Big 5 Personality Traits Analysis

MAT4103 – Statistical Inference

Abstract

Psychologists today agree that the Big Five Model is the most commonly accepted personality theory. According to the idea, personality may be broken down to five main variables that can be classified into binary categories, with each personality attribute being a spectrum. The Five Factor Model is a statistically inferred research-based model of personality; the five top level qualities of openness, conscientiousness, extroversion, agreeableness, and neuroticism are made up of 30 subscale personality traits grouped into these five main dimensions. People's Big Five scores are mostly steady for the majority of their lives, with occasional minor variations from childhood to maturity. The Big Five, like all personality theories, is impacted by both nature and nurture.

The Big Five Broad Traits each have two DeYoung Facets, each of which have 3 NEO Sub-Facets, totaling to 30 personality tests. In this project, we look for independent variables using mosaic plots and the area of similar traits within our dataset to draw out certain hypotheses on which our aim would be to implement Z-test for our marked criteria.

Introduction

Extraversion (sometimes called extroversion), agreeableness, openness, conscientiousness, and neuroticism are the Big Five personality qualities. Each attribute is a part of a larger continuum. For each trait, people can fall anywhere along the spectrum. Throughout most of one's life, the Big Five stay quite steady. With an estimated heritability of 50%, they are influenced extensively by both genes and the environment. They've also been shown to predict major life events like schooling and health. Each of the Big Five personality traits is a large category that encompasses a wide range of personality-related terminology. Each attribute incorporates a wide range of additional characteristics. The Big Five Model's technique to gauging personality is another essential feature. It emphasizes the concept of qualities as a spectrum rather than binary groupings. It understands that the majority of people fall somewhere in between the extremes of the spectrum.

Conscientiousness

Conscientiousness refers to a person's capacity to manage their impulses and engage in goal-directed behaviours. It assesses aspects of behaviour such as control, inhibition, and persistence. Facets of conscientiousness include the following:

High

- Competence
- Organized
- Dutifulness
- Achievement striving
- Self-disciplined
- Deliberation

Low

- Incompetent
- Disorganized
- Careless
- Procrastinates
- Indiscipline

Conscientiousness vs. Lack of Direction:

Those with a high conscientiousness score are structured, disciplined, detail-oriented, deliberate, and cautious. They also have good impulse control, allowing them to finish tasks and achieve their objectives.

Those who score low on conscientiousness may have trouble controlling their impulses, making it difficult to complete tasks and achieve goals. They have a tendency to be unorganised and despise structure. They may also be more impulsive and careless in their actions.

Agreeableness

The term "agreeableness" refers to how people approach interpersonal relationships. Unlike extraversion, which is concerned with relationships, agreeableness is concerned with people's attitudes and interactions with others. Facets of agreeableness include the following:

High

- Trust (forgiving)
- Straightforwardness
- Altruism (enjoys helping)
- Compliance
- Modesty
- Sympathetic
- Empathy

Low

- Sceptical
- Demanding
- Insults and belittles others
- Stubborn
- Show-off
- Unsympathetic
- Doesn't care about how other people feel

Agreeableness vs. Antagonism:

Soft-hearted, trusting, and well-liked people have a high level of agreeableness. They are considerate of others' needs, as well as helpful and cooperative. They are seen as trustworthy and altruistic by the general public. Those with a low level of agreeability may be seen as suspicious, manipulative, or uncooperative. When interacting with others, they may be hostile, making them less likely to be loved and trusted.

Extraversion

The inclination and intensity with which a person wants engagement with their surroundings, particularly socially, is referred to as extraversion. It includes people's levels of comfort and aggressiveness in social situations. It also indicates the energy sources from which someone gets their energy. Facets of extraversion include the following:

High

- Sociable
- Energized by social interaction
- Excitement-seeking
- Enjoys being the center of attention
- Outgoing

Low

- Prefers solitude
- Fatigued by too much social interaction
- Reflective
- Dislikes being the center of attention
- Reserved

Extraversion vs. Introversion:

Extraverted people are forceful, gregarious, fun-loving, and extroverted in general. They thrive in social circumstances and are unafraid to express themselves. Being among people causes them to gain energy and get excited. Introverts are people who score low on the extraversion scale. These individuals are more reserved and silent. Rather of needing to be heard, they prefer to listen to others. Introverts typically require

periods of alone to re-energize, as attending social activities can be exhausting. It's worth noting that introverts don't always despise social gatherings; rather, they find them exhausting.

