Major reasons accounting for the school dropout rate of Turkish women*

Turkey women's right to education has not been realized in 1998

Qiuhan Wang Yuxuan Yang Zihan Zhang

02 April 2022

Abstract

Education is vital for both national economic growth and personal lives, but in some areas, especially backward regions, the importance of education is ignored, and sexual discrimination in education influences women's opportunities as well. In this paper, we use the Turkish Demographic and Health Survey data in 1998 to analyze the main reasons for dropping out of school among Turkish young women (15-24 years old) with different levels of education. We find that the most important reason for their leaving is that their parents lack the awareness of the importance of education. As the public awareness of education gets increased through govenment promotion, we expect the women dropout rate in Turkey would drop and the education efficacy gets imporved.

*Code and data are available at: https://github.com/zihanjasmine/Turkey-DHS.

1 Introduction

Education is a key essence of modern society, which helps countries grow economically. Education cultivates talents for the country, and investing in education is considered a kind of way developing country because the citizens with a higher level of education are more likely to create higher GDP. Also, people with high educated can easily differentiate between right and wrong; thus, the crime rate can be reduced, which is beneficial to keeping society's stability. However, in some developing countries, the importance of education still be ignored, and sexual discrimination still exists. It's difficult for people, especially females, to have opportunities to accept higher education. In this paper, we aim to analyze why females leave school in Turkey and give suggestions to the government to reduce the dropout rate.

The paper "The relationship between economics and education" discussed the relationship between GDP per capita and average level of education. In addition, the negative impacts of unequal treatment between males and females are investigated. (Banton 1994) In many developed countries, such as the USA, the reasons for dropout are concluded, but it hasn't been down in Turkey before. (Holsinger 2009)

A short survey was created and conducted in 1998. Women aged 15-24 who were not currently attending school were asked about the main reason for giving up further studying in Turkey. The survey was conducted through the Turkish Demographic and Health Survey (TDHS-98). The data is authoritative and representative. From the responses, we find that fewer females in rural have a chance to complete primary school compared with urban ones. For females in both rural and urban, their parents didn't send them to school, becoming the most significant reason. Those women who have a chance to accept secondary education face a big challenge in passing the exam.

The article is structured as follows. First, we introduce the survey, including the sample size, population size, and survey method. Second, we calculate the percentage of females in Turkey attending school. Analyzing the various reasons females leave schools in rural and urban areas separately and finding out the most important reasons that cause a high dropout rate are necessary. Finally, some advice on improving females' educational opportunities is suggested.

2 Data

2.1 Data collection

All the data used in this paper was retrieved from "Turkey DHS, 1998, Final Report" on The Demographic and Health Surveys Program (DHS program).(Ankara 1999) The DHS Program collaborates with governments to gather and share critical data about individuals, their health, and their healthcare systems. DHS program assists developing countries to collect primary data through several types of questionnaires. This Turkey DHS report we chose illustrates a transaction in the structure of Turkey's population. To be more specific, it is a survey designed to provide information on fertility levels and trends, infant and child mortality rates, families, and maternal and child health with a nationally representative sample.

2.2 Data processing

All analyses were done with a statistic programming language R. (R Core Team 2020) First, we use the R package "pdftools" (Ooms 2022) to import the pdf file of "Turkey DHS, 1998, Final Report". A full-page table on page 46 interested us. Then we extracted data from table 2.11 on page 46 and split each character string into text lines using the R package "stringi".(Gagolewski 2021) R packages "tidyverse"(Wickham et al. 2019), "janitor"(Firke et al. 2021) were also used in data manipulation and cleaning. Then we separated the data into columns and stored it as raw data. The raw data was messy, so we corrected some typos in the raw data and stored it as cleaned data.

