Is Ignorence Truly Bliss? Relationship between Education, Gender & Happiness

Ece Cavusgil

Abstract

Although we think we live in a developed society, today the distinction between women and men still shows itself in many areas and places in Turkey. Undoubtedly, the best solution to this is to educate individuals in order to create awareness about the position of women in society. Thus, a society that educates itself will be able to live the “happy” life it deserves. In this study, the behavior of happiness levels depending on education level and gender throughout the country was examined from this perspective, and Turkey was mapped in this context. According to these two factors, estimated regression models were defined for the levels of individuals feeling happy.

1.Introduction

The source of motivation for human beings in terms of life is subjective and can change over time, but for most people, this is “happiness”. One of the factors that causes a person to give this answer and provides the subjectivity here may actually be their level of education. At this point, how an increase in the level of education will affect a person's perspective on life is an important question. The question of whether ignorance is bliss can actually be used as a tool to find an alternative answer to this question outside the context it expresses. In this motivation we are talking about, we can assume that gender has a great effect, especially considering the living conditions such as the social environment and culture to which the person belongs.

The fact that individuals' life satisfaction may vary depending on gender and education level factors brings with it important questions such as how and to what extent these interactions occur. The aim of this study is to reveal whether there is a relationship between education level, gender and life satisfaction.

2.Data

In order to perform the analysis within the scope of the study, three different datasets were used from the Life Satisfaction Survey [1] and Population Statics Portal [2] sources created by Turkstat. The datasets were subjected to preprocessing processes in order to scan for missing data (NA) and to ensure information equality between them. The datasets created after this step are “education”, “byeducation” and “bygender”.

2.1. “education” Dataset

This dataset includes the total number of people, the number of women/men and their percentages at different education levels in each province in Turkey between the years 2008-2023. The variables (columns) and definitions of the dataset are as follows.

- *Year*: The year of the study (ranging from 2008 to 2023).
- *Province*: Name of the province (81 provinces in total).
- *Educational_Status*: Education level. Categorical variable that has levels of *"No School Completed," "Primary School," "Primary Education or Junior High School," "High School or Equivalent," "Higher Education"*.
- *Total*: Total number of individuals in a given year, province, and education level.
- *Male/Female*: Number of males/females in a given year, province, and education level.
- *Percentage_Male/Female*: Percentage of males/females in a given year, province, and education level.

Descriptive statistics for variables are presented in Figure 1.

```
     Year             Province
 Length:6480        Length:6480
 Class :character   Class :character
 Mode  :character   Mode  :character


                                    Educational_Status     Total
 No School Completed                        :1296      Min.   :    681
 Primary School                             :1296      1st Qu.:  42114
 Primary Education or Junior High School:1296      Median :  86103
 High School or Equivalent                  :1296      Mean   : 190227
 Higher Education                           :1296      3rd Qu.: 183118
                                                       Max.   :16154476
      Male              Female         Percentage_Male Percentage_Female
 Min.   :    257   Min.   :    424   Min.   : 0.00   Min.   : 0.00
 1st Qu.: 21807   1st Qu.:  19350   1st Qu.:12.70   1st Qu.:11.80
 Median : 43320   Median :  41185   Median :19.80   Median :18.70
 Mean   : 95128   Mean   :  95099   Mean   :19.45   Mean   :19.44
 3rd Qu.: 91336   3rd Qu.:  91346   3rd Qu.:25.82   3rd Qu.:25.80
 Max.   :7930608   Max.   :8223868   Max.   :54.80   Max.   :79.50
```

Figure 1. Descriptive statistics for “education” dataset.

2.2.“byeducation” Dataset

This dataset contains the percentage of general happiness levels by educational status between 2008 and 2023. The variables (columns) and definitions of the dataset are as follows.

- *Year*: The study year (ranging from 2008 to 2023).
- *Happiness_Level*: Levels of happiness ("Unhappy," "Neither Happy nor Unhappy," "Happy"). (Categorical Variable)
- Columns that indicates The percentage of individuals with respected educational level *(No School Completed, Primary School, Primary Education or Junior High School, High School or Equivalent and Higher Education)* for a given year and happiness level.

Descriptive statistics for variables are presented in Figure 2.

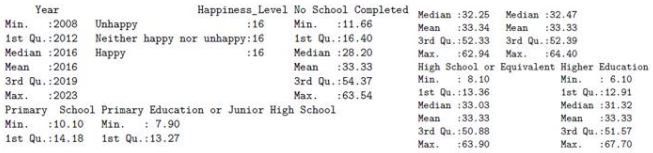


Figure 2. Descriptive statistics for “byeducation” dataset.