Openness

The readiness to try new things as well as engage in imaginative and intellectual activities is referred to as openness to experience. It entails being able to "think outside the box." Facets of openness include the following (John & Srivastava, 1999):

High

- Curious
- Imaginative
- Creative
- Open to trying new things
- Unconventional

Low

- Predictable
- Not very imaginative
- Dislikes change
- Prefer routine
- Traditional

Openness vs. Closeness to Experience:

Those with a high receptivity to experience are seen to be more creative and artistic. They cherish independence and prefer variety. They appreciate travelling and learning new things and are curious about their environment. Those with a low openness to new experiences favour routine. They are averse to change and attempting new things, thus they choose the known to the unknown. They have a hard time thinking creatively or abstractly since they are practical individuals.

Neuroticism

Neuroticism is a personality trait that indicates a person's overall emotional stability based on how they view the world. It considers how likely a person is to perceive circumstances as dangerous or challenging. It also includes a person's proclivity for unpleasant feelings. Facets of neuroticism include the following (John & Srivastava, 1999):

High

- Anxious
- Angry hostility (irritable)
- Experiences a lot of stress
- Self-consciousness (shy)
- Vulnerability
- Experiences dramatic shifts in mood

Low

- Doesn't worry much
- Calm
- Emotionally stable
- Confident
- Resilient
- Rarely feels sad or depressed

Neuroticism vs. Emotional Stability:

Neurotic people are prone to feeling uncomfortable, insecure, and self-pitying. They are frequently misunderstood as temperamental and unpleasant. They have a tendency to be depressed and have low self-esteem. Those with a low neuroticism score are more likely to be calm, confident, and self-satisfied. They're less likely to be labelled as jittery or moody. They are more likely to have a high sense of self-worth and to be resilient.

Methodology

The Big 5 Personality Traits analysis requires a dataset of people from various groups classified into various attributes; and the hypothesis tests we plan to conduct needed us to collect data from different age groups, gender, people of different income level and sexual orientation. Real time data was collected (as of 24th December, 2021) with the help of social fill-up form (Google Form) and was exported to a spreadsheet to work upon. The data was imported into our work area (Jupyter Notebook) as a Comma Separated Values file. Junk data and null values were cleaned out using classical Data Cleaning and Wrangling methods with PANDAS library.

The questionnaire had 44 personality trait bonded questions along with each individuals Age, Gender, Annual Income and Sexual Orientation. The 5 Personality Traits had 8-10 questions divided among each from the total of 44 questions where each individual had to mark the priority of how they feel about the questions from 'Disagree a lot' to 'Agree a lot'.

While calculating the total score of each individual concerning to each of the 5 Personality Traits, a few question remarks had to be reversed as the questions framed were in negation. In the Questionnaire data set, some particular question should be reversed scored according to the scoring instruction otherwise they define the opposite meaning. For example, for the question: "I notice other people's weak points."- Low score for this question means that person is helpful. High score for this question means that person is uncooperative. But according to the standard scoring system used for the questions altogether, it opposes this fact. So, the scale was reversed which then concludes that high score means the person is helpful and low score means the person is uncooperative.

