A short introduction to circular data analysis

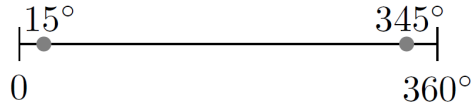
Jolien Cremers

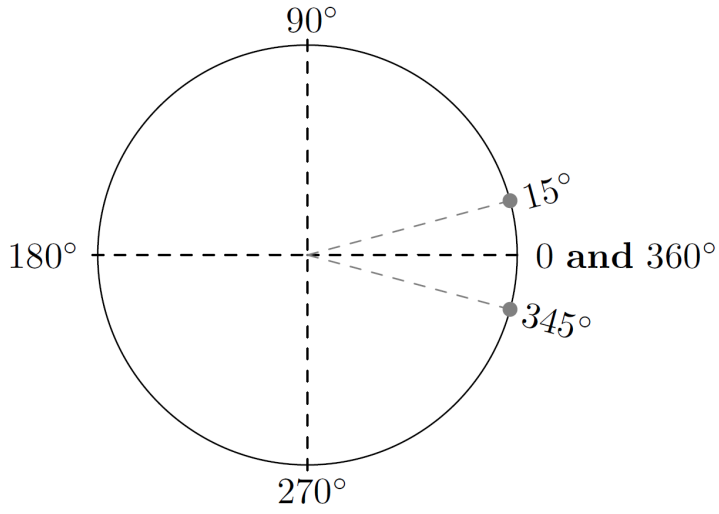
Section of Biostatistics, University of Copenhagen
Statistics Denmark
jcre@sund.ku.dk

February 8th, 2020

What is circular data?

A variable θ defined on $[0, 2\pi)$ radians or $[0, 360)$ degrees:





Three different types of circular data

- Directions in 2D space / Compass data
- Time of day / clock times
- Circular measurement scales

Directions in 2D space / compass data

Spatial cognition:

- Brunyé, T. T., Burte, H., Houck, L. A., & Taylor, H. A. (2015). The map in our head is not oriented north: Evidence from a real-world environment.
- Postma, A., Zuidhoek, S., Noordzij, M. L., & Kappers, A. M. (2008). Keep an eye on your hands: on the role of visual mechanisms in processing of haptic space.

Eye movements:

- Mulder, K., Klugkist, I., van Renswoude, D., & Visser, I. (2020). Mixtures of peaked power Batschelet distributions for circular data with application to saccade directions.

Time of day / clock times

Phase difference:

- Puglisi, G., Leonetti, A., Landau, A., Fornia, L., Cerri, G., & Borroni, P. (2017). The role of attention in human motor resonance.

Time of day:

- Rocchi, M. B., & Perlini, C. (2002). Is the time of suicide a random choice? A new statistical perspective.
- Brunsdon, C., & Corcoran, J. (2006). Using circular statistics to analyse time patterns in crime incidence.

Circadian rhythms:

- Peeters, F., Berkhof, J., Delespaul, P., Rottenberg, J., & Nicolson, N. A. (2006). Diurnal mood variation in major depressive disorder.

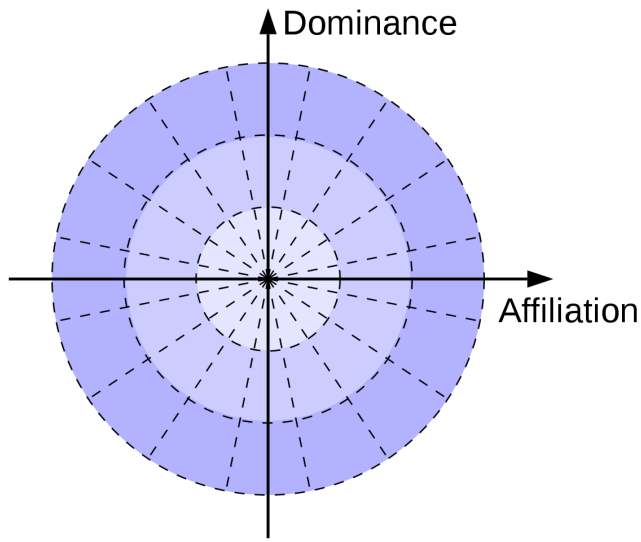
Circular measurement scales

Rose of Leary:

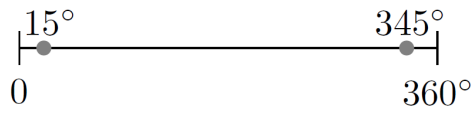
- Leary, T. (1958). Interpersonal diagnosis of personality.