2.3.“bygender” Dataset

This dataset contains the percentage of general happiness levels by gender 2008 and 2023. The variables (columns) and definitions of the dataset are as follows.

- *Year*: The study year (ranging from 2008 to 2023).
- *Happiness_Level*: Levels of happiness ("Unhappy," "Neither Happy nor Unhappy," "Happy"). (Categorical Variable)
- Columns that indicates the percentage of total individuals and respected with gender.

Descriptive statistics for variables are presented in Figure 3.

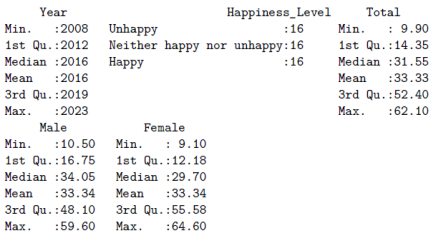


Figure 3. Descriptive statistics for “bygender” dataset.

3.Analysis

3.1. Exploratory Data Analysis

In this section, in order to get to know the data in more detail, visualizations with different variables were made and more detailed information about the variables were tried to be obtained. In order to make the process more efficient, the datasets were examined in order.

For the detailed examination of the “education” dataset, the first 10 provinces with the highest number of individuals with higher education and the highest number of people without education were examined on a yearly basis (Figure 4). In these education levels, there are similar provinces such as Istanbul, Ankara and Izmir and the number of people increases over time.

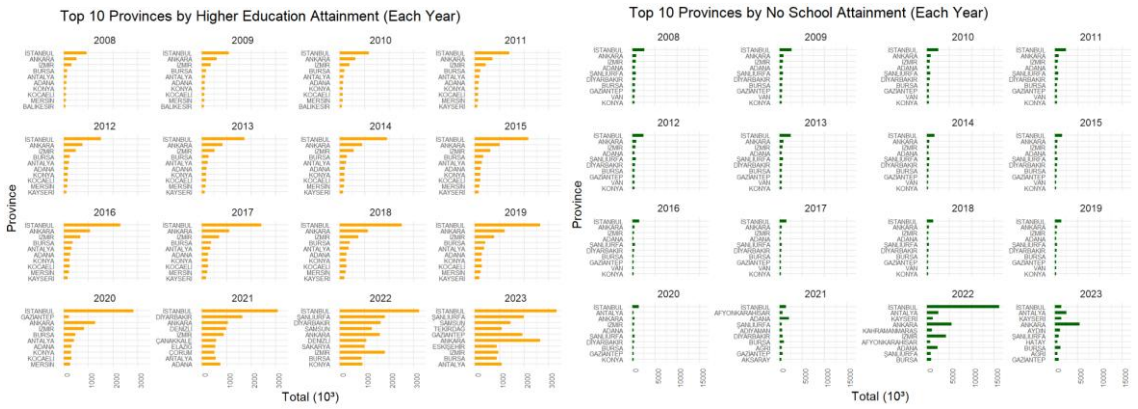


Figure 4. Top 10 provinces by higher education or none.

The numbers of women and men who do not have any education in the provinces are compared (Figure 5). Although this number seems to be higher for men in the provinces, when the ratios of women and men in terms of education levels are examined, it is seen that this is not exactly the case, indicating that the absolute magnitude between the gender is greater.

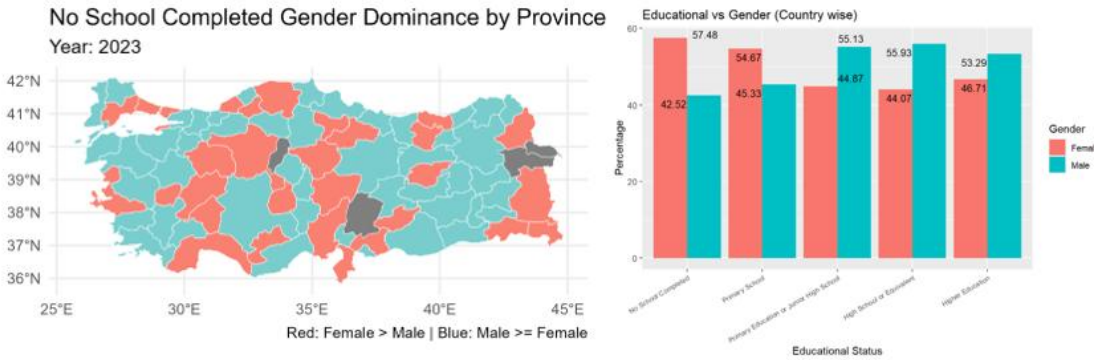


Figure 5. Analysis for "education" dataset by gender.