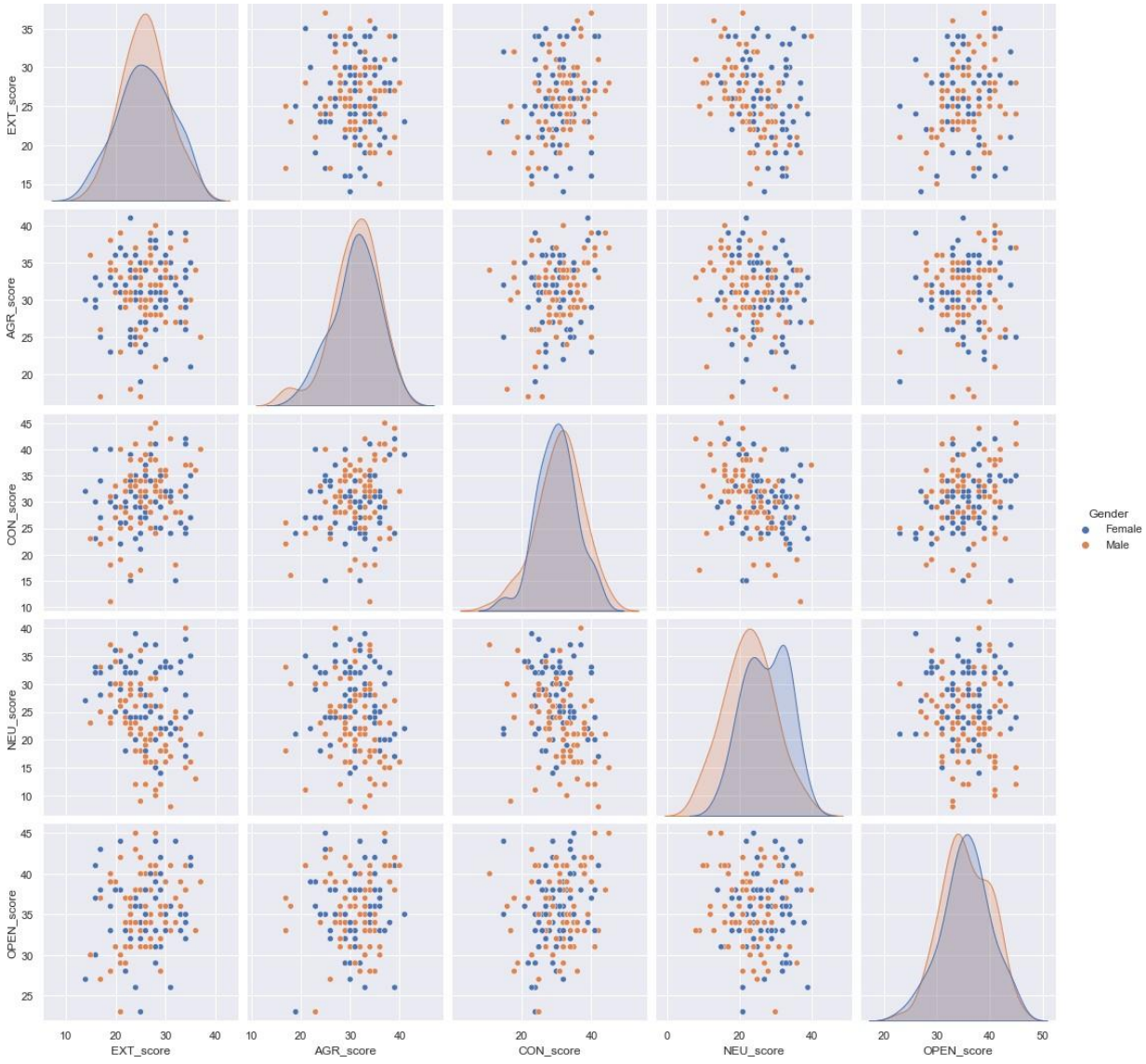
Extraversion	Forward Scale	EXT1, EXT3, EXT4, EXT6, EXT8
	Reverse Scale	EXT2, EXT5, EXT7
Aggreeableness	Forward Scale	AGR2, AGR4, AGR5, AGR7, AGR9
	Reverse Scale	AGR1, AGR3, AGR6, AGR8
Conscientiousness	Forward Scale	CON1, CON3, CON6, CON7, CON8
	Reverse Scale	CON2, CON4, CON5, CON9
Neuroticism	Forward Scale	NEU1, NEU3, NEU4, NEU6, NEU8
	Reverse Scale	NEU2, NEU5, NEU7
Openness	Forward Scale	OPEN1, OPEN2, OPEN3, OPEN4, OPEN5, OPEN6, OPEN8, OPEN10
	Reverse Scale	OPEN7, OPEN9

