**How demographics affect risk propensity: A data analytics study**

Sarra El Housseine, Ahmed Sulaiman Saif Al Abdulsalam

*Department of Analytics in the Digital Era*

*College of Business and Economics*

*United Arab Emirates University, Al Ain, UAE*

*201850013@uaeu.ac.ae, 202009528@uaeu.ac.ae*

**Abstract:** The objective of this research is to identify and understand how different demographic attributes are associated with risk propensity of people. The research method is the application of data analytics techniques and tools including statistical scoring, normality tests, and comparison of means. The source data was a survey conducted in an earlier research by 12 conductors and with 2,278 respondents from different educational and work backgrounds, as well as different universities. Gender, English level, and German level were found to be attributes that result in statistically significant differences in risk propensity scores. However, it was also found that the conductor of the survey may also be affecting the scores. Specifically in our analysis, we found that one conductor had much lower scores than others. So, as future research, the analysis can be conducted again, excluding the 237 survey responses & data points from that outlier conductor.

**1. Introduction**

This paper presents a research study that investigates how different demographics may be affecting a person’s propensity (inclination) to take financial risk.

To begin with, *risk management* is the process of analyzing, assessing and controlling threats to a certain company or an individual (Cole, 2020). Knowing how to manage risk can really benefit a person or individual. How likely or the tendency of someone to take a risk is called risk propensity (Wang *et al.,* 2016).

It is important to understand risk propensity because it helps analyze the risk behavior of individuals and offer the most suitable financial services to that person. To this end, in order to provide better guidance to the development and offering of financial products and their parameters, we obtained results in research on how willing a person is in taking risks in our research study.

In our study, we used risk scores of individuals as a measure of their risk propensity. Risk scoring is a methodology that is used to calculate how likely an individual is to take risks based on selected number of factors.

When analyzing the results, it is shown that different demographic attributes are associated with different risk scores, which also means different risk propensity with statistical significance. Higher risk scores refer to increased propensity for risk taking.



**2. Literature**

Every business or individual needs to learn how to manage and maintain risk, it is a crucial aspect and skill to learn because it reduces the damage of risk on the firm or individual, with efficient risk management we can prepare for the unexpected and minimize our costs.

There are multiple research studies that investigate the relation between risk propensity and demographics. (Alleyne & Broome, 2011) analyzes the factors that affect the how likely an investor is going to invest. The research conducted a survey to implement the analyzation, it has been found that different views, subjective social customs, risk propensity, and recognized behavioral control were important aspects in finding out the intentions of investment. It is reported from the results that subjective norms is the significant indicator of investment intention for an investor, as investors are keen to look for others for the approval of decision.

Wang et *al.* (2016) develop the study of how construction project managers of different personalities look at the risk of projects. The Big Five personality model. It was found from a sample of 152 participants that Extraversion, Agreeableness, and Conscientiousness have a huge effect on risk propensity. Adding to that, risk propensity has a negative consequence on risk perception. Moreover, Extraversion, Agreeableness, and Conscientiousness have important effects on risk perception. The findings were found by qualitative proof from five semi-structured interviews with project managers.

Yordanova & Alexandrova‐Boshnakova (2011) reports the role that gender plays on risk propensity, risk perception and risk behavior of entrepreneurs which includes direct and indirect gender outcomes. The risk behavior hypotheses is used on 382 entrepreneurs from Bulgaria. The results from the findings include that business men and women have similar risk perceptions, however, female entrepreneurs are more prone to have a lower risk propensity than business men. The risk propensity which is effected by gender is attributed to risk preference, outcome history, and age.

**3. Methodology**

***3.1. Data collection***

*Subjects*

The survey was conducted in Turkey, in Turkish language. The number of respondents was 656, with 346 working people vs. 250 undergraduate students and 60 graduate students. And 283 females vs. 373 males. A balanced sample from different educational and work backgrounds, as well as different universities. Attribute value distributions are available in Appendix B of the supplementary document of the earlier research (Ertek *et al.,* 2012b). The data was assembled in a spreadsheet software, cleaned and anonymized (Ertek et el., 2012).

