Evaluating Drug-Use Risk in a Theorized Drug-Use Profile of Users and Non-Users; A Five Factor Model of Personality Approach

Pablo Bendiksen

Boston College: Lynch School of Education

Abstract

It is well established that drug use and abuse has myriad adverse effects on individuals and society, from health-related factors behind the risk associated with drug use to increased criminal activity. Evaluating the risk of individual drug consumption is a difficult task given that its been argued that use, abuse, and dependency should not be conceptualized as general term that cuts across substances. Whereas most analyses in the literature seek to classify individuals as users or non-users of a single drug at a time; we opted for a more involved approach that attempts to find a natural grouping of drugs via drug-related items; after all, drug use is correlated to personality types. Theorizing and measuring drug profiles based on patterns of shared use of similar drugs can prove insightful in better understanding the complex relationships involving substance use and personality trait. Using a predominant factor that explained about 40 percent of the variance associated with 16 drug frequency use items, one drug per item, we developed a metric for dichotomized grouping as either a non-user/non-participant in a predominant drug profile or a user therein. The metric was defined by the highest loading drug items of the factor analysis: Amphetamines, Cannabis, Cocaine, Magic Mushrooms, and Ecstasy. We then performed a linear discriminant analysis in order to evaluate the efficiency of a categorization model that could find applicability in preemptive diagnostics for at-risk drug users -under the context of a predominant latent drug profile. The context surronding the agent of use is imperative. Ultimately, focusing on efficient drug use characterization techniques, that are modeled by personality traits, can literally save lives. From preemptive diagnostics and intervention to outreach that may target at-risk populations.

Evaluating Drug-Use Risk in a Theorized Drug-Use Profile of Users and Non-Users; A Five Factor Model of Personality Approach

Introduction

Defining drug abuse or drug dependency does not fall into a quick catch all terminology that is accepted by medical professionals or scholars. In general terms, drug dependence may be viewed as a state of psychic or physical dependence, or both, on a drug; coming about from use of that drug on a periodic or continuous basis; Eddy, Halbach, Isbell, and Seevers (1965) elaborate that it is neither possible nor is it desirable to define the term drug dependency independently of the agent involved. This is because the characteristics of a state of drug dependency vary with the substance involved. For example, drug dependency of morphine type is distinct from drug dependency of an amphetamine type, and a cannabis type.

Because drug dependence as a state is seen more as a concept for clarification and not a specific definition, this general term indicates a relationship that overlaps with the idea of drug abuse, but it nevertheless will differentiate specific characteristics according to the nature of the agent involved(Eddy et al., 1965).

Although this argument that specific drugs take on different roles in how they confer dependency to the user, it is also well documented that usage of some drugs are significantly correlated (Heyman, Dunn, & Mignone, 2014). This correlation denotes that drug-using individuals have preferences over substances they ingest and, more importantly, that they may often times seek pleasurable effects from some drugs but not from others. That personality traits can relate to drug using behavior of course may provide insight into why certain people abuse certain drugs; the motivation becomes to better target the complex web of factors that lead to problematic drug use behavior, and its connections to health adversity and even criminal activities, that plague societies.

The potential for personality indicators of drug use is in fact the subject of many research studies (Boogar, Tabatabaee, & Tosi, 2014; Dubey, Gupta, & Kumar, 2010; Terracciano, Lockenhoff, m Crum, & Bienvenu, 2008). Taking a broad orientation by focusing on determining trait profiles for substance abusers, without contextualizing

drug use by substance, Dubey et al. (2010) showed that non-substance abusers scored significantly higher on Openness and significantly lower Neuroticism and Extraversion as compared to substance abusers; they specifically cite Extraversion as a strong indicator of drug abusers. Furthermore, a low score of Conscientiousness, and high score of Extraversion or high score of Neuroticism correlate strongly with multiple risky health behaviours. Considering the context of drug use; by performing a MANCOVA with user status as the independent variable, and while controlling for demographic variables, Terracciano et al. (2008) found significant personality differences among and across marijuana and cocaine/heroin status groups. Whereas, compared to never users, current marijuana users scored higher on Openness and lower on Agreeableness and Conscientiousness, cocaine/heroin users scored higher on Neuroticism and lower on Conscientiousness.

The personality traits referenced above are arguably the most relied upon metrics in personality trait measurements. It is important to note, by this point, that psychologists have largely agreed that the personality traits of the Five Factor Model (FFM), the so-called *Big Five*, are the most adaptive and comprehensive measure for understanding person-to-person differences (Costa & MacCrae, 1992). We therefore choose to restrict our personality-related inquiries under the scope of the Big Five.

The Big Five personality traits are as follows: Neuroticism, a higher risk towards unpleasant emotions such as anger, anxiety, depression or vulnerability; Extraversion, a tendency to seek stimulation in the company of others; Openness, a willingness and openness to experiences; Agreeableness, the tendency to be compassionate and cooperative towards others; and Conscientiousness, the tendency towards self-discipline, achievement, and dutifulness.

The literature demonstrates that, when it comes to drug abuse and, more aptly considered, drug use/abuse in the context of specific substances, the big five personality traits have been consistently linked to groupings of drugs or to patterns of risky or impulsive behavior that correlates to drug abuse. However, such analyses either test drugs under consideration on a one-by-one basis, or develop global terminology intended

to cut across drug categories (something we have already been warned against.)

A more sophisticated approach behind associating personality traits to drug use capitalized on the use of correlation pleiades, or clusters of correlated traits. By relying on this concept to distinguish the core and peripheral elements of correlation pleades, **?** identified three groups of correlated drugs centered around heroin, ecstasy, and benzodiazepines.

The present study takes on a similar focus on groupings of drugs to the end of evaluating consumption risk. Selective drug use by groupings of drug-use frequencies may suggest personality trait correlates to a pattern of use across multiple drugs. Such a pattern of similar use (in terms of frequency) across multiple drugs may provide a "drug profile" to categorize people as either belonging in, or not belonging in. The present study enacted a three stage analysis in order to 1) develop a meaningful grouping of drugs (from drug use responses) and thereby avoid individual drug evaluations; 2) provide a binary classification by union of part of classes from drug response categories into 2 classes that were then transformed to values of the composite grouping in 1); and, 3) classify cases across these two classes by Linear Discriminant Analysis modeling with measures of the Big Five personality traits as predictors.