Then we ran some tests to confirm that the data has no obvious errors. The "pointblank" (Iannone and Vargas 2022) agent allows us to run a few validation tests. We ran three validation functions, and the agent can get some intel about whether the category of all the variables is correct. The first validation test checked whether the column "Reasons stopped attending school" was a characteristic variable. The second validation

Table 1: Glimpse: Percent of women aged 15-24 in urban areas by reasons for leaving school, Turkey in 1998

reason_stopped_attending_school	primary_incomplete	primary_complete	secondary_incomplete
Currently_attending	3.6	0.6	52.2
Got_pregnant	0.0	0.0	0.0
Got_married	0.0	2.3	4.1
Take_care_of_children	0.0	0.1	0.9
Family_needed_help	3.2	1.2	0.6
Could_not_pay_school_fees	4.3	7.4	1.4
Needed_to_earn_money	1.9	2.1	0.4
Graduated	0.0	0.6	0.1
Did_not_pass_exams	2.4	0.0	0.3
Did_not_like_school	16.3	23.9	17.6
School_not_accessible	1.3	3.7	0.4
Parents_did_not_send_to_school	30.5	36.2	9.1
Other	36.5	19.1	11.8
Missing	0.0	2.8	1.0
Total	100.0	100.0	100.0
Number	55.0	829.0	482.0

function ensured all answers to “Reasons stopped attending school” was one of “currently attending”,” got married”, “got pregnant”, “take care of children”, “family needed help”, ”could not pay school fees”, “needed to earn money”, “graduated”, “did not pass the exam”, “did not like school”, “school not accessible”, “parents did not send to school”, “other”, “missing”, “total”,” number”. The third validation function checked if the rest columns were all numeric variables.

2.3 Survey method

2.3.1 Sampling method

The target population of the DHS survey was divided into married women, single women, and husbands. These household populations completed different questionnaires. The questionnaires were designed in English and then translated into Turkish to collect information about their socioeconomic status. All types of questionnaires collected information about background characteristics, reproduction, marriage, sexually transmitted diseases and AIDS, knowledge, and use of family planning, fertility preference. Questionnaire for single women covered work and status, anthropometry: questionnaire for married women covered work and status, husband’s background, immunization and health, maternal and child anthropometry; questionnaire for husband covered attitudes specifically.

The Turkey DHS report used a weighted, multistaged, and stratified sampling method to select the sample. Multistaged sampling is always considered a more advanced version of cluster Sampling. The researchers first divided the target population into different clusters, which were West, South, Central, North, and East regions of Turkey in this report. Then the researchers broke one cluster into smaller and smaller pieces at each stage. The five major regions were divided into different numbers of sub-regions, for example, east region contained three sub-regions, while north region contained two sub-regions. Each sub-region was consist of several provinces.

Then stratified sampling method was used to divide sub-regions. The stratified sampling method is a method in which we divide our population into mutually exclusive strata or sub-populations. (Alexander, R., 2022). The sub-regions were stratified into two parts, urban and rural. Sub-regions with a population of more than 1000, such as provincial centers or district centers, were considered urban regions, while places with a population of less than 10,000 were considered rural regions.

2.3.2 Strengths

When a population has distinct internal stratification, stratified sampling can improve the representativeness of the sample. What's more, when the sample size is the same, the results of stratified sampling and cluster sampling are more accurate than the results of random sampling, so it is more accurate to use the conclusions drawn from stratified to infer the population data. Due to its high precision, we only need a small sample, so stratified sampling and cluster sampling are cheaper.

2.3.3 Weaknesses

A downside of stratified sampling and cluster sampling is that they make the evaluation and interpretation of the results more complicated. Clustering methods require the same representative size from samples, however, the size of representation is difficult to discern, so it is prone to significant sampling errors.(Gaille 2020)

2.4 Data characteristics

The dataset contains 50 observations in total, showing the percentage of women's highest education levels and the reasons they stopped attending school in 1998. The target population is women aged 15-24 in Turkey in 1998. We created Table 1 to get a sense of what our dataset looks like. From Table 1, we could observe different reasons for leaving school for women aged 15-24 in Turkey in 1998, together with percent of women who have ever attended school but didn't attend anymore for each reason. Each row displays a specific reason for leaving school, and each column represents the highest education level. For example, primary_incomplete refers to women who have attended primary school but didn't finish it; primary_complete refers to women who have completed primary school but didn't pursue higher education. We also have a row called 'number' to record the total number of women by highest education attained. (percent; numeric (%))

Table 2: Summary Statistics of percent women leaving school, per highest education level, urban areas

highest_edu	avg	min	max	median
higher	7.15	0	67.5	0.65
primary_complete	7.14	0	36.2	2.20
primary_incomplete	7.14	0	36.5	2.15
secondary_complete	7.14	0	35.8	3.65
secondary_incomplete	7.14	0	52.2	0.95