The Forward Scale and Reverse Scale Calculation for Questions

Visualization

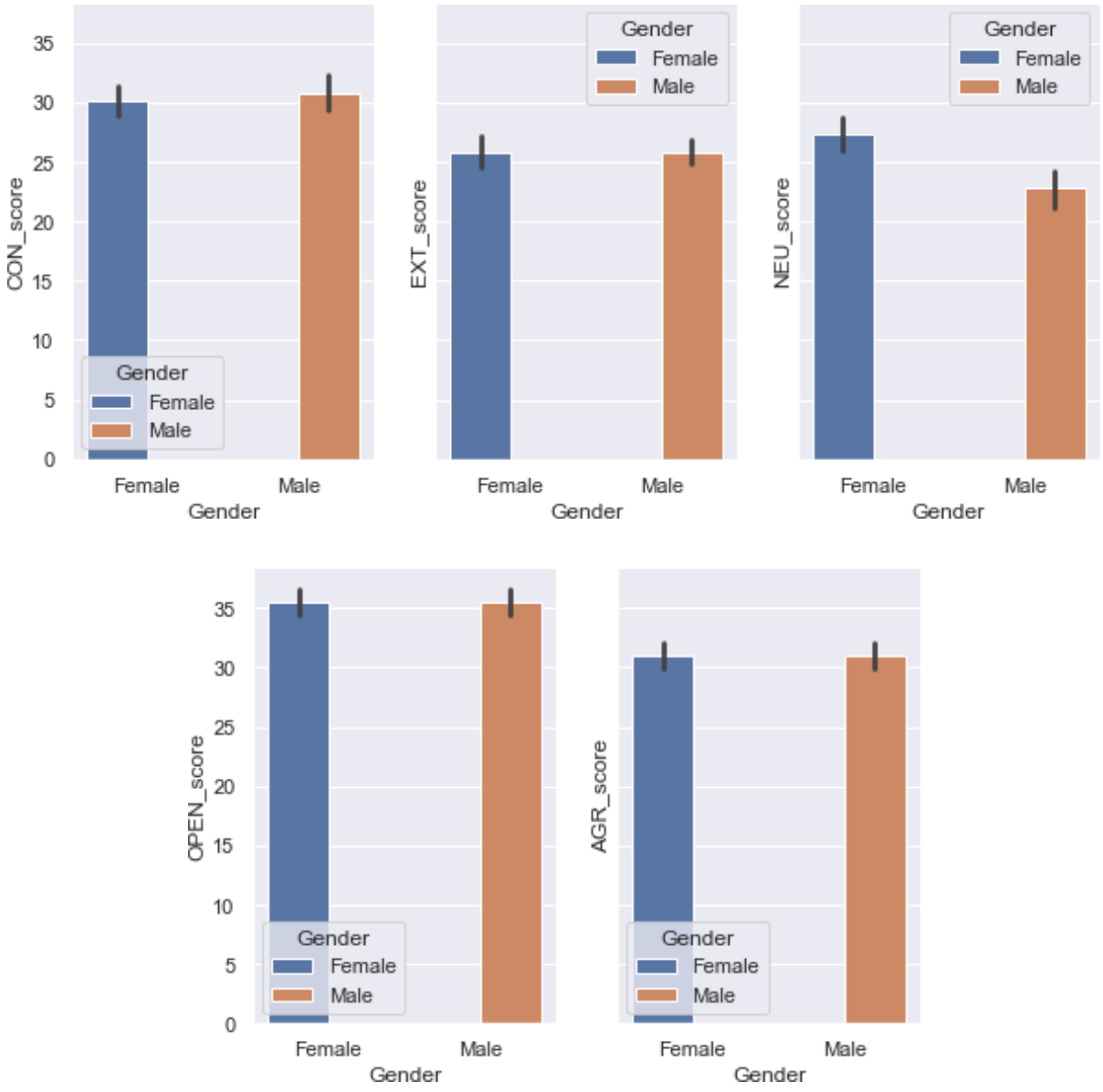
All the data to be worked upon had to be visualized in several methods to find out interdependence and independence of the trait variables for each group of population.

We have used pair plot among the 5 scorelines to find out correlation and patterns.



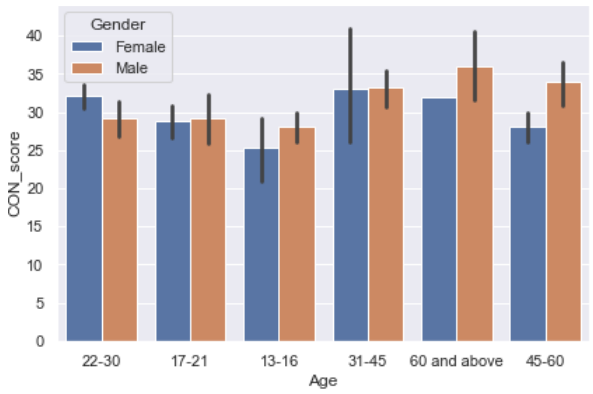
Interpretation: No correlation was found out between the score parameters for the 5 Personality traits. Hence, we could conclude that they were independent.

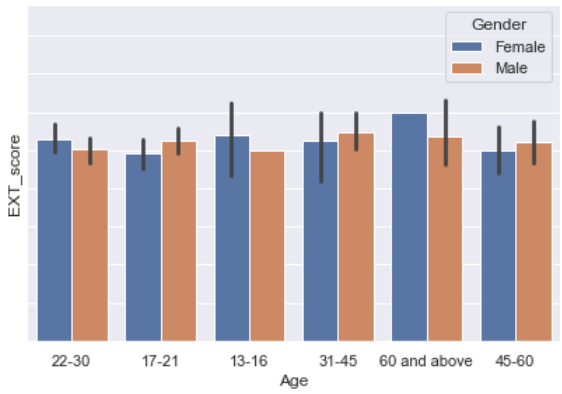
Bar plots were used to visualize the gender based score relations for all the 5 scores.

