

Catalog Study Report

MSSP Consulting Group: Qian Xu, Zichun Liu, Zijian Han

January 24, 2016

0. Tasks we have completed and outline of the report

1. Recategorized 38 measurements into 6 variables
2. Compute a table of proportions, for each gender, for each country, for each composite variable. (the “boy_score.csv” and the “girl_score.csv”, attached with the report)
3. We computed two-sample t-test statistics, and permutation tests p-values between genders for each of the 6 coding variables. Then we converted both of them into z-scores by the inverse probability transformation. After that we visualised the results with heat maps.
4. Visualized the the correlations between z-scores from p-values (permutation test and t test) and GGI.

1. The two-sample t-statistics and converted z scores.

According our clients’ advice, we created composite scores by grouping similar variables together. Initially we have 38 variables for each gender, after grouping we have 3 composite variables for each gender, 6 in total: Feminine emotional expression, Masculine emotional expression, Feminine nonverbal behavior, Masculine nonverbal behavior, Feminine objects in picture and Masculine objects in picture. Please see the attachments “boy_new.csv” and “girl_new.csv”.

2. Compute a table of proportions

Please see the attachments “boy_score.csv” and “girl_score.csv”. For coding, please see the rmd file “combined.rmd”.

3.1 The two-sample t-statistics and converted z scores.

We computed the two-sample t-statistics between genders for each country and each composite variable.

To demonstrate, here we take country ‘Australia’, composite variable ‘Feminine nonverbal behavior’ as an example. 5 Australian boys and 4 Australian girls were in the experiment. The composite variable ‘Feminine nonverbal behavior’ is created by grouping 7 sub variables: ‘smiling at viewer of image’, ‘smiling at someone in picture’, ‘looking at the viewer or image’, ‘gazing or smiling at someone in nurturant manner’, ‘standing or sitting with legs close together’, ‘arms close to trunk of body’, and ‘head cant’; each of the sub variables has 5 reponses from Australian boys and 4 reponses from Australian girls. Therefore the composite variable ‘Feminine nonverbal behavior’ has 35 Australian boys’ responses and 28 Australian girls’ responses in total.

What we did is to run a t test for 2 vectors, one contains the 35 responses from Australian boys regarding ‘Feminine nonverbal behavior’, the other contains the 28 responses from Australian girls regarding ‘Feminine nonverbal behavior’. Since we have 20 countries and 6 composite variables, we have conducted 120 t tests totally and a t-statistic matrix with 20 columns(countries) and 6 rows(composite variables) was computed. These t tests have different degrees of freedom due to different sample size, so we converted the all the 120 t statistics into standard normal distribution scale by inverse probability transformation, for more convenient illustration.

3.2 The Permutation test p values and converted z scores.

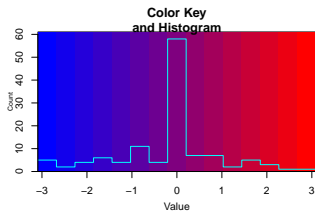
For formal testing, permutation tests are more reliable for our experiment compared with t-tests; as t-test requires normality of samples. So we conducted permutation tests between genders for each country and each composite variable, following the same logic as the t tests described above. Again, since we have 20 countries and 6 composite variables, we have conducted 120 permutation tests totally, and constructed a matrix of permutation tests p values, with 20 columns(countries) and 6 rows(composite variables). P values of permutation tests measures the probability of the 2 vectors being tested coming from the same distribution, and the values range from 0 to 1. We then converted all the 120 p values into standard normal distribution scale(converted z scores), by inverse probability transformation.

Generally, fixing one country and one composite variable: if there is no significant difference between genders, the converted z scores of that country and that composite variable would be close to 0; if boys in that country showed more tendency regarding that composite variable, the corresponding z score would be positive; if girls in that country showed more tendency regarding that composite variable, the corresponding z score would be negative. For example: fix a country 'Australia' and one composite variable 'Feminine nonverbal', the z score is 0; this means that Australian girls and boys are showing almost the same amount of feminine nonverbal behaviors during the experiment. For Australia, this time we fix the variable 'Feminine Emotion', the z score is negative, hencing that Australian girls actually showed more feminine emotion than boys in the experiment. Also, for Australia, if we fix another composite variable 'Masculine Emotion', the z score is positive, this means Australian boys showed more masculine emotions than Australian girls.

