P0001_checkup

1 Imports

```
library(devtools)
library(tidyverse)
library(readr)
library(readr)

load_all(
  path="~/cc/dev/c2023a/c0512_apaTablesCoCoR/apaTablesCoCoR",
  reset=TRUE,
  recompile=TRUE,
  export_all=TRUE,
  helpers=TRUE,
  quiet=FALSE)

options("width" = 222)
options(cli.width=222)
paste(getwd())
```

```
/home/jiko/cc/dev/c2023a/c0501_bertagent_devel/code/p0001_valid-01-part-008-correlations
```

2 Load data

```
dir0 <- "../../data/d0042_gold-standard"
if0 <- "gold-std-x0002_20230513T203848_bertagent-clean.csv"
if0 <- "gold-std-x0002_bertagent-clean.csv"
df0 <- read_csv(file=paste(dir0, if0, sep="/"))
# df0 = read.csv(file=paste(dir0, if0, sep="/"))
paste(str(df0))
```

```
indexing gold-std-x0002_bertagent-clean.csv [=============================================

Rows: 300 Columns: 24
  Column specification
Delimiter: ","
```

```
chr  (4): sents, text, ba0, ba4
dbl (20): idx0, SENT, HumEvalN, HumEvalSD, HumEval, PietA, PietB, PietC, NicoPos, NicoNeg,

 Use `spec()` to retrieve the full column specification for this data.
 Specify the column types or set `show_col_types = FALSE` to quiet this message.
spc_tbl_ [300 × 24] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ idx0       : num [1:300] 0 1 2 3 4 5 6 7 8 9 ...
 $ SENT       : num [1:300] 1 2 3 4 5 6 7 8 9 10 ...
 $ HumEvalN   : num [1:300] 30 29 30 31 30 29 30 31 29 31 ...
 $ HumEvalSD  : num [1:300] 1.196 0.953 1.326 0.653 1.224 ...
 $ HumEval    : num [1:300] -0.867 -2.138 -1.633 -2.323 -2.133 ...
 $ PietA      : num [1:300] 0 0 0 0.1667 0.0667 ...
 $ PietB      : num [1:300] 0.0714 0 0 0 0 ...
 $ PietC      : num [1:300] 0 0 0 0.1667 0.0667 ...
 $ NicoPos    : num [1:300] 0 0 0 0 0 ...
 $ NicoNeg    : num [1:300] 0 0.0909 0 0 0 ...
 $ NicoCom    : num [1:300] 0 -0.0909 0 0 0 ...
 $ sents      : chr [1:300] "['This has led to frustration and anger after the many years
 $ text       : chr [1:300] "This has led to frustration and anger after the many years of
 $ sents_count: num [1:300] 1 1 1 1 1 1 1 1 1 1 ...
 $ ba0        : chr [1:300] "[0.06802511215209961]" "[-0.8474809527397156]" "[-0.647848486
 $ ba4        : chr [1:300] "[0.04629361256957054]" "[-0.5952255725860596]" "[-0.405694216
 $ baTot_sum  : num [1:300] 0.0463 -0.5952 -0.4057 0.1863 0.0891 ...
 $ baPos_sum  : num [1:300] 0.0463 0 0 0.1863 0.0891 ...
 $ baNeg_sum  : num [1:300] 0 0.595 0.406 0 0 ...
 $ baAbs_sum  : num [1:300] 0.0463 0.5952 0.4057 0.1863 0.0891 ...
 $ BAPos      : num [1:300] 0.0463 0 0 0.1863 0.0891 ...
 $ BANeg      : num [1:300] 0 0.595 0.406 0 0 ...
 $ BATot      : num [1:300] 0.0463 -0.5952 -0.4057 0.1863 0.0891 ...
 $ BAAbs      : num [1:300] 0.0463 0.5952 0.4057 0.1863 0.0891 ...
 - attr(*, "spec")=
  .. cols(
  ..   idx0 = col_double(),
  ..   SENT = col_double(),
  ..   HumEvalN = col_double(),
  ..   HumEvalSD = col_double(),
  ..   HumEval = col_double(),
  ..   PietA = col_double(),
  ..   PietB = col_double(),
  ..   PietC = col_double(),
  ..   NicoPos = col_double(),
  ..   NicoNeg = col_double(),
  ..   NicoCom = col_double(),
  ..   sents = col_character(),
  ..   text = col_character(),
  ..   sents_count = col_double(),
  ..   ba0 = col_character(),
  ..   ba4 = col_character(),
  ..   baTot_sum = col_double(),
  ..   baPos_sum = col_double(),
```

```
  ..   baNeg_sum = col_double(),
  ..   baAbs_sum = col_double(),
  ..   BAPos = col_double(),
  ..   BANeg = col_double(),
  ..   BATot = col_double(),
  ..   BAAbs = col_double()
  .. )
 - attr(*, "problems")=<externalptr>
character(0)
```

