Investigation the relationship between personality traits, six basic emotions and gender with respect to Age

Mohamed Mostafa[1], Tom Crick[2], Ana C. Calderon[2], and Giles Oatley[3]

[1] Cardiff University, UK
mostafam1@cardiff.ac.uk
[2] Cardiff Metropolitan University, UK
{tcrick,acalderon}@cardiffmet.ac.uk
[3] Murdoch University, Australia
g.oatley@murdoch.edu.au

Abstract. Abstract here...

Keywords: TBC

1 Introduction

The previous experiments suggested a strong correlation between personality traits and emotions, furthermore, the attempt of modelling server status, suggested a strong and potenial method to model the user behaviour in different complex system behaviour. This analysis to explore the big personality traits and emotions assosication and correlation, further correlation between Gender/Age and Personality traits - Emotions. According to [**?**] Gender and Age correlate with the personality traits however, in same study the emotion features was not included. Therefore, this experiment is essential in cross-validating the methodology used in the [**?**] and add more features to the model to check if emotions can play a positive role in the equation.

1.1 Data set

The same data set will be used in the analysis as Personality Traits retrieved from the Motivation letter and emotions retrieved from different platform that was used in communication (i.e: Facebook, HelpDesk). The basic information will be retrieved to include the gender and age as extra parameters and the analysis will be running separately.

1.2 Binomial Logistic Regression

The data set combination suggested to use Binomial Logistic Regression, as the dataset similar to **??**, with Gender instead of Stages ID. The experiment is to

investigate the probability of being able to predicate the gender based on Big Five traits and Emotions, and according to the result, it will be decided whether to include the Gender as controlling variable in the conceptual model. Adding the Age variable alongside with Big Five Traits and Emotions to investigate if it would improve the model or not.

In order to apply, Binomial Logistic Regression, the data needs to pass the following assumptions:

- Linear relationship between the Big Five Traits, Emotions and logit transformation of the gender variable.
- Data must not show multicollinearity
- There should be no significant outliers, high leverage points or highly influential points

Linear relationship between the Big Five Traits, Emotions and logit transformation of the ***gender*** *variable* The first part of the Box-Tidwell (1962) procedure requires that all continuous independent variables are first transformed into their natural logs, this means that we need to perform natural log transformations on our continuous independent variables: Big Five Traits and Emotions.The second part of the Box-Tidwell (1962) procedure requires that you create interaction terms for each of your continuous independent variables and their respective natural log transformed variables. Since we have three continuous independent variables in our example, this means that we have to create Big Five Trait and Emotions - interaction terms: `ln_sadness` * sadness (i.e., the product of `ln_sadness` by sadness then need to be entered into the binomial logistic regression procedure, together with the gender and age.

According to [**?**], to calculate the new alpha () level (i.e., p-value) for current dataset , it is by dividing the alpha level (p ¡ .05) by the number of terms in your model. Formulaically, this is:

$$adjustedalphalevel = \frac{OriginalAlphaLevel}{numberofComparisons} \tag{1}$$

The new adjusted alpha level in this case is *0.002*, (i.e., 0.05 / 23= 0.002). Linearity of the Big Five Traits and Emotions with respect to the logit of the *Gender* variable was assessed via the Box-Tidwell (1962) procedure. A Bonferroni correction was applied using all twenty-one terms in the model resulting in statistical significance being accepted when p ¡ .002 [**?**]. According to the table 1 , all continuous independent variables were found to be linearly related to the logit of the dependent variable.

Data must not show multicollinearity , next step to investigate if the data shows or does not show multicollinearity to validate the possibility of applying binomial logistic regression. According to table 2 there was one studentized residual with a value of -2.376743 standard deviations, which was kept in the analysis.