Table 3: Summary Statistics of percent women leaving school, per highest education level, rural areas

highest_edu	avg	min	max	median
higher	7.14	0	57.1	0.00
primary_complete	7.14	0	44.6	1.85
primary_incomplete	7.14	0	40.4	1.65
secondary_complete	7.14	0	36.1	3.00
secondary_incomplete	7.14	0	45.6	2.10

3 Results

From Table 2 and Table 3 where we generated some summary statistics, we could observe the minimum, maximum, mean, and median percent of women leaving school at each education level. We got a rough idea that in both rural and urban areas, women aged 15-24 experienced similar sets of reasons for leaving school in 1998, yet the percent distribution of women leaving school is not even for each reason at each education level.

From Figure 1, we observed that among all of our respondents (females aged 15-24) in the rural areas in 1998, the proportion of females who left school after completing primary school is approximately 70%; in contrast, this proportion is around 40% for females in urban areas. We also observed that the distribution of the time stages where females left school is very similar for both rural and urban regions. Regardless of the fact that for both rural and urban women, the most common stage where they left school was after completing primary school, there was also a sharp contrast between rural and urban women in the way that the percent of women leaving school in rural areas due to failure to continue to secondary school almost doubled compared to those in urban areas. This indicates that girls in rural areas experienced a more severe loss of basic education compared with those in cities.