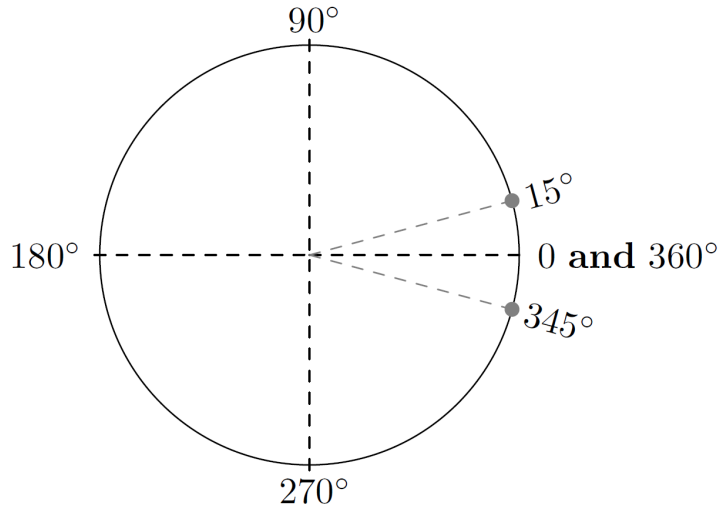
Interpersonal circumplex:

- Gurtman, M. B. (2009). Exploring personality with the interpersonal circumplex.

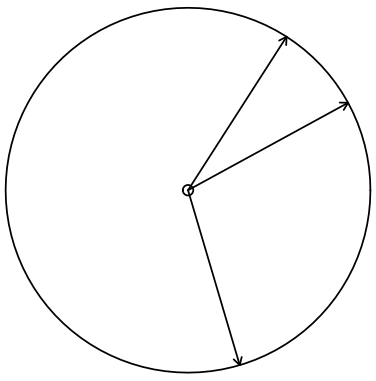


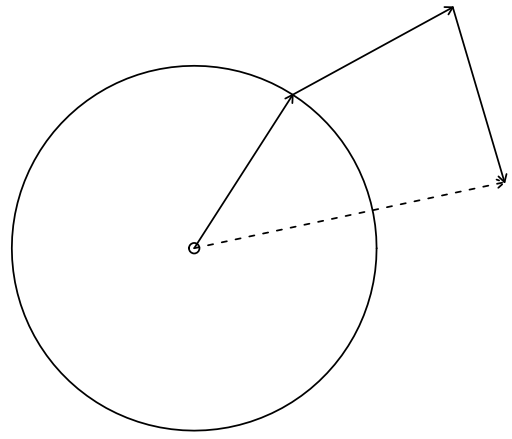
Why do we need different analysis methods?

