $p_c$ (hint: use the `rescale`-function in sensR)?

**Answer to the exercise:**

If we use the `rescale` function we get

```
rescale(d.prime=1, method="threeAFC")
```

```
##
## Estimates for the threeAFC protocol:
##          pc        pd d.prime
## 1 0.6337021 0.4505531       1
```

Here $p_d = 0.45$ meaning that 45% of the population is able to detect the difference between the products.

Also $p_c = 0.63$, so we would expect 63% correct answers if we conducted an experiment.

# Exercise 3

You have conducted a Triangle test (test 1) with 73% correct answers. Your colleague tested the same products with the same test (test 2) and she found that 37% of the participants could detect the difference between the products.

1. Calculate $p_c$, $p_d$ and $d'$ for both tests using the `rescale`-function in sensR
2. Which of the two tests indicate the largest product difference?

**Answer to the exercise:**

1. In test 1 there were 73% correct answers, so $p_c = 0.73$ in a Triangle test. Using `rescale`, we can find the corresponding values of $p_d$ and $d'$:

```
rescale(pc=0.73, method="triangle")

##
## Estimates for the triangle protocol:
##     pc    pd  d.prime
## 1 0.73 0.595 2.676407
```

If 37% can detect the difference, it means that the proportion of discriminators is $p_d = 0.37$. The corresponding values of $p_c$ and $d'$ are:

```
rescale(pd=0.37, method="triangle")

##
## Estimates for the triangle protocol:
##     pc   pd  d.prime
## 1 0.58 0.37 1.874403
```

2. Since $d'$ in test 1 is 2.68, but only 1.87 in test 2, test 1 indicates the largest product difference.

# Exercise 4

You are reading about a sensory difference test of two products A and B in a report from another group in your company. They obtained 70% correct answers in a 2-AFC test and concluded that there was a large difference between the products.

1. Do you agree with their finding? (Hint: compute $d'$ using `rescale` and interpret how big it is)
2. Later in the report you read about another difference test between product A and a new product C. Here the group used a Duo-Trio test and obtained only 60% correct answers. Consequently the group concludes that the sensory difference between A and C is considerably smaller than between A and B. Do you agree with their conclusion?

**Answer to the exercise:**

1. First we compute $d'$:

```
rescale(pc=0.7, method="twoAFC")

##
## Estimates for the twoAFC protocol:
##    pc  pd   d.prime
## 1 0.7 0.4 0.7416143
```

so $d' = 0.74$ which is rather small, so we disagree that there is a large difference between the products.

2. First we compute $d'$ from the second difference test:

```
rescale(pc=.6, method="duotrio")

##
## Estimates for the duotrio protocol:
##    pc  pd d.prime
## 1 0.6 0.2 1.11521
```

Here $d' = 1.12$, which is larger than from the 2-AFC test, so contrary to the conclusions from the report, the difference between A and C is probably larger than the difference between A and B, though the difference between the two $d'$s is not that large.

# Exercise 5

A company has contacted you and offers to provide an ingredient you need for your products but at a lower price than your current supplier. You are worried, however, that using the new supplier will result in important changes to your product, so you decide to test if there is a difference with the Triangle test before adopting the new supplier.

1. Write up the null and alternative hypotheses: Express them in words and subsequently express them in terms of the proportion of correct answers, the proportion of discriminators and in terms of $d'$.

2. You planned and conducted the experiment and obtained 82 correct answers out of 200 consumer evaluations in a Triangle test. What is the probability of obtaining 82 or more correct answers out of 200 if there is no difference between the products? Does that provide evidence in favour of a difference between the products?

3. Is it plausible that the product difference is as large as $d' = 2$?

4. Is it plausible that as little as 2% of consumers can detect the difference between the products?

**Answer to the exercise:**

1. The null hypothesis is that the products are not different and the alternative hypothesis is that the products are different. The null hypothesis can be expressed as

$$H_0: \; p_c = 1/3, \quad p_d = 0 \quad d' = 0$$

and the alternative hypothesis can be expressed as

$$H_0: \; p_c > 1/3, \quad p_d > 0 \quad d' > 0$$

2. The probability of obtaining at least 82 correct answers out of 200 is exactly the $p$-value for the test. This $p$-value is produced by the `discrim` function:

```
discrim(82, 200, method="triangle")

##
## Estimates for the triangle discrimination protocol with 82 correct
## answers in 200 trials. One-sided p-value and 95 % two-sided confidence
## intervals are based on the 'exact' binomial test.
##
##          Estimate Std. Error  Lower  Upper
## pc         0.4100    0.03478 0.3411 0.4816
## pd         0.1150    0.05217 0.0117 0.2224
## d-prime    0.9475    0.23146 0.2923 1.3684
##
## Result of difference test:
## 'exact' binomial test:  p-value = 0.01401
## Alternative hypothesis: d-prime is greater than 0
```

   The probability of obtaining 82 or more correct answers out of 200 if there is no difference between the products is therefore $p = 0.014$. Since the $p$-value is less than 5%, it provides evidence that the products are different.

3. The plausible range of values for $d'$ given the data is provided by the confidence interval. The 95% CI for $d'$ is (0.29; 1.37) and since $d' = 2$ is outside that interval it is *not* a plausible value for $d'$ in light of these data.

4. The 95% CI for $p_d$ is (0.01; 0.22). Since 0.02 is included in this interval, it is in fact plausible that as little as 2% of the consumers can detect the difference between the products.

# Exercise 6

You are asked by the management in your company to conduct a 2-AFC test with n=50 and alpha=5%. You expect the products are easily confused with a $d'$ around 0.5 to 0.7.

1. Use the `d.primePwr` function to find out what the probability of detecting a difference around 0.5 and 0.7 is (i.e. estimate the power of this test).