For a detailed examination of the "byeducation" dataset, the relationship between education levels and happiness levels was examined, and for the "bygender" dataset, the relationship between gender and happiness levels was examined (Figure 6).

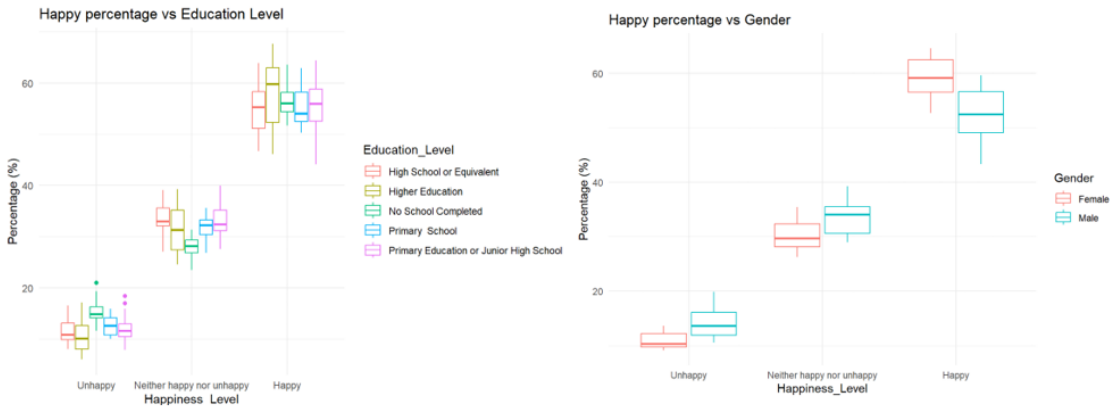


Figure 6. Happy Percentage by education level and gender.

According to the results obtained, while people with higher education feel happier, this situation is the opposite for people who have no education. This also happened in a parallel way in the case where the happiness level is "Unhappy". The majority of people who define themselves as happy are women, while the majority of people who define themselves as unhappy are men.

3.1. Trend Analysis

In this section, the behavior of the variables depending on time is examined. As in the previous sections, the datasets will be considered separately.

After determining the provinces with the highest level of education determined by looking at the “education” dataset, the provinces where the number of people with this level of education increased the fastest were examined as the next step (Figure 7). According to the results obtained, Istanbul and Izmir provinces continued their increases stably over time, while Ankara and later Konya and Bursa provinces experienced a very steep rise after 2020. The reason for this may be the increasing number of universities in recent years and the effect of migration between provinces.

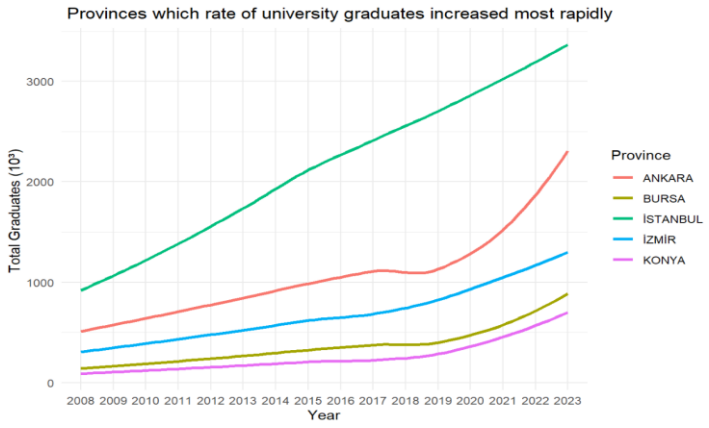


Figure 7. Provinces which rate of university graduates increased most rapidly.

In the previous sections of the analysis, in the studies conducted according to the “bygender” dataset, there were findings indicating that there were significant differences in the happiness levels of women and men. In this direction, an analysis can be made on the significance of the difference between men and women (Figure 8). It was observed that this difference was significant at the “happy” and “unhappy” levels (p-value = 0.046, p-value = 0.0268), and that it decreased by 0.24 for those who were happy over time, and increased by 0.17 for those who were unhappy.