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1a	Anger	1.248	2.244	0.309	1	0.578
Disgust	-0.332	23.43	0	1	0.989	0.717
Fear	1.667	2.822	0.349	1	0.555	5.294
Joy	1.961	1.496	1.719	1	0.19	7.107
Sadness	-1.251	0.854	2.145	1	0.143	0.286
Openness	-2.494	1.152	4.691	1	0.03	0.083
Conscientiousness	2.303	1.266	3.311	1	0.069	10.008
Extraversion	0.484	0.911	0.282	1	0.595	1.622
Agreeableness	0.155	0.954	0.026	1	0.871	1.167
Neuroticism	0.166	0.832	0.04	1	0.842	1.181
Age	0.289	0.584	0.245	1	0.621	1.335
Anger by ln_anger	1.945	3.032	0.412	1	0.521	6.992
Disgust by ln_disgust	2.757	11.198	0.061	1	0.806	15.75
Fear by ln_fear	-0.623	3.262	0.036	1	0.849	0.536
Joy by ln_joy	4.399	2.49	3.121	1	0.077	81.354
Sadness by ln_sadness	-1.914	2.128	0.81	1	0.368	0.147
Openness by ln_openness	7.291	3.366	4.693	1	0.03	1466.912
Conscientiousness by ln_conscientiousness	1.096	2.669	0.169	1	0.681	2.992
Extraversion by ln_extraversion	-3.599	2.282	2.487	1	0.115	0.027
Agreeableness by ln_agreeableness	-0.662	2.413	0.075	1	0.784	0.516
Neuroticism by ln_neuroticism	1.957	2.622	0.557	1	0.456	7.076
Age by ln_age	-0.053	0.129	0.17	1	0.68	0.948
Constant	-0.39	4.891	0.006	1	0.936	0.677

Table 1. Variables in the Equation - Gender

Casewise Listb						
Case	Selected Statusa	Observed	Predicted	Predicted Group	Temporary Variable	
gender					Resid	ZResid
7	S	F**	0.85	M	-0.85	-2.377

Table 2. Casewise Diagnostics

Bionomial Findings This aim of this experiment is to investigate which variable were statistically significant of the Big Five Traits, Emotions and Age with respect to the Gender, only three were statistically significant: Openness (p¡0.072), Conscientiousness and Age (as shown in Table 3). The result reported does not give enough accuracy regarding the correlation between Big Five Traits, Emotions and Age to predict the Gender. Therefore, another form of analysis is applied next to explore and investigate potential association between the above variables.

Variables in the Equation								
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Anger	0.329	1.518	0.047	1	0.828	1.39	0.071	27.209
Disgust	-5.201	7.187	0.524	1	0.469	0.006	0	7219.283
Fear	1.405	1.635	0.739	1	0.39	4.075	0.165	100.404
Joy	0.169	0.951	0.031	1	0.859	1.184	0.184	7.631
Sadness	-0.567	0.731	0.6	1	0.438	0.567	0.135	2.379
Openness	-1.556	0.866	3.229	1	0.072	0.211	0.039	1.152
Conscientiousness	1.261	0.764	2.722	1	0.099	3.529	0.789	15.786
Extraversion	0.389	0.78	0.249	1	0.618	1.476	0.32	6.801
Agreeableness	0.262	0.778	0.113	1	0.736	1.3	0.283	5.974
Neuroticism	0.234	0.644	0.132	1	0.717	1.263	0.358	4.46
Age	0.059	0.031	3.664	1	0.056	1.061	0.999	1.127
Constant	-1.395	1.175	1.41	1	0.235	0.248		

Table 3. Binomial Log - Variables in the Equation

1.3 Pearson's partial correlation

As the Binomial Logistic Regression, suggested a correlation between *Openness*, *Conscientiousness* and *Age* to predict the *Gender*, the Pearson's partial correlation was run to assess the relationship between Big Five Traits, Emotions, Age and Gender and to confirm the output of the Binomial or include more variable as strong association.

According to analysis performed in **??**, there were linear relationships between Big Five Traits and Emotions, as assessed by scatterplots and partial regression plots. There was univariate normality, as assessed by Shapiro-Wilk's test (p ¿ .05), and there were no univariate or multivariate outliers, as assessed by Mahalanobis Distance respectively. **?? ??**

Findings and Discussion The above tables shows the output of Pearson's Partial Correlation. In table 5, the controlling variable is *Gender*, a bivariate Pearson's correlation established that there was a strong, statistically significant linear relationship between *Conscientiousness* and *Anger*, r(204) = -.141, p ¡ .05 , *Neuroticism* and *Fear* r(204) = -.166, p ¡ .05. Pearson's partial correlation showed that the strength of this linear relationship was improved when *Gender* was controlled for *Conscientiousness* and *Anger* rpartial(203) = -.139, p=0.47 and it is still the same between *Neuroticism* and *Fear* rpartial(203) = -.166 - p=.017 and still statistically significant. In table **??**, the controlling variable is *Age*. Pearson's partial correlation showed that the strength of this linear relationship was improved when *Age* was controlled, in respect to the relationship between *Conscientiousness* and *Anger* rpartial(203) = -.138 - p=0.49 and between *Neuroticism* and *Fear* rpartial(203) = -.152 - p=0.030 and still statistically significant. In table 6, the controlling variable is *Gender* and *Age*. Pearson's partial correlation showed that the strength of this linear relationship

