Bayesian Statistics short course – week 1

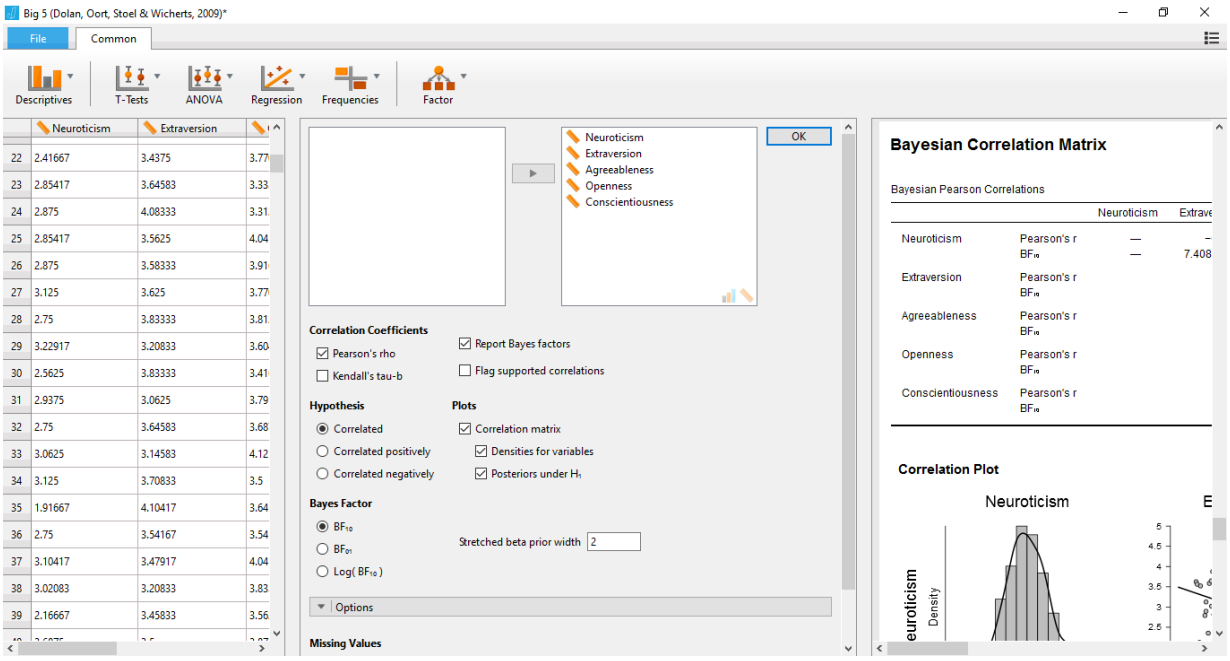
Jonathan J Forster and Jakub Bijak, University of Southampton

JASP – basic information

> > https://jasp-stats.org/ < <

- Open source, SPSS-style point-and-click software for statistical analyses
- Developed at the University of Amsterdam by the team led by Prof. Eric-Jan Wagenmakers, started by a grant of the European Research Council (ERC)
- Accepts data files in tabular format (`csv`, `txt`, `sav` and its own - `jasp`)
- Bayesian analyses: t-tests and binomial tests, ANOVA (analysis of variance), correlation and regression, analysis of contingency tables and log-linear models for counts. A lot of inference is presented in the form of Bayes Factors (BF), measuring the strength of evidence (*more on this in week 2*).
- All these have frequentist counterparts, plus there are options of performing descriptive analysis and factor analysis / principal components analysis

Key advantages: Very intuitive and easy-to-use interface; Bayesian analysis enabled without the need to code / program or worry about the numerical algorithms; aesthetic outputs – graphs and tables, ready to be copied to publications or reports. Can be very useful (and quick) for well-defined applied problems.



Key limitations: Limited number of functions in the present version, specification of some Bayesian aspects of the existing functions (e.g. priors) not always formal enough. At present, only a few key analyses are available (mainly those useful to psychologists), but software is still under development. Not (yet) suitable for more advanced or bespoke statistical problems. Documentation is very scarce.

Example 1 – T-test and ANOVA

A JASP example on a psychological experiment involving turning kitchen rolls (`Kitchen rolls.csv`)
From Wagenmakers et al. (2015) http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00494/full

T-Tests ⇒ Bayesian Independent Samples T-Test
Dependent variable: mean_NEO (test result); Grouping variable: Rotation

- Descriptive statistics
- Main test result (Bayes factor)
- Prior-posterior plot with additional information
- Robustness check – different priors (here: Cauchy distribution)
- Sequential analysis – different priors and subsamples

Was the experiment conclusive? Do prior assumptions play a role?

ANOVA ⇒ Bayesian ANOVA
Dependent variable: mean_NEO (test result); Fixed factors: Rotation, Age, Sex

- Model selection based on the weight of evidence (probability of a hypothesis given the data)

Which variables are important in this example?

Example 2 – Correlation and regression analysis

Analysis of the 'big five' psychological traits (`Big 5 (Dolan, Oort, Stoel & Wicherts, 2009).csv`)
After: CV Dolan, FJ Oort, RD Stoel, JM Wicherts (2009) in: *Structural Equation Modeling*, 16 (2), 295-314.

Regression ⇒ Bayesian Correlation Matrix (include all variables)

- Correlation matrix (but why the asterisks?)
- Plots: densities and posteriors for the correlation coefficient
- Impact of different priors (here: stretched Beta distribution)

Which of the five traits are related with one another? Do prior assumptions play a role?

Regression ⇒ Bayesian Linear Regression
Dependent variable: Neuroticism; Covariates: Everything else

- Model selection and analysis of individual effects
- Sensitivity to prior distribution for correlations

Which variables are important as covariates of neuroticism in this example?

Example 3 – Log-linear regression for counts

A made up migration flow table for 6 x 6 imaginary regions (`Migration.csv`).

Frequencies ⇒ Bayesian log-linear regression
Counts: Migrants; Factors: Origin, Destination

- Posterior characteristics of the estimated effects – (saturated) model with interactions
- Posterior characteristics of the estimated effects – (unsaturated) model without interactions

Which effects are important in both models?