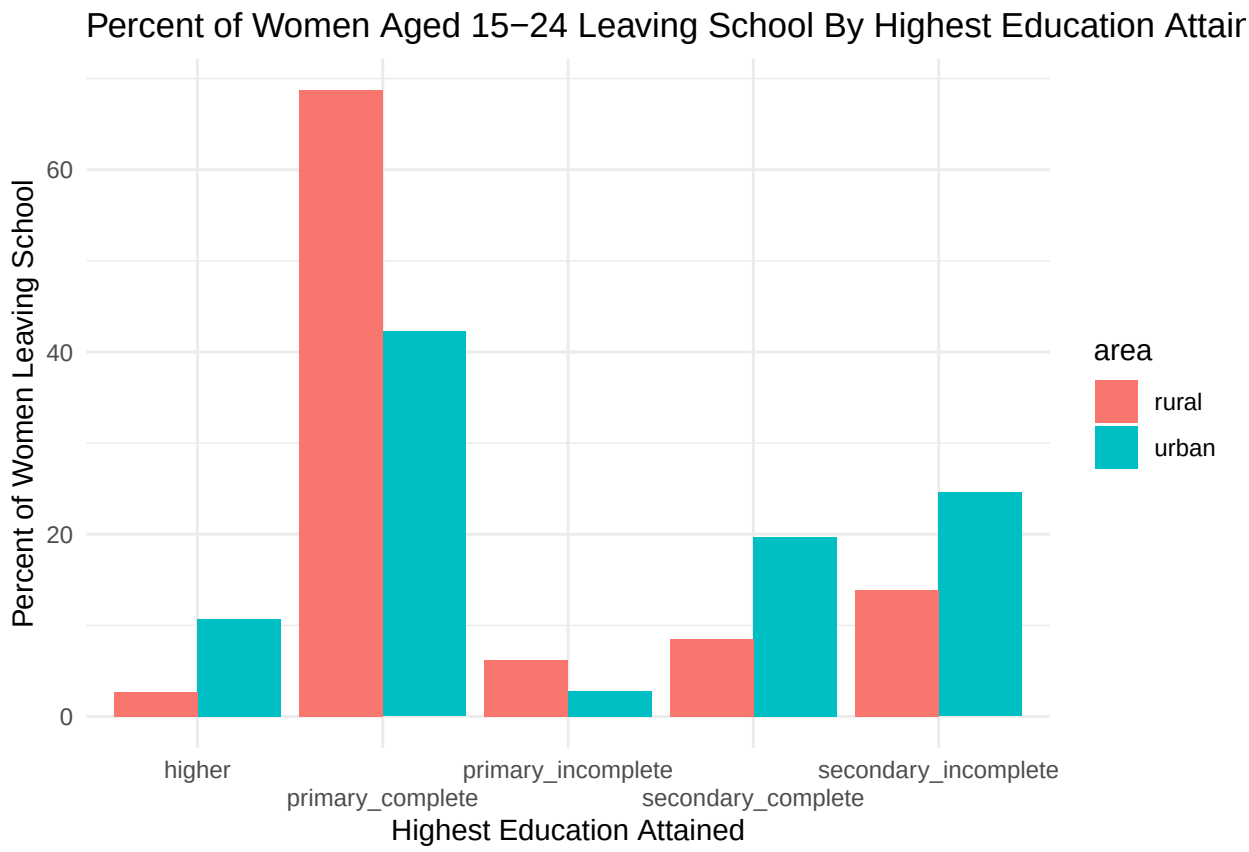

Figure 1: Percent of Women Leaving School In Each Education Level, By Area

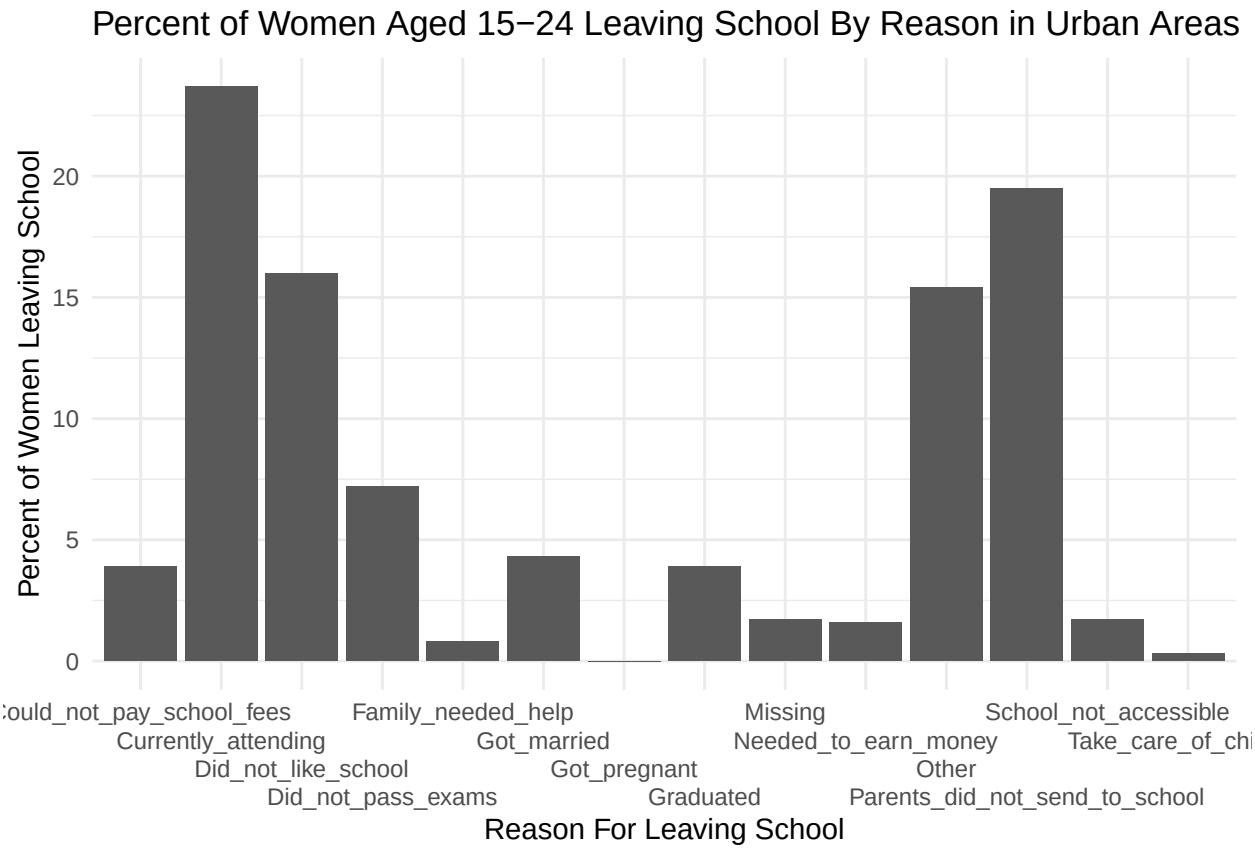

Figure 2: Percent of Women Leaving School By Reason, Urban Areas

From Figure 2, we observed that in urban areas, despite women who were currently attending school, the percent of women who left school due to "parents did not send to school" takes up more than 20% of all females aged 15-24 who left school, and "parents did not send to school" is the most common reason why these females left school. In Figure 3, we observed that in rural areas, women who left school due to "parents did not send them to school" take up more than 30% of our sampled females; the most common reason