# Contextualising Discrimination: Investigating the Effect of Gender and Poverty on the Outcome of a Driving Test

Lewis Buttle - U1804747 - CS430

## 1 Abstract

**This project aimed to investigate the existing disparity in driving test success rates, and investigate the possibility that success is affected by a person's gender and financial situation. The process of acquiring data relevant to this investigation is explained, such as acquiring data indexing locational poverty, and the specific data of how males and females perform on driving tests. Specifically how this data is handled, cleaned, integrated together and analysed is detailed throughout this document, as well as references to the tools utilised to perform this data management efficiently. When relevant, data has been appropriately plotted and visualized to help communicate some noticeable behaviours in the data. In conclusion, the project is successfully able to model data to provide substantial evidence for two rationalized hypotheses:**

- **Poverty damages a female's chances of passing a driving test more than a man's.**
- **The more impoverished an area a test centre resides in, the lesser its overall pass rates will be.**

## 2 Introduction

Discrimination diminishes a person's rights simply because of who they are, or what they believe. Without a doubt, discrimination exists both intentionally and unintentionally within many areas of society and ultimately, people suffer from the consequences.

To eliminate discrimination is a monumental task as there is a countless number of factors which produce a prejudiced environment. However, Amnesty International: a UK based organization focused on defending human rights, claims that the solution to removing discrimination is to identify and subsequently tackle the root causes of discrimination in all aspects of society, however small[9]. This suggested solution from Amnesty serves as the main inspiration for this project.

This project aims to investigate if poverty is a root cause of gender and financial discrimination in the context of passing a driving test in the UK. The project aims to achieve this by using existing data and research linking poverty to discrimination rates, and testing if an area having characteristics of poverty affects its driving test centre results. By identifying any root causes, this project could be used as evidence to strengthen the existing arguments of the factors that cause discrimination.

The context of driving test pass rates was deemed novel to investigate, as the female pass rate for driving tests has been substantially lower (around 6-7%) than the male's pass rate every year between 2010 and 2020[10]. No other substantial research has been made into this large disparity, hence this poses the idea that some form of discrimination may be involved, and would be worth investigating.

## 3 Background

### 3.1 Discrimination in the UK

Amnesty International identifies two main forms of discrimination: *direct* and *indirect*[9].

*Direct* discrimination is when a perpetrator explicitly denies someone's rights directly because of their identity group. For example, the BBC reports that in the UK, around twice as many females to males are victims of stalking[1]. This is *direct* discrimination because it appears that females are being targeted by stalkers specifically because of their gender.

*Indirect* discrimination is more complicated. This occurs from an act which is neutral and does not distinguish by group: however, it unintentionally disadvantages specific groups due to underlying issues. For example, the Royal Society of Arts reports that UK students residing in more deprived neighbourhoods are much less likely to have access to studying 'more difficult' subjects[14], thus denying students the ability to further their formal education simply because of their financial situation. The Royal Society of Arts found that the schools in poorer neighbourhoods were more likely underachieving, and removed access to these difficult subjects to better meet the grade expectations set by the Department of Education.

This example is *indirect* discrimination, as there is no active perpetrator enforcing that poorer students should not have access to more difficult subjects, but rather the discrimination stems from a chain of smaller underlying issues. Hence, the issue lied within the correlation between poorer areas having underachieving schools, which subsequently could be broken down into even more smaller causes.

There are many more examples of discrimination occurring in the UK. A major example being workplace discrimination. In 2021, 36% of UK adults reported experiencing workplace discrimination[15], which could include harassment by peers or denial of a job completely because of identity aspects such

as age, gender and ethnicity.
These examples show how *indirect* discrimination issues are difficult to solve due to the lack of a perpetrator, and the fact that the issue is formed of many smaller causes, which all need to be identified and solved to tackle the overarching issue. Additionally, these examples highlight the many ways people can be discriminated against, though this project has chosen only to focus on financial and gender discrimination due to the project's limited scope.