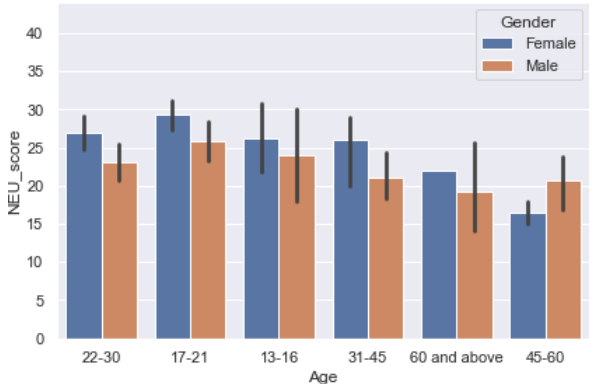


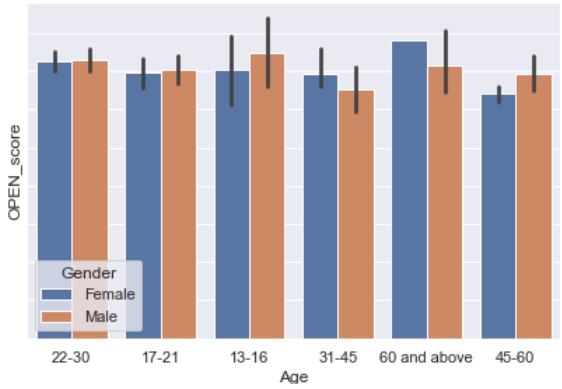
Interpretation: There is no significant difference in personality scores for male and female categories except the case that female NEU_score (Neuroticism) is higher than that of male category. This implies that women are more tend to possess unstable emotions than men.

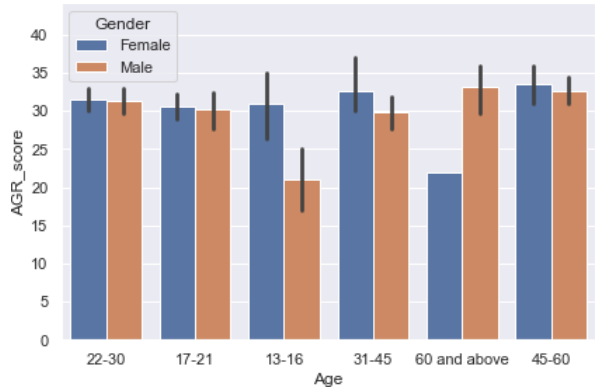
Bar plots were again used to visualize the age based score relations for all the 5 scores.