females left school in rural areas is also the fact that their parents didn't send them to school. Therefore, we could conclude that the biggest reason for women not attending school is the negative attitudes from parents towards sending girls to school.

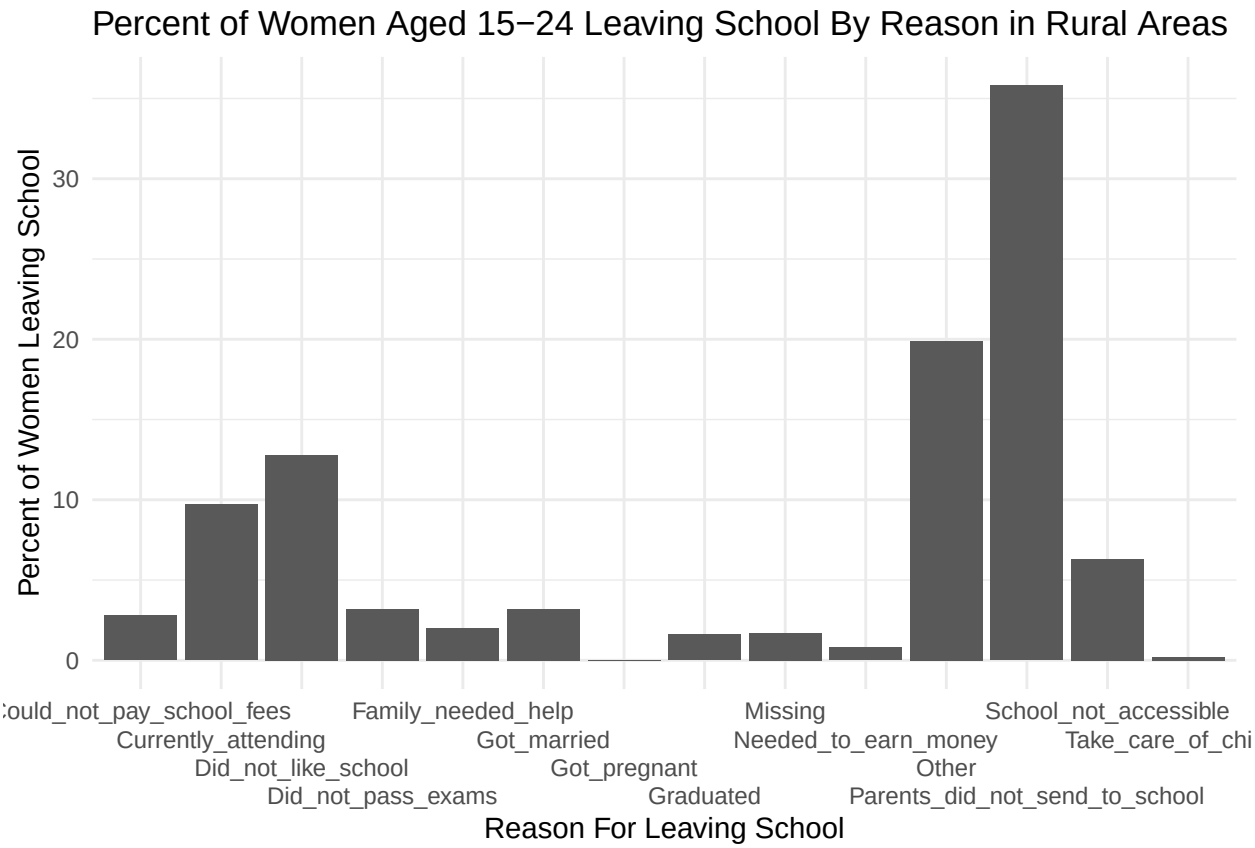

Figure 3: Percent of Women Leaving School By Reason, Rural Areas

Table 4: Highest percent of women leaving school and corresponding reason for each education level

reason_stopped_attending_school	highest_edu	percentage
Parents_did_not_send_to_school	primary_incomplete	35.9
Parents_did_not_send_to_school	primary_complete	40.1
Currently_attending	secondary_incomplete	50.6
Did_not_pass_exams	secondary_complete	35.9
Currently_attending	higher	66.2