### 3.2 Poverty

Reports describe both discrimination and poverty as two closely linked types of injustice.
Regarding gender, it is shown that females are more likely to be in poverty, hence this economic inequality experienced by females contributes to a societal view in poorer areas that females are of lower standing, and hence this prejudice negatively affects how females are treated in society[3]. Additionally, females in poverty exclusively have a higher level of anxiety amongst them, as poorer males do not have this correlation[12].
Similarly, poverty is also a major cause of financial discrimination. People from poorer, more poverty-stricken backgrounds are likely to have lower self-esteems and have generally less confidence, as well as less time and money to spend on non-essential services[2]. Hence, it can be seen how poverty negatively affects a person's access to many services simply because of their economic situation.
Both of these points highlight how poverty can be the root cause for gender and financial based *indirect* discrimination.

### 3.3 The Driving Test

A number of test centres exist in England which allow a person to pay to be examined to acquire their driving license. To pass, the DVSA recommends 45 hours of practice time, in which a person would usually pay for lessons. To pass the test and acquire a license, it is estimated to cost **£1551**[13], assuming a first time pass.
As shown in figure 1, test centres are spread somewhat evenly across the country, with an expected higher density around London. Additionally, through visualising the pass rates, there seems to be a notably higher pass rate around the most northern areas of England. In terms of pass rates, the total pass rate hovers around 50%, however, females consistently underperform quite substantially in comparison to males[10]. Additionally, the test can be a stressful experience for many: there is a link between high levels of anxiety negatively affecting a person's driving test result[8].

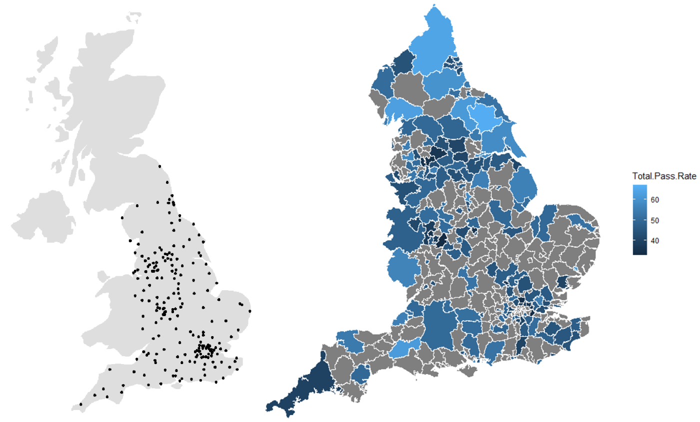


Figure 1: Every driving test centre in England(Left), with the average total pass rate of each test centre's local authority area(Right)(2018)

## 4 Hypotheses

This project's rationale is based upon the substantial, interesting links found between the damage caused by discrimination and the statistics surrounding driving tests:

- Females in poverty develop a higher rate of anxiety unlike males, where anxiety also negatively effects driving test performance.
- People in poverty have less time and money, where passing a driving test requires a substantial amount of time practicing and money for lessons.

These links were deemed appropriate to become a justification to the hypotheses of this project, in which either hypotheses being proven could evidence indirect discrimination taking place in this context. The formal hypotheses to test both gender and financial discrimination is as follows:

1 A positive correlation between the impoverishment of the location of a test centre and its female pass rate exists, and should be **stronger** than the correlation between its impoverishment and its male pass rate.

2 The more impoverished a test centre's location, the lower its total pass rate.

## 5 Tools

Many data tools were utilised to manipulate the data correctly and effectively:

### 5.1 Microsoft Excel

Microsoft Excel was used as a fundamental tool for accessing and visualising specific data records. Its user friendliness made it useful for adding small manual adjustments to datasets with ease, such as removing redundant columns and rows during cleansing.

## 5.2 R & RStudio

R served as a primary tool for manipulating and visualising the data. The power of R allowed for the completion of jobs that could not feasibly be done manually, such as normalising data, joining datasets, and plotting data. Additionally, RStudio served usefully as a UI for R, adding an element of user friendliness to the tool that made smaller jobs much easier to accomplish. Additionally, the *maps* library allowed for the construction of effective geographical data visualisations.

## 5.3 Weka

Weka allowed for the construction of regression models on a dataset, detailing useful information about the model for detailed analysis.

## 5.4 Postcodes.io

A free, open source postcode and geolocation API for the UK. This was utilised within R scripts to lookup and store location based details for the test centres quickly and effectively. The tool helped solve very specific location based problems throughout the project.