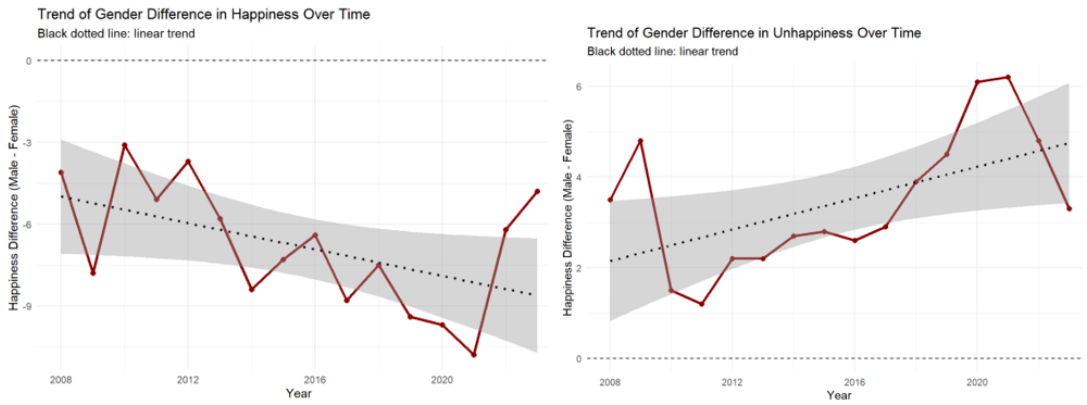
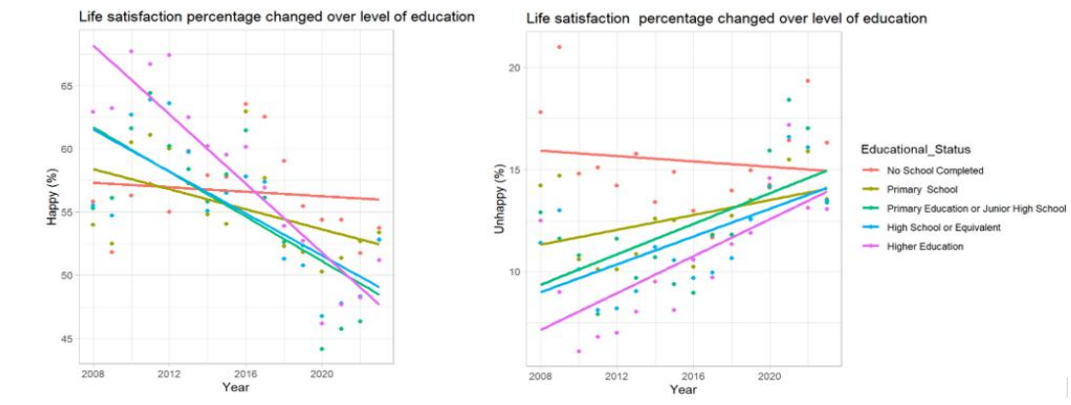


Figure 8. Trend of gender differences in happiness and unhappiness over time.

The "byeducation" dataset was used to examine the behavior of the percentages of individuals with different levels of education feeling happy and unhappy over time (Figure 9). The percentage of happiness decreased for all levels of education. The lowest level of decrease was for those with no education, with an annual average of 0.0089, and the highest level of decrease was for those with higher education, with an annual average of 1.369. The percentage of those who were unhappy decreased only for those without education, with an annual average of 0.45. The highest annual average was for those with higher education, with an annual average of 0.065, and it also increased for other levels of education.




```
# Groups:   Educational_Status [5]
  Educational_Status                  term  estimate std.error statistic p.value
  <fct>                               <chr>    <dbl>     <dbl>     <dbl>   <dbl>
1 No School Completed                 Year   -0.0883     0.189    -0.468 6.47e-1
2 Primary  School                     Year   -0.397      0.195    -2.04  6.11e-2
3 High School or Equivalent           Year   -0.832      0.213    -3.91  1.58e-3
4 Primary Education or Junior High S… Year   -0.884      0.231    -3.82  1.86e-3
5 Higher Education                    Year   -1.37       0.171    -7.99  1.38e-6
```

```
# Groups:   Educational_Status [5]
  Educational_Status                  term  estimate std.error statistic p.value
  <fct>                               <chr>    <dbl>     <dbl>     <dbl>   <dbl>
1 Higher Education                    Year    0.452     0.124      3.66  0.00257
2 Primary Education or Junior High S… Year    0.373     0.129      2.90  0.0116
3 High School or Equivalent           Year    0.340     0.112      3.03  0.00894
4 Primary  School                     Year    0.182     0.0967     1.89  0.0802
5 No School Completed                 Year   -0.0657    0.132     -0.498 0.626
```

Figure 9. The percentages of people with different levels of education feeling happy and unhappy over time.