*Sample characteristics*

The survey investigated 23 direct, risk-related questions and 9 indirect (demographic) questions. All are popular questions in risk scoring surveys provided on the Web by several financial institutions. Following are the nine demographic attributes (Ertek *et al.,* 2012b), followed by the RiskScore calculated using a statistical algorithm (Ertek *et al.*, 2012a).

**Table 1.** List of selected demographic attributes in the dataset.

|  |  |
| --- | --- |
| **Attribute** | **Possible Values** |
| *Gender* | Male, female. |
| *Conductor* | Name of survey conductor. |
| *IsStudent* | Student, working. |
| *StudentLevel* | Masters, undergrad, NA. |
| *IncomeType* | Fixed salary, incentive based, or a combination of both. |
| *SoccerTeam* | The soccer team supported by the person. |
| *HighSchoolType* | a) public high school for which the medium of instruction is in a foreign language[[1]](#footnote-1), b) public science high school, c) private science high school, d) private high school, or e) other. |
| *EnglishLevel* | a.) advanced, b.) intermediate, c.) beginner, d.) none. |
| *GermanLevel* | a.) advanced, b.) intermediate, c.) beginner, d.) none. |
| *FrenchLevel* | a.) advanced, b.) intermediate, c.) beginner, d.) none. |

***3.2. Risk scoring***

The risk scoring algorithm, coded in Matlab computing environment, determines direct attributes appropriate for the risk score, their weights, and the calculated risk score for all respondents. Consequently, a certain top percentage are labeled as risk-seeking and the rest as risk-averse. The weights obtained for the direct risk-related attributes are demonstrated in Fig. \_. And five out of the 23 direct attributes (Q20, Q21, Q22, Q38, Q40) were eliminated by the algorithm—assigned a weight of 0, statistically insignificant. The range of the positive weights is (0.2792, 1.6320). Among all attributes, the hypothesized directions of choice ranks are correct (Γj = 1,∀ j ∈ J). The histogram of the calculated risk scores labels 20% of the respondents as risk-seeking. (Fig. \_) Shapiro-Wilk normality test, conducted in R statistical package, resulted in p = 3.2E − 7 0.05, implying a non-normal distribution.

***3.3. Nonparametric statistics***

*Shapiro-Wilk normality test*

The Shapiro-Wilk normality test, developed by Samuel Sanford Shapiro and Martin Wilk, is the fittest and most omnibus test in so many situations. It is based on the correlation between the data and the corresponding normal scores. For a significant W static, or concordance W coefficient, the researcher rejects their hypothesis that the distribution is a normal one. Test statistic: W = , where is the i-th largest order statistic, x- is the sample mean, and n is the number of observations.

*Comparison of means*

A t-test is an inferential statistic, hypothesis testing tool, that tells whether there is significant difference between the means of two groups that have a potential relationship. A t-test is usually used when the data sets are normally distributed and might have hiding variances.

T-Test assumptions:

1. The scale of measurement follows a continuous or ordinal scale, e.g. IQ scores.
2. In “a simple random sample, the data is collected from a representative, randomly selected portion of the total population.”
3. Data plotted is normally distributed.
4. Variance is homogeneous; samples have similar standard deviations.

To determine whether the difference is inherent and related to the study, a t-value is calculated, obtained by the t-test, using the mean differences, the standard deviations of each group, and how many data values each group has. The t-value is a ratio of the mean difference to the variation. The higher the t-value (t-score) the more difference there is between groups.

In order to determine if two independent groups come from the same population, the Mann–Whitney U test, a distribution-free test, is used. The test statistic U is calculated by comparing each two values, with a score of 1 or 0 depending on which group observation is higher or lower. It equals the resulting scores from all pairs. It is then used in the appropriate table of critical values to determine whether the null hypothesis of no difference between the two data sets can be rejected. Mann–Whitney U can be more practical than the t-test at times, for it needs less assumptions and conditions.