Pearson's partial correlation						
		Anger	Disgust	Fear	Joy	Sadness
Controlling Variable: None						
Openness	Correlation	.044	-.008	-.038	.024	-.015
	Significance (2-tailed)	.529	.913	.586	.729	.833
	df	204	204	204	204	204
Conscientiousness	Correlation	-.141	-.093	-.099	.057	-.040
	Significance (2-tailed)	.043	.183	.155	.418	.566
	df	204	204	204	204	204
Extraversion	Correlation	.040	.056	-.035	-.079	.003
	Significance (2-tailed)	.567	.421	.616	.262	.960
	df	204	204	204	204	204
Agreeableness	Correlation	.010	.041	.026	-.094	.053
	Significance (2-tailed)	.881	.556	.715	.178	.447
	df	204	204	204	204	204
Neuroticism	Correlation	-.038	-.058	-.166	-.006	.030
	Significance (2-tailed)	.585	.407	.017	.937	.664
	df	204	204	204	204	204
Age	Correlation	-.042	-.082	-.187	.170	-.085
	Significance (2-tailed)	.551	.241	.007	.015	.222
	df	204	204	204	204	204
Controlling Variable: Gender						
Openness	Correlation	.041	-.014	-.039	.029	-.020
	Significance (2-tailed)	.558	.844	.579	.681	.772
	df	203	203	203	203	203
Conscientiousness	Correlation	-.139	-.086	-.100	.052	-.034
	Significance (2-tailed)	.047	.219	.155	.463	.632
	df	203	203	203	203	203
Extraversion	Correlation	.041	.058	-.035	-.080	.005
	Significance (2-tailed)	.560	.410	.617	.256	.946
	df	203	203	203	203	203
Agreeableness	Correlation	.013	.046	.026	-.098	.057
	Significance (2-tailed)	.855	.514	.711	.163	.413
	df	203	203	203	203	203
Neuroticism	Correlation	-.036	-.054	-.166	-.008	.034
	Significance (2-tailed)	.606	.439	.017	.904	.627
	df	203	203	203	203	203
Age	Correlation	-.038	-.075	-.188	.166	-.080
	Significance (2-tailed)	.587	.282	.007	.017	.257
	df	203	203	203	203	203

Table 4. Pearson's Partial correlation - Controlling Variable - Gender

Pearson's partial correlation						
		Anger	Disgust	Fear	Joy	Sadness
Controlling Variable: None						
Openness	Correlation	.044	-.008	-.038	.024	-.015
	Significance (2-tailed)	.529	.913	.586	.729	.833
	df	204	204	204	204	204
Conscientiousness	Correlation	-.141	-.093	-.099	.057	-.040
	Significance (2-tailed)	.043	.183	.155	.418	.566
	df	204	204	204	204	204
Extraversion	Correlation	.040	.056	-.035	-.079	.003
	Significance (2-tailed)	.567	.421	.616	.262	.960
	df	204	204	204	204	204
Agreeableness	Correlation	.010	.041	.026	-.094	.053
	Significance (2-tailed)	.881	.556	.715	.178	.447
	df	204	204	204	204	204
Neuroticism	Correlation	-.038	-.058	-.166	-.006	.030
	Significance (2-tailed)	.585	.407	.017	.937	.664
	df	204	204	204	204	204
Gender	Correlation	.030	.056	.005	-.041	.050
	Significance (2-tailed)	.668	.423	.947	.558	.473
	df	204	204	204	204	204
Age	Correlation	-.042	-.082	-.187	.170	-.085
	Significance (2-tailed)	.551	.241	.007	.015	.222
	df	204	204	204	204	204
Controlling Variable Age						
Openness	Correlation	.050	.003	-.014	.002	-.004
	Significance (2-tailed)	.475	.964	.842	.978	.960
	df	203	203	203	203	203
Conscientiousness	Correlation	-.138	-.084	-.080	.038	-.031
	Significance (2-tailed)	.049	.229	.257	.592	.664
	df	203	203	203	203	203
Extraversion	Correlation	.041	.058	-.032	-.083	.005
	Significance (2-tailed)	.559	.407	.650	.235	.940
	df	203	203	203	203	203
Agreeableness	Correlation	.001	.023	-.018	-.057	.035
	Significance (2-tailed)	.990	.743	.796	.413	.622
	df	203	203	203	203	203
Neuroticism	Correlation	-.034	-.051	-.152	-.022	.039
	Significance (2-tailed)	.623	.471	.030	.753	.580
	df	203	203	203	203	203
Gender	Correlation	.025	.046	-.020	-.019	.040
	Significance (2-tailed)	.724	.513	.772	.785	.574
	df	203	203	203	203	203
Cells contain zero-order (Pearson) correlations.						