# 6 Datasets

Multiple datasets were required to test the project's hypotheses. The ability to fairly and legally use each dataset was validated under the *Open Government License v3.0*.

## 6.1 A - Car Pass Rates by Gender, Month and Test Centre[5]

This dataset contained the percentage pass rates for each gender, per month for each test centre in the UK. The file contained different sheets for each year between 2008 and 2021, although, as explained in section 9, only the 2018 sheet was of relevance for this project. The dataset also summarized the total pass rate, and male and female pass rate for the year for each test centre, which incidentally was the only data of interest from this dataset.

## 6.2 B - Driving Test Centres[6]

Regarding specific geographical data of the test centres, dataset **A** only stored the test centre's city name, which was insubstantial for the project's goals as the poverty measure database used local authority codes as means of referencing geography, which could not be effectively matched to test centre names. Fortunately, the DVSA provide an open dataset including detailed information about each driving test centre. This dataset provided the address, postcode and coordinates of each test centre, as well as miscellaneous data such as disabled access. The only data of interest in this dataset was the test centre names, postcodes and coordinates, which could be linked with database **A** to give more geographical information about the test centres, and thus allow the dataset to be joined with more complex geographical codes from datasets **C**, **D** and **E**. This dataset is mapped visually in figure 1.

## 6.3 C - Regional Gross Disposable Household Income[4]

This dataset stored two sheets for *GDHI* and *GDHI Per Head* information for each LAU3 region in the UK, for the years 1997 to 2019. For this project, only the 2018 columns were relevant from both sheets.

## 6.4 D - Mapping Income Deprivation[11]

This dataset stored information about income deprivation per local authority district, including rankings and the district's rurality. The specific data used by this project was the 'average score' of income deprivation for each district, which served as a relative scale for comparing income deprivation poverty between areas.

## 6.5 E - English Indices of Deprivation[7]

This dataset was important to the project in giving accurate, relative indices for both crime and education, for each English local authority district. The dataset also provided indices for factors such as health and living, but these were not deemed relevant to the project's background.

## 6.6 Data Limitations

There was no available data measuring poverty for each local authority, so varying measures of poverty were used. The alternative measures used were:

- Gross Household Disposable Income - The disposable income the average household has after paying taxes. A measure of material wealth.
- Gross Household Disposable Income Per Head - The same as above, but divided by population. Accounts for the population of local authorities.
- Income Deprivation - The score of people in receipt of some form of economic support. A measure of monetary poverty.
- Crime - A score of the crime committed in the area. A measure of impoverished areas.
- Education - A score for the average level of education. A non-monetary measure of poverty.

Breaking down poverty into different measures benefited the investigation, by being able to construct more complex models using combinations of the measures to hence develop a better understanding of specifically which measures were most important to correlation, rather than just a single poverty measurement.

# 7 Data Cleaning

Each dataset had to be cleaned to ensure each dataset was robust enough to be processed and analysed more reliably, with the goal of the cleaning stage to produce a single, combined dataset which will store for each test centre: pass rate data, geographical data, and each chosen measure of poverty. By having this single dataset and ensuring it is robust with no missing values, it will be able to be loaded into *Weka* and processed effectively.

## 7.1 Data Transformation to CSV

To ensure each dataset would be compatible when joining, the first goal was that each dataset was reformed to resemble a simple *CSV* file, by removing redundant columns and rows. A few of the datasets had notes stored in certain cells to include additional government information for example, but this had to be removed to allow for the datasets to be compatible. *Excel* was used to manually remove these unnecessary cells and sheets from each dataset, until each file was in the standard format.

A challenging conversion was dataset A, as the original dataset was systematically saturated with unnecessary data for each test centre, hence manually cleaning this would be inefficient. Instead of *Excel*, *R* was used to clean this dataset by using the command *A[grep("2018", A$TestCentre, invert = TRUE) ,]*, which searched every row in the first column for the monthly pass rate records and deleted them. A similar command was used to delete the redundant test centre name rows. This more difficult cleaning is shown in figure 2.

