Researchers in Schools

Predicting the unpredictable? Let's assess a prediction model!

Key Stage 4 Programme

Pupil Name

Coursebook Designed by Dr Deborah Agbedjro De Gregorio

Contents

Timetable and Assignment Submission 1
Timetable – Tutorials 1
Timetable – Homework Assignments 1
Assignment Submission – Lateness and Plagiarism 1
Course Rationale 2
Uni Pathways Mark Scheme 2020 – statistics-based assignments 4
Baseline Assignment (essay based): Pupil Feedback Report 6
Final Assignment (problem-set): Pupil Feedback Report 7
Launch Event 8
Subject Vocabulary 9
Tutorial 1 – Introduction to prediction modelling 11
What is the Purpose of Tutorial 1? 11
TASK 0 – What is the purpose of the Uni pathways course? 11
TASK 1 – Are these research questions leading to predictive (P) or explanatory (E) models? Why? 12
TASK 2 – What is a prediction model? 12
TASK 3 – Why do we experiment to assess a medical treatment and why do we develop prediction models on experiments? 13
TASK 4 – Randomised controlled trails (RCTs) 14
BASELINE ASSIGNMENT 15
Tutorial 2 – From developing to publishing a prediction model 16
What is the Purpose of Tutorial 2? 16
STARTER ACTIVITY – Recap with a QUIZ 16
TASK 1 – Type of treatment outcome 17
TASK 2 – Seven steps for developing a prediction model (Steyerberg and Vergouwe 2014) 18
TASK 3 – How to write a research article presenting a prediction model 19
TASK 4 – HINGE POINT QUESTION 19
HOMEWORK 20
Tutorial 3 – Discrimination and Calibration for prediction models 21
What is the Purpose of Tutorial 3? 21
STARTER - Recap 21
TASK 1 - Matching activity 22
TASK 2 - Give a measure of discrimination for the following prediction model on the data used to develop it. 23
TASK 3 - Is the model calibrated on the new data? 24
HOMEWORK 25
Tutorial 4 – Validation for prediction models 27
What is the Purpose of Tutorial 4? 27
STARTER – Recap 27
TASK 1 - True or False? 27
TASK 2 - Calculate the validated discriminative performance 28
TASK 3 - Assessing the model performance 29

HOMEWORK 29
Tutorial 5 – Calibration plots 31
What is the Purpose of Tutorial 5? 31
STARTER – Recap 31
TASK 1 - Creating a calibration plot 31
TASK 2 -Analysing calibration plots 33
TASK 3 - Assessing calibration performance 35
Final assignment 36
Essay: What is and how do we develop and assess a prediction model? 36
Problem set: Assess the performance of a prediction model for treatment outcome 36
Tutorial 6 – Feedback tutorial 39
What is the Purpose of Tutorial 6? 39
Final assignment feedback from your Uni Pathways Teacher 39
Tutorial 7 – Final tutorial 41
Appendix 1 – Referencing correctly 43
How do I reference? 44
Notes 45

Timetable and Assignment Submission

Timetable – Tutorials

Tutorial	Date	Time	Location
1			
2			
3			
4			
5			
6 (Feedback)			
7 (Feedback)			

Timetable – Homework Assignments

Homework Assignment	Description	Due Date
Tutorial 1		
Tutorial 2		
Tutorial 3		
Tutorial 4		
Tutorial 5		

Assignment Submission – Lateness and Plagiarism

Lateness	
Submission after midnight on 9th August	10 marks deducted
Plagiarism	
Some plagiarism	10 marks deducted
Moderate plagiarism	20 marks deducted
Extreme plagiarism	Automatic fail

Course Rationale

What is prediction? Prediction is a statement about an uncertain event. For example, it answers to the questions: "What is the weather like tomorrow?", "How many will the new Covid-19 cases be in the next week approximately?"

"*Prediction is difficult, especially about the future.*"

Niels Bohr (1885-1962), Nobel Prize in Physics for quantum theory.

Many things are unpredictable. *Prediction modelling* aims to predict events or outcomes based on the pattern of available data using a mathematical algorithm (rule, process).
What is so exciting about prediction modelling? Over the last two decades, prediction modelling has flourished with application in all fields: climate change, finance, Google data use, disease risk and treatments, just to name a few! The demand for good statisticians able to develop prediction models has rocketed. In particular, clinical prediction models have always been so useful (especially during pandemic times), as they inform health care providers and service users about the risk of developing a disease, the risk of the presence of a disease and about the future course of an illness, all based on currently available information about the patient!

It is easy to develop prediction models, but it is not so easy to develop prediction models that actually work. Most prediction models only work on the data used to develop them, but then when they are used to predict the outcome on new subjects, they miserably fail. How can you be sure that prediction models work in general? How do you assess the performance of a prediction model? This is what this course is all about. This course is meant to be a very partial introduction to prediction modelling, and I look forward to being your teacher!

In the first tutorial we will learn about the difference between explanatory and prediction models, we will introduce prediction models and learn about experiments like randomised controlled trails on which most clinical prediction models are built.

The second tutorial will explore the different types of outcomes and variables you can have in your models. Then we will study the different steps needed to develop a prediction model and the different sections of an academic paper presenting a prediction model.

In the third tutorial I will introduce you to measuring model performance in terms of discrimination and calibration.

The fourth tutorial will tackle the most important concept of model validation, which is an essential process to assess a prediction model.

Finally, in the fifth tutorial we will enjoy plotting calibration lines and learn to assess validated calibration for prediction models with continuous outcomes by looking at the calibration plot.