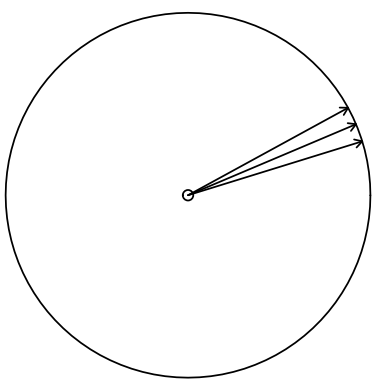


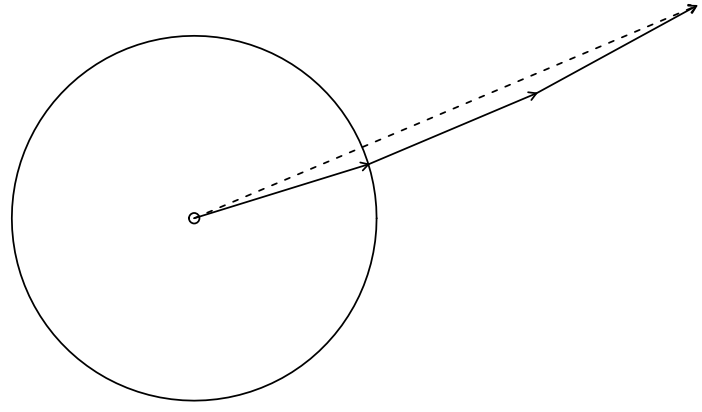


Circular summary statistics









Circular mean direction

$$\bar{\theta} = \begin{cases} \tan^{-1}\left(\frac{\sum_{i=1}^{n}\sin\theta_i}{\sum_{i=1}^{n}\cos\theta_i}\right) & \sum_{i=1}^{n}\sin\theta_i > 0, \sum_{i=1}^{n}\cos\theta_i > 0 \\ \tan^{-1}\left(\frac{\sum_{i=1}^{n}\sin\theta_i}{\sum_{i=1}^{n}\cos\theta_i}\right) + \pi & \sum_{i=1}^{n}\cos\theta_i < 0 \\ \tan^{-1}\left(\frac{\sum_{i=1}^{n}\sin\theta_i}{\sum_{i=1}^{n}\cos\theta_i}\right) + 2\pi & \sum_{i=1}^{n}\sin\theta_i < 0, \sum_{i=1}^{n}\cos\theta_i > 0 \end{cases}$$

Mean resultant length

$$R = \sqrt{\left(\sum_{i=1}^{n} \cos\theta_i\right)^2 + \left(\sum_{i=1}^{n} \sin\theta_i\right)^2}$$

$$\bar{R} = R/n$$

$1 =$ All datapoints have the same value

$0 \neq$ uniform, e.g. bimodality at exactly opposite parts of the circle (0°, 180°) also leads to a zero value.

Other summary statistics

- circular variance, $V = 1 - \bar{R}$
- circular standard deviation $v = -2\log(1 - V)^{0.5}$
- circular median direction
- circular modal direction

see Fisher (1995) for an overview.

Circular distributions

Three classes:

- Intrinsic distributions (e.g. von Mises)
- Wrapped distributions ($\theta \sim N(\mu, \sigma) \bmod 2\pi$)
- Projected distributions
  ($\theta = (\sin\theta, \cos\theta)^t = \boldsymbol{y} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})/||\boldsymbol{y}||$)

Intrinsic distributions

Directly defined on the circle, e.g.:

