Statistical methods in the opair package

Rune Haubo B Christensen

October 6, 2013
file: `stat_methods.Rnw`

Contents

1 A sensory discrimination protocol for ordinal paired comparisons

This document describes the Thurstonian model for a sensory discrimination method for paired comparisons that leads to ordinal ratings. The `opair` package implements these methods and makes them available for d' estimation and further analyses. An accompanying tutorial shows how various functions in the `opair` package can be applied to analyze ordinal paired comparisons.

In this protocol, two products are considered; a test product and a reference product. In each trial an assessor receives two samples, either two reference samples (a placebo pair or placebo trial) or a reference sample and a test sample (a test pair or a test trial). One sample is presented to the assessor as 'reference', and the other is presented 'test' whether or not it is a test trial or a placebo trial. The assessor is asked to compare the test sample with the reference sample with respect to some attribute, for example sweetness. The assessor is then asked to provide a rating response as the result of the comparison. Note that two test samples are not compared in this version of the protocol.

The question posed to the assessor could be something like: "How would you characterize the (sensory) intensity of attribute X in the test sample compared to the reference sample?"

1.1 Response scale

An assessor can evaluate the comparison of the test sample with the reference sample with respect to some attribute on a *symmetric* and *directional* degree of difference scale ranging from, for example, *much less* to *much more* with a number of intermediate categories as exemplified in Figure˜1.

much more	more	equal	less	much less
☐	☐	☐	☐	☐

Figure 1: Response scale for ordinal paired comparisons.

1.2 Decision rule

It is assumed that an assessor adopts a set of $J - 1$ thresholds (τ_j, $j = 1, \ldots, J - 1$), where J is the number of response categories. The assessor will answer $Y = j$ if the intensity of the test sample t relative to the reference sample r is between τ_{j-1} and τ_j, where we assume that $\tau_0 = -\infty$ and $\tau_J = \infty$.

1.3 Characterization of the protocol

The protocol can be characterized by the following aspects:

- This discrimination protocol can be used with an equal number of response categories (Forced Choice) or with an uneven number of categories ('same' or 'equal' is a possible response).

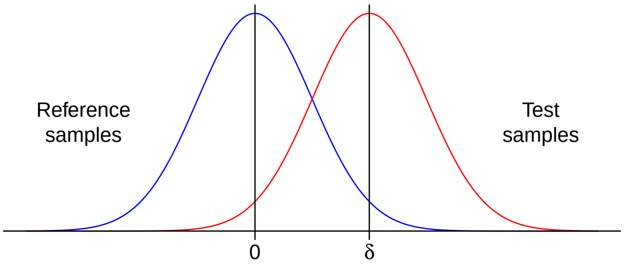

Figure 2: Thurstonian distributions

- The protocol is attribute specific and therefore an alternative to 2-AFC, 3-AFC, 2-AC, specified Tetrads etc.
- The protocol results in rating responses just like the A-not A with sureness protocol and the degree of difference protocol.

2 Thurstonian model for ordinal paired comparisons

The following exposition will assume a rating scale with five categories as in Figure˜1, and we will assume observations have been observed in each of the five categories for either placebo pairs or test pairs. The basic Thurstonian model where normal distributions represent the perceptual intensity of reference and test samples is shown Figure˜2.

The Thurstonian model for the ordinal paired comparison protocol is illustrated in Figure˜3 for placebo pairs and test pairs respectively. In this model, reference and test products are assumed to be normally distributed as

$$R \sim N(0,1) \qquad T \sim N(\delta,1)$$

and hence the differences $T - R$ and $R - R$ are then distributed as

$$(T - R) \sim N(\delta,2) \qquad (R - R) \sim N(0,2)$$

as shown in Figure˜3. Here we assume independence of the R and T random variables. The thresholds, τ are ordered and increasing:

$$\tau_1 < \tau_2 < \tau_3 < \tau_4$$

Without further restrictions imposed on the model, a total of five parameters (four thresholds and δ) are estimated from a 2×5 frequency table.

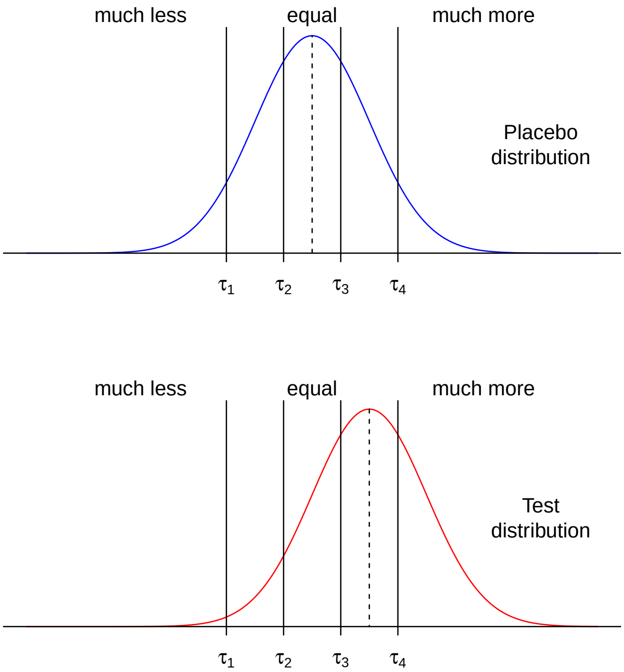

Figure 3: Difference distributions (RR: placebo and TR: trial) for the Thurstonian model for the ordinal paired comparison protocol.