In this manner, the present study attempts to observe the importance of drug use in the context of the substances that undergirths its manifestation; while extending past the variance modeling of one drug-related item at a time. With composite values representing disproportionate contributions across drug categories, we hope to better capture a drug use "profile" and, more specifically, categorize persons across the absence or existence of this profile. In this manner, we neither subscribe to generalized or blanketed terms that lose substantive meaning by cutting across all drug substances, nor do we encumber ourselves with overly-narrow scopes that can only model single drug-related outcomes.

The impact of analyses related to ours is gleamed in its real-world value. Modeling drug use profiles and building precise categorization models with personality-trait inputs can literally save lives. Preemptive interventions and the targeting of at-risk

youths are a couple of examples of the big gains that result from more nuanced conceptualization of patterns of drug use in persons and the modeling to identify such persons or persons at-risk of becoming drug users and (by extension) abusers.

Therefore, those with lower conscientiousness, agreeableness, and openness had higher tendency toward substance abuse than those with increased ones. Agreeableness, conscientiousness, and openness are healthy personality traits that positively affect the health tendencies for engagement in health behaviors. Therefore, these traits had negative correlation with tendency toward substance abuse.

Data Description and Definition of Variables

The data for our study was collected in 2011 and 2012 thanks to an approved research proposal by the University of Leicester School of Psychology. Over a 12-month period the study recruited a useable sample of 1885 participants. For each participant 12 attributes were measured: Personality measures which included the Big Five of interest to us. Each of these traits was already transformed into a single variable value per case within the dataset we accessed. Originally, they all stemmed from a 60 item survey with a five-point Likert scale that ranged from "Disagree" to "Strongly Agree". Other attributes included, but were not limited to, level of education, age, gender, and ethnicity. All input attributes were originally categorical but were quantified so that they may be treated as real numbers. For example, the Neuroticism item varied from 12 to 60 before quantification; upon accessing the data set, the variable (as well as all other attributes) was transformed such that it varied from -3.46436 to 3.27393.

A part from the attributes, participants were questioned about their use of 18 legal and illegal drugs (with one drug, Semeron, being a fictitious drug designed to identify over-claimers). For each drug, respondents selected one of seven answer: never used the drug, used it over a decade ago, used it in the last decade, year, month, week, or day. Table 1 of the Appendix lists all 30 items for ease of access. The data was gathered via an online survey tool from Survey Gizmo and the data can be found online through UCI's Machine Learning repository.

The data set contained no missing data of its 1885 cases. Data cleaning was still necessary, however.

Methodology

Following an exploratory factor analysis of drug-use items, a linear discriminant analysis was employed with the use of a binary transformed outcome space that itself originated from a continously distributed reduced-factor score distribution. All data cleaning, preparation, and analysis was done via SPSS.

Principal axis factoring permitted the extraction of a primary dominant factor that maximized eavery item loading on a factor while minimizing its loadings on other factors. The first extracted factor was inspected for optimal to meritorious factor loadings. Factor scores for drug-use items that corresponded to factor loadings above .7 were saved; a new computed variable representing this "reduced" linear combination of highly-loading drugs (factor component) was justified as necessary in order to capture the most dominant drug "profile of the data set. We justify the use of exploratory factor analysis as an attempt to extract this latent trait of drug-use pattern.

A categorization criteria was set such that drug-use response classes of 2 or less (0,1,2) were binned and labeled as "non-users", and drug-use response classes of 3 or more (3,4,5,6) were binned and labeled as "yearly users".

Having determined the the value differentiating between non-users and users, this cutoff was then transformed to equal a value along the reduced factor component distribution, thereby providing a cutoff for the binary transformation of this distribution. Having binned and labeled the two groups of the reduced factor component, our data was ready for classification modeling.

We evaluated drug consumption risk via a classification model wherein five personality measures were used to discriminate between two groups considered to be either non-users or users along a continuum of disproportionately weighted drug contributions of a grouping of drugs argued to represent a drug profile. Having used a classification technique, it's clear that a number of other classification techniques could

have been proposed alternatively; decision tree, random forest, logistic regression and naive Bayes. For example, Huskinson, Naylor, Rowlett, and Freeman (2014) employs a decision-tree approach in examining abuse potential, recommending that the same approach be used to examine a drug being evaluated for abuse potential against multiple known drugs of abuse.

Having checked normal Q-Q plots as well as having generated histograms for our five predictors, we provide strong evidence for univariate normality of all predictors, which may in this case translate to mulitvariate normality; having started with standardized predictor variables (means approximately zero and standard deviations approximately one), we noted near-zero user and non-user group means across predictors. We believe we've met the requirements of LDA sufficiently well such that this technique may classify better than binary logistic regression (the LDA model is asymptotically more efficient).

We were curious to see what magnitudes and directions were estimated for each of the Big Five discriminating variable's coefficients in the discriminant function. Research has suggested that high risk takers, substance abusers, marijuana users, and cocaine and heroin users are rated lower on Conscientiousness. Becuase we see this trait decrease across multiple contexts, from general tendencies down to the use of a number of distinct drug, we theorize Conscientiousness will carry a negative weight in our discriminant function. We believe Conscientiousness generally decreases with illicit drug use as illicit drug use is usually considered risky behavior; but even such a claim is open to wide interpretation depending on the substance under consideration. Similarly, becuase Extraversion was found to increase with substance abusers and marijuana users, we like to believe that this trait will carry a moderate positive coefficient in the discriminant function. An increase in Extraversion will have a positive impact on the discriminant score, increasing the likelihood of predicting a drug user under our theorized drug profile of similar drug use patterns across five drugs. And though marijuana users scored higher on Openness and lower on Agreeableness, cocaine users scored higher on Neuroticism. But what of a measure that is contributed upon by both

substances? Again, we highlight the fact that we are trying to learn more about the classification of grouped multi-drug users and non-users. We hope this research to be informative in elucidating personality trait relations to grouped and profiled drug users. We present our model for the discriminant function we will estimate. We will only yield one function because we binned across two groups:

$$Z_{D.S} = .020NEUROTIC + .202EXTRAVERSION + .714OPENNESS - .357AGREEABLE - .560CONSCIENTIOUSNESS \quad (1)$$

Results

The initial data set provided the drug-use items as string variables. We duplicated these variables, and then converted one set to numerics. Table 1 of the Appendix lists all items of the data set, with the understanding that they are treated as continuous. Items labeled as "Real" had response values converted to real numbers.