- von Mises distribution

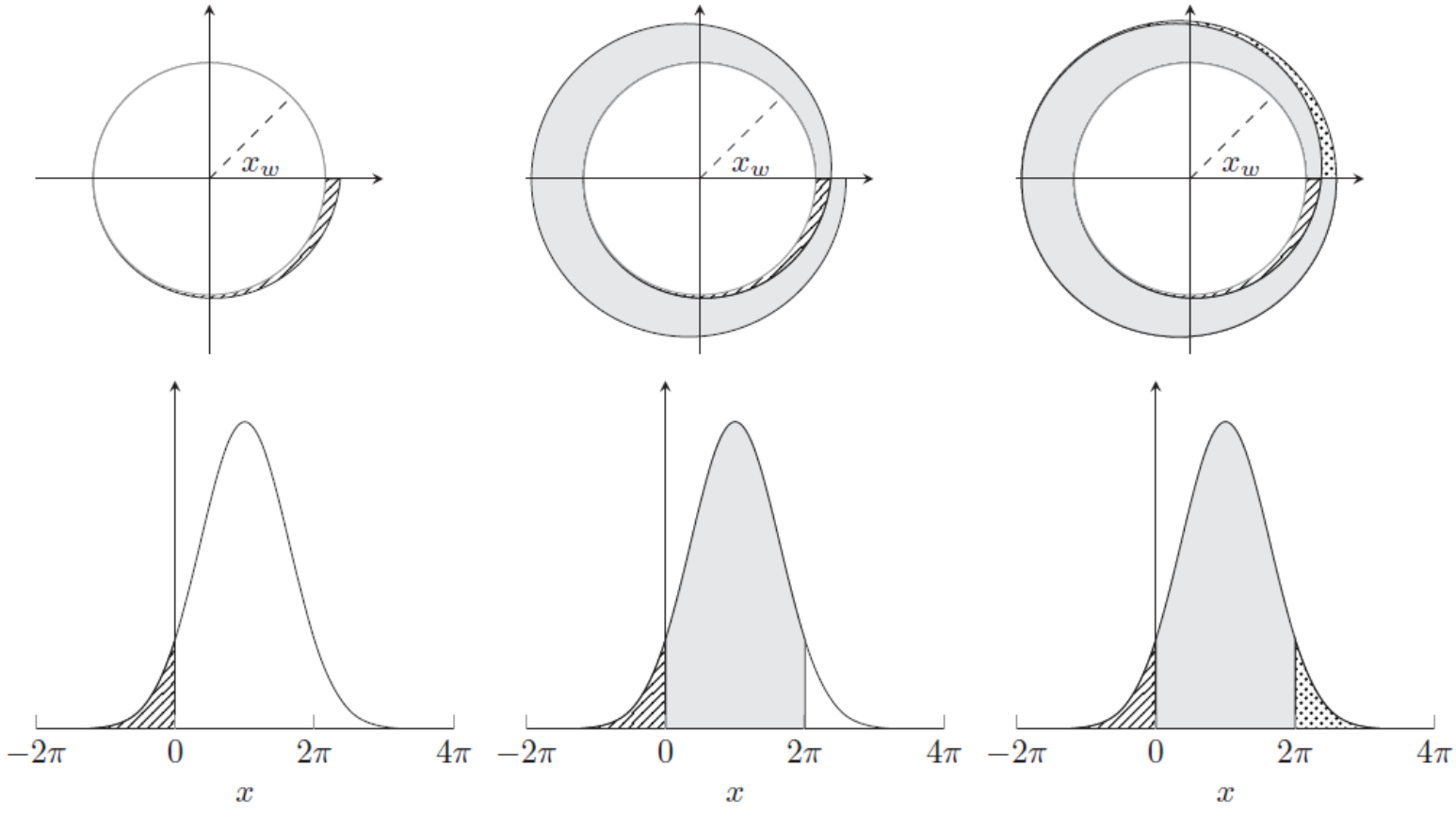
$$f(\theta) = [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(\theta - \mu)]$$