From Table 4, we observed that for women aged 15-24 who didn't go to secondary school (either completed or not completed primary school), the most common reason they failed to continue to school is "parents did not send to school"; specifically, 35.9% of women who didn't complete primary school was leaving school because of their parents, and 40.1% of females who completed primary school but failed to go to secondary school were leaving due to parents' attitudes. In contrast, for those who have completed secondary education but failed to go to college, the most common reason is they didn't pass the exams. Thus we could conclude from table 4 that specifically for females who experienced the loss of basic education, the major reason that they stopped attending school was the fact that their parents didn't send them to school.

4 Discussion

4.1 Purpose and summary of the survey

Nowadays, education becomes a hot topic because a good education has the power to change a country's destiny and a person's life. Governments are investing more than before in educating their citizens, so it's beneficial to social development. Having well education helps people think, feel, and behave in a way that contributes to their success, and improves not only their personal satisfaction but also their community. Although more and more people are well educated in developing countries, gender bias still exists. People in some backward regions have insufficient awareness of education for women. That's why the rate of women leaving school before they accept higher education.

In the report, we use a table from page 46 of Turkish Demographic and Health Survey (TDHS-98) to find the most important reason causing females' to leave. Our target population is all the Women aged 15-24 who were not currently attending school were asked about the main reason for giving up further studying in Turkey. The results from respondents were collected by TDHS interview directly. In order to make sure objectivity and authority, we selected the respondents randomly from rural and urban areas. Also, the females who stopped attending schools at different stages (incomplete primary school, complete primary school, incomplete secondary school, complete secondary school, higher) are covered as well. In the survey, we provided several reasons to explain their leaving, including getting pregnant, getting married, taking care of children, their family needed help, unaffordable school fees, disliking school, needing to earn money, being unsupported by family, and others.

In this paper, we discussed the percentage of women are currently attending school and the percentage of them who stop attending school at each education level in rural and urban separately. The main reasons for their leaving both regions are analyzed. In addition, we display the highest percentage of women leaving school and corresponding reasons for each education level. The research can provide guidance to programs seeking to improve educational opportunities for women.

4.2 Learn about the world

From the survey, we learn that only 18% of females in Turkey have the opportunity to continue their studying, which is a pretty low percentage. In urban, the rate of females are currently attending school is 23.7, which is higher than that in rural (9.7%). This observation means that the government must pay more attention to improving females' education levels, especially those females in the rural region. We suggest that the government can spend more money investing in schools and educational institutions to make citizens realize the importance of attending school. Also, some policies related to education can be published, such as compulsory education to enforce everyone attends school.

51.57% of females drop out after completing primary school. Their parents refusing to send them to school to accept further education is the biggest reason. This observation means that most elder generations cannot realize the importance of education for females. To figure out this problem, the government can encourage citizens to join in some social activities about education and convert their traditional opinions through video, newspapers, and other public media.

14.8% of females drop out of school because they don't like school. The courses should be well set to stimulate students' interest in learning. For example, teachers are encouraged to change the tedious teaching method to a funny one; they can play some games with students in the class to motivate students. Also, a good studying/school environment should be created because students' physical and mental health is the most important to their growth. Some of them dislike school because they cannot adapt to the environment or face some abuse. There are four main types of abuse: physical abuse, sexual abuse, emotional abuse, and neglect. Teachers must be aware of signs of abuse and pay attention to the change in the external environment and students' emotions. Teachers should encourage students to study rather than blame them. Students who are currently accepting primary education are sensitive and immature. If they face lots of pressure and psychological problems, they may be unwilling to go to school anymore. Isolation in the school must be prevented. When they have some issues, teachers should listen actively without giving emotional solid reactions or judgments.