**DVSA0201 Practical car pass rates by gender, month and DTC**

**Time Period Captured** **April 2018 to March 2019**

| | Male tests | | |
|---|---|---|---|
| | **Conducted** | **Passes** | ***Pass rate (%)*** |
| **Aberdeen North** | | | |
| April 2018 | 103 | 66 | *64.1* |
| May 2018 | 139 | 90 | *64.7* |
| June 2018 | 121 | 86 | *71.1* |
| July 2018 | 83 | 62 | *74.7* |
| August 2018 | 135 | 80 | *59.3* |
| September 2018 | 135 | 80 | *59.3* |
| October 2018 | 159 | 84 | *52.8* |
| November 2018 | 180 | 115 | *63.9* |
| December 2018 | 105 | 67 | *63.8* |
| January 2019 | 151 | 100 | *66.2* |
| February 2019 | 126 | 61 | *48.4* |
| March 2019 | 140 | 72 | *51.4* |
| **Aberdeen North** | 1577 | 963 | *61.1* |
| **Aberdeen South (Cove)** | | | |
| April 2018 | 114 | 66 | *57.9* |
| May 2018 | 163 | 98 | *60.1* |
| June 2018 | 149 | 73 | *49.0* |
| July 2018 | 194 | 122 | *62.9* |
| August 2018 | 151 | 93 | *61.6* |
| September 2018 | 114 | 69 | *60.5* |
| October 2018 | 171 | 91 | *53.2* |
| November 2018 | 138 | 74 | *53.6* |
| December 2018 | 71 | 35 | *49.3* |
| January 2019 | 85 | 53 | *62.4* |
| February 2019 | 115 | 61 | *53.0* |
| March 2019 | 123 | 48 | *39.0* |
| **Aberdeen South (Cove** | 1588 | 883 | *55.6* |
| **Abergavenny** | | | |

| Name (character) | Male tests (double) | Female tests (double) | Total tests (double) |
|---|---|---|---|
| Aberdeen North | 61.1 | 51.4 | 55.5 |
| Aberdeen South (Cove) | 55.6 | 48.1 | 51.7 |
| Abergavenny | 62.0 | 53.1 | 57.3 |
| Aberystwyth (Park Avenue) | 55.5 | 45.1 | 49.7 |
| Airdrie | 48.4 | 42.8 | 45.5 |
| Alness | 59.1 | 57.0 | 58.2 |
| Alnwick | 70.7 | 59.3 | 64.3 |
| Arbroath | 69.6 | 60.7 | 64.7 |

Figure 2: The original, bloated dataset A (Left) with the cleaned version (Right).

## 7.2 Aligning Geographical Data

To merge datasets together, a similar data column is required in both datasets, to serve as the pivot for merging. For this project, the pivot column would be the geographical data of the test centres: to ensure that the local poverty measures were being assigned to the correct test centre. A difficulty encountered during this process was the difference in geographical data for each dataset, as A had a name, B had a postcode, C used LAU3, and D and E used local authority district codes. A means of finding a universal geographical identity between the datasets had to be found.

An ideal dataset was found which stored the LAU3, district and postcode for every address in the UK, however this was too large to be workable with *R*, and no smaller datasets existed which stored this information. Alternatively, the *postcodes.io* API was able to solve this issue, as it provided a lookup function in *R* which, when given a postcode, would return LAU3, district codes and more. To run a lookup for every test centre, a *R* script was utilised to automate the process and store the results: shown in figure 3.

```r
#Load in the AB joined dataset
AB <- read.csv("ABJoined.csv")

#Load postcode Lookup Library
library(PostcodesioR)


#For each record...
for(i in 1:nrow(AB)) {
  #Store the postcode of the record
  postcode <- AB[i,]$Postcode

  #Lookup the postcode using PostcodesIO
  lookup <- postcode_query(postcode)[[1]]

  #Retrieve the NUTS value for the postcode
  nuts <- lookup$codes$nuts

  #Append it to the record in the AB table, if the result is not null.
  if (!is.null(nuts)){
    AB[i,8] <- nuts
  }
}
```

Figure 3: The R script used to store all the LAU3 values in Dataset A.

## 7.3 Data Integration

*R* was used to merge the datasets. The datasets A and B were merged first via the 'Name' column, as that is what they had in common. This merge was done by using the command: $AB <- merge(A,B,by='Name')$. With AB, I used variations of the *R* script from figure 3 to add columns for LAU3 and LAD codes within the *AB* dataset, so that the remaining datasets would have a column to join by. This merging process continued until every dataset had been combined.