Uni Pathways Mark Scheme 2020 – statistics-based assignments

	Subject Knowledge	Critical Thinking	Written Communication
1st	The work shows a depth of knowledge and understanding of key concepts and statistical methods, drawing on sources of evidence where appropriate. Knowledge is used to build and support highly effective statistical arguments and explanations.	Analyses the available data/evidence and identifies all the relevant statistical information. Identifies and critically evaluates available statistical approaches and evidence, deciding on their credibility, strength and relative significance. Makes connections between different pieces of information and puts them together to form accurate solutions and verify answers and assumptions.	Statistical arguments are complete, and the reasoning behind each step has been clearly communicated to the reader. The work has a coherent flow and is well structured. The writing style is appropriate; statistical language and key statistical terms are used accurately and effectively to support the arguments and explanations made. There are no, or very few, errors in spelling or grammar. Consistent referencing, appropriate paragraphing and use of correctly labelled tables and figures; matching the style taught in the course.
2:1	The work shows an understanding of key concepts and mathematical methods, drawing on sources of evidence where appropriate. Knowledge is used to build and support effective statistical arguments and explanations.	Analyses the available data/evidence and identifies all the relevant statistical information. Engages with multiple approaches and critically evaluates their utility, forming judgements and using reasoning and evidence to inform problem solving strategy. Shows some understanding of the relative value of evidence and data.	Statistical arguments are clearly made, with few gaps in logical reasoning. The work is well-structured. The writing style is appropriate; statistical language and key terms are used correctly. There are few errors in spelling or grammar. Mostly consistent referencing and use of tables and figures; matching the style taught in the course.
2:2	The work shows an understanding of key concepts and statistical methods, with no major misconceptions. Beginning to apply this knowledge to build and support effective statistical arguments and explanations.	Identifies and uses basic statistical evidence and reasoning. Evaluates the quality of statistical reasoning and evidence, forming judgements. Not yet showing understanding of the relative value of evidence and reasoning.	Statistical arguments have been constructed, but there is little explanation from step to step. The work has some structure. The writing style can sometimes be informal; occasionally key terms are not used when it would be appropriate to do so. There are some errors in grammar and spelling, but these do not get in the

			way of communicating the content. Referencing has some consistency; matching the style taught in the course. Limited use of tables and figures.
3rd	Shows a developing understanding of key concepts and statistical methods, with some misconceptions. Not yet applying this knowledge to build and support statistical arguments and explanations.	Beginning to analysis statistical evidence and reasoning. Describes or attempts to solve the problem without any/or little evaluation of whether the approach was reasonable or accurate.	Statistical arguments to solve problems are poorly sequenced, with little or no explanation. The grammar, spelling, style, and structure of the work need improving in order to communicate ideas to the reader. Key terms are not always used correctly. Limited, or no use of referencing, tables and figures.

Baseline Assignment (essay based): Pupil Feedback Report

Name of Pupil	
Name of School	
Name of RIS teacher	
Title of Assignment	

How your assignment is graded:

Grade	Marks	What this means
1st	70+	Performing to an excellent standard at A-level
2:1	60-69	Performing to a good standard at A-level
2:2	50-59	Performing to an excellent standard at GCSE
3rd	40-49	Performing to a good standard at GCSE
Working towards a pass	0-39	Performing below a good standard at GCSE
Did not submit	DNS	No assignment received by The Brilliant Club

Lateness	
Any lateness	10 marks deducted
Plagiarism	
Some plagiarism	10 marks deducted
Moderate plagiarism	20 marks deducted
Extreme plagiarism	Automatic fail

Marks

OVERALL MARK / 100		FINAL MARK / 100 (including any deductions)	
DEDUCTED MARKS		FINAL GRADE	

If marks have been deducted (e.g. late submission, plagiarism) the teacher should give an explanation in this section:

Mark Breakdown and Feedback

Subject knowledge	
	mark
Critical thinking	
	mark
Written communication	
	mark

Final Assignment (problem-set): Pupil Feedback Report

Name of Pupil	
Name of School	
Name of RIS teacher	
Title of Assignment	

How your assignment is graded:

Grade	Marks	What this means
1st	70+	Performing to an excellent standard at A-level
2:1	60-69	Performing to a good standard at A-level
2:2	50-59	Performing to an excellent standard at GCSE
3rd	40-49	Performing to a good standard at GCSE
Working towards a pass	0-39	Performing below a good standard at GCSE
Did not submit	DNS	No assignment received by The Brilliant Club

Lateness	
Any lateness	10 marks deducted
Plagiarism	
Some plagiarism	10 marks deducted
Moderate plagiarism	20 marks deducted
Extreme plagiarism	Automatic fail

Marks

OVERALL MARK / 100		FINAL MARK / 100 (including any deductions)	
DEDUCTED MARKS		FINAL GRADE	

If marks have been deducted (e.g. late submission, plagiarism) the teacher should give an explanation in this section:

Mark Breakdown and Feedback

Subject knowledge	
	mark
Critical thinking	
	mark
Written communication	
	mark
Problem Set	
Comment	*mark*

Launch Event

The launch event is your opportunity to learn more about this course and Uni Pathways. You will watch some videos, listen to your teacher and have some discussion. Below is some room to make notes for some of the videos you will watch during the event.