3.1. Model Fitting

In this section, according to the problem addressed in the study, the variables affecting the happiness levels of individuals and their relationships with their happiness levels were statistically examined and the model that would predict their future happiness percentages was developed according to the results obtained.

First, the Pearson`s Chi Square [3] test was used to test whether there was a relationship between two categorical variables and the relationship between happiness levels and education levels and gender. As seen in Figure 10, there is a significant (p-value < 0.05) relationship between happiness levels and education levels, and as seen in Figure 11, there is a significant (p-value < 0.05) relationship between happiness levels and gender.

```
        Pearson's Chi-squared test

data:  contingency_matrix
X-squared = 277.05, df = 8, p-value < 2.2e-16
```

Figure 10. Pearson`s Chi-Squared Test for Happiness Level vs Education Level

```
        Pearson's Chi-squared test

data:  contingency_matrix1
X-squared = 170.61, df = 2, p-value < 2.2e-16
```

Figure 11. Pearson`s Chi-Squared Test for Happiness Level vs Gender

It was observed once again that the level of happiness ("Happy") was affected by education level and gender. In the next stage, separate prediction models were worked on according to education and gender variables to predict happiness levels.

Happiness Rate Prediction Model According to Education Levels;

This prediction model has been established in three different ways, Model 1 is a model that assumes that the effect of happiness rate is only affected by education levels. Model 2 also includes the effect of time in the prediction model. Model 3 is a model in which the interaction between education level and time is added to the prediction model. According to the results obtained, when the AIC (Akaike Information Criterion)[4] values of the models are compared, it is seen that the model that best explains the data is Model 3 with the lowest AIC value. Thus, it can be commented that the effect of the interaction between the year variable and education level should also be included in the prediction model to be created. Accordingly, in the prediction model to be created to find the percentage of happiness, the significant primary education or junior high school, high school and higher education levels and the interactions of these education levels with the year variable should be added. (In the created models, the reference category selection approach used when working with categorical data was used [5]. In this direction, the reference category is the first level for education level, the category in which no education was received.)

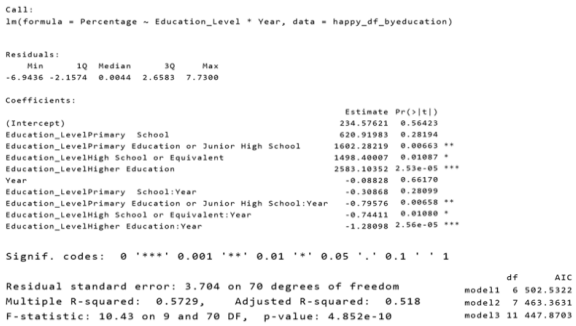


Figure 12. Choosen regression model for predict happiness percentage with educational level.

Happiness Rate Prediction Model According to Gender;

As in the previous prediction model, this model was established in three different ways and then the AIC values were compared to decide on the model that best explains the data. The approach of the three different models established is similar to the previous model. It was seen that the model that best explains the data was Model 3 with the lowest AIC value. Accordingly, the variables GenderFemale (0 for Male, 1 for Female) and Year, which are found to be significant in the prediction model to be created for the happiness percentage prediction, should be added. (The reference category selection approach used when working with categorical data was used in the created models [5]. In this direction, the reference category is the first level male categorical variable for gender.)

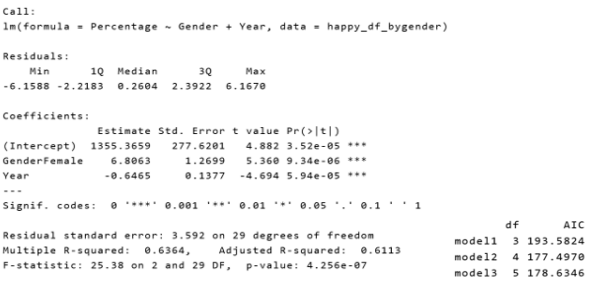


Figure 13. Choosen regression model for predict happiness percentage with gender.