We generated histograms for all 18 drug-use items as a first step in building our composite outcome. We note that a disproportionate amount of items have positively skewed distributions, that a number of remaining items have negatively skewed items, and that a couple of items exhibited bi-modality (see Figure 1 of the Appendix). An item bias towards positive skewness is expected, as the lowest response categories (e.g. "Never Used", "Used Over a Decade Ago") signify lack of experience with a drug. Certain illicit substances are not commonly consumed for numerous reasons; penality, lack of accessibility, fear of effects, and cultural taboos, to name a few.

We ran a correlation matrix in order to check all bivariate correlations between these 18 items. Table 2 and Table 3 of the Appendix depict all such correlations. Based on correlation coefficients that were no larger than .35, we decided to eliminate the following variables from a subsequent factor analysis: VSA (volatile substances), Chocolate, Caffeine, Alcohol, and Semer; we eliminated Semer because it represented a fictitious drug. Our goal was to determine the strongest pattern of drug use across

substances. Because Factor Analysis derives its model from the decomposition of the correlation matrix, we felt this a justifiable first step.

We ran an exploratory factor analysis using Principal Axis Factoring estimation, in order to maximize single-factor loadings per item. Table 4 of the Appendix informs us of a meritorious KMO indicator and a non-significant Bartlett's Test, thereby suggesting sufficient partial correlations for the model to be estimated. Our final model extracted three dominant factors. We tested a number of rotations but settled for the unrotated model, as it's extracted primary factor explained the most amount of total variance than any rotated solution. Table 5 of the Appendix depicts the fact that the first extracted factor accounted for nearly 40 percent of the total variance. Furthermore, Table 6 of the Appendix depicts the factor loadings unto each item by each of the three dominant factors; with the second and third factor performing poorly (loadings lower than approximately .5). We highlight loadings greater than .7 on the first factor to signify the highest loading substances across all factors. Similarly, Table 7 of the Appendix denotes the factor scores (coefficients) associated with each drug substance. From Table 6 and 7 we recognize that the substances of Amphetamine, Cannabis, Cocaine, Mushrooms, and Ecstasy are suggested to most strongly represent or manifest an underlying construct. The factor is represented most strongly in them.

Because factors 2 and 3 were neglible in their expression as loadings, we worked with the saved factor scores of factor 1, only. Furthermore, we computed a new variable value for all cases, consisting of the factor component *with regards to the bolded drug items and their respective coefficients only*; we titled this new variable "Reduced Factor Score".

After having computed "Reduced Factor Score", we proceeded by making a categorization criterion; we needed to make a binary outcome from this variable but were unsure of what cutoff to select. Substantively, a cutoff wherein categories lower than monthly use (used drug in last year, decade, etc) would be labeled as "non-users" and the monthly-use category as well as higher categories would be labeled as "users", would have been ideal. However, this binning criterion lead to largely unequal groups;

we therefore opted for a "within-year use" criteria, wherein categories 0,1,2 (never used, used, over a decade ago, used in last decade) were binned as the drug non-user group, and categories 3,4, 5, and 6 (used in last year, used in last month, used in last week, used in last day) were binned as the drug user group. We quickly transformed this dichotomous metric to the Reduced Factor Score distribution. In other words, we dichotomized the Reduced Factor Score variable along the equivalent cutoff on its distribution; by applying the cutoff value of 3 into equation 1, we obtained the cutoff value of 1.48 for the ReducedFactorScore variable. We set all values above (and including) this to represent the new user category, and set all other values to equal the new non-user category. Figure 2 of the Appendix elucidates this process for the two previously competing dichotomized variables, after transformation. We note that the (transformed) "yearly" criteria variable has a much more balanced grouping (a roughly 60/40 proportion).

With a dichotomized grouping variable representing a binning of a former-linear combination of various substances in terms of frequency of their use, we finally began our discriminant analysis.

We began by checking the assumptions of linear discriminant analysis:

1. *Multivariate normality* As mentioned previously, univariate distributions and Q-Q Plots suggested normality of our five discriminating variables (see Figure 3 in the Appendix). These measures are well-established in the literature and are generally regarded to be normally distributed in the population. And though univariate normality does not necessitate multivariate normality, a monotonically increasing linear combination of these variables is generally regarded to maintain normality in higher-dimensional space.

2. *Equality of within-group variance-covariance matrices* SPSS allows for a test of the assumption of equal within-group variance-covariance matrices. This is assessed through Box's test of equality of covariances matrices (see Table 12 in the Appendix). A significant p-value ($p < .001$) provides us evidence to reject the null hypothesis that the two groups have equal variance-covariance matrices of the

independent variables. We recognize that this is often times a difficult assumption to meet, as distinct groups in most contexts are likely to have variation in their multivariate distributions across the predictors. We continue with our analysis but recognize that this is a limitation.

3. *Balanced Design; relatively equal group sample size* Table 8 of the Appendix informs us that our two groups constitute a roughly 60 to 40 ratio drug non-users to users. They can be considered to be roughly equal.

4. *Multicollinearity Check Assumption* We can quickly survey a scatterplot matrix of our discriminating variables to assess if they have bivariate relationships. Figure 4 of the Appendix illustrates scatterplots between all pairs of the Big Five measures, these scatterplots look weakly related, at best. Surveying a correlation matrix yields an upper limit pearson-correlation value of .431 between Extraversion and Neuroticism. All other correlations are markedly lower than this. We do not believe that these measures are strongly related to each other and continue with the analysis.