Virtual Campus Tours

Notes…

The Scholars Programme Alumni

Notes…

Student Ambassadors

Notes…

Subject Vocabulary

Word	Definition	In a sentence
Treatment outcome	The result or consequence of applying a treatment to a patient.	The *treatment outcome* for the treatment of a disease was given by three categories: "improved", "stable", and "worsened".
Prediction model	A prediction model is a tool (or algorithm) developed to forecast future events (or outcomes) using known (previous) information (or data).	The doctor used a *prediction model* for cancer treatment outcome to understand whether the patient would benefit of the treatment or not.
Control (as a noun)	A group that is part of an experiment and does not receive the treatment (or intervention) in the experiment in order to check if the treatment works by comparing it with the treatment group.	To understand if the new medicine works for constipation in an experiment, we need a group taking the new medicine to be compared with a *control* receiving the previous medicine.
Randomized Controlled Trail (RCT)	An experiment (or study) that measures if a new intervention or medical treatment works.	The data used to develop a prediction model for treatment outcome need to be measured in an *RCT*.
Discrimination	The ability of a prediction model to distinguish a patient with one category of the outcome from a patient with another category of the outcome.	Assessing *discrimination* of a prediction model is a way of assessing an important part of his performance.
Calibration	The ability of a prediction model to make the predictions agree with the known (previous or observed) data.	Checking *calibration* of a prediction model allows you to assess whether the model is predicting fairly or not.
Validation	The process of assessing if a developed prediction model performs well on new/unseen data, different form the data used to develop the model.	*Validation* of performance for a prediction model is key to be able to say that the model is predicting well or not.
Generalizable	Fairly applicable to different samples of the same population (in statistical terms) or to similar populations.	A prediction model is *generalizable* if predicts new data from a similar population well.
Mean Squared Error (MSE)	Measure of discrimination for prediction models for continuous outcomes. The lower the value of MSE, the better the model. $$MSE = \frac{1}{n}\sum(y - \hat{y})^2$$ where n is the sample size, y is the observed outcome and $\hat{y}$ is the predicted outcome.	The *MSE* of the prediction model for cancer outcome applied to new data was fairly low.

Population variance	The population variance of a variable y measures its spread around the mean: $$Var(y) = \frac{1}{n}\sum(y-\bar{y})^2$$ where n is the sample size, y is the observed outcome and $\bar{y}$ is the mean of the observed outcome.	The variable recording the salaries of the people in London has a very large *variance*, meaning that there are very low and very high salaries compared to the mean salary.
R-squared	Measure of discrimination for prediction models for continuous outcomes. It is the proportion of variance explained in the outcome by the model: $$\begin{aligned} R\ squared &= 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} \\ &= 1 - \frac{MSE}{\sum(y-\bar{y})^2} \times n \\ &= 1 - \frac{MSE}{Var(y)} \end{aligned}$$ where n is the sample size, y is the observed outcome, $\hat{y}$ is the predicted outcome and $\bar{y}$ is the mean of the observed outcome.	When the validated *R-squared* of a prediction model is high implies that the validated MSE is low.
Calibration slope	The ability of a prediction model to make the predictions agree with the known (previous or observed) data.	If the validated *calibration slope* of a prediction model is larger than 1, low predictions will be too high and high predictions will be too low.
Continuous, Binary, Ordinal and Categorical variables	**Binary**: variable with only two numerical values or with two nominal values (two categories) **Categorical**: with a limited number of categories (usually less than 10), which have no natural order. **Ordinal**: with a limited number of values (usually less than or equal 10), which are ordered. **Continuous**: with many values (usually 10 or more), which are ordered.	The variable taking the numerical values "1" and "0" and the variable taking the nominal values "yes" and "no" are *binary variables*. A variable recording the possible colours for a newly designed car, with the categories "green", "white" and "purple", is a *categorical variable*. The variables used in a questionnaire, answering questions about symptoms severity for an illness, which takes the values "no symptoms", "mild", "moderate", "severe", is an *ordinal variable*. A variable recording the age of the participants of a clinical trial is a *continuous variable*.

Tutorial 1 – Introduction to prediction modelling

What is the Purpose of Tutorial 1?

- Students can recall the aims and purpose of Uni pathways.
- Students can recall what a prediction model is and why it is useful.
- Students can recall why we do experiments to assess medical treatment and why we develop prediction models on experiments.
- Students can explain what a randomized controlled trial is and why it is necessary to develop a prediction model for treatment outcome.

TASK 0 – What is the purpose of the Uni pathways course?

Write down the purpose of Uni pathways and prepare to cold calling.

TASK 1 – Are these research questions leading to predictive (P) or explanatory (E) models? Why?

Think Pair Share (TPS): 1 minute to think, 1 minute to share in pairs and 2 minutes to share with the class. Tick in the correct column whether the research question will lead to predictive (P) or explanatory (E) modelling, explain why when called out and identify outcome (dependent variable) and independent variable or predictor.

Research question	P	E	Outcome-dependent variable	Independent variable/predictor
1. Being able to tell the probability/risk of death for a person knowing their characteristics like age, gender, weight and height, lifestyle (smoking and exercise), physical and mental health.				
2. Trying to understand if alcohol consumption has an effect on memory.				
3. Does the treatment for depression work on average?				
4. Deciding if a patient suffering from cancer will benefit from a treatment or not.				
5. What are the patients' characteristics which best predict recovery?				

Key concepts

- **Prediction modelling:** aims to find the group of factors (characteristics) that best predict a current diagnosis, future event or treatment outcome. The focus is predicting well.
- **Explanatory modelling:** aims to identify factors that can cause an outcome. The focus is explaining causality.
- **Outcome or dependent variable**: variable that we observe and measure, whose value will change depending on the value of another variable, called the independent variable. The dependent variables are the outcomes of the experiments determining what was caused or what changed as a result of the study.
- **Independent variable vs predictor variable**: An **independent variable** isn't changed by the other variables you are trying to measure: it is independent from other variables, but it changes the dependent variable. The independent variable is a variable that we change in an experiment to understand what happens to the dependent variable which depend on it. An independent variable is called a **predictor variable** in prediction modelling.

TASK 2 – What is a prediction model?

Complete this True/False quiz about prediction modelling and prepare to cold calling:

Statement	T	F
A prediction model is a mathematical tool used to predict future events.		
Prediction modelling also aims to explain if the predictors cause the outcome.		
Prediction models are developed using already known data.		