3 Estimation of the Thurstonian model

The probability that the response for RR and TR trials fall in or below the jth category can be written as

$$P(Y \leq j|\text{pair} = RR) = \text{˜}\Phi\left(\frac{-\tau_j}{\sqrt{2}}\right)$$
$$P(Y \leq j|\text{pair} = TR) = \text{˜}\Phi\left(\frac{-\tau_j - \delta}{\sqrt{2}}\right)$$

We can write this as

$$\gamma_{ij} = \Phi\left(\frac{-\tau_j - \delta \cdot x_i}{\sqrt{2}}\right) \tag{1}$$

where x_i is a dummy variable being zero for RR trials and one for TR trials; $\gamma_{ij} = P(Y \leq j|x_i)$ is the cumulative probability conditional on x_i.

Writing model˜(1) as

$$\gamma_{ij} = \Phi(\theta_j - \beta \cdot x_i)$$

where $\theta_j = -\tau_j/\sqrt{2}$ and $\beta = \delta/\sqrt{2}$ clarifies that the model˜(1) has the form of a cumulative link model (CLM) (McCullagh, 1980; Agresti, 2002; Christensen, 2012b,a; Christensen and Brockhoff, 2013). The Thurstonian model for the ordinal paired comparison protocol can therefore be estimated as a CLM and d' (used here to denote the estimator of δ) can be estimated with

$$d' = \sqrt{2} \cdot \hat{\beta}$$

Similarly the standard error of d' is given by

$$\text{se}(d') = \sqrt{2} \cdot \text{se}(\hat{\beta})$$

The Thurstonian model outlined above is not the only possible Thurstonian model for this kind of data. One possible additional assumption could be that the thresholds are symmetric such that $\tau_1 = -\tau_4$ and $\tau_2 = -\tau_3$ in the version with five response categories. A likelihood ratio test could be used to assess how reasonable this assumption is.

One advantage of this assumption is that the resulting Thurstonian model would not need placebo trials to estimate d' as the version outlined above does. The three-category version of this model would be identical to the Thurstonian model for the 2-AC protocol described in Christensen et˜al. (2012).

4 Hypothesis tests

The hypotheses of the conventional difference test are

$$H_0 : \text{˜}d' = 0 \qquad \text{versus} \qquad H_A : \text{˜}d' \neq 0$$

The p-value for this test can be computed with

$$p = 2(1 - \Phi(|w_0|))$$

where $w_0 = d'/\text{se}(d')$ is the Wald statistic.

The corresponding $(100 - \alpha)\%$ confidence interval is given by

$$d' \pm z_{1-\alpha/2} \cdot \text{se}(d')$$

An equivalence test involves the following hypotheses

$$\begin{aligned} H_0 &: \tilde{}\, d' < -d'_0 \quad \text{or} \quad d' > d'_0 \\ H_A &: \tilde{}\, -d'_0 \leq d' \leq d'_0 \end{aligned}$$

where d'_0 defines the equivalence region.

Using the two one-sided tests procedure by Schuirmann (1981, 1987), the test can be conducted by performing two separate one-sided tests:

$$\begin{aligned} H_{0a} &: \tilde{}\, d' < -d'_0 \qquad \text{versus} \qquad H_{Aa} : \tilde{}\, d' > -d'_0 \\ H_{0b} &: \tilde{}\, d' > d'_0 \qquad \text{versus} \qquad H_{Ab} : \tilde{}\, d' < d'_0 \end{aligned}$$

The overall p-value is then the largest of the p-values for the two separate one-sided tests.

The Wald test statistic and p-values are computed as

$$\begin{aligned} p_a &= 1 - \Phi(w_{0a}) \quad \text{where} \quad w_{0a} = \frac{d' - (-d'_0)}{\text{se}(d')} \\ p_b &= \Phi(w_{0b}) \quad \text{where} \quad w_{0b} = \frac{d' - d'_0}{\text{se}(d')} \end{aligned}$$

Here large values of w_{0a} results in small p-values and small values of w_{0b} result in small p-values.

5 Examples of theory in action

In this section, the theory outlined in previous sections are applied to the NV data set using R. The NV data set is available in the opair package. The following examples are meant to illustrate how the computations in the `opair` function can be carried out 'by hand'.