3 Check correlations

- Diedenhofen and Musch (2015)
- Diedenhofen (2022)
- https://f-santos.gitlab.io/2020-04-01-comparing-correlation-coefficients.html

```
df2 <- df0 %>% select(
    HumEval,
    PietA,
    PietB,
    PietC,
    NicoPos,
    NicoNeg,
    NicoCom,
    # BAOTot,
    # BAOPos,
    # BAONeg,
    # BAOAbs,
    BATot,
    BAPos,
    BANeg,
    BAAbs,
)

devtools::unload(package="apaTables", quiet=FALSE)
load_all(path="~/cc/dev/c2023a/c0512_apaTablesCoCoR/apaTablesCoCoR",reset=TRUE,recompile=TRUE,export
tab0 <- apa.cocor.table(
  df2,
  filename="p0001_gold-standard-correlations.doc",
  table.number=1,
  common="HumEval"
)
print(tab0)
```

```
 Loading apaTables
```

```
Table 1
```

Means, standard deviations, and correlations with confidence intervals

Variable	M	SD	1	2	3	4	5	6
1. HumEval	0.12	1.54						
2. PietA	0.05	0.05	.17** [.06, .28]		-1.25	0.28	0.05	5.35
3. PietB	0.02	0.03	.25** [.14, .35]	.40** [.30, .49]		1.27	1.16	6.58
4. PietC	0.05	0.05	.17** [.06, .28]	.99** [.99, 1.00]	.40** [.30, .49]		0.03	5.34
5. NicoPos	0.03	0.04	.17** [.05, .27]	.18** [.07, .29]	.23** [.12, .34]	.17** [.06, .28]		5.49
6. NicoNeg	0.01	0.03	-.28** [-.38, -.17]	-.10 [-.21, .01]	-.01 [-.12, .11]	-.10 [-.21, .02]	-.03 [-.14, .09]	
7. NicoCom	0.02	0.05	.30** [.19, .40]	.20** [.09, .31]	.19** [.08, .30]	.19** [.08, .30]	.82** [.78, .85]	-.60 [-.6
8. BATot	0.09	0.35	.78** [.73, .82]	.21** [.10, .31]	.24** [.13, .34]	.20** [.09, .31]	.22** [.11, .33]	-.42 [-.5
9. BAPos	0.19	0.21	.72** [.66, .77]	.22** [.11, .32]	.34** [.23, .43]	.21** [.10, .32]	.31** [.20, .41]	-.16 [-.2
10. BANeg	0.10	0.19	-.61** [-.67, -.53]	-.14* [-.25, -.02]	-.06 [-.17, .06]	-.13* [-.24, -.02]	-.06 [-.17, .05]	.58* [.50
11. BAAbs	0.29	0.21	.17** [.06, .28]	.09 [-.02, .21]	.29** [.18, .39]	.09 [-.02, .20]	.26** [.15, .36]	.37* [.27

Note. M and SD are used to represent mean and standard deviation, respectively.
Values in square brackets indicate the 95% confidence interval.
The confidence interval is a plausible range of population correlations
that could have caused the sample correlation (Cumming, 2014).
 * indicates p < .05. ** indicates p < .01.

4 Manual checkups

```
df2 <- as.data.frame(df2)
cc0 <- cocor::cocor(~HumEval+BATot|HumEval+BAPos, data=df2)
cc0 <- cocor::cocor(~HumEval+BAAbs|HumEval+PietB, data=df2)
```

```r
cc0 <- cocor::cocor(~HumEval+PietB|HumEval+BAAbs, data=df2)
cc0 <- cocor::cocor(~HumEval+BATot|HumEval+PietA, data=df2, test=c("hittner2003"))

cc0
cc0@hittner2003$statistic
cc0@hittner2003$p.value
gtools::stars.pval(cc0@hittner2003$p.value)
```

```
  Results of a comparison of two overlapping correlations based on dependent groups

Comparison between r.jk (HumEval, BATot) = 0.778 and r.jh (HumEval, PietA) = 0.1699
Difference: r.jk - r.jh = 0.6081
Related correlation: r.kh = 0.2091
Data: df2: j = HumEval, k = BATot, h = PietA
Group size: n = 300
Null hypothesis: r.jk is equal to r.jh
Alternative hypothesis: r.jk is not equal to r.jh (two-sided)
Alpha: 0.05

hittner2003: Hittner, May, and Silver's (2003) modification of Dunn and Clark's z (1969) u
  z = 10.9466, p-value = 0.0000
  Null hypothesis rejected
[1] 10.9466
[1] 0
[1] "***"
attr(,"legend")
[1] "0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1"
```

References

Diedenhofen, B. (2022, June 28). *Cocor: Comparing Correlations* (Version 1.1-4). (Cit. on p. 3).

Diedenhofen, B., & Musch, J. (2015). Cocor: A Comprehensive Solution for the Statistical Comparison of Correlations (J. Olivier, Ed.). *PLOS ONE*, *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945 (cit. on p. 3)