A test for equality of means between non-users and users across all discriminating variables is presented in Table 11 of the Appendix. We can see that both groups differ significantly in all variables at the $\alpha < .05$ level, with Extraversion carrying the largest p-value ($p = .045$). Furthermore, Figure 5 of the Appendix further suggests that the two groups do not have equal means, a necessary condition for discrimination between them to be successful. Wilk's Lambda gives a measure of variance unexplained between the two groups, it is a ratio of within group variance over total (between) group variance. The Wilk's λ's approach one, suggesting they do not discriminate strongly between groups.

The eigenvalue of our discriminant analysis ($\lambda = .041$) is an indicator of the function's ability to discriminate between the two groups. Our function may not differentiate very well (see Table 13 of the Appendix).

At this point, we are able to interpret our standardized discriminant function via

the discriminating weights (see Table 14 of the Appendix):

$$Z_{D.S} = .020NEUROTIC + .202EXTRAVERSION + .714OPENNESS - .357AGREEABLE \quad (2)$$
$$-.560CONSCIENTIOUSNESS$$

Because the coefficients are standardized, we can see that Openness and Conscientiousness have a disproportionately large effect in separating the groups and, consequently, in affecting classification for any given case. As theorized, we see that Conscientiousness has a moderate and negative impact in discriminating; with increased Conscientiousness lowering the function value, thereby increasing the likelihood of classifying an individual as a non-user (and non-match) for our latent drug use profile! Openness is our biggest discriminating variable, with the markedly largest magnitude in its coefficient. In terms of the contribution each variable makes to the function, Table 15 (analagous to factor loadings) of the Appendix informs us that Extraversion is the disproportionately largest contributor to the discriminant score.

We use a classification table (see Table 16 of the Appendix) to assess the classification probability (percentage of observations correctly classified). The first part of the table applies to the training data set (n = 1165); By summing predicted cases and dividing by sample size, we see that 69 percent of the cases in our training data set were correctly classified. That is 82.4 percent correctly classified non-users and 48.5 percent correctly classified users. Because we know that 62.1 percent of cases in the training dataset were non-users, and that 37.9 percent were users (see Table 8 of the Appendix), by the law of total probability, we can calculate that:

$1165 * .621 = 723.465$

$1165 * .379 = 441.535$

$723.465 * .621 = 449.27$ $441.435 * .379 = 167.30$

$167.30 + 449.27 = 616.57$

$441.535 + 723.465 = 1165$

$\frac{616.57}{1165} = .52925$

Therefore, with a random correct classification rate of 52.93, we have increased the classification about 16 percent above the random classification rate.

Surveying the Cross-validated part of the table we see that using the cross-validation method, 68.5 percent of the observations were correctly classified. This is very similar from the original classification results.

Also, doing a classification of the testing dataset gives a classification rate of 68.5, again around the value of the original classification results, that is still 16 percent increase from the rate due to chance!

Finally, by incorporating the training versus testing data sets, we essentially test the classification on a brand new data set. We get a value of 69.4., which is similar to our original findings. We settle for a model that classifies at 16 percent above the rate due to chance.

Discussion

Evaluating the risk of drug use, which many times preambles drug abuse, can allow for preemptive intervention; which itself may ultimately prove more effective than acute treatment on individuals who have already developed dependencies.

The continual development of higher-specificity classifying models, can inform and be informed by other research to ultimately find applications throughout the mental health sector; youths and teenagers seeking (or referred) to therapists, for example, may be more efficiently screened for drug use/abuse potential in order for measures to be taken to mitigate the environmental factors that increase their probability of using and abusing drugs.

We advocate for continual categorization and predictive modeling of drug users across varied definitions of drug profiles or otherwise grouped drugs. Individual drug modeling may not be capturing the complex webbed relations that underly patterns of drug use.

For example, Extraversion's largest correlation with the discriminant function signifies that our latent trait is comprised of Extraversion more than it is any other

discriminant variable! Yet, this variable was only third largest in discriminating between the two groups; Openness, on the other hand, had the largest (as well as positive) weight in the function (.714), we recognize it discriminates the best amongst our independent variables. Yet it was the third largest contributor to the function. Prior literature search already suggested an increase in Openness with regards to marijuana users (but not substance abusers). Perhaps this warrants further research into Openness with regards to the drugs that make up our latent profile: Amphetamines, Cannabis, Cocaine, Mushrooms, and Ecstasy. Interpreting Openness' large discriminating ability may be accomplishable by considering what these drugs have in common. Perhaps this group of drugs are associated with a youth party culture. Young person's who seek to socialize may be at a bigger risk to use any of these substances; this is especially noteworthy considering that the above mentioned drugs are party favorites at social events such as festivals and night clubs. It is in retrospect, then, that the disproportionate discriminating ability of Openness, with increased scores increasing the likelihood of a drug user categorization, has substantive meaning. This corroborates previous findings about marijuana users but not about substance abusers. These drugs are nevertheless categorized as scheduled substances due to their potential for being abused. It is strange then, how increased Openness may be associated with Marijuana users, as well as with users falling into our latent drug profile, but not necessarily drug abusers. These interesting- yet potentially conflicting- findings highlight the need for further research into the conceptualization of drug-use profiles and grouped drug latent traits. Such theorization needs work before setting out to measure further with models.

Appendix

References

Boogar, I. R., Tabatabaee, S., & Tosi, J. (2014). Attitude to substance abuse: Do personality and socio-demographic factors matter? *International Journal of High Risk Behaviors and Addiction*, *3*(3).

Costa, P., & MacCrae, R. (1992). Revised neo-personality inventory (neo-pi-r) and neo five-factor inventory (neo ffi). *Odessa, FL: Psychological Assessment Resources*.

Dubey, C., Gupta, S., & Kumar, B. (2010, January). Five factor correlates: A comparison of substance abusers and non-substance abusers. *Frontiers in Psychiatry*, *36*(1), 107-114.

Eddy, N. B., Halbach, H., Isbell, H., & Seevers, M. H. (1965). Drug dependence: its significance and characteristics. *Bulletin of the World Health Organization*, *32*(5), 721-733.