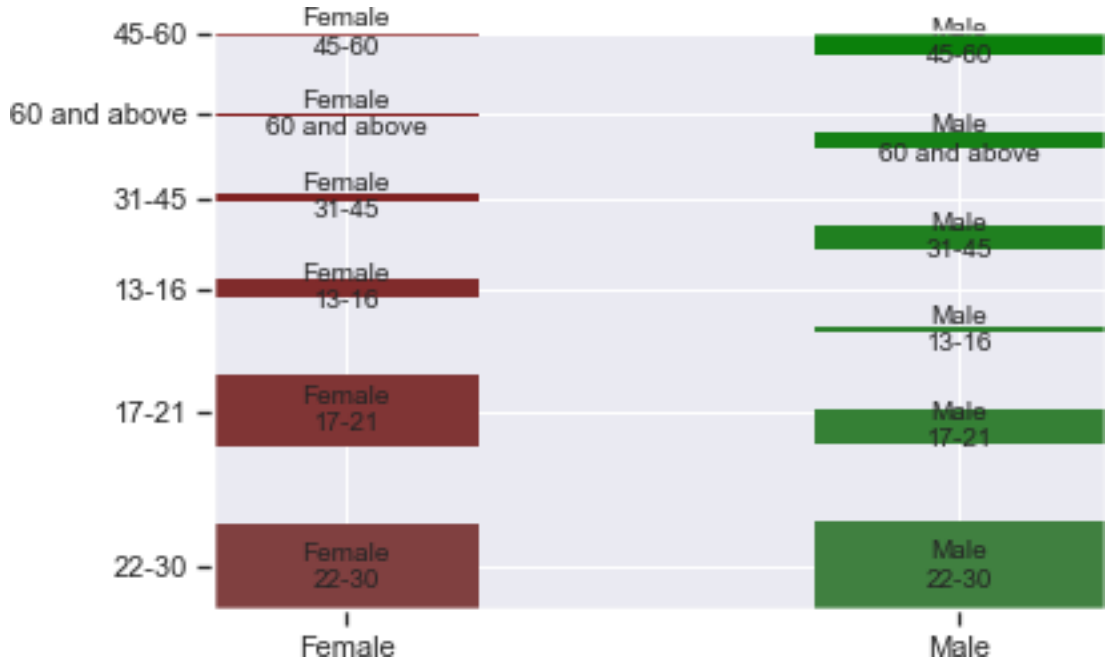






Interpretation: No such correlation or dependencies were found out between the 5 score parameters for the different age groups. Hence, we could conclude that they were independent as well.

Finally, we used mosaic plots to tests the independencies of the scores with age group and gender.



Interpretation: Since there is no overlapping of data there is no dependency between the gender and age of the personalities. We can move forward with Z – test for our hypothess with our datasets from these inferences.

Z-test is any statistical test for which the distribution of the test statistic under the null hypothesis can be approximated by a normal distribution. Z-tests test the mean of a distribution. For each significance level in the confidence interval, the Z-test has a single critical value (for example, 1.96 for 5% two tailed) which makes it more convenient than the Student's t-test whose critical values are defined by the sample size (through the corresponding degrees of freedom).

From the above visualization, we were able to formulate the following relations and perform Z-test for the preferred hypotheses.

Age Group	Gender	Annual Income
17-21	Female	
22-30		
17-21	Male	
22-30		
17-21	Male	
	Female	
22-30	Male	
	Female	
	Male	Less than 3,00,000
	Female	

Results and Analysis

We have taken the following facts into consideration while performing all the 'difference in mean' hypotheses for each of the 5 Personality traits individually.

- Level of significance: 5 % (0.05)
- Alpha value: 0.05
- If p value is less than level of significance, we can conclude that we have sufficient evidence to reject H_0.
- If p value is greater than level of significance, we can conclude that we don't have enough evidence to reject H_0.
- When there is enough evidence to reject null hypothesis we can conclude there is significant difference between the means but we still don't know which mean is greater.
- In such case, we have to proceed with a one sided test.

Z-test for the mean of teenager's (17-21) data:

H_0: There is no difference between teenager's (17-21) mean score and mean score for all age groups.

H_A: There is a significant difference between teenager's (17-21) mean score and mean score for all age groups.

Results:

	Extraversion	Agreeableness	Conscientiousness	Openness	Neuroticism
Z Value	-0.790229308	-0.719984144	-1.411036031	-0.640508116	2.70249302
P Value	0.429393863	0.471534758	0.15823399	0.521842315	0.006882163
Inference	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to accept null hypothesis

There is enough evidence to reject null hypothesis for **Neuroticism** score and we can conclude there is significant difference between the means but we still don't know which mean is greater. Hence, we have to proceed with a one sided test.

One-sided Test:

H_0: Mean Neuroticism score of teenager's (17-21) greater than or equal to the mean Neuroticism score of all age groups.

H_A: Mean Neuroticism score of teenager's (17-21) lesser the mean Neuroticism score of all age groups.

Result:

- P value is 0.996558918526744
- We failed to reject the null hypothesis.
- Mean Neuroticism score of teenager's (17-21) greater than or equal to the mean Neuroticism score of all age groups.

Z-test for the mean of females between age 17-21 and females of ages 22-30:

H_0: There is no difference between the mean of females between ages 17-21 and females of ages 22-30.

H_A: There is a significant difference between mean of females between ages 17-21 and females of ages 22-30.

Results:

	Extraversion	Agreeableness	Conscientiousness	Openness	Neuroticism
Z Value	-1.425835369	-0.828796964	1.514502108	-2.403026475	-1.18756277
P Value	0.153915869	0.407219306	0.12989854	0.01626001	0.235005713
Inference	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to accept null hypothesis	Failed to reject null hypothesis

There is enough evidence to reject null hypothesis for **Openness** score and we can conclude there is significant difference between the means but we still don't know which mean is greater. Hence, we have to proceed with a one sided test.

One-sided Test:

H_0: Mean Conscientiousness score of females between ages 17-21 is greater than or equal to the Mean Conscientiousness score of females between ages 22-30.

