Statistical Case Studies 2025/26 Semester 1
Part 2: Forensic Modelling of Blood Alcohol Evidence

Deadline: 21 November 2025 16:00

1 Background

In the UK it is a criminal offence to drive a motor vehicle with a blood or breath alcohol concentration above the prescribed limit. In this project we will look at the testing of alcohol in blood. In the UK, the majority of testing is done using the concentration of alcohol in breath, but this is then converted to a measurement of concentration in blood, so the same calculations apply in both cases.

When a person is arrested for driving under the influence of alcohol it is not usually possible to perform an accurate test of the level of alcohol in the blood or breath immediately. Breath tests can be used as an initial screening tool at the scene, but these are not sufficiently accurate for prosecution. Instead, people are taken to a police station or hospital, where the test can be carried out using proper laboratory protocols. As the body clears alcohol from the blood through time this means that if the individual was over the limit, the measured level of alcohol in the blood will be lower at the time of measurement than it was when the person was driving a motor vehicle. To deal with this situation, forensic scientists use an equation to back calculate what the concentration of alcohol would have been in the blood at the time the person was driving.

In this project you will look at the calculations used by forensic scientists to assess whether individuals are above the drink driving limit based on their measured blood alcohol concentration.

2 Task

Forensic services at Police Scotland have contacted you for advice on their blood alcohol computations for criminal cases associated with drink driving. They have provided you with some data (more details below) and want you to answer the following questions:

1. The β elimination rate is used to estimate what a person's blood alcohol concentration would have been at the time the person was driving. It measures how fast the body eliminates alcohol in the blood. If the blood alcohol concentration after time t (hours) is measured as C_t, the blood alcohol concentration at time 0 is estimated as $C_0 = C_t + \beta t$.

   Forensic scientists currently use a range for β to assess whether the suspect may have been over the drink driving limit at time 0. They construct a distribution for β using available data and take the 2.5th and 97.5th percentiles of this distribution as the limits of this range. The range used is the same for all individuals. If an individual has a blood alcohol concentration above the limit, calculated using the 2.5th percentile of the distribution of β, the forensic laboratory will report that they are above the limit.

   Police Scotland want to know whether this is a reasonable approach to take in assessing the probability that a person is over the limit and if not, whether there is an alternative approach you would recommend? If you recommend an alternative approach you should clearly specify how you think Police

Scotland should report their blood alcohol testing results in their expert witness reports. You should describe the advantages and disadvantages of your proposed approach (either the current method or your new method) and any limitations. Note that if you believe the current method is the best available, it is not necessary to recommend a new method, but you should clearly argue why you believe this to be the case.

2. Police Scotland want you to demonstrate your method on the following case study.

   A 70 year old female (weight: 70kg, height: 160cm) is arrested after being stopped by the police while driving. She provides a blood sample to the police 2 hours after her arrest which gives a reading of $C_t = 0.15g/kg$ (i.e. grams of alcohol per kilogram of blood). The legal limit is $x = 0.47g/kg$. The police wish to charge her with the criminal offence of driving while over the legal limit.

   Use the approach you have recommended (either current or new) to give your opinion on whether the individual was over the drink driving limit. Note you can use the dataset to compute the empirical 2.5th percentile of the distribution of β if required.

3. Sometimes, it is too late to use a blood or breath test and the only information available is eyewitness testimony of the quantity of alcohol consumed. While this might not be sufficient evidence on its own for a drink driving conviction (depending on the reliability of the eyewitness testimony), forensic scientists can carry out this sort of analysis as part of a broader collection of evidence in a case – for example in assessing whether someone was drunk when committing another offence. To assess eyewitness testimony in this way, forensic scientists need to use Widmark's equation:

$$C_t = \frac{A}{\text{Weight} \times V_d} - \beta t, \tag{1}$$

   where this time, C_t is the blood alcohol concentration at the time of interest to the court (where t is measured in hours that have elapsed since drinking), A is the dose of alcohol consumed, Weight is the weight of the individual, and V_d is a parameter known as the volume of distribution.

   You can use your dataset to compute estimates of V_d for each individual using the following equation:

$$C_o = \frac{A}{\text{Weight} \times V_d},$$

   where C_o is the blood alcohol concentration (given in your dataset as C_o).

   When using the equation above to estimate the level of alcohol in blood, forensic scientists use the same approach for assessing the range of V_d as they use for β (e.g. they construct a distribution for V_d from data and take the 2.5th and 97.5th percentiles as the limits of this range. The same range is used for all individuals). They also assume independence between β and V_d. In practice, this means that they compute the ranges for β and V_d separately and use the 97.5th percentile for each parameter in Equation (1) to estimate C_t. If C_t is above the prescribed limit, they report the individual as being over the limit at time t.