After finding what groups (attributes in this case) have significant difference, Dunn’s test, a post hoc non parametric test, determines which specific means are significant from the others. It is used for multi sample comparison i.e. three or more means. Its null hypothesis is that there is no difference between groups—groups can differ in size.

***3.4. Data analytics workflow***

While earlier research of Ertek et al. (2012a) developed a risk scoring method and calculated the risk scores for the collected data, there was not an analysis of the relation between the demographic attributes and the risk scores. This was the research gap that we wanted to fill in our research.

First, we would check for normality using Shapiro-Wilk tests for each attribute value. If the attribute had two values (e. g. male vs. female) we would test the differences using t-tests if both values were normal, and Mann-Whitney U tests if one at least was not normal. All via Microsoft Excel. If the attribute had more than two values, we would carry the comparisons by Dunn tests, using R statistical package. [Kruskal–Wallis?]

**4. Analysis**

We analyzed the risk score of individuals that were computed based on the methodology of Ertek *et al.* (2012a) and survey data from the same earlier research. The questions in the survey can be found in Ertek *et al.* (2012b).

The survey included attributes related to financial risk taking, such as “What proportion of your wealth would you feel comfortable allocating to stock or bonds markets?”, as well as demogrphic attributes, such as gender, favorite soccer team, English level and German level. The full list of demographic attributes is as follows:

* Gender
* IsStudent
* StudentLevel
* IncomeType
* SoccerTeam
* HighschoolType
* EnglishLevel
* GermanLevel
* FrenchLevel

Furthermore, for each survey response, the Conductor was also noted. Later on in our research, we realized that survey data from a particular conductor, referred to as “NE” may not be correct, as the risk scores of respondents from “NE” were significantly lower than the scores from other conductors.

While earlier research of Ertek *et al.* (2012a) developed a risk scoring method and calculated the risk scores for the collected data, there was no analysis of the relation between the demographic attributes and the risk scores. This was the research gap that we wanted to fill in our research.

Through our analysis, we found that even indirect attributes that seem insignificant such as gender, play a huge role in the risk propensity of people, as we analyzed and came to the conclusion that males have higher risk scores than females using the Mann-Whitney test, as provided in Table 3.

The factors that affect risk propensity can either be direct factors like the question from the previous paragraph mentioned “what proportion of your wealth would you feel comfortable allocating to stock or bonds markets?” in which the risk propensity can be directly attributed, as opposed to indirect attributes such as German level or English level, as we collected from our survey and analyzed in our research, we found out that such attributes can be associated with risk propensity. For example, we found that people with higher English levels are more risk seeking than those with lower English levels, as shown in Figure 3.

For each survey response, the Conductor was also noted. Later on in our research, we realized that survey data from a particular conductor, referred to as “NE” may not be correct, as the risk scores of respondents from “NE” were significantly lower than the scores from other conductors.

The factors that affect risk propensity can either be direct factors like the question from the previous paragraph mentioned “what proportion of your wealth would you feel comfortable allocating to stock or bonds markets?” in which the risk propensity can be directly attributed, as opposed to indirect attributes such as German level or English level, as we collected from our survey and analyzed in our research, we found out that such attributes can be associated with risk propensity. For example, we found that people with higher English levels are more risk seeking than those with lower English levels, as shown in Figure 3

Through our analysis, we found that even indirect attributes that seem insignificant such as gender, play a huge role in the risk propensity of people, as we analyzed and came to the conclusion that males have higher risk scores than females using the Mann-Whitney test, as provided in T