H_A: Mean Conscientiousness score of females between ages 17-21 is lesser than the Mean Conscientiousness score of females between ages 22-30.

Result:

- P value is 0.008130005084477997
- We failed to accept null hypothesis.
- Mean Conscientiousness score of females between ages 17-21 is lesser than the Mean Conscientiousness score of females between ages 22-30.

Z-test for the mean of males between ages 17-21 and males of ages 22-30:

H_0: There is no difference between the mean of males between ages 17-21 and males of ages 22-30.

H_A: There is a significant difference between mean of males between ages 17-21 and males of ages 22-30.

Results:

	Extraversion	Agreeableness	Conscientiousness	Openness	Neuroticism
Z Value	0.828005309	-0.768530435	1.24601762	0.045344051	-0.939914355
P Value	0.407667493	0.442172116	0.212757926	0.963833076	0.347261493
Inference	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis

Hence, there is no difference between the mean of males between ages 17-21 and males of ages 22-30.

Z-test for the mean of males and females between ages 17-21:

H_0: There is no difference between the mean of males between ages 17-21 and females of ages 17-21.

H_A: There is a significant difference between mean of males between ages 17-21 and females of ages 17-21.

Results:

	Extraversion	Agreeableness	Conscientiousness	Openness	Neuroticism
Z Value	1.129618538	-0.209141391	0.224172623	0.237082846	-2.002698781
P Value	0.258636996	0.834337864	0.822622985	0.81259252	0.04520963
Inference	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to accept null hypothesis

There is enough evidence to reject null hypothesis for **Neuroticism** score and we can conclude there is significant difference between the means but we still don't know which mean is greater. Hence, we have to proceed with a one sided test.

One-sided Test:

H_0: Mean Neuroticism score of females between ages 17-21 is lesser than or equal to the Mean Neuroticism score of males between ages 17-21.

H_A: Mean Neuroticism score of females between ages 17-21 is greater than the Mean Neuroticism score of males between ages 17-21.

Result:

- P value is 0.022604814842892257
- We failed to accept the null hypothesis.
- Mean Neuroticism score of females between ages 17-21 is greater than the Mean Neuroticism score of males between ages 17-21.

Z-test for the mean of males and females between ages 22-30:

H_0: There is no difference between the mean of males between ages 22-30 and females of ages 22-30.

H_A: There is a significant difference between mean of males between ages 22-30 and females of ages 22-30.

Results:

	Extraversion	Agreeableness	Conscientiousness	Openness	Neuroticism
Z Value	1.146559838	0.101763238	1.959327158	-0.243525782	2.185942823
P Value	0.251563583	0.918944606	0.050074485	0.807598115	0.028819789
Inference	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to accept null hypothesis

There is enough evidence to reject null hypothesis for **Neuroticism** score and we can conclude there is significant difference between the means but we still don't know which mean is greater. Hence, we have to proceed with a one sided test.

One-sided Test:

H_0: Mean Neuroticism score of females between ages 22-30 is greater than or equal to the Mean Neuroticism score of males between ages 22-30.

H_A: Mean Neuroticism score of females between ages 22-20 is lesser than the Mean Neuroticism score of males between ages 22-30.

Result:

- P value is 0.9855901056371492
- We failed to accept the null hypothesis.
- Mean Neuroticism score of females between ages 22-20 is lesser than the Mean Neuroticism score of males between ages 22-30.

Z-test for the mean of males and females having annual income of less than 3,00,000:

H_0: There is no difference between the mean of males and females having annual income of less than 3,00,000.

H_A: There is a significant difference between mean of males and females having annual income of less than 3,00,000.

Results:

	Extraversion	Agreeableness	Conscientiousness	Openness	Neuroticism
Z Value	-0.10126863	-0.177615103	0.670711084	-0.592540353	2.337701877
P Value	0.919337218	0.859025263	0.5024046	0.553488807	0.019402719
Inference	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to reject null hypothesis	Failed to accept null hypothesis

There is enough evidence to reject null hypothesis for **Neuroticism** score and we can conclude there is significant difference between the means but we still don't know which mean is greater. Hence, we have to proceed with a one sided test.

One-sided Test:

H_0: Mean Neuroticism score of females having annual income of less than 3,00,000 is greater than or equal to the Mean Neuroticism score of males having annual income of less than 3,00,000.

H_A: Mean Neuroticism score of females having annual income of less than 3,00,000 is lesser than the Mean Neuroticism score of males having annual income of less than 3,00,000.