mean direction $= \mu$

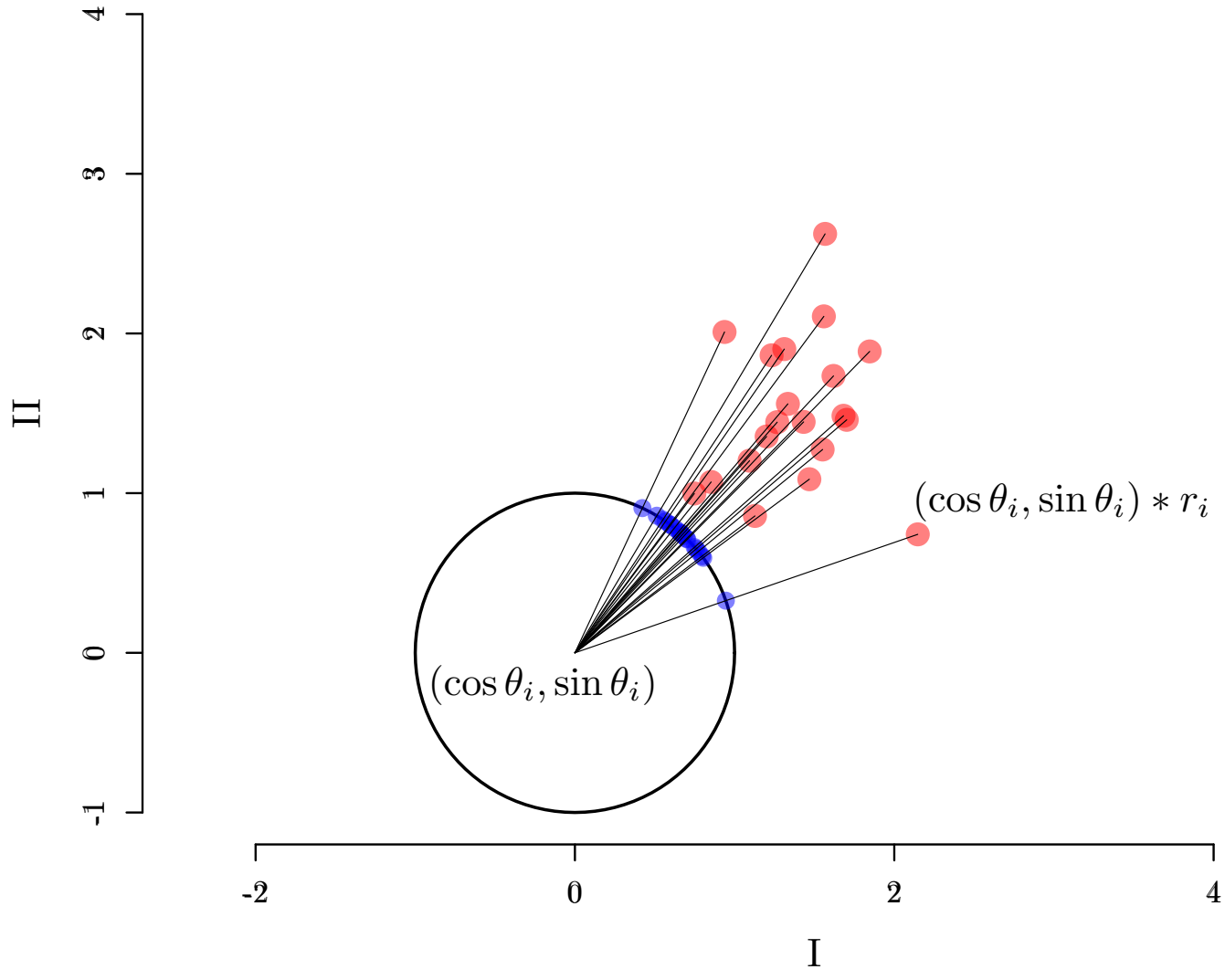
concentration parameter $= \kappa$

mean resultant length $= A_1(\kappa) = I_1(\kappa)/I_0(\kappa)$, see Fisher (1995)