Table 5. Pearson's Partial correlation - Controlling Variable - Age

was improved when *Age* was controlled, in respect to the relationship between *Conscientiousness* and *Anger* and *Neuroticism* and *Fear* , rpartial(203) = -.138 -

Pearson's partial correlation						
Correlation		Anger	Disgust	Fear	Joy	Sadness
Controlling Variable: None						
Openness	Correlation	.044	-.008	-.038	.024	-.015
	Significance (2-tailed)	.529	.913	.586	.729	.833
	df	204	204	204	204	204
Conscientiousness	Correlation	-.141	-.093	-.099	.057	-.040
	Significance (2-tailed)	.043	.183	.155	.418	.566
	df	204	204	204	204	204
Extraversion	Correlation	.040	.056	-.035	-.079	.003
	Significance (2-tailed)	.567	.421	.616	.262	.960
	df	204	204	204	204	204
Agreeableness	Correlation	.010	.041	.026	-.094	.053
	Significance (2-tailed)	.881	.556	.715	.178	.447
	df	204	204	204	204	204
Neuroticism	Correlation	-.038	-.058	-.166	-.006	.030
	Significance (2-tailed)	.585	.407	.017	.937	.664
	df	204	204	204	204	204
Age	Correlation	-.042	-.082	-.187	.170	-.085
	Significance (2-tailed)	.551	.241	.007	.015	.222
	df	204	204	204	204	204
Gender	Correlation	.030	.056	.005	-.041	.050
	Significance (2-tailed)	.668	.423	.947	.558	.473
	df	204	204	204	204	204
Controlling Variable: Age and Gender						
Openness	Correlation	.047	-.003	-.012	.004	-.009
	Significance (2-tailed)	.501	.969	.870	.950	.902
	df	202	202	202	202	202
Conscientiousness	Correlation	-.136	-.079	-.083	.036	-.026
	Significance (2-tailed)	.053	.260	.239	.614	.714
	df	202	202	202	202	202
Extraversion	Correlation	.042	.059	-.032	-.084	.006
	Significance (2-tailed)	.554	.399	.646	.233	.930
	df	202	202	202	202	202
Agreeableness	Correlation	.004	.028	-.021	-.060	.039
	Significance (2-tailed)	.958	.686	.770	.394	.576
	df	202	202	202	202	202
Neuroticism	Correlation	-.033	-.048	-.153	-.023	.041
	Significance (2-tailed)	.639	.495	.029	.741	.557
	df	202	202	202	202	202

a Cells contain zero-order (Pearson) correlations.

Table 6. Pearson's Partial correlation - Controlling Variable - Gender and Age

p=0.49 , and between and *Neuroticism* and *Fear* rpartial(203) = -.152 - p=0.030 and still statistically significant. The above findings suggests that *Gender* and *Age* as controlled variable combined 6 would improve the linear relationship be-

tween Big Five and Emotions variables specially *Conscientiousness*, *Neuroticism* , *Anger* and *Fear* and improve strength of linear relationship between *Extraversion* and *Anger*, *Disgust*, *Fear*, *Joy* and *Sadness* although the linear relationship was not statistically significant. Those findings are aligned with the output from the Binomial Logistic Regression 1.2, in the correlation of the *Conscientiousness* and *Age* and impact of *Gender* in improving the association between variables.