## 7.4 Final Cleaning

There was no available data measuring poverty for each local authority, so varying measures of poverty were used. The alternative measures used were:

- Gross Household Disposable Income - The disposable income the average household has after paying taxes. A measure of material wealth.
- Gross Household Disposable Income Per Head - The same as above, but divided by population. Accounts for the population of local authorities.
- Income Deprivation - The score of people in receipt of some form of economic support. A measure of monetary poverty.

- Crime - A score of the crime committed in the area. A measure of impoverished areas.

- Education - A score for the average level of education. A non-monetary measure of poverty.

Breaking down poverty into different measures benefited the investigation, by being able to construct more complex models using combinations of the measures to hence develop a better understanding of specifically which measures were most important to correlation, rather than just a single poverty measurement.

Some missing data was present in the merged dataset, specifically three test centres that had null pass rates. Through research, I found that these test centres were closed during 2018 and hence no data had been collected. Replacing the null value with the mean pass rates was considered, but it was deemed necessary to remove the records as it was only three test centres, and the dataset still had 211 other test centres, meaning relatively not much data had been lost, and the project outcomes would likely not be effected.
The large education values had to be cleaned to have commas removed, as otherwise the data would be interpreted as nominal and not numeric in *Weka*. This was done using *R*, with the command: *Education< −as.numeric(gsub(”,”,””,Education))* to replace all commas with a blank space.
As the range of the education attribute was quite large in comparison to other attributes, it was thought to be good practice to normalize every attribute, to create an environment where correlation can be better and more reliably identified during analysis.

# 8 Analysis

## 8.1 Data Visualisation

Mapping the data served as an initial analysis of the data. Using *R* and *GGPlot*, it was possible to plot the test centre data to display two factors at once. By modelling the total pass rate as size, and income deprivation of the area as colour (shown in figure 4), this gave a very initial indicator for hypothesis 2. By eye, it is hard to discern an immediate correlation, however, near the bottom of the figure, a small trend shows multiple test centres with the lowest pass rates also having less poverty. This small trend does not follow the financial discrimination hypothesis, but it is only apparent in a small, grouped subset of the data, so will require further testing to prove.

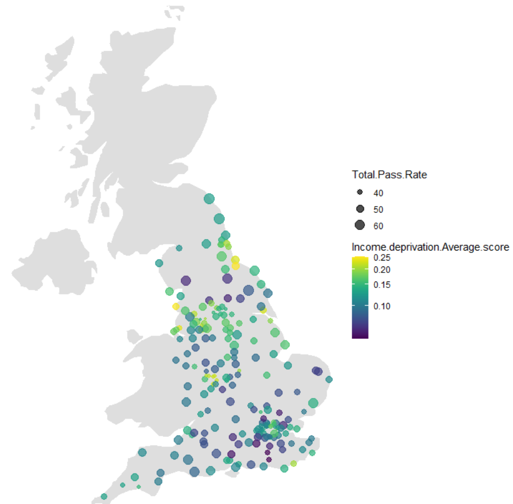


Figure 4: Every test centre in the final dataset, where dot size represents pass rate and colour represents the area's income deprivation.

## 8.2 Modelling

Using *Weka*, both hypotheses were tested by constructing different regression models to predict a certain pass rate given the poverty data. The models produced were then analysed.

### 8.2.1 Hypothesis 1.

This hypothesis predicts that models that predict female's pass rates using the poverty data will outperform models that predict male's pass rates, using the same explanatory variables. The direct results of this analysis are shown in table 1.
Firstly, two linear regression models were constructed to predict both female and male rates, using all five poverty measures as explanatory variables. Both models had correlation coefficients between **0.4-0.6**, showing a reasonable correlation between the poverty in a test centre's area, and the pass rate in general. Additionally, examining the difference in the models highlights which poverty measures have a greater impact on either gender's pass rates. Both models fail to include *GDHI per head* as an effective variable, which infers that there is little to no correlation between these poverty measures and the pass rate of either gender. Interestingly, only the female model utilises the *GDHI total* factor whilst the male model does not, inferring that *GDHI* exclusively affects only female's pass rates: although as the gradient for this factor is relatively low in the model, it shows that this correlation is quite weak and thus likely not too important of a difference in reality.
Both models have strong correlations with *income deprivation*, *education* and *crime*, which means that these factors indiscriminately affect pass rate as a whole: and will likely appear in the investigation of the 2nd hypothesis. The gradient for *crime* is noticeably higher in the male model compared to the female's, meaning that amount of crime has a greater impact on the prediction of a male's score than a female's in these models.
The female pass rate model succeeded in demonstrating a rea-