Wrapped distributions



Projected distributions



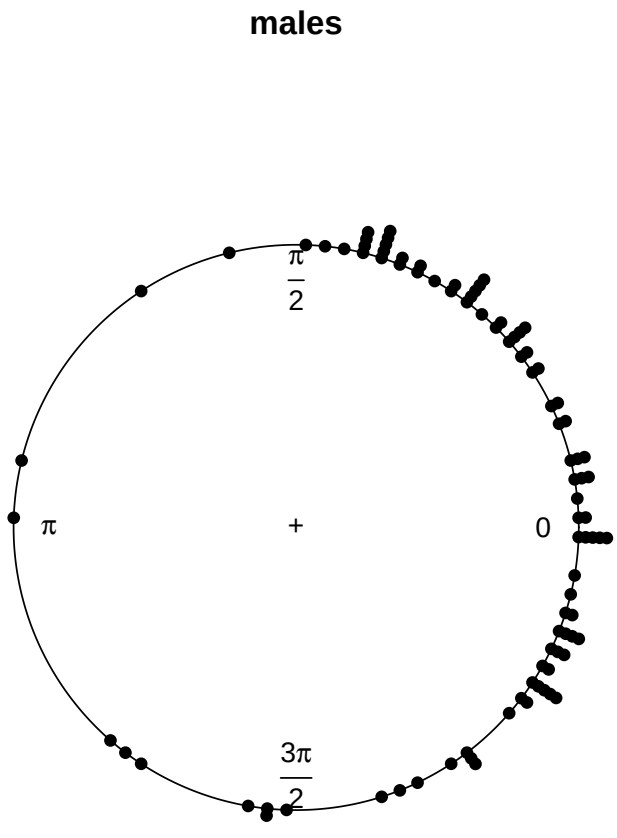
Projected distributions

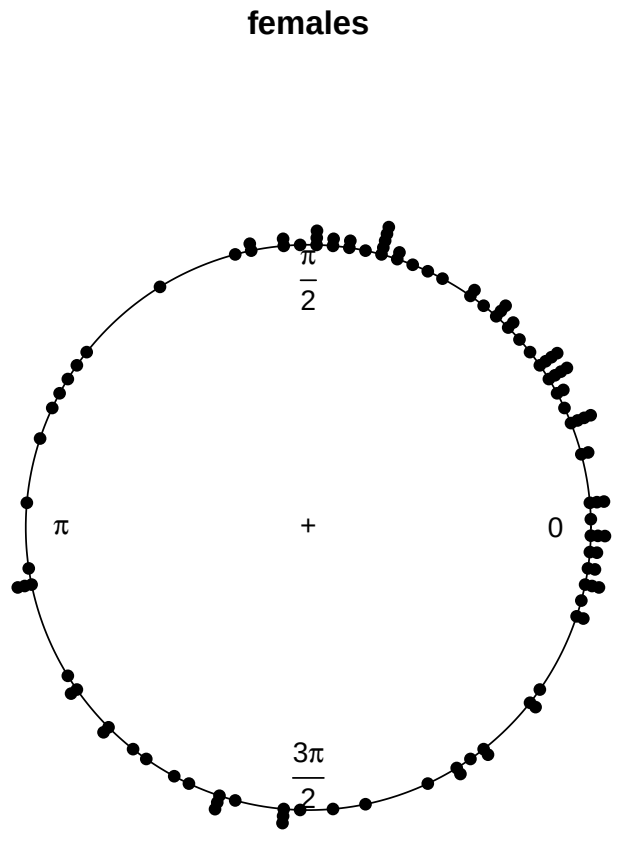
Instead of inference on the circular outcome, θ, directly we perform inference on underlying bivariate data, $\mathrm{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{I})$.

- Advantage: **flexible** approach, make use of already existing models for bivariate linear data.
- Disadvantage: Y and r are latent, need special inference methods (EM algorithm (Presnell, Morrison, & Littell, 1998) or **data augmented MCMC sampling** (Nunez-Antonio & Gutierrez-Pena, 2005)).
- Disadvantage: **interpretation** on two components instead of on circle.

A short example analysis using the R-package bpnreg

Brunyé, T. T., Burte, H., Houck, L. A., & Taylor, H. A. (2015). The map in our head is not oriented north: Evidence from a real-world environment. PLoS One, 10(9), e0135803.





Fit a circular regression model

```
fit <- bpnr(theta ~ sex + age + experience + sbsod,
            data = data,
            its = 10000, burn = 100,
            seed = 101)
```

Results

Estimated circular means for males and females with average age, experience and Santa Barbara Sense of Direction (SBSOD) score:

```
coef_circ(fit, type = "categorical", units = "degrees")
```

```
##                mean  mode   sd    LB    UB
## (Intercept) 10.92 12.52 7.29 -3.88 24.71
## sexfemale   32.04 35.23 9.90 12.42 51.39
```