The adequacy of both models was verified by drawing a Q-Q Plot to analyze whether the residuals were suitable for a normal distribution and by performing the Anderson-Darling Normality [6] test.

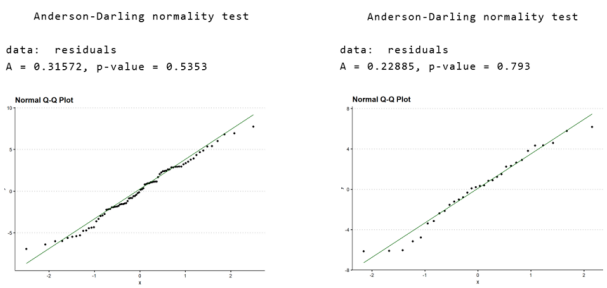


Figure 14. Choosen regression models for predict happiness percentage with educational level and gender (respectivley)

4.Conclusion

In this study, 3 different datasets were studied in order to examine how individuals' happiness is affected by factors such as gender and their level of education. From these datasets, information was obtained on the total number of people in different provinces in Turkey according to their level of education over the years, the ratio of women to men, the happiness levels that individuals defined themselves over the years according to different levels of education and gender, and the study was carried out in line with this information. In the analysis phase, as a first step, the structures of the variables and the changes they showed were observed and more detailed information about the variables was tried to be obtained. Then, this information obtained was further enriched by examining the behaviors of the variables over time. Finally, the effects of variables on happiness percentages were discussed and models based on gender and education level were established in order to estimate the happiness percentages of individuals.

The important results obtained as a result of this analysis are as follows.

• Among the cities where the number of people with university and higher education and also people without any education increased the most over time are Istanbul, Ankara and Izmir. While the increase in people with higher education was more stable in Istanbul and Izmir, this increase in Ankara increased very rapidly, especially after 2020. The majority of individuals with no education or primary school education in the country are women. The opposite is true for other education levels.

• It has been observed that individuals who define themselves as happy are mostly those with higher education, while those who define themselves as unhappy are mostly those who are not educated. However, the percentage of those who feel happy has decreased over time according to each level of education. The smallest decrease was 0.0089 in individuals with no education, and the largest decrease was 1.369 in individuals with higher education. The percentage of those who feel unhappy has decreased by 0.065 over time for individuals with no education. The percentage of feeling unhappy has increased over time in other levels of education. The largest increase was 0.45 in individuals with higher education.

• It has been observed that individuals who define themselves as happy are mostly women, while those who define themselves as unhappy are mostly men. However, the difference between men who define themselves as happy and women who define themselves as happy decreases from an average value of 0.24. The difference between men and women in individuals who define themselves as unhappy increases by 0.17 every year.

• Different regression models were created according to gender and education levels to estimate the percentage of happiness. Models created with the main effects and interactions of variables for different factors were compared, and the model that best explains the data was determined with the AIC value. Accordingly,

in the model created according to gender, the effect of time as well as gender was included in the model. In the model according to education levels, the main effects of primary school, high school and higher education levels and their interactions with the time variable were included.

According to these results, it is seen that educational disadvantage, especially in recent years, affects not only women but also men. The reasons for this should be considered from different perspectives and steps should be taken towards a solution. The level of education that individuals have creates a difference in their perspective and expectations on life and the level of happiness they define themselves in this way changes to a great extent. The fact that the percentage of people who do not receive any education feeling unhappy has increased over time, even if only slightly, has actually proven the saying that ignorance is bliss, the meaning of which we examine from a different perspective. Although women seem happier in general in the country, it should be approached with skepticism as to whether this reflects the truth in the living conditions we live in. When approached with this suspicion, the real truth is that the difference in happiness between men and women decreases over time, and women feel unhappier day by day.

References

[1] Turkish Statistical Institute (TURKSTAT). "Life Satisfaction Survey, 2023". Retrieved from https://data.tuik.gov.tr/ (accessed May 15, 2025)

[2.] Turkish Statistical Institute (TURKSTAT). (2024). "Population Statics Portal, 2024". Retrieved from https://nip.tuik.gov.tr/ (accessed May 15, 2025)

[3] K. Pearson, “On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[4] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[5] D. Firth and R. Menezes, “Quasi-variances,” *Biometrika*, vol. 91, no. 1, pp. 65–80, 2004.

[6] T. W. Anderson and D. A. Darling, “A test of goodness of fit,” *Journal of the American Statistical Association*, vol. 49, no. 268, pp. 765–769, 1954.