*Results of normality tests*

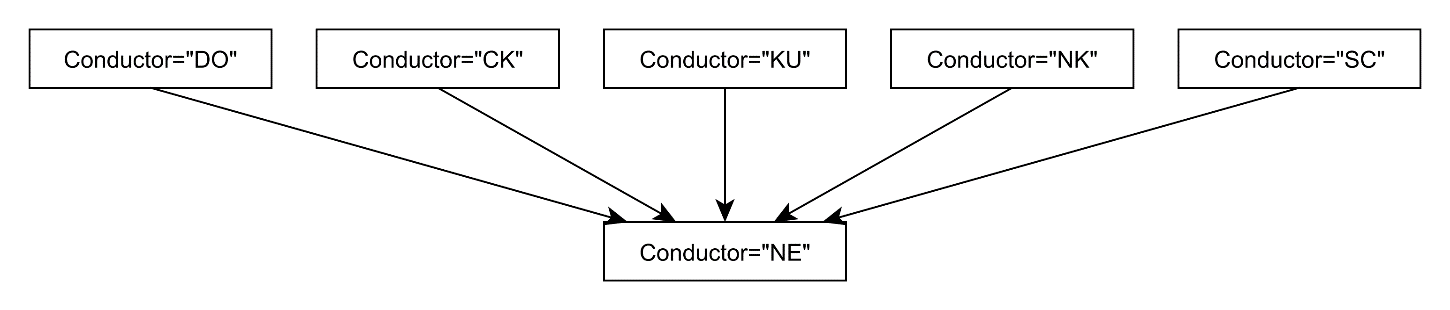
**Table 2.** Results of Shapiro-Wilk normality test for groups within selected attributes.

|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute** | **AttributeValue** | **p-value** | **TestResult** |
| Gender | M | 0.3274 | Normal |
|  | F | 0.0029 | NotNormal |
| Conductor | SC | 0.0303 | NotNormal |
|  | NE | 0.1317 | Normal |
|  | CK | 0.8307 | Normal |
|  | KU | 0.2154 | Normal |
|  | NBO | 0.6090 | Normal |
|  | OG | 0.0040 | NotNormal |
|  | TY | 0.2473 | Normal |
|  | USK | 0.3817 | Normal |
|  | NK | 0.3087 | Normal |
|  | HEG | 0.2709 | Normal |
|  | DC | 0.1088 | Normal |
|  | DO | 0.0522 | Normal |
| IsStudent | Student | 0.0000 | NotNormal |
|  | Working | 0.0638 | Normal |
| StudentLevel | Master | 0.0529 | Normal |
|  | Undergrad | 0.0000 | NotNormal |
|  | NA | 0.0870 | Normal |
|  | Commission | 0.0134 | NotNormal |
|  | Salary | 0.3722 | Normal |
|  | SalaryComm | 0.0207 | NotNormal |
| SoccerTeam | Galatasaray | 0.0012 | NotNormal |
|  | Besiktas | 0.3225 | Normal |
|  | Fenerbahce | 0.5309 | Normal |
|  | Other | 0.2356 | Normal |
|  | Trabzon | 0.5309 | Normal |
| HighschoolType | a | 0.0530 | Normal |
|  | b | 0.3207 | Normal |
|  | c | 0.2132 | Normal |
|  | d | 0.0222 | NotNormal |
|  | e | 0.0042 | NotNormal |
| EnglishLevel | a | 0.0009 | NotNormal |
|  | b | 0.0002 | NotNormal |
|  | c | 0.0459 | NotNormal |
|  | d | 0.0009 | NotNormal |
| GermanLevel | a | 0.0721 | Normal |
|  | b | 0.3072 | Normal |
|  | c | 0.1311 | Normal |
|  | d | 0.0002 | NotNormal |
| FrenchLevel | a | 0.0474 | NotNormal |
|  | b | 0.0778 | Normal |
|  | c | 0.0778 | Normal |
|  | d | 0.0010 | NotNormal |