Results

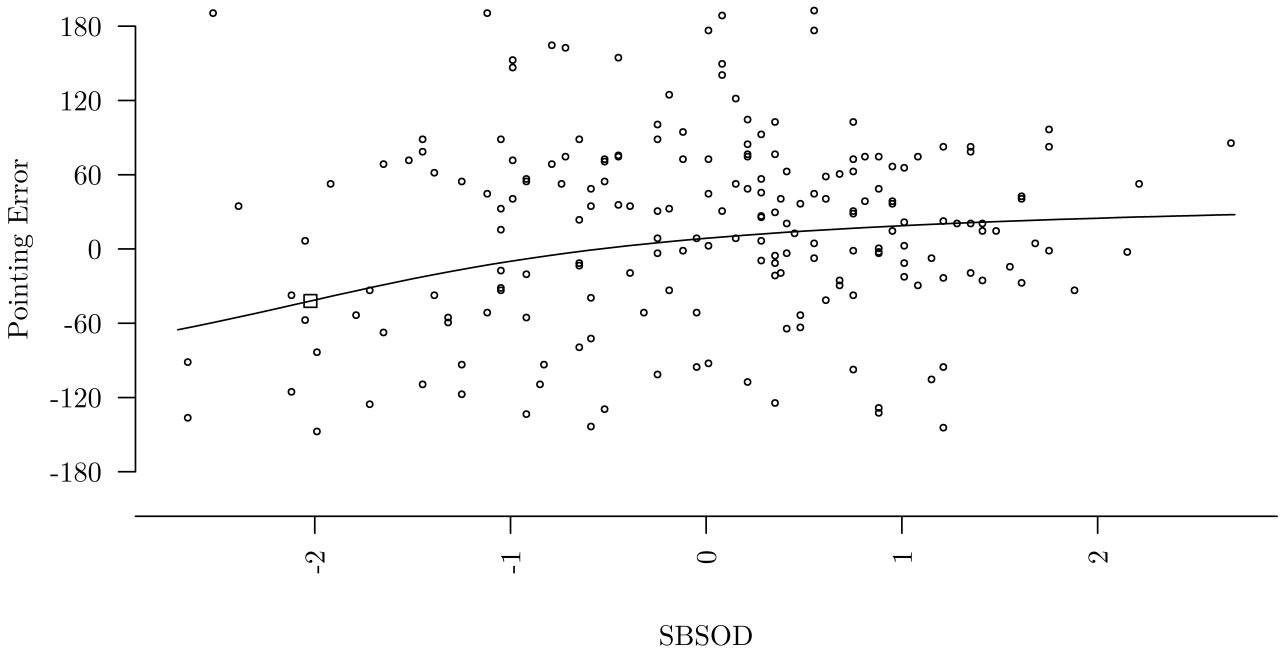
Estimated mean resultant length for males and females with average age, experience and Santa Barbara Sense of Direction (SBSOD) score:

```
##             mean mode   sd   LB   UB
## (Intercept) 0.55 0.56 0.06 0.43 0.65
## sexfemale   0.38 0.38 0.06 0.26 0.50
```

Results

	mean	mode	sd	LB HPD	UB HPD
age ax	166.00	187.14	815.97	-788.58	1026.88
experience ax	-198.54	-148.36	225.27	-590.93	45.05
sbsod ax	-140.80	-114.51	61.76	-250.45	-51.63
age ac	13.76	85.00	60.89	-76.79	100.48
experience ac	39.38	84.91	71.39	-82.12	102.05
sbsod ac	-44.74	-47.95	22.42	-74.67	-15.71
age bc	1.36	8.65	13.75	-23.30	24.56
experience bc	-5.90	-16.54	21.30	-37.20	34.70
sbsod bc	23.70	22.74	11.11	9.02	43.29
age AS	10.58	-5.24	5148.08	-179.94	143.21
experience AS	-11.55	8.68	1453.90	-341.68	372.51
sbsod AS	18.99	12.49	605.77	4.19	27.17
age SAM	-35.50	-5.21	1599.56	-206.92	162.68
experience SAM	0.33	8.87	2131.26	-398.87	477.85
sbsod SAM	13.85	11.81	32.38	5.03	21.35
age SSDO	-6.58	-53.40	51.71	-66.49	65.19
experience SSDO	21.29	67.12	66.73	-84.29	87.65
sbsod SSDO	-69.12	-72.24	24.94	-92.14	-54.15