Finally, those females who have opportunities to accept higher education face challenges in passing the exam. Failure to pass the exam becomes another non-negligible reason for their leaving. Therefore, to increase the number of females who have the chance to study further, the government and schools should adjust the difficulties of the exam so that more citizens can attend school successfully.

4.3 Weakness and learn in the future

However, there are several limitations in our research so our data may not be accurate and we need to make some improvements.

Firstly, over 63000 thousand people lived in Turkey in 1998, and 50.75% were female. The sample size is only 3031, which is too small relative to the population of females aged 14-25. The small sample size gets decreasingly representative of the entire population. The small sample size can also lead to biased cases, such as non-response, which may influence data precision. Therefore, we want to survey as large a sample size as possible in further research.

Secondly, the survey is conducted by interview and completing questionnaire, which means that these respondents have to answer the questions face to face, and they have the probability of concealing the real reason for their leaving because of environmental pressure or public opinion. For example, some females drop out of school when they get pregnant, but they are unwilling to say that and randomly choose another reason. In order to solve this problem, we suggest using an online anonymous questionnaire to make sure the privacy of respondents in the future.

Thirdly, we observe that 17% of females choose "other" as the reason for stopping attending school, which is a high percentage. It reminds us that we may still ignore some main reasons causing their leave. Therefore, we suggest that the respondents answering "others" are encouraged to write down the specific reasons in order to help us collect data accurately. This extra information is beneficial to us by giving more specific suggestions to improve the education level of females in Turkey.

Finally, in the future survey, how family income and national GDP influence females' education levels can be investigated. We should learn more about the relationship between citizens' educational attainment and the development of a country. In addition, how parental factors influence children literacy development should be discussed. Also, sexual discrimination is considered a core issue in many fields, not just in education. If we want to fundamentally improve females' total education levels, analyzing the reasons causing discrimination is another topic necessary.

Appendix

Motivation

1. *For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*
   - The dataset was created to study why Turkey stopped attending school in 1998. No specific gap needs to be filled.
2. *Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*
   - Qiuhan Wang, Yuxuan Yang, Zihan zhang created the dataset on behalf of University of Toronto.
3. *Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*
   - No funding was received for this project.
4. *Any other comments?*
   - No

Composition

1. *What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*
   - The answers to the "HACETTEPE UNIVERSITY INSTITUTE OF POPULATION STUDIES 1998 TURKISH DEMOGRAPHIC AND HEALTH SURVET" questionnaire comprise the dataset represented.
2. *How many instances are there in total (of each type, if appropriate)?*
   - There 51 observations in total. Education level had 6 options and reasons stopped attending school had 16 options.
3. *Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*
   - The dataset is a sample of instances from a larger set. The larger set is the education level and reasons why they stopped attending school for all Turkey women in 1998. Stratified sampling and cluster sampling were used to choose samples, so we believe the sample is representative of the larger set.
4. *What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*
   - The data each instance consist of reasons stopped attending school and the highest education level. Reasons include currently attending, got pregnant, got married, take care of children, Family needed help, could not pay school fees,needed to earn money, graduated, did not pass exams, did not like schools,school not accessible,parents did not end to school, Other, Missing, Total, Number. Education level includes primary incomplete, primary complete, secondary incomplete, secondary complete, higher,total.
5. *Is there a label or target associated with each instance? If so, please provide a description.*
   - Yes, there are reasons stopped attending school and the highest education level.
6. *Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*
   - The dataset shows the percentage of women's education level in Turkey in 1998. Individual instances are confidential, so we don't know if there is any missing information from individual instances.
7. *Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*
   - No, the dataset only shows percentages.

8. *Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*
    - No
9. *Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*
    - No
10. *Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*
    - The dataset is a summary of the DHS Turkey dataset. The dataset will exist and remains constant over time. The official archival versions of the complete dataset can be obtained from Hacettepe University, Institute of Population Studies, 06100 Ankara, Turkey. You need to send requests to Hacettepe University for the complete dataset.
11. *Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*
    - No
12. *Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*
    - No
13. *Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*
    - The dataset identify sub-populations by gender. The dataset shows the percentage of women's education level.
14. *Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*
    - No
15. *Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*
    - NO
16. *Any other comments?*
    - No

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*
    - The data was derived from "Turkey DHS, 1998, Final Report". We used R package "pointblank" to run some validation tests.
2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*
    - The dataset was collected by a survey using stratified sampling and cluster sampling method.
3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*
    - stratified sampling and cluster sampling method were used.