First we load the opair package to get access to the NV data set:

```
R> library(opair)
```

By fitting a cumulative link model (using `clm` from the ordinal package), we can extract d' and its standard error. Here we use the attribute `Thickness`:

```
R> library(ordinal)
R> fm1 <- clm(factor(Thickness) ~ Samples, link="probit", data=NV)
R> summary(fm1)

formula: factor(Thickness) ~ Samples
data:    NV

 link   threshold nobs logLik AIC   niter max.grad cond.H
 probit flexible  33   -42.81 97.61 4(0)  7.51e-09 2.2e+01

Coefficients:
```

```
           Estimate Std. Error z value Pr(>|z|)
Samples568   0.1941     0.4544   0.427   0.6693
Samples841   1.3383     0.4861   2.753   0.0059 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
      Estimate Std. Error z value
-2|-1  -1.6348     0.5116  -3.195
-1|0   -0.3268     0.3515  -0.930
0|1     0.8342     0.3618   2.306
1|2     1.7271     0.4311   4.006
```

d's and standard errors for the two products 568 and 841 relative to the reference 432 are then

```
R> b <- coef(summary(fm1))
R> (b2 <- b[5:6, 1:2] * sqrt(2))
```

```
            Estimate Std. Error
Samples568 0.2745231  0.6426666
Samples841 1.8926681  0.6874585
```

5.1 Computing confidence intervals

The 95% CI can be computed with

```
R> alpha <- .05
R> z <- qnorm(1 - alpha/2)
R> ## lower and upper for prod 568:
R> b2[1, 1] +  c(-z, z) * b2[1, 2]
```

```
[1] -0.9850803  1.5341266
```

```
R> ## lower and upper for prod 841:
R> b2[2, 1] +  c(-z, z) * b2[2, 2]
```

```
[1] 0.5452743 3.2400620
```

```
R> ## Both:
R> b2[, 1] +  b2[, 2] %*% t(c(-z, z))
```

```
           [,1]     [,2]
[1,] -0.9850803 1.534127
[2,]  0.5452743 3.240062
```

5.2 Difference test

To compute the p-value of the difference test, first compute the Wald statistic:

```
R> (W <- b2[, 1] / b2[, 2])
```

```
Samples568 Samples841
 0.4271625  2.7531381
```

Now the p-value is given by

```
R> 2 * pnorm(abs(W), lower.tail=FALSE)

 Samples568  Samples841
0.669260942 0.005902699
```

5.3 Equivalence test

First choose a value for d'_0, e.g.~1.5:

```
R> dp0 <- 1.5
```

The Wald statistic and p-value for the a-hypothesis is then:

```
R> (Wa <- (b2[1,1] + dp0)/b2[1, 2])

[1] 2.761188

R> (pa <- pnorm(Wa, lower.tail=FALSE))

[1] 0.00287958
```

and similarly the Wald statistic and p-value of the b-hypothesis is

```
R> (Wb <- (b2[1,1] - dp0)/b2[1, 2])

[1] -1.906862

R> (pb <- pnorm(Wb))

[1] 0.0282692
```

The resulting p-value for the equivalence test is then the largest of the two p-values:

```
R> (pval <- max(pa, pb))

[1] 0.0282692
```

References

Agresti, A. (2002). *Categorical Data Analysis* (Second ed.). Wiley.

Christensen, R. H.~B. (2012a). *Analysis of cumulative link models — estimation with the R package ordinal*. Package vignette for ordinal version 2012.01-19.

Christensen, R. H.~B. (2012b). ordinal—regression models for ordinal data. R package version 2012.09-12 http://www.cran.r-project.org/package=ordinal/.

Christensen, R. H.~B. and P.~B. Brockhoff (2013). Analysis of replicated categorical ratings data from sensory experiments. *Journal of the French Statistical Society, Accepted*.

Christensen, R. H.~B., H.-S. Lee, and P.~B. Brockhoff (2012). Estimation of the Thurstonian model for the 2-AC protocol. *Food Quality and Preference~4*, 119–128.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B~42*, 109–142.

Schuirmann, D.~J. (1981). On hypothesis testing to determine if the mean of a normal distribution is contained in a known interval. *Biometrics~37*, 617.

Schuirmann, D.~J. (1987). A comparison of the two one-sided tests procedures and the power approach for assessing the equivalence of coverage bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics~15*, 657–680.

A SessionInfo

```
R> sessionInfo()

R version 3.0.2 Patched (2013-10-04 r64027)
Platform: x86_64-unknown-linux-gnu (64-bit)

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=C
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods
[7] base

other attached packages:
[1] ordinal_2013.9-30 opair_0.2-0

loaded via a namespace (and not attached):
[1] MASS_7.3-29     Matrix_1.0-14   grid_3.0.2      lattice_0.20-23
[5] tools_3.0.2     ucminf_1.1-3
```