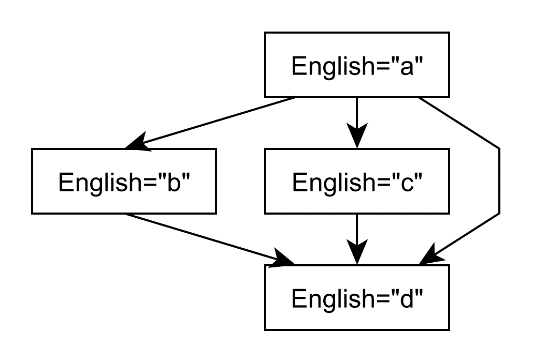
*Results of comparison of means*

**Table 3.** Results of comparison of means tests for groups within selected attributes.

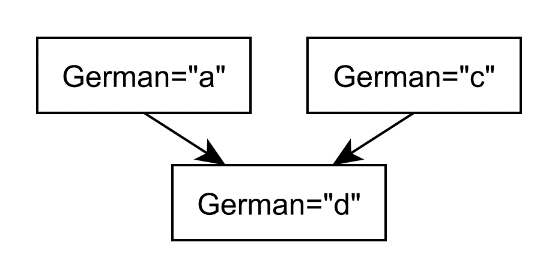
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **VariableName** | **Results** | **IsConfirmed** | **TestApplied** | **p-value** |
| **Gender** | **Males have higher mean risk score than females.** | **Yes** | **MannWhitneyU** | **0.000** |
| IsStudent | Students and working professionals risk scores are similar. | Yes | MannWhitneyU | 0.688 |
| IncomeType | People with different stated sources of income have similar risk scores. | Yes | KruskalWallis | 0.328 |
| **Conductor** | **People interviewed by different conductors have different risk scores.** | **Yes** | **KruskalWallis** | **0.000** |
| **EnglishLevel** | **Respondents with different English levels have different risk scores.** | **Yes** | **KruskalWallis** | **0.000** |
| HighSchoolType | People from different high school types have similar risk scores. | Yes | KruskalWallis | 0.283 |
| FrenchLevel | Respondents with different French levels have similar risk scores. | Yes | KruskalWallis | 0.233 |
| **GermanLevel** | **Different German levels among respondents also resulted in different risk scores** | **Yes** | **KruskalWallis** | **0.000** |
| SoccerTeam | The variety in soccer team preference does not affect risk scores. | Yes | KruskalWallis | 0.997 |



**Figure 1.** Statistically significant differences between mean risk scores with respect to Conductor.



**Figure 2.** Statistically significant differences between mean risk scores with respect to English level.



**Figure 3.** Statistically significant differences between mean risk scores with respect to German level.

**Conclusions**

In this research we analyzed the risk scores of individuals which were surveyed by Ertek *et al.* (2012a) and also followed the methodology which was developed along with the same data which was obtained. We used a risk scoring methodology to find out the risk propensity of individuals based on different demographics, to study the effect of the various demographics on the risk propensity of people, thus calculating the risk score, higher risk score means higher tendency to take risks.

The methodology of tests that were used are as follows (we analyzed the indirect attributed data only and eliminated direct attributed data from the survey conducted) :

* The normality test to find out if the distribution of data is normal
* Mann-Whitney test for data of two sets such as gender, and KruskalWallis and Dunn test for more than two data sets such as English level. We used all of these tests to compafe between the data sets in regards to risk scores, as we did with gender and found that males are more likely to take risks than females.

When the results were analyzed it is concluded that different demographic backgrounds do have a statistical effect on the risk propensity of people, as we found that males for example, are more likely to take risks than females as it is shown in table 3. So indirect attributes like gender can affect the tendency to take risks

**Acknowledgement**

This project would have been impossible without the guidance and support of our mentor Dr. Gurdal Ertek. We would like to express our sincere gratitude to him for his continuous motivation and patience, and for his knowledge that paved the way for us.

