



Andrew Ho <kironide@gmail.com>

Week 2 Day 3: Afternoon

1 message

Jonah Sinick <jsinick@gmail.com>

Wed, Feb 24, 2016 at 2:00 PM

To: Ali Bagherpour <ali.bagherp@gmail.com>, Andrew Ho <Kironide@gmail.com>, Chad Groft <clgroft@gmail.com>, David Bolin <david@bolin.at>, Jacob Pekarek <jpekarek@trinity.edu>, Jaiwithani <jaiwithani@gmail.com>, James Cook <cookjw@gmail.com>, Linchuan Zhang <email.linch@gmail.com>, Matthew Gentzel <magw6270@terpmail.umd.edu>, Olivia Schaefer <taygetea@gmail.com>, Sam Eisenstat <sam.eisenst@gmail.com>, Tom Guo <tomguo4@gmail.com>, Trevor Murphy <trevor.m.murphy@gmail.com>

I have attached:

- A zip file with a presentation that I've made on principal component analysis and factor analysis, using answers to the *Multidimensional Sexual Self-Concept Questionnaire*.
- A zip file with ~20k survey responses to a big 5 personality inventory (BIG5 under http://personality-testing.info/_rawdata/).

After the presentation, you'll be doing a similar analysis of the Big 5 personality inventory.

Some questions:

- What principal components do you get if you do PCA on the 50 questions? How about factor analysis with 5 factors?
- By focusing on the 10 questions corresponding to a trait, for each trait, can you replicate the subfactors discussed in the wikipedia article https://en.wikipedia.org/wiki/Hierarchical_structure_of_the_Big_Five? How about the superfactors "alpha" and "beta"?
- Create a logistic regression predictive model of gender in terms of the 50 questions, then in terms of the 5 factors derived from them. Compare the coefficients of the two models. Explore what happens if you replace the 5 factors by the subfactors or super-factors discussed in the article.
- Explore other datasets on http://personality-testing.info/_rawdata/ using dimensionality reduction.

2 attachments**MSSCQDimensionalityReduction.zip**

1481K

**BIG5.zip**

512K