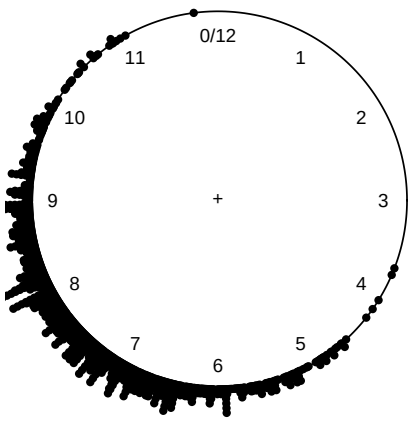
Results



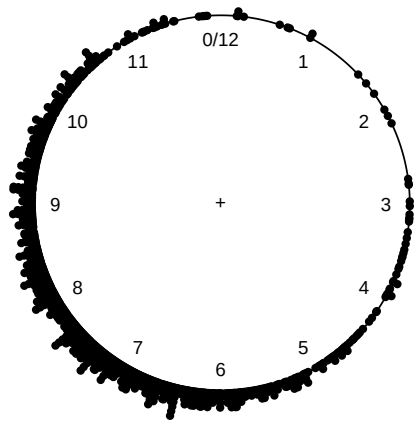
Interpretation of coefficients, see Cremers, Mulder, & Klugkist (2018).

Some thoughts on circular statistics and the Libet clock

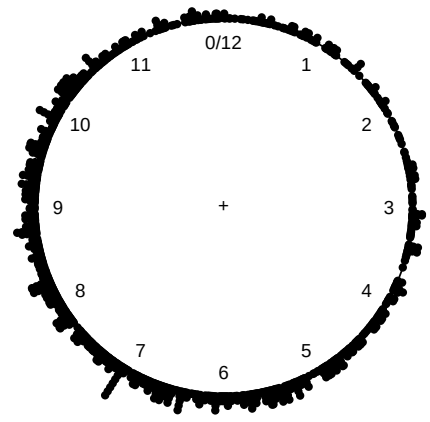
Group 1



Group 2



Group 3



Some thoughts on circular statistics and the Libet clock

Location

Linear analysis:

- Group 1 mean: 7.623
- Group 2 mean: 7.601
- Group 3 mean: 8.026

Circular analysis:

- Group 1 mean: 7.625
- Group 2 mean: 7.613
- Group 3 mean: 7.632

Some thoughts on circular statistics and the Libet clock

Accuracy

Linear analysis:

- Scores converted to 0-180 degree scale, deviance from correct time
- linear regression
- heterogenous regression models

Circular analysis:

- Models that allow for simultaneous modelling of mean and variance (heterogenous)
- Projected normal regression (available in R-package ‘bpnreg‘)
- Heterogenous von Mises models

Some thoughts on circular statistics and the Libet clock

Linear analysis:

- Careful with scale and results
- Analysis methods easily available

Circular analysis:

- Often quite similar to linear analysis
- Scale is correct
- Analysis methods not always readily available

Software

- R-packages:
  - `circular`: general purpose package
  - `bpnreg`: Bayesian circular regression and mixed-effects models based on projected normal distribution
  - `circglmbayes`: Bayesian circular regression based on von Mises distribution
  - `CircSpaceTime`: Bayesian spatial and spatio-temporal models.
- Other:
  - `CircStat`: general purpose MATLAB toolbox

Suggested literature

- Introductory
  - Batschelet (1981). Circular statistics in biology.
  - Fisher (1995). Statistical analysis of circular data.
  - Pewsey, Neuhäuser & Ruxton (2013). Circular statistics in R.
  - Ley & Verdebout (Eds.). (2018). Applied directional statistics: Modern methods and case studies.
  - Cremers & Klugkist (2018). One direction? A tutorial for circular data analysis using R with examples in cognitive psychology.
- Advanced (requires mathematical background)
  - Pewsey & García-Portugués (2020). Recent advances in directional statistics.
  - Mardia & Jupp (2000) Directional statistics.
  - Ley & Verdebout (2017). Modern directional statistics.

References

Cremers, J., Mulder, K., & Klugkist, I. (2018). Circular interpretation of regression coefficients. *British Journal of Mathematical and Statistical Psychology*, *71*(1), 75–95. doi:10.1111/bmsp.12108

Fisher, N. I. (1995). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.

Nunez-Antonio, G., & Gutierrez-Pena, E. (2005). A Bayesian analysis of directional data using the projected normal distribution. *Journal of Applied Statistics*, *32*(10), 995–1001. doi:10.1080/02664760500164886

Presnell, B., Morrison, S. P., & Littell, R. C. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, *93*(443), 1068–1077. doi:10.1080/01621459.1998.10473768