Key concept:

- **Prediction model definition:** A prediction model is a mathematical process/equation/formula/algorithm/function that seeks to predict future events or outcomes by analysing patterns that are likely to forecast future results from known, previously observed data.

TASK 3 – Why do we experiment to assess a medical treatment and why do we develop prediction models on experiments?

Answer the questions in the workbook and prepare to cold calling:

1) Why are experiments useful to understand if a medical treatment work?

_____

_____

_____

2) Give an example of experiment of your choice.

_____

_____

_____

3) How can experiments for medical treatment be used to create a prediction model?

_____

_____

_____

Key concept:

- **Statistical experiment**: In an experiment, a researcher tests an intervention on some subjects. By noting how the intervention affects the outcome (response variable), the researcher can test whether a causal relationship exists between the intervention and the response variable. When the intervention is a medical treatment and the outcome is the subject response to an illness, then the researcher can understand how this medical treatment affects the outcome when the experiment is correctly set up. When the statistical experiment is correctly planned, the data measured form the results of the experiment can be used to develop a prediction model to forecast the outcome of an intervention.

TASK 4 – Randomised controlled trails (RCTs)

Answer one question at a time about RCTs on the mini whiteboards:

1) List the three characteristics of an RCT?
2) Why do you need RCT data to develop a prediction model for treatment outcome?
3) A study was conducted to assess the efficacy of a psychological treatment for depression and the treatment was given to all the participants. Is the data from this study suitable to develop a prediction model for treatment outcome? Why?

Key concepts:

- **RCT**: sufficiently large experiment usually comprising at least 2 groups of subjects/people with the illness, randomly drawn from the population. One group receives the treatment (treatment group), the other group receives a placebo (control), which is a fake treatment (giving a sweet instead of a medicine, by making the patient believe they are receiving a medicine). Then a comparison is made between the treatment outcome of the two groups to see if the treatment works. To summarise, an RCT should be a **large** study, **controlled** (having a control), and **randomised** (the participants need to be randomly allocated to treatment or control in order to have unbiased/fair results).
  To measure a treatment outcome, we need to set up an RCT because only this study/experiment returns an unbiased (fair) estimate of the treatment effect and therefore of the treatment outcome.
- **Prediction models for treatment outcome** are algorithms (tools, processes, functions) developed to forecast future treatment outcomes for patients with a certain illness using known data about them. Prediction models for treatment outcome are needed to decide whether giving the treatment to a particular patient or not, considering their characteristics (predictor variables) such as "age", "gender", "symptoms before treatment", or any other characteristic relevant to the specific illness, measured before treatment.

BASELINE ASSIGNMENT

What are prediction models, why are they useful? Why do we experiment and develop prediction models on experiments?

- **Details of the assignment:**
  Write a small essay of 400 to 500 words introducing prediction models according to what you learnt in tutorial 1 for next week.

- **Success criteria:**
  By structuring the essay with an introduction, a main body, and a conclusion, you will need to include the following:
  1. What prediction models are and why we develop them.
  2. Why we do experiments to assess medical treatment and what randomised controlled trails (RCTs) are.
  3. Why we need RCTs to develop prediction models for treatment outcomes.

  The text should be correctly referenced.

- **Show me that you understand:**
  - The difference between prediction modelling and explanatory modelling
  - Prediction models for treatment outcome

- **Show me that you are able to:**
  - Distinguish between different kinds of prediction models.
  - Communicate your ideas clearly and correctly.

Tutorial 2 – From developing to publishing a prediction model

What is the Purpose of Tutorial 2?

- Students can list different types of treatment outcomes.
- Students can name and recall explanation for the 7 steps to develop a prediction model.
- Students can recall what features of a developed prediction model are presented in an academic paper.

STARTER ACTIVITY – Recap with a QUIZ

1) Tick the correct answer.

   In an RCT,

   a. all participants receive the intervention. ☐

   b. there are always exactly 2 group of participants, one receiving treatment and the other receiving the usual treatment or a placebo. ☐

   c. there needs to be a large enough number of patients. ☐

   d. the participants are not randomly allocated to the treatment group and the control group. ☐

2) The research team of the biostatistics department of King's College London University investigated on the effect of lockdown on the productivity of women at work. Is this research question leading to prediction modelling? Why?

   _____

   _____

   _____

TASK 1 – Type of treatment outcome

Tick the correct box for each example of treatment outcome:

Treatment outcome	binary	continuous	categorical	ordinal
"Recovered" or "not recovered"				
Fever temperature				
Scale 1-6 indicating the severity (i.e., the degree of pain) of an illness with '1' being low pain and '6' extremely high pain.				
"Improved", "Not improved", "Not measurable"				

Key concept:

- **Treatment outcomes** can be binary, ordinal, categorical or continuous variables. The **binary** outcome only has two values and, when the values are nominal, they are usually the statements: "The patient has improved/recovered" or "The patient has remained stable with their illness/has worsened", summarised in the binary outcome "YES" or "NO", answering the question "Has the patient improved/recovered?". An **ordinal** treatment outcome has more than 2 possible values (almost always less than 10 values), but these values (or categories) are ordered (ranked), for example: "recovered", "improved, but not recovered", "worsened"; or just a rank of numbers, like from 1 to 8, according to the degree of improvement or disease severity. A **categorical** treatment outcome is like an ordinal outcome, but the categories are not ordered, for example the result of a test for the illness symptoms, which sometimes is void, like: "presence of blood clots", "no blood clots detected", "null test"). A **continuous** outcome has many values (usually 10 or more), which are ordered, for example the measurement of symptoms of an illness according to a scale ranging from 1 to 30.