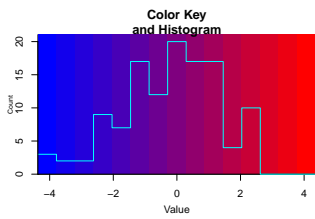
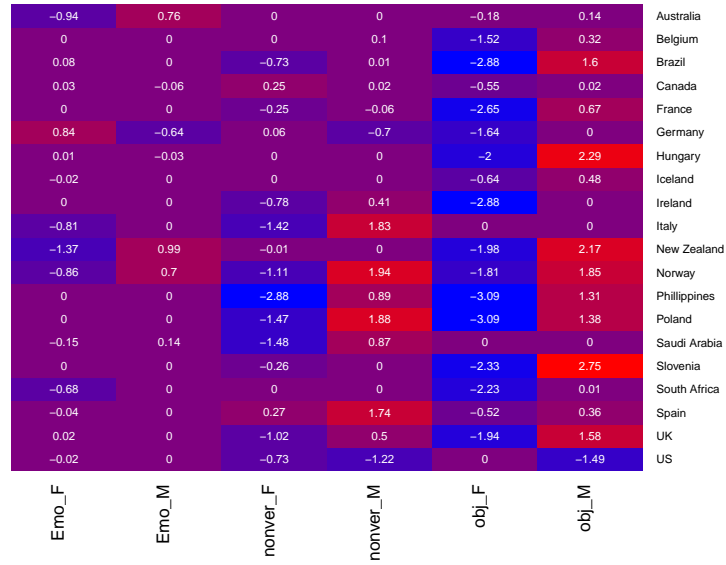
3.3 Data Visualisation and interpretation of heatmaps.

Even though 'T tests Z-scores Heatmap' gives us similar results, we focus on the interpretation of 'Permutation Z-score Heatmap'; since permutation tests results are more reliable in our experiment. Generally, we can see 3 kinds of colors in the heatmap. The color purple means 'no significant difference', color blue means 'girls are doing this more than boys', color red means 'boys are doing this more than girls'.

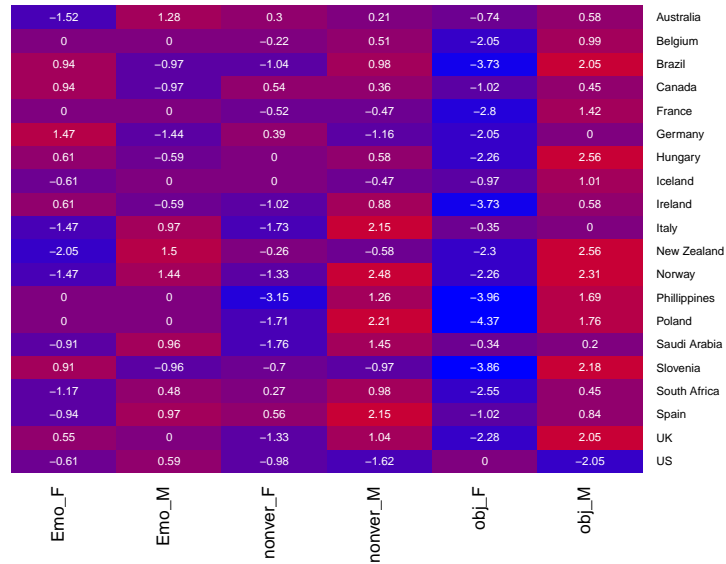
We have some really interesting findings. Regarding 'Feminine Emotion Expression' and 'Masculine Emotion Expression', we can see that there is not much difference across the countries. Nonverbal behavior differences between genders are more obvious, while objects in picture differences are significant cross most of the countries.



Permutation Z-score Heatmap



T test Z-score Heatmap



4. Correlation between permutation&t tests z-scores and GGI

In this section, we calculated the correlation between GGI of each country and visualised it. To interpret this plot, differences in feminine nonverbal behaviors, feminine emotion expression and masculine objects are positively correlated with GGI; the difference in feminine nonverbal behaviors has the largest positive correlation with GGI. While differences in feminine objects, masculine nonverbal behaviors and masculine emotion expression are negatively correlated with GGI.