Heyman, G., Dunn, B., & Mignone, J. (2014). Disentangling the correlates of drug use in a clinic and community sample: A regression analysis of the associations between drug use, years-of-school, impulsivity, iq, working memory, and psychiatric symptoms. *Frontiers in Psychiatry*, *5*(70).

Huskinson, S. L., Naylor, J. E., Rowlett, J. K., & Freeman, K. B. (2014, December). Predicting abuse potential of stimulants and other dopaminergic drugs; overview and recommendations. *Neuropharmacology*, *0*(1), 66-80.

Terracciano, A., Lockenhoff, C. E., m Crum, R., & Bienvenu, O. J. (2008, April). Five-factor model personality profiles of drug users. *BMC Psychiatry*, *8*(22).

Item	Item
ID	Benzos
Age (Real)	Caffeine
Gender (Real)	Cannabis
Education (Real)	Chocolate
Country (Real)	Cocaine
Ethnicity (Real)	Crack
Neuroticism (Real)	Ecstasy
Extraversion (Real)	Heroin
Openness (Real)	Ketamine
Agreeable (Real)	Legalh
Conscientiousness(Real)	LSD
Impulsive (Real)	Meth
SensationSeeing (Real)	Mushrooms
Alcoho	Nicotine
Amphetamine	Semer
VSA	Amyl

Table 1

Table of all items for the University of Leicester's School of Psychology Trait and Drug Use Data Set

ρ	Alcoh.	Amph.	Amyl	Benz	Caff	Canna	Choc	Cocain	Crack	Heroin
Alcoh	1.00	-.087	.083	-.012	.104	.068	.036	.137	-.008	-.047
Amph.	-.087	1.000	.119	.371	.065	-.022	.031	.168	.116	.217
Amyl	.083	.119	1.000	.088	.091	-.149	.044	.228	.036	.009
Benzo	-.012	.371	.088	1.000	.047	.024	-.029	.241	.278	.391
Caff	.104	.065	.091	.047	1.000	-.005	.053	.101	.037	.045
Canna.	.068	-.022	-.149	.024	-.005	1.000	.057	-.116	-.014	-.003
Choc	.036	.031	.044	-.029	.053	.057	1.000	-.029	-.144	-.079
Cocaine	.137	.168	.228	.241	.101	-.116	-.029	1.000	.219	.267
Crack	-.008	.116	.036	.278	.037	-.014	-.144	.219	1.000	.444
Heroin	-.047	.217	.009	.391	.045	-.003	-.079	.267	.444	1.000

Table 2

An example table.

ρ	Keta	Legah	LSD	Meth	Mushroom	Nicotin	Semer	VSA	Ecstasy
KetamineC	1.000	.204	.186	.010	.150	.083	.041	.054	.261
LegalhC	.204	1.000	.162	.103	.188	.050	-.044	.165	.202
LSDC	.186	.162	1.000	.073	.452	.009	.034	.126	.226
MethC	.010	.103	.073	1.000	.064	.094	-.032	.178	-.059
MushroomsC	.150	.188	.452	.064	1.000	.001	.076	.050	.044
NicotineC	.083	.050	.009	.094	.001	1.000	-.005	.151	.039
SemeronC	.041	-.044	.034	-.032	.076	-.005	1.000	.053	-.018
VSAC	.054	.165	.126	.178	.050	.151	.053	1.000	.010
EcstasyC	.261	.202	.226	-.059	.044	.039	-.018	.010	1.000

Table 3

An example table.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.910
Bartlett's Test of Sphericity	Approx. Chi-Square	11191.596
	df	91
	Sig.	.000

Table 4

An example table.

Factor	Initial Eigenvalues			Extracted Eigenvalues		
	Total	Pct. of Variance	Cumulative Pct.	Total	Pct. of Variance	Cumulative Pct.
1	5.914	42.243	42.243	5.443	38.875	38.875
2	1.503	10.734	52.977	1.005	7.178	46.053
3	1.058	7.554	60.531	.529	3.779	49.832
4	.862	6.156	66.687			
5	.772	5.511	72.198			
6	.584	4.170	76.368			
7	.541	3.863	80.231			
8	.510	3.644	83.875			
9	.469	3.352	87.227			
10	.433	3.091	90.318			
11	.410	2.926	93.244			
12	.345	2.462	95.707			
13	.328	2.343	98.050			
14	.273	1.950	100.000			

Table 5

An example table.

	Factor Loadings		
	1	2	3
AmphetC	**.712**		
AmylC	.397		.419
BenzosC	.606		
CannabisC	**.710**		
CocaineC	**.734**		
CrackC	.456	.398	
HeroinC	.525	.521	
KetamineC	.568		
LegalhC	.682		
LSDC	.672		
MethC	.529	.425	
MushroomsC	**.711**		
NicotineC	.522		
EcstasyC	**.768**		

Table 6

Table of Factor Loadings following PAF Factor Analysis. Loadings of less than .3 have been suppressed. loadings of greater than .7 have been highlighted

	Factor Score Coefficients		
	1	2	3
AmphetC	**.115**	.047	.089
AmylC	.051	-.061	.315
BenzosC	.094	.179	.003
CannabisC	**.130**	-.140	-.150
CocaineC	**.150**	.094	.358
CrackC	.058	.200	.008
HeroinC	.100	.391	-.049
KetamineC	.067	-.051	.144
LegalhC	.104	-.116	-.069
LSDC	.103	-.136	-.279
MethC	.086	.285	-.193
MushroomsC	**.143**	-.188	-.333
NicotineC	.053	-.026	.021
EcstasyC	**.171**	-.257	.197

Table 7

Table of Factor Scores for 14 drug-use related items. Coefficients associated with Factor Loadings greater than .7 have been highlighted.