sonably stronger correlation, as the female pass rate's model's correlation coefficient is **0.0904** larger than the male pass rate's model. This evidences that the hypothesis has been proven: poverty has a substantially stronger correlation with female's pass rates than male's. Additionally, examining the linear regression model for female's pass rates shows how the hypothesis has been fully met, as *income deprivation* and *crime* have negative gradients, meaning the models show the more crime and income deprivation within the test centre's location, the **lower** the test scores attained: which agrees with the hypothesis predicting a correlation. To better visualize this gradient, the *income deprivation x female pass rate* was plotted for this model using *R*, shown in figure 5. As shown in the graph, the trend follows that the more deprived an area, the lower the test results will be: and as shown, this appears to correlate more strongly with females more than males as hypothesised. Interestingly, the female model shows a negative gradient between *GDHI* and the *female pass rate*, which does not follow the trend of poverty correlating with less pass rates, however, due to this gradient being relatively small, it can be considered anomalous.

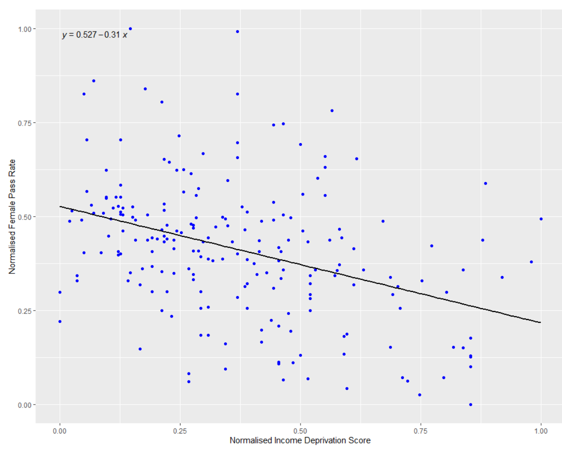


Figure 5: The linear relationship between female pass rate and income deprivation: the more deprivation, the lower the pass scores.

To extend this investigation, the table also shows the result of a multi-layer perceptron model for each gender for all 5 explanatory poverty variables. This was done to check how high of a correlation coefficient can be achieved with the data, and if a suitable predictive model could be derived. Additionally, the models were fine tuned through *Weka*'s parameters to try and maximize their correlation coefficient: in which this investigation is also shown in the table. For the female model, the highest coefficient that was attained was **0.5935**, which is relatively high and thus shows a firm correlation between the poverty measures and the female pass rate, and infers that this model could potential have worth in being used in a predictor. Once again, this model had a better correlation coefficient than the male model; which was maximized to be **0.4884**. Whilst this value is still reasonably high and shows some correlation, the female model continues to highlight that these poverty characteristics are certainly more likely to impact a female's pass rate than a man's.

| | **Female** | **Male** |
|---|---|---|
| **Linear Regression Correlation Coefficient** | 0.543 | 0.4526 |
| **Linear Regression Model** | Female.Pass.Rate =<br>-0.0883 * GDHI.Total +<br>-0.3601 * Income.deprivation.score +<br>0.3136 * Education.Score +<br>-0.3452 * Crime.Score +<br>0.6088 | -0.2905 * Income.deprivation.score +<br>0.352 * Education.Score +<br>-0.3467 * Crime.Score +<br>0.5527 |
| **Multilayer Perceptron Correlation Coefficients + Fine Tuning Learning Rate(a) and Momentum(b)** | a = 0.3 b = 0.2 ->0.5898<br>a = 0.25 b = 0.2->0.5893<br><br>a = 0.3 b = 0.35 ->0.5931<br>a = 0.3 b = 0.4 ->0.5933<br>a=0.3 b=0.45 ->0.5935<br>a=0.3 b=0.45 ->0.5933 | a = 0.3 b = 0.2 ->0.4884<br>a = 0.25 b = 0.2->0.4853<br>a = 0.35 b = 0.2 ->0.4847<br><br>a = 0.3 b = 0.25 ->0.4875<br>a = 0.3 b = 0.15 ->0.4876 |

Table 1: The results of the models used to predict both female and male pass scores, using all five poverty measures.