**Appendix A. Selected Survey Questions**

**Q2.** (**Gender**) What is your gender?

*Male  
Female*

**Q4.** (**Conductor**) Name of conductor:

**Q5.** (**IsWorking**) Are you a student or are you working full time?

*Student  
Working*

**Q7.** (**StudentLevel**) If you are a student, what is your level of study?

*Undergrad  
Masters  
Ph.D.   
NA (Not applicable)*

**Q10.** (**IncomeType**) If you are working full time, what is the structure of your income?

*Commission  
Salary  
Both salary and comission   
NA (Not applicable)*

**Q12.** (**HighSchoolType**) What type of high school have you graduated from?

*a. Public high school with medium of instruction in a foreign language  
b. Public science high school  
c. Private science high school   
d. Private high school   
e. Other*

**Q13.** (**SoccerTeam**) Which soccer team do you support?

*Besiktas  
Fenerbahce  
Galatasaray   
Trabzonspor  
Other*

**Q15.** (**EnglishLevel**) What is the level of your English language?

*a. Advanced  
b. Intermediate  
c. Beginner   
d. None*

**Q16.** (**GermanLevel**) What is the level of your German language?

*a. Advanced  
b. Intermediate  
c. Beginner   
d. None*

**Q17.** (**FrenchLevel**) What is the level of your French language?

*a. Advanced  
b. Intermediate  
c. Beginner   
d. None*

**Appendix B. Detailed Results**

*Average Risk Scores for Gender Groups*

|  |  |
| --- | --- |
| Gender | Average of RiskScore |
| F | 137.42 |
| M | 140.17 |

*Average Risk Scores for Conductor Groups*

|  |  |
| --- | --- |
| Conductor | Average of RiskScore |
| *Blank* | 140.96 |
| CK | 140.83 |
| KU | 140.78 |
| SC | 139.81 |
| DO | 139.14 |
| USK | 138.78 |
| NBO | 138.69 |
| NK | 138.61 |
| TY | 138.55 |
| HEG | 138.39 |
| DC | 138.26 |
| OG | 138.24 |
| NE | 136.54 |

*Average Risk Scores for English Level Groups*

|  |  |
| --- | --- |
| English Level | Average of RiskScore |
| a | 139.78 |
| b | 138.63 |
| c | 137.77 |
| d | 134.27 |

*Average Risk Scores for German Level Groups*

|  |  |
| --- | --- |
| German Level | Average of RiskScore |
| a | 140.3109 |
| b | 138.7968 |
| c | 140.1987 |
| d | 138.1688 |

**References**

1. Alleyne, P. & Broome, T. (2011). Using the theory of planned behaviour and risk propensity to measure investment intentions among future investors. *Journal of Eastern Caribbean Studies*, *36*(1), 1-21.
2. Cole, B. (2020, April 7). *risk management*. SearchCompliance. Available under <https://bit.ly/3tIfwCO>
3. Ertek, G., Kaya, M., Kefeli, C., Onur, Ö., Uzer, K. (2012a) “Scoring and predicting risk preferences” in Behavior Computing: Modeling, Analysis, Mining and Decision. Eds: Longbing Cao, Philip S. Yu. Springer.
4. Ertek, G., Kaya, M., Kefeli, C., Onur, Ö., and Uzer, K. (2012b) Supplementary Document for “Scoring and Predicting Risk Preferences”. Accessed on March 17, 2021. Available under <https://ertekprojects.com/gurdal-ertek-publications/wp-content/uploads/2014/07/03.pdf>
5. Wang, C. M., Xu, B. B., Zhang, S. J., & Chen, Y. Q. (2016). Influence of personality and risk propensity on risk perception of Chinese construction project managers. *International Journal of Project Management*, *34*(7), 1294-1304. <https://doi.org/10.1016/j.ijproman.2016.07.004/>
6. Yordanova, D. I., & Alexandrova‐Boshnakova, M. I. (2011). Gender effects on risk‐taking of entrepreneurs: evidence from Bulgaria. *International Journal of Entrepreneurial Behavior & Research*. Vol. 17 No. 3, pp. 272-295. <https://doi.org/10.1108/13552551111130718>

1. [↑](#footnote-ref-1)