TASK 2 – Seven steps for developing a prediction model (Steyerberg and Vergouwe 2014)

Match each step of prediction model development with the correct description:

1. Problem definition	a. Computing the parameters of the model. For example, in the line of best fit with equation y=3x+5, 3 and 5 are the parameters.
2. Data pre-processing	b. Exhibiting the model as a formula in a research article or as a user-friendly software (e.g., in Excel or as a web-based calculator).
3. Model specification	c. Using the developed prediction model to predict new/unseen data, different form the data used to develop the model, to assess performance fairly.
4. Model estimation	d. Measuring the abilities of the model to predict well.
5. Model performance	e. The research question is stated and what is currently known about the potential predictors is searched for. The outcome is defined and how the treatment effect is dealt with in the predictive analysis is explained (e.g., is it included in the model or ignored?). Information about the selection criteria of the patients and the available data is provided such as reliability and completeness of the dataset (missing data problem).
6. Model validation	f. Deciding which method to use to develop the model, which can be a line of best fit (also called regression), or a more complicated algorithm. And determining which predictors to include in the model according to the academic literature.
7. Model presentation	g. Categorical variables to include in the model as potential predictors need to be inspected since categories with small frequency can be merged with others to have a more accurate model. The dichotomisation of continuous variables can be done if clinically sensible and user-friendly only in a later phase of the model development depending on the quantity of predictive information lost. In this step, averages and measures of spread for each variable are provided to describe the data.

TASK 3 – How to write a research article presenting a prediction model

Group activity:

In groups of two, write here below, in order, all the standard parts of a research paper and briefly summarise what happens in each of them if you're reporting the research about a developed prediction model. The names of the first two sections of an academic paper have been written down for you:

0. Abstract: _____

_____

_____

1. Introduction/ background: _____

_____

_____

2._____: _____

_____

_____

3._____: _____

_____

_____

4._____: _____

_____

_____

TASK 4 – HINGE POINT QUESTION

Think (1 minute), pair (2 minute), share (2 minutes)

- In a research article reporting a developed prediction model, the authors decided to write the descriptive statistics of the data under investigation in the methods section. Do you think they are correct? Why?

HOMEWORK

Read the following abstract of a published article about the development and validation of a prediction model for mortality. I have removed the titles to the four different abstract sections, which divided the abstract into paragraphs. I will provide you with the title and the reference of this paper in the next session.

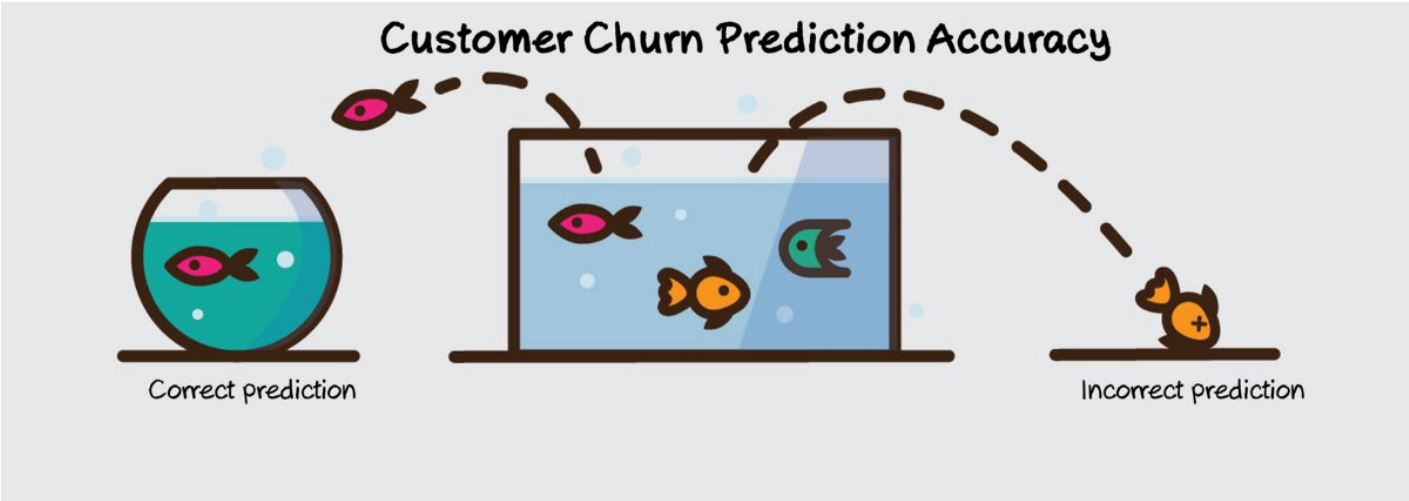
1) You're not expected to understand everything, just identify, and add the titles to the different sections, by dividing the text into paragraphs.
2) Try to highlight any sentence referring to the seven different steps of prediction model development and specify which step it refers to.

In increasingly ageing populations, there is an emergent need to develop a robust prediction model for estimating an individual absolute risk for all-cause mortality, so that relevant assessments and interventions can be targeted appropriately. The objective of the study was to derive, evaluate and validate (internally and externally) a risk prediction model allowing rapid estimations of an absolute risk of all-cause mortality in the following 10 years. For the model development, data came from English Longitudinal Study of Ageing study, which comprised 9154 population-representative individuals aged 50–75 years, 1240 (13.5%) of whom died during the 10-year follow-up. Internal validation was carried out using Harrell's optimism-correction procedure; external validation was carried out using Health and Retirement Study (HRS), which is a nationally representative longitudinal survey of adults aged ≥50 years residing in the United States. Cox proportional hazards model with regularisation by the least absolute shrinkage and selection operator, where optimisation parameters were chosen based on repeated cross-validation, was employed for variable selection and model fitting. Measures of calibration, discrimination, sensitivity and specificity were determined in the development and validation cohorts. The model selected 13 prognostic factors of all-cause mortality encompassing information on demographic characteristics, health comorbidity, lifestyle and cognitive functioning. The internally validated model had good discriminatory ability (c-index=0.74), specificity (72.5%) and sensitivity (73.0%). Following external validation, the model's prediction accuracy remained within a clinically acceptable range (c-index=0.69, calibration slope β=0.80, specificity=71.5% and sensitivity=70.6%). The main limitation of our model is twofold: 1) it may not be applicable to nursing home and other institutional populations, and 2) it was developed and validated in the cohorts with predominately white ethnicity. A new prediction model that quantifies absolute risk of all-cause mortality in the following 10-years in the general population has been developed and externally validated. It has good prediction accuracy and is based on variables that are available in a variety of care and research settings. This model can facilitate identification of high risk for all-cause mortality older adults for further assessment or interventions. Keywords: Mortality, Survival, Prognostic factors, Statistical learning, Absolute risk, Population-based longitudinal study.