Result:

- P value is 0.9902986406174367
- We failed to reject the null hypothesis.
- Mean Neuroticism score of females having annual income of less than 3,00,000 is greater than or equal to the Mean Neuroticism score of males having annual income of less than 3,00,000.

Inference

The following inferences could be sketched out from the hypotheses tests conducted:

- Neuroticism of teenager's between the age group of 17-21 greater than or equal to the Neuroticism of all other age groups.
- Conscientiousness of females between ages 17-21 is lesser than the Conscientiousness of females between ages 22-30.
- There is no difference between the traits of males between ages 17-21 and males of ages 22-30.
- Neuroticism of females between ages 17-21 is greater than the Neuroticism of males between ages 17-21.
- Neuroticism of females between ages 22-20 is lesser than the Neuroticism of males between ages 22-30.
- Neuroticism of females having annual income of less than 3,00,000 is greater than or equal to the Neuroticism of males having annual income of less than 3,00,000.

Conclusion

Many health problems, including depression, schizophrenia, diabetes, asthma, irritable bowel syndrome, and heart disease, appear to be linked to neuroticism. People who have a high level of neuroticism are more prone to mood disorders like depression. Low agreeability has also been connected to an increased risk of health issues. Neuroticism is greater within the age group of 17-21 with respect to other age groups, and within the mentioned age group, females are more prone to neuroticism. Though for the age group of 22-30, it can be inferred to be greater in male than female. Females having an annual income of less than 3,00,000 are also more subjected to neuroticism than males of the same income group.

Females of age group of 17-21 are less conscientious than females of age group of 22-30. Conscientiousness appears to be a preventive factor against sickness, according to research. People with a high conscientiousness score have been shown to have greater health and longevity. Conscientious people, according to researchers, have regular and well-structured lifestyles, as well as the impulse control to stick to diets, treatment regimens, and so on.

References

- Ackerman, C. (2017, June 23). *Big Five Personality Traits: The OCEAN Model Explained*. PositivePsychology.com. https://positivepsychology.com/big-five-personality-theory
- Cherry, K. (2019, October 14). *What Are the Big 5 Personality Traits? Verywell Mind*. Retrieved 12 June 2020, from https://www.verywellmind.com/the-big-five-personality-dimensions-2795422
- Costa, P., Terracciano, A., & McCrae, R. (2001). Gender Differences in Personality Traits Across Cultures: Robust and Surprising Findings. *Journal of Personality and Social Psychology, 81*(2), 322-331. https://doi.org/10.1037/0022-3514.81.2.322
- Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology, 44*(3), 329-344. https://doi.org/10.1037/h0057198
- Grohol, J. M. (2019, May 30). *The Big Five Personality Traits*. PsychCentral. Retrieved 10 June 2020, from https://psychcentral.com/lib/the-big-five-personality-traits
- Gurven, M., von Rueden, C., Massenkoff, M., Kaplan, H., & Lero Vie, M. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager-farmers in the Bolivian Amazon. *Journal of personality and social psychology, 104*(2), 354–370. https://doi.org/10.1037/a0030841
- Jang, K. L., Livesley, W. J., & Vernon, P. A. (1996). Heritability of the Big Five Personality Dimensions and Their Facets: A Twin Study. *Journal of Personality, 64*(3), 577–592. https://doi.org/10.1111/j.1467-6494.1996.tb00522.x
- Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology, 74*(6), 1556–1565.
- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). New York: Guilford Press.

- Lahey B. B. (2009). Public health significance of neuroticism. *The American psychologist, 64*(4), 241–256. https://doi.org/10.1037/a0015309
- Loehlin, J. C., McCrae, R. R., Costa, P. T., & John, O. P. (1998). Heritabilities of Common and Measure-Specific Components of the Big Five Personality Factors. *Journal of Research in Personality, 32*(4), 431–453. https://doi.org/10.1006/jrpe.1998.2225
- McCrae, R. R. (2002). Cross-Cultural Research on the Five-Factor Model of Personality. *Online Readings in Psychology and Culture, 4*(4). https://doi.org/10.9707/2307-0919.1038
- Myers, David G. (2011). *Psychology (10th ed.)*. Worth Publishers.
- Neal, A., Yeo, G., Koy, A., & Xiao, T. (2012). Predicting the form and direction of work role performance from the Big 5 model of personality traits. *Journal of Organizational Behavior, 33*(2), 175–192. https://doi.org/10.1002/job.742