	Frequency	Percent	Valid Percent	Cumulative Percent	
Valid	.00	1171	62.1	62.1	62.1
	1.00	714	37.9	37.9	100.0
	Total	1885	100.0	100.0	

Table 8

Counts and Percentages of drug non-users (0) and users (1), as defined by a Factor Score cutoff of 1.418

Cases	N	Percent
Valid	1165	61.8
Unselected	720	38.2
Total	1885	100

Table 9

Table of descriptives for valid cases used in Discriminant Analysis (training data set) and unused cases (testing data set)

0 - 1.418, 1 - 1.418+		Mean	Std. Deviation	Valid N (listwise)	
			Unweighted	Weighted	
.00	Neuroticism	-.0651346	.98996910	705	705.000
	Extraversion	-.0599340	.99056678	705	705.000
	Openness	-.2254101	.97258402	705	705.000
	Agreeableness	.1504718	.96841098	705	705.000
	Conscientiousness	.2039990	.98155378	705	705.000
1.00	Neuroticism	.0913515	.99868071	460	460.000
	Extraversion	.0629573	1.06680747	460	460.000
	Openness	.3856495	.90141077	460	460.000
	Agreeableness	-.1978659	1.02707945	460	460.000
	Conscientiousness	-.2855137	.99820884	460	460.000
Total	Neuroticism	-.0033461	.99593404	1165	1165.000
	Extraversion	-.0114104	1.02266521	1165	1165.000
	Openness	.0158667	.99086257	1165	1165.000
	Agreeableness	.0129307	1.00608021	1165	1165.000
	Conscientiousness	.0107150	1.01633058	1165	1165.000

Table 10

Counts and Percentages of drug non-users (0) and users (1), as defined by a Factor Score cutoff of 1.418, by independent variable category

	Wilks' Lambda	F	df1	df2	Sig.
Neuroticism	.994	6.907	1	1163	.009
Extraversion	.997	4.030	1	1163	.045
Openness	.909	116.359	1	1163	.000
Agreeableness	.971	34.325	1	1163	.000
Conscientiousness	.945	68.312	1	1163	.000

Table 11

Table depicting results for an equality of group means; and independent samples t test, given two groups only

Box's M		39.421
F	Approx.	2.615
	df1	15
	df2	3862604.582
	Sig.	.001

Tests null hypothesis of equal population covariance matrices.

Table 12

Table depicting results for Box's Test of Equality of Covariance Matrices

Function	Eigenvalue	% Variance	Cumulative %	Canonical Correlation
1	.041	100.00	100.0	.198

Test of Function	Wilks' λ	χ^2	df	Sig.
1	.961	22.277	5	.000

Table 13

Table of summaries for a single yielded discriminant function (Top); Table of Wilk's Lambda test of function discriminating significance (Bottom)

Discriminating Vars	Function Coefficients
Neuroticism	.020
Extraversion	.202
Openness	.714
Agreeableness	-.357
Conscientiousness	-.560

Table 14

Table of Standardized Canonical Discriminant Function Coefficients

Discriminating Vars	Correlation
Neuroticism	.604
Extraversion	.599
Openness	.432
Agreeableness	-.335
Conscientiousness	.088

Table 15

Table of pooled within-groups correlations between discriminating variables and standardized canonical discriminant function

Cases	Reduced Factor Score (Dich)	ted (0)		Predicted (1)	Total
Training	Count (0)	581	124	705	
Training	Count(1)	237	223	460	
Training	% (0)	82.4	17.6	100	
Training	%(1)	51.5	48.5	100	
CrossVal	Count (0)	578	127	705	
CrossVal	Count(1)	241	219	460	
CrossVal	% (0)	82.0	18.0	100	
CrossVal	%(1)	52.4	47.6	100	
Testing	Count (0)	383	83	466	
Testing	Count(1)	122	132	254	
Testing	% (0)	82.2	17.8	100	
Testing	%(1)	48.0	52.0	100	

Table 16

Table of LDA Classification Results; from use of training data, from use of training data under leave-one-out cross validation technique, and from applying the LDA to a the testing data set. Percent correctly classified cases across three methods are: a) 68.9; b) 69.4) and c) 68.5

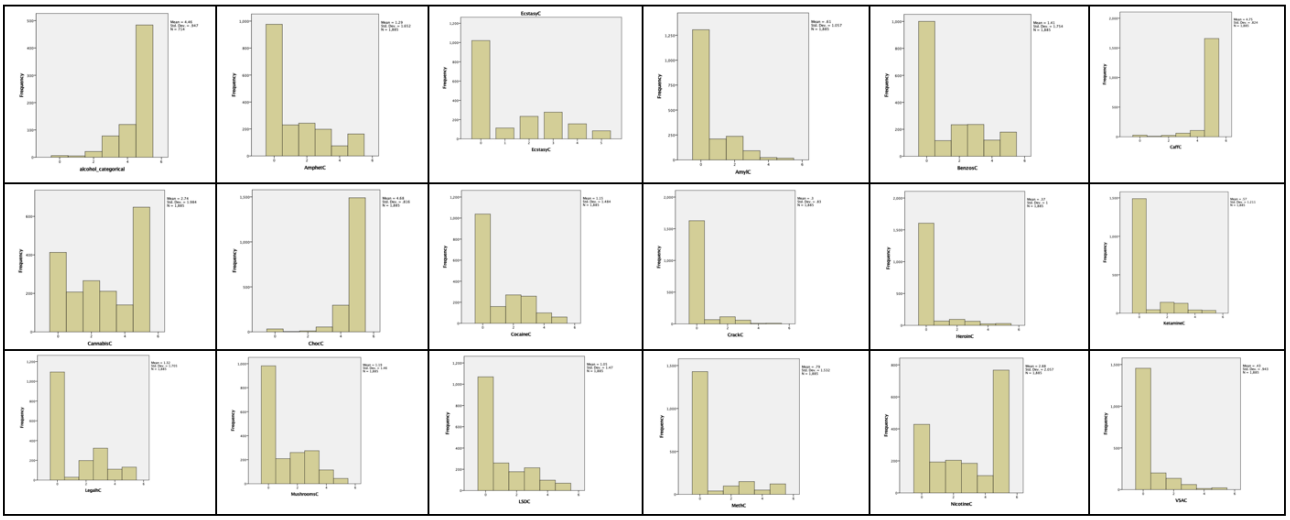

Figure 1. Histograms of all drug-use items where each item corresponds to one substance.