REFERENCES

Steyerberg, EW and Y Vergouwe (2014). 'Towards better clinical prediction models: seven steps for development and an ABCD for validation'. In: European Heart Journal 35, pp. 1925–1931.

Tutorial 3 – Discrimination and Calibration for prediction models



What is the Purpose of Tutorial 3?

- Students can define discrimination and calibration as abilities of a prediction model.
- Students can name two measures of performance used to assess discrimination and one measure for calibration for continuous treatment outcomes.
- Students can assess the fit of a prediction model with continuous outcome by looking at the value of these three measures of performance

STARTER - Recap

Answer the following questions:

1) Briefly, explain what you do in the model specification step to develop a prediction model.

   _____

   _____

   _____

2) A variable records level of salary defined with Low, Medium, and High categories. Is this variable continuous, ordinal, categorical or binary?

   _____

3) List the 4 sections of an academic paper.

   1-_____

   2-_____

   3-_____

   4-_____

TASK 1 - Matching activity

In pairs, match the statements on the right with the measure of performance they refer to:

1. Discrimination

2. Calibration

a) It is the reliability, the 'correctness' of the predictions according to the observed outcome

b) In presence of a continuous outcome, it's the ability to separate or rank-order observations with different outcomes

c) It is better for a model which predicts risk with more variability

d) It is the ability of the model to make unbiased predictions

e) It's related to whether or not the outcome predictions are in any way associated to the observed outcome

Key concept – Prediction model performance is expressed in terms of discrimination and calibration:

- **Discrimination:** is the ability of a prediction model to distinguish between a patient with the outcome and a patient without the outcome, to separate between high and low risk. Discrimination is associated to whether or not the predictions are related to the outcome. Discrimination is the most important measure of performance for prediction models. If a prediction model has poor discrimination, no adjustment can be done to improve it.
- **Calibration**: is the ability of a prediction model to return fair (unbiased) predictions, i.e., the ability to return predictions which agree with the observed data. A model with good discrimination can have poor calibration, which can be corrected with recalibration. Discrimination of a model won't change after recalibration. Good calibration doesn't imply good discrimination, but bad calibration implies lower discrimination. Perfect calibration is possible with poor discrimination.

TASK 2 - Give a measure of discrimination for the following prediction model on the data used to develop it.

A model to predict cognitive abilities knowing the severity of symptoms was developed using the data of 10 people with schizophrenia (SCZ, mental illness). The model has this formula:

$$y = -4.2x + 30$$

where x measures the **symptom score** for schizophrenia (the higher the score, the worse the symptoms) and the outcome y measures the **cognitive ability score**. The scores of each of the participants are in the table below.

1) Complete the table.

Individuals $n = 10$	SCZ symptom score (x)	Cognitive ability score (y)	Predicted cognitive ability score ($\hat{y}$)	Error $y - \hat{y}$	Squared Error $(y - \hat{y})^2$
Mark	1	28			
Jahied	2	29			
Cloe	6	14			
Ilias	4	18			
Olesya	2	21			
Diana	4	12			
Omar	1	31			
Rukhshana	6	9			
Sakiba	6	1			
Jenny	7	5			

2) Calculate the Mean Squared Error (MSE) of this model, writing down your working outs.

3) How is the model discrimination on this data, knowing that the population variance of the outcome y is 97.56?

Key concepts – see definitions for Mean-squared-error (MSE), variance of a variable and R-squared in the Subject Vocabulary at pages 10-11.

TASK 3 - Is the model calibrated on the new data?

The same model in Task 2 was used to predict the cognitive abilities of other 10 patients with SCZ as a test, and the calibration slope of the model applied on the new data was 1.8. Can we say that the model is well-calibrated on the new patients' data? Why?

Key concept

- A calibration slope of **1** gives you perfect calibration, if it is too far from one (more than 1.5 or less than 0.8), then calibration is poor. However, we can always correct a model for calibration and this process is called recalibration.

HOMEWORK

A randomised controlled trial was run to predict cognitive abilities on 20 people with schizophrenia offered either a psychological treatment called Cognitive Remediation (CR) or treatment as usual. The model has this formula:

$$y = -3.61t - 0.62y_0 + 21.31$$

Where t is the binary variable indicating **treatment randomisation** (1 for CR and 0 for usual treatment), y_0 measures the **baseline cognitive abilities scores** for schizophrenia (outcome measured before randomisation) and the outcome y measures the **cognitive ability score after treatment**. The scores of each of the participants are in the table below.

1) Complete the table.