### 8.2.2 Hypothesis 2.

This hypothesis predicted that the more impoverished a test centre's area is, the lower its overall pass rate will be. To test this, another linear regression model was formed using the five poverty measures as explanatory variables, and the total pass rate as the dependant variable. The results of this modelling process is shown in table 2.

As shown, *income deprivation, education, and crime* form the linear regression model for predicting the total pass rate with a respectable correlation coefficient of **0.5051**, inferring that these three factors have a definite effect within impoverished areas on effecting a driving test's outcome. Unlike the gender models, the total pass model has similar gradients ranges for its three comprising factors, meaning each factor is similarly responsible for the test outcome. This model supports the set hypothesis, as *income deprivation* and *crime* are both negative gradients, this supports that the more impoverished an area, the lesser the average pass rate will be, similar to the graph shown in figure 5.

Additionally, the *education* gradient is positive which also supports the hypothesis, as in this context a higher education score means an area is less impoverished. This evidences a correlation between more educated people, and a higher pass rate, which is an interesting addition to proving this hypothesis.

Additionally, to extend this hypothesis investigation: a multi-layer perceptron was trained on the data. Performing the same fine tuning as described in hypothesis 1, it was found that **0.5549** was the highest correlation coefficient that a model could have when predicting the total pass rate. Comparing this to the gender-related perceptron models, the coefficient falls below the female model but above the male model. This is expected, as male and female pass rates directly form the total pass rate, so the perceptron model for the total pass rate is an *average* of both gender models, hence the total pass rate model correlation coefficient falls roughly in between the coefficients for the male and female models.

# 9 Limitations & Assumptions

Many limitations were recognised throughout this project's development, and assumptions had to be taken into account to deal with them:

| | Total Pass Rate |
|---|---|
| **Linear Regression Correlation Coefficient** | 0.5051 |
| **Linear Regression Model** | Pass.Rate = -0.3402 * Income.deprivation.score + 0.3577 * Education.Score + -0.3755 * Crime.Score + 0.5668 |
| **Multilayer Perceptron Correlation Coefficients + Fine Tuning Learning Rate(a) and Momentum(b)** | a = 0.3 b = 0.2 ->0.5549<br>a = 0.25 b = 0.2->0.5541<br>a = 0.35 b = 0.2 ->0.5518<br><br>a = 0.3 b = 0.4 ->0.5541<br>a=0.3 b=0.15 ->0.5534 |

Table 2: The results of the models used to predict the total pass rate, using all five poverty measures.

- There was no appropriate data or means of getting perfect accuracy on impoverishment of a test centre's userbase, hence it is assumed that the local authority the test centre resides in is its entire userbase.

- The data used is not up to date, and is taken from 2018. This is due to the data between 2019-2021 to be much more limited due to the COVID-19 Pandemic. Specifically, there was significantly less tests taken during these years and it was thought that more valid, relevant models for the task at hand could be obtained from non-pandemic data, as it is unknown how the pandemic may have interfered with the project's conclusions.

- Data for the local authority poverty measures could only be obtained for England, meaning the project has ignored test centres in Wales and Scotland. Realistically, this should not affect the project's conclusions as England still had a large enough number test centres to be able to derive meaningful models.

- The *Postcodes.io* tool was unable to find local authority codes for about three test centres that were in extremely remote areas, and even manually there was no data available on their local authority code. Hence, they were dropped from the dataset. As it was only three removed, this likely did not affect the result of the investigation.

- The driving test pass rate per ethnicity was available as a dataset, however, it was not deemed worthwhile to investigate this data, as the dataset showed that about 95% of people reported to have no preference for their ethnicity, meaning the distribution for each test centre was extremely skewed to this category, and the remaining 5% could not reliably present a reasonable argument due to its size.