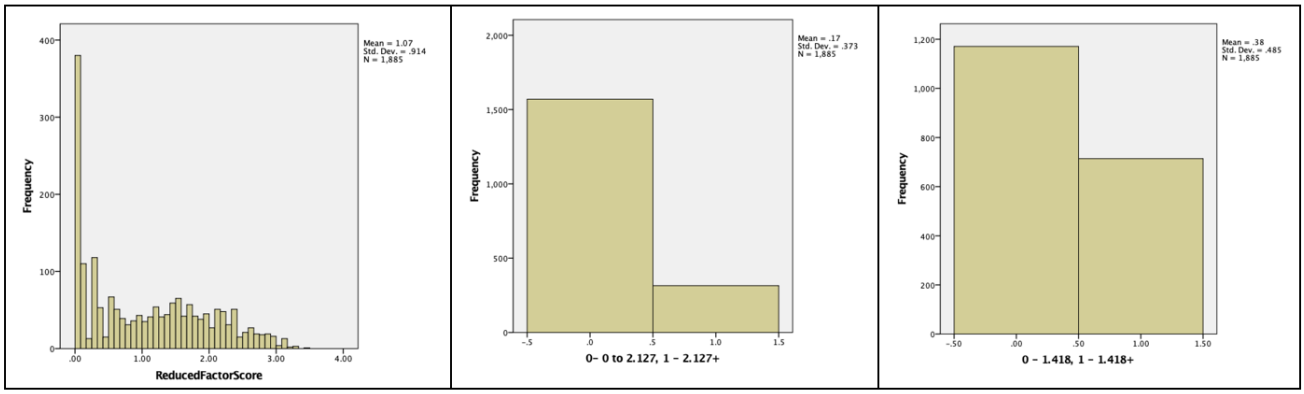

Figure 2. Images depicting two binning strategies for the Reduced Factor Component Distribution (Left). Partitioning along a yearly use of substance (Middle) did not provide as balanced groups as did partitioning along a decade-long use of a substance (Right).

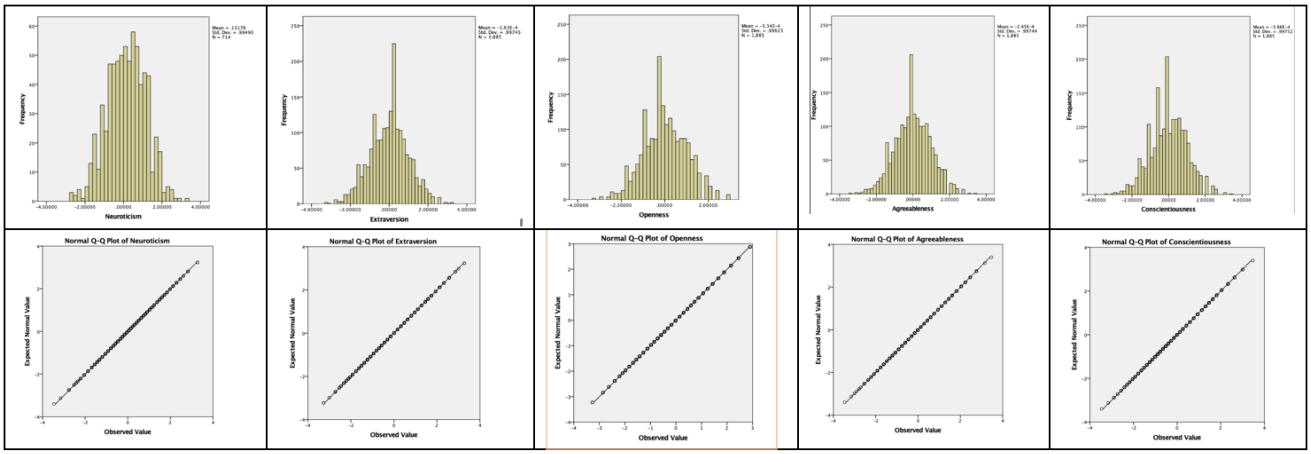

Figure 3. Histograms of all five personality trait measures (Top) and corresponding Q-Q Normal Plots (Bottom).