Individuals $n = 10$	Treatment allocation (t)	Baseline cognitive ability score (y_0)	Cognitive ability score post treat. (y)	Predicted cognitive ability score ($\hat{y}$)	Error $y - \hat{y}$	Squared Error $(y - \hat{y})^2$
1	0	7	17			
2	0	9	15			
3	0	18	7			
4	0	14	14			
5	0	8	16			
6	0	14	11			
7	0	16	18			
8	0	20	17			
9	0	19	9			
10	0	16	8			
11	1	20	11			
12	1	19	1			
13	1	16	5			
14	1	20	9			
15	1	19	9			
16	1	16	8			
17	1	20	5			
18	1	19	12			
19	1	16	13			
20	1	20	11			

2) Calculate the Mean Squared Error (MSE) of this model, writing down your working outs.

3) How is the model discrimination on this data, knowing that the population variance of the outcome y is 19.66?

4) The same model was used to predict the cognitive abilities of other 20 patients with SCZ as a test, and the calibration slope of the model applied on the new data was 0.92. Can we say that the model is well-calibrated on the new patients' data? Why?

Tutorial 4 – Validation for prediction models

What is the Purpose of Tutorial 4?

- Students can recall and explain why we need to validate the model performance to assess a prediction model.
- Students can recall and summarize the process of validation.
- Students can assess the performance of a prediction model with continuous outcome, given the values of the validated performance measures.

STARTER – Recap

Think Pair Share:

Aryan calculated a prediction model R-squared on the data used to develop the model. He obtained an R-squared of 0.85 and he stated that the model performs well in discrimination in general. Is he right?

TASK 1 - True or False?

Statement	T	F
The performance of a prediction model on the data used to develop the model is valid and reliable.		
Prediction models need to be tested on new data in order to assess their true performance.		
The performance of a prediction model on the data used to develop the model is called apparent performance.		
The apparent performance is usually lower than the validated performance.		

Key concept

- **Apparent performance** of prediction models: it is the performance of the model on the development data (i.e. the data used to develop/train the model). It is usually too optimistic about the real/general performance of the model.
- **Validation** of prediction models: it is the process of testing the model on new data to obtain valid measures of performance.
- **Validated performance** of prediction models: it is the performance of the model when applied to new data. It is a good estimate of the real/general performance of the model.

TASK 2 - Calculate the validated discriminative performance

Consider the prediction model in Tutorial 3, Task 2:

$$cognition = -4.2 \times symptoms + 30$$

The model was then tested on 10 different SCZ patients, whose data are in the table below.

1) By completing the table, work out the validated MSE and R-squared, knowing that the population variance of the cognitive ability scores for the 10 new patients is 129.84.

New Individuals $n = 10$	SCZ symptom score (x)	Cognitive ability score (y)	Predicted Cognitive ability score $(\hat{y})$	Error $y - \hat{y}$	Squared Error $(y - \hat{y})^2$
1	3	25			
2	4	23			
3	8	2			
4	2	30			
5	7	7			
6	5	17			
7	9	1			
8	3	26			
9	2	30			
10	1	33			

2) Compare the apparent and validated measures of discrimination (the apparent R-squared was 0.75, see Tutorial 3, Task 2). Did you expect a difference? Why?

TASK 3 - Assessing the model performance

Consider the model in Task 2 above and consider the same new 10 patients. The calibration slope of the model on this new data is 1.03. Suppose this model was only tested on the 10 new patients in the table above. Assess the model performance in terms of discrimination (see your working outs above) and calibration. Would this model be good for clinical practice? Hint: think of the measure of validated performance and sample size.

Key concept

- In order to be good for clinical practice, not only has a model to have good validated measures of performance in discrimination and acceptable calibration measures (which can be corrected through recalibration), but it also has to be developed and validated on a substantial sample size, to be credible.

HOMEWORK

Reconsider the model studied in the homework of Tutorial 3: a randomised controlled trial run to predict cognitive abilities on 20 people with schizophrenia offered either a psychological treatment called Cognitive Remediation (CR) or treatment as usual. The model had this formula:

$$y = -3.61t - 0.62y_0 + 21.31$$

Where t is the binary variable indicating **treatment randomisation** (1 for CR and 0 for usual treatment), y_0 measures the **baseline cognitive abilities scores** for schizophrenia (outcome measured before randomisation) and the outcome y measures the **cognitive ability score after treatment**.

The model was then tested on 20 different SCZ patients to validate its performance. The data for the validation sample are in the table below.

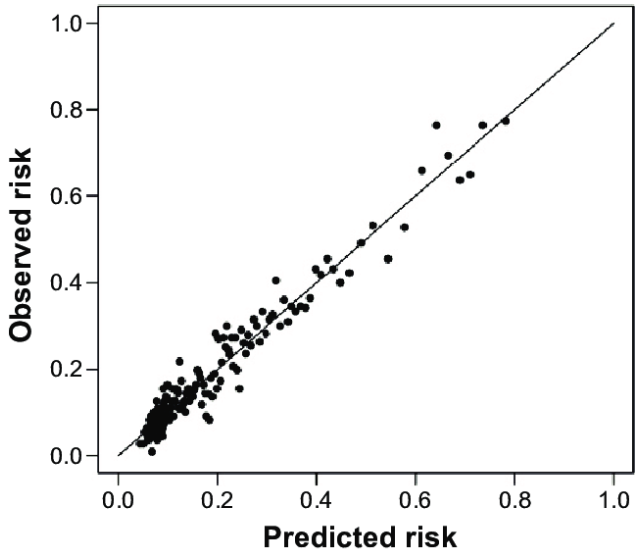
1) By completing the table, work out the validated MSE and R-squared, knowing that the population variance of the cognitive ability scores after treatment for the 20 new patients is 15.49.

2) Compare the apparent and validated measures of discrimination (the apparent R-squared was 0.70, see Homework of Tutorial 3). Assess the discriminative ability of the model.

3) The calibration slope of the model on this new data is 0.9. Assess the model performance in terms of calibration. Considering the model validated discrimination (answer to question 2), would this model be good for clinical practice?