4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*
    - Contractors were involved in the data collection process. Every group consisted of one manager, five females, and one male.
5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*
    - The data was a recent crawl of old data in 1998.
6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*
    - No
7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*
    - The dataset was obtained from third parties. The complete dataset can be obtained from Hacettepe University, Institute of Population Studies, 06100 Ankara, Turkey.
8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*
    - The individuals in question were notified about the data collection. Detailed information can be found from Hacettepe University, Institute of Population Studies, 06100 Ankara, Turkey.
9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*
    - The individuals in question consent to the collection and use of their data. The questionnaires are available at https://dhsprogram.com/publications/publication-fr108-dhs-final-reports.cfm.
10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*
    - The questionnaire did not mention whether participants can revoke their consent to the collection and use of data.
11. *Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*
    - No
12. *Any other comments?*
    - No

Preprocessing/cleaning/labeling

1. *Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*
    - The dataset was split into text lines using the R package "stringi". Then we separated the data into columns and stored it as raw data.
2. *Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*
    - The raw data was saved in addition to the cleaned data. It is available thorugh Github.
3. *Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*
    - R was used.
4. *Any other comments?*
    - No

Uses

1. *Has the dataset been used for any tasks already? If so, please provide a description.*
   - The dataset was used to analyze the reasons for dropping out of school for women with different levels of education in Turkey in 1998.
2. *Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*
   - No
3. *What (other) tasks could the dataset be used for?*
   - The dataset can also be used to study the relationship between education level and other aspects like salary if supplementary datasets are provided.
4. *Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*
   - No
5. *Are there tasks for which the dataset should not be used? If so, please provide a description.*
   - No
6. *Any other comments?*
   - No

Distribution

1. *Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*
   - The dataset and report are available through Github. Code and data are available at:https://github.com/zihanjasmine/Turkey-DHS.
2. *How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*
   - The dataset is available through Github. The dataset does not have a DOI.
3. *When will the dataset be distributed?*
   - The dataset is available through Github.
4. *Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*
   - The dataset is not distributed under a copyright, IP license, and ToU. The dataset is licensed under the MIT License.
5. *Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*
   - No
6. *Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*
   - No
7. *Any other comments?*
   - No

Maintenance

1. *Who will be supporting/hosting/maintaining the dataset?*
   - Zihan Zhang
2. *How can the owner/curator/manager of the dataset be contacted (for example, email address)?*

- zhanzihan.zhang@mail.utoronto.ca

3. *Is there an erratum? If so, please provide a link or other access point.*
   - No
4. *Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*
   - No
5. *If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*
   - No. Individual answers are confidential.
6. *Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*
   - No
7. *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*
   - If others want to extend/augment/build on/contribute to the dataset, they can pull requests on Github. We will credit their contributions.
8. *Any other comments?*
   - No

References

Ankara. 1999. “Hacettepe University, Institute of Population Studies/Turkey and Macro International.”

Banton, Michael. 1994. “Discrimination.” *Open University Press* 1 (3): 86.

Firke, Sam, Bill Denney, Chris Haid, Ryan Knight, Malte Grosser, and Jonathan Zadra. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://cran.r-project.org/package=janitor.

Gagolewski, Marek. 2021. *Stringi: Fast and Portable Character String Processing in R.* https://stringi.gagolewski.com/.

Gaille, Louise. 2020. “16 Key Advantages and Disadvantages of Cluster Sampling.” Vittana.org.

Holsinger, W.James, Donald B;Jacob. 2009. “Inequality in Education Comparative and International Perspectives.” *Comparative Education Research Center* 2 (2): 168. https://doi.org/https://doi.org/10.1007/978-90-481-2652-1.

Iannone, Richard, and Mauricio Vargas. 2022. *Pointblank: Data Validation and Organization of Metadata for Local and Remote Tables.* https://CRAN.R-project.org/package=pointblank.

Ooms, Jeroen. 2022. *Pdftools: Text Extraction, Rendering and Converting of Pdf Documents.* https://CRAN.R-project.org/package=pdftools.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.