   Is this a reasonable approach for reporting blood alcohol results in expert witness reports? Describe the advantages, disadvantages and any limitations of the approach.

2.1 Data

You have been provided with a dataset of measurements associated with blood alcohol concentration (BAC) and breath alcohol concentration (BrAC). Note that you have only been asked to consider blood alcohol measurements. The measurements were taken from a PhD thesis which you have also been provided with. The data are taken from Table 8.1 in the thesis and from the graphs in Section 8.3.

The variables in the dataset are:

- Page: the page number of the corresponding graph
- Participant number
- FigureOnPage: the order of the graph on the page
- Sample Type: whether the data are for blood alcohol concentration (BAC) or breath alcohol concentration (BrAC).
- Beta60 (g/kg/h): the negative of the β elimination rate, which describes how quickly alcohol is eliminated from the blood or breath. This is the slope of the fitted line on the graph.
- Co (g/kg): the concentration of alcohol in the blood/ breath at time zero. This is the intercept of the fitted line on the graph. Note that in reality this would always be zero, but C_o is an artificial construct used to compute the alcohol dose.
- Sex: whether the individual is male or female.
- Age (years)
- Weight (kg)
- Height (cm)
- Amount of Alcohol Consumed (g): the alcohol dose.
- Drinking Time (h). Period of time over which the alcohol was consumed.
- Maximum BAC/BrAC (g/kg). The maximum measured BAC or BrAC over time.
- BAC Peak time (min). The time at which the maximum occurred.

2.2 Format

You should submit a written report detailing your findings. Your report must contain no more than 4000 words (not including references, tables, figures and their captions). Please state the word count under the title of your report. You can structure your report as you like, but:

(a) It should contain an executive summary that could be presented in a business context to people without a background in mathematics. This should present and explain your results. This section of the report should be no longer than 500 words and should be readable without specialist statistical knowledge.

(b) The rest of your report should be aimed at forensic scientists (e.g. those with a scientific background but without specific statistical expertise), and should explain exactly what you did, why you did it, and what the conclusions were, in a manner that would allow the analysis to be repeated. By this, I mean that the statistical structure of what you have done should be clear, so that it could be repeated using any statistical software the reader wanted to use. This means that models and results should be presented mathematically and with graphs or tables as necessary, and not using computer code and output.

(c) You should include a description of the data used along with any exploratory data analysis.

(d) You can assume that the reader of your report has read the executive summary.

(e) The report should include no computer commands and no raw output.

(f) You should also submit well-documented R code, either plain or as markdown so that your analysis is reproducible.

The work must be completed in your group of 3, which you must have arranged and registered on Learn. If you have any difficulties with this or have not managed to find yourself a group, please email me or speak to me in the workshops. The first paragraph of your report must list your names with university user names.

Contributions from different team members never end up completely equal, but you should aim for rough equality, with team members each making sure to 'pull their weight', as well as not unfairly dominating. It is the responsibility of everyone in the group to make sure that everyone has a chance to contribute.

3 Mark Scheme

There is no single correct analysis for this type of project, so you will not be marked on the basis of how close you get to some particular model answer. The marks are not subdivided, but will be allocated on a combination of statistical approach and justification, interpretation of results in context and presentation.

- 80 – 100% A report that could be presented to the client or collaborator with little or no revision. Analysis is sound so that conclusions are well-supported statistically. Interpretation is reasonably mature. The project should demonstrate a clear overview of the work, without getting lost in details, and be free of all but minor statistical errors. The work is to a publishable standard.
- 70-79% A report that could be presented to the client or collaborator with little or no revision. Analysis is sound so that conclusions are well-supported statistically. Interpretation is reasonably mature. The project should demonstrate a clear overview of the work, without getting lost in details, and be free of all but minor statistical errors.
- 60 – 69% A project that could be presented after a round of revision, but without having to re-do much of the actual analysis. Some flaws in the analysis or presentation (or minor flaws in both), but basically sound. A good grasp of the statistics and context, so that interpretation is reasonable.
- 50 - 59% Major re-working required before the project could be presented, but containing some sound statistics demonstrating understanding of statistical modelling and its application. Reasonable presentation and organisation.
- 40 – 49% Major flaws in analysis and presentation, but demonstrating some understanding of statistics, and a reasonable attempt to present the results.
- Fail (below 40%) Flawed analysis demonstrating little or no understanding of statistics, and/or incomprehensible or very badly organised presentation.