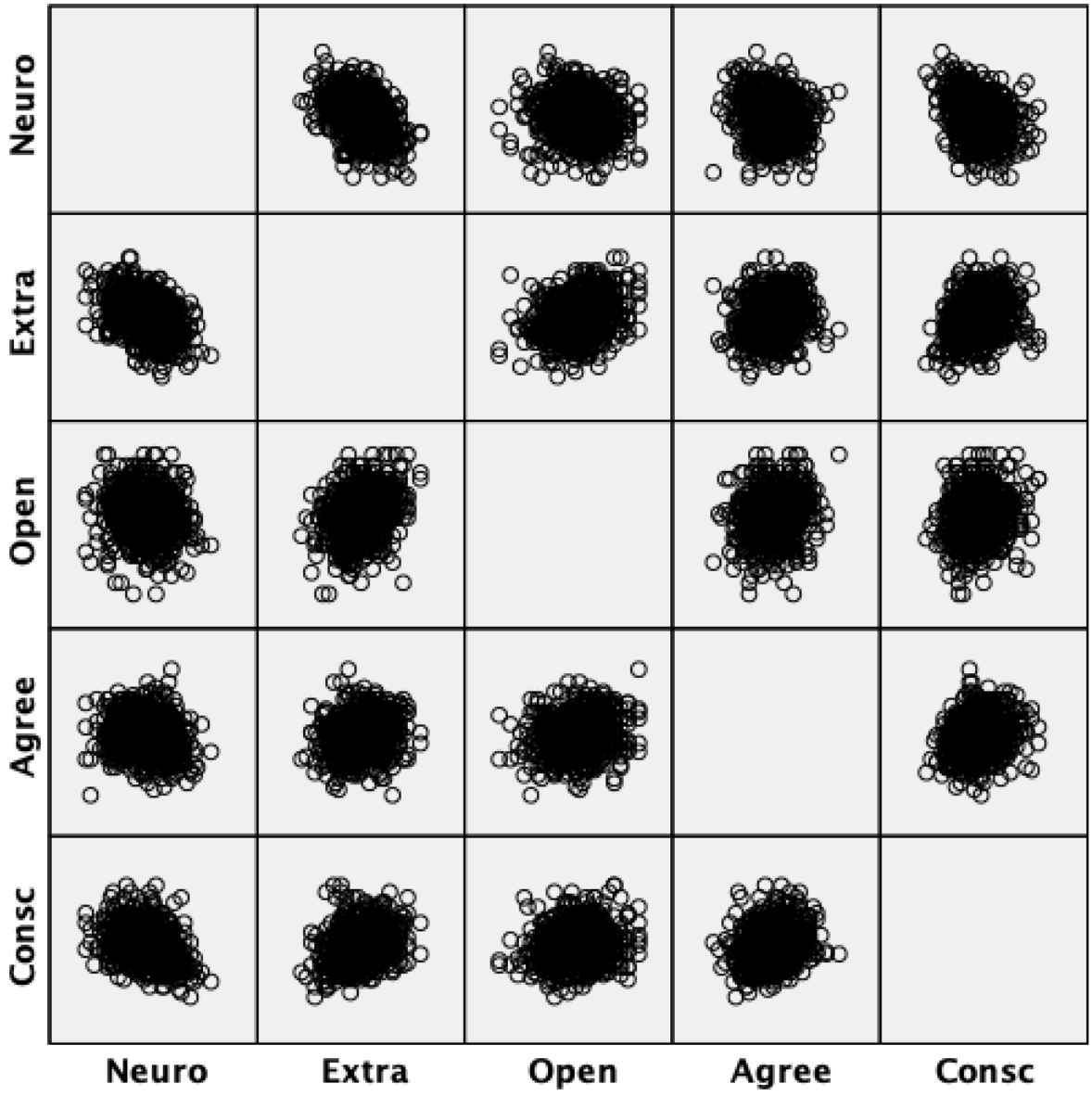

Figure 4. Matrix of Bivariate Scatterplots for all five independent variables

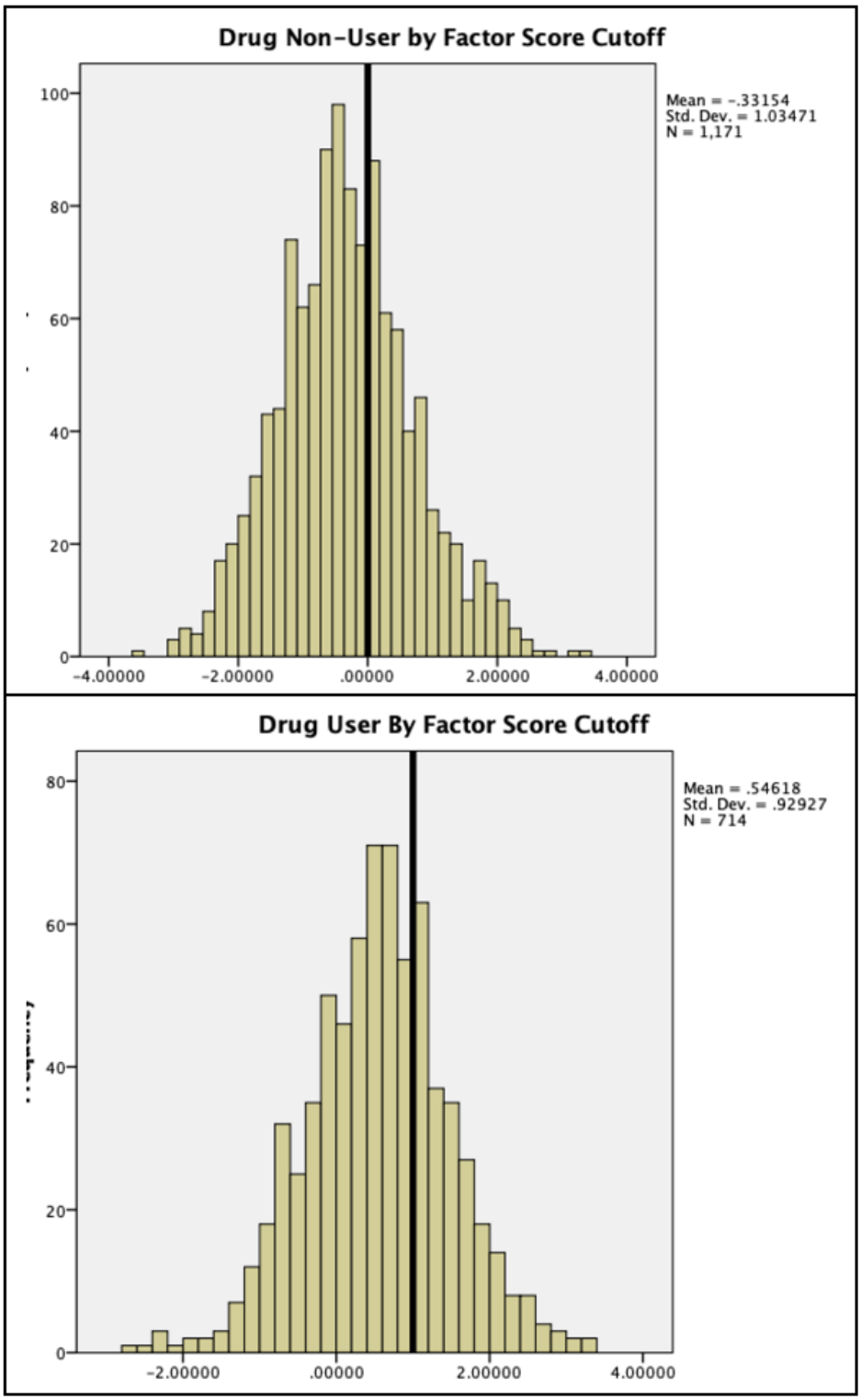

Figure 5. Histogram Distributions for Dichotomized Reduced Factor Score Groups, Drug Non-Users (Top) and Drug Users (Bottom). Mean values are highlighted in black to depict difference in respective values.