2. Are 50 subjects enough to detect a difference of these magnitudes with a reasonable power (around 80%)?

3. Management informs that it is important to find out if these products really are different, so you are asked to conduct the experiment with as many subjects as you need. Management requires that you have at least 90% power in detecting the difference. Use the `d.primeSS` function to compute the number of subjects that you need (alpha = 5% and $d'$ between 0.5 and 0.7)

4. Compute both the first-exact and stable-exact sample sizes. How many samples are enough in the most optimistic case? How many would you need in the most cautious case? How many samples do you suggest to management that you use?

**Answer to the exercise:**

1. Using `d.primePwr`, we find that

```
d.primePwr(d.primeA=0.5, sample.size=50, alpha=0.05, method="twoAFC")

## [1] 0.5530559

d.primePwr(d.primeA=0.7, sample.size=50, alpha=0.05, method="twoAFC")

## [1] 0.8201104
```

   So power is around 55% if $d'$ is 0.5 and around 82% if $d'$ is 0.7

2. 50 subjects are enough if $d'$ is 0.7, but if it is as low as 0.5, 50 is far from enough. The probability of finding a difference (i.e. the power) depends heavily on the size of the difference we want to detect.

3. 
```
d.primeSS(d.primeA=0.7, target.power=0.9, alpha=0.05, method="twoAFC")

## [1] 60

d.primeSS(d.primeA=0.5, target.power=0.9, alpha=0.05, method="twoAFC")

## [1] 113
```

   So 60 subjects is enough if $d'$ is 0.7, but we need 113 if $d'$ is only 0.5 to have a power of 90%.

4. To compute both the first-exact and the stable-exact sample sizes, we need to use the `statistic = "both"` argument to `d.primeSS`:

```
d.primeSS(d.primeA=0.7, target.power=0.9, alpha=0.05, method="twoAFC",
          statistic="both")

## [1] 60 65

d.primeSS(d.primeA=0.5, target.power=0.9, alpha=0.05, method="twoAFC",
          statistic="both")

## [1] 113 122
```

   In the most optimistic case, we only need 60 samples — in the most cautious case we need twice as many.

# Exercise 7 (difficult)

Your boss has asked you to carry out a 3-AFC test with 30 observations to determine if there is a difference between two slightly different recipes of the same product. Your company could save money by using the cheaper recipe, but you don't want to send a different product on the market.

1. What is the value of $d'$ under the null and alternative hypotheses? Adopt a significance level of 5% and work with a scenario where the difference between the products is $d' = 0.8$.
2. What is the power of this test?
3. What is the probability of
   (a) Correctly claiming that the products are different?
   (b) Wrongly accepting that the products are not different?
   (c) Wrongly claiming that the products are different?
   (d) Correctly accepting that the products are not different?
4. Use the `d.primePwr` function and experiment with different values of `d.primeA` to find out which value of $d'$ you would be able to detect with a probability of 80%.

**Answer to the exercise:**

1. We want to perform a sensory difference test. Our null hypothesis is that there is no difference between the products and our alternative hypothesis is that there is a difference between the products corresponding to $d' = 0.8$. We therefore have:

$$H_0 : \quad d' = 0, \qquad\qquad H_A : \quad d' = 0.8$$

2. The power of this test can be computed with

```
d.primePwr(d.primeA=0.8, d.prime0=0, sample.size=30, alpha=0.05,
           method="threeAFC")

## [1] 0.8436441
```

   So the power of this test is 84%. This means that the probability that the hypothesis tet will yield a $p$-value less than 5% is 0.84 if the true $d'$ is in fact 0.8.

3. The four different decisions we can make are summarized in table 1 below.

Table 1: Decisions in hypothesis testing.

| | Decision | |
|---|---|---|
| | Accept $H_0$ | Reject $H_0$ |
| $H_0$ is true | ✓ | Type I error ($\alpha$) |
| $H_A$ is true | Type II error ($\beta$) | ✓ |

(a) Power is the probability of finding a difference, if it is really there, that is, if $H_A$ is true. In other words, the power of the test is the probability of correctly claiming a product difference. In this case the probability of correctly claiming a product difference is 84%.

(b) If we accept that products are not different, we accept $H_0$. When we are wrong in accepting $H_0$, it means that we are committing a Type II error, cf. Table 1. Since power $= 1 - \beta = 0.84$, then $\beta = 1 - 0.84 = 0.16$. In conclusion: the probability of wrongly accepting that products are not different is 16%.

(c) If we are claiming that products are different, we are believing in $H_A$ and therefore rejecting $H_0$. When we do so wrongly, it means that we are committing a Type I error, cf. Table 1. We specified $\alpha = 0.05$ so the probability of committing a Type I error is at most 0.05. The critical value for our test is

```
findcr(sample.size=30, alpha=0.05, p0=1/3)

## [1] 15
```

and the probability of obtaining at least 15 correct answers out of 30 if $H_0$ is true (i.e. $p_c = 1/3$) is $Pr(X \geq 15) = 1 - Pr(X \leq 14)$, which we compute with

```
1 - pbinom(14, size=30, prob=1/3)

## [1] 0.04348228
```

so the actual probability of a Type I error is 4.35%, just below the 5% we specified. In conclusion the probability of wrongly claiming that the products are different is 4.35%.

(d) If we are accepting that the products are not different, we are believing in $H_0$, and if we are correct in doing so, then $H_0$ must be true. The probability of correctly accepting that the products are not different is therefore $1 - \alpha = 1 - 0.0435 = 0.9565$.

4. After some experimentation, we find that we have 80% power of detecting a difference of $d' = 0.75$ with 30 subjects in a 3-AFC test:

```
d.primePwr(d.primeA=.75, sample.size=30, method="threeAFC")

## [1] 0.7994836
```