# 10 Conclusion

This project sought to identify any root causes of gender and financial discrimination, by investigating the disparity in driving test pass rates across England. The project concludes that elements of poverty can correlate with indirect discrimination, specifically in the case of the financial status and gender of a person taking a driving test. The project was able to provide substantial evidence toward proving both of the proposed hypotheses, specifically that poverty indirectly has a greater negative effect on females compared to males in this driving test context, and that generally elements of poverty in an area can make it more difficult to achieve on a test which was never designed to consider the wealth or gender of the person taking the test. Additionally, whilst this project was based upon only a limited number of poverty factors, it did show that income deprivation, education, and crime are factors that are important in affecting a location's driving test pass rates.

## 10.1 Extensions

These conclusions could be used as evidence for more generalized research into the causation of discrimination. More specifically, this project could be expanded upon to further prove the original hypotheses, by testing if they hold for more years than just 2018. By applying the hypotheses to other years, the disparity in how poverty affects pass rates could be compared throughout the years, to perhaps find that driving test indirect discrimination has been getting better or worse the past decade: or that the proposed hypotheses coincidentally only held for the year 2018 that this project chose. To extend upon this project's context, gathering more data surrounding driving tests could identify more correlations with pass rates, such as the gender of the examiner, the type of car used during the test, tests taken in other countries, or even weather and days of the week the test is taken on. Gathering data like this could also be an extension, as this project found that substantial, openly available data involving UK driving tests is extremely limited.

# References


[1] BBC. *How many violent attacks and sexual assaults on women are there?* URL: https://www.bbc.co.uk/news/explainers-56365412 (visited on 09/01/2022).

[2] Matt Davies. *Poverty Exclusion Discrimination.* URL: https://www.jrf.org.uk/sites/default/files/jrf/migrated/files/2271-poverty-exclusion-discrimination.pdf (visited on 09/01/2022).

[3] Equinet. *Addressing poverty and discrimination: two sides of the one coin.* URL: https://equineteurope.org/wp-content/uploads/2019/02/poverty_opinion_2010_english.pdf (visited on 09/01/2022).

[4] Trevor Fenton. *Regional gross disposable household income: all ITL level regions.* URL: https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/datasets/regionalgrossdisposablehouseholdincomegdhi (visited on 09/01/2022).

[5] GOV.UK. *Car driving test data by test centre.* URL: https://www.gov.uk/government/statistical-data-sets/car-driving-test-data-by-test-centre (visited on 09/01/2022).

[6] GOV.UK. *Driving test centres.* URL: https://data.gov.uk/dataset/fe19beff-5716-4ca9-be58-027e56856b48/driving-test-centres (visited on 09/01/2022).

[7] GOV.UK. *English indices of deprivation 2019.* URL: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019 (visited on 09/01/2022).

[8] Stephen H.Fairclough. *Anxiety and performance in the British driving test.* URL: https://www.sciencedirect.com/science/article/abs/pii/S1369847805000707 (visited on 09/01/2022).

[9] Amnesty International. *Discrimination Overview.* URL: https://www.amnesty.org/en/what-we-do/discrimination/ (visited on 09/01/2022).

[10] Perrys. *Driving test pass rates: Men vs Women.* URL: https://www.perrys.co.uk/blog/driving-test-pass-rates-men-vs-women/96876 (visited on 09/01/2022).

[11] Richard Prothero. *Mapping income deprivation at a local authority level.* URL: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/datasets/mappingincomedeprivationatalocalauthoritylevel (visited on 09/01/2022).

[12] Olivia Remes. *The anxiety puzzle: why are women in deprived areas more likely to suffer?* URL: https://theconversation.com/the-anxiety-puzzle-why-are-women-in-deprived-areas-more-likely-to-suffer-46966 (visited on 09/01/2022).

[13] ROC. *How much does it cost to learn to drive?* URL: https://www.rac.co.uk/drive/advice/learning-to-drive/how-much-does-it-cost-to-learn-to-drive/ (visited on 09/01/2022).

[14] RSA. *Students in deprived neighbourhoods being denied access to 'tough subjects'.* URL: https://www.thersa.org/press/releases/2015/students-in-deprived-neighbourhoods-being-denied-access-to-tough-subjects (visited on 09/01/2022).

[15] Daniel Whitmore. *Workplace discrimination statistics in 2021.* URL: https://www.ciphr.com/advice/workplace-discrimination-statistics/ (visited on 09/01/2022).