New Individuals $n = 20$	Treatment allocation (t)	Baseline cognitive ability score (y_0)	Cognitive ability score post treat. (y)	Predicted cognitive ability score $(\hat{y})$	Error $y - \hat{y}$	Squared Error $(y - \hat{y})^2$
1	0	17	4			
2	0	13	13			
3	0	17	7			
4	0	13	16			
5	0	22	9			
6	0	17	10			
7	0	16	6			
8	0	19	8			
9	0	14	8			
10	0	13	11			
11	1	14	10			
12	1	15	10			
13	1	7	17			
14	1	15	4			
15	1	24	4			
16	1	19	1			
17	1	19	8			
18	1	13	8			
19	1	18	4			
20	1	17	7			

Tutorial 5 – Calibration plots



What is the Purpose of Tutorial 5?

- Students can recall and explain how a calibration plot for a prediction model with continuous outcome, is obtained.
- Students can analyze a calibration plot.
- Students can assess calibration of a prediction model by looking at the model calibration plot.

STARTER – Recap

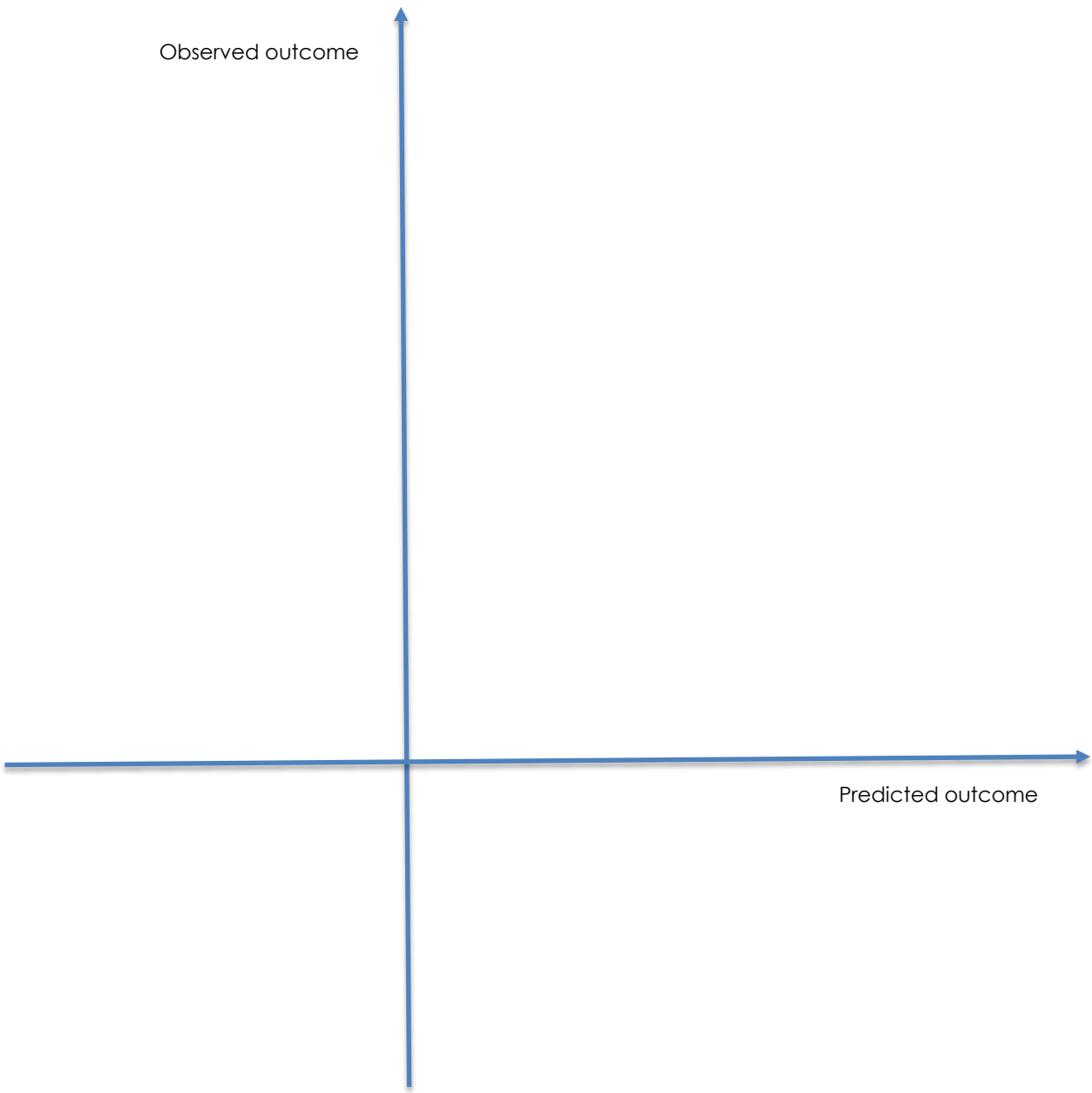
True or false?

Statement	T	F
If the apparent performance of a prediction model is good, then the model is generalizable.		
The validated performance is usually higher than the apparent performance.		
To validate a prediction model, you test it on new data and then measure the performance of it on the new data.		

TASK 1 - Creating a calibration plot

Let's consider the model predicting cognitive abilities knowing the illness symptoms, which was tested on 10 new people with SCZ (see Tutorial 4, Task 2). Using the table below, draw a calibration plot.

New Individuals $n = 10$	Observed Cognitive ability score (y)	Predicted Cognitive ability score $(\hat{y})$
1	25	17.4
2	23	13.2
3	2	-3.2
4	30	21.6
5	7	0.6
6	17	9.0
7	1	-7.8
8	26	17.4
9	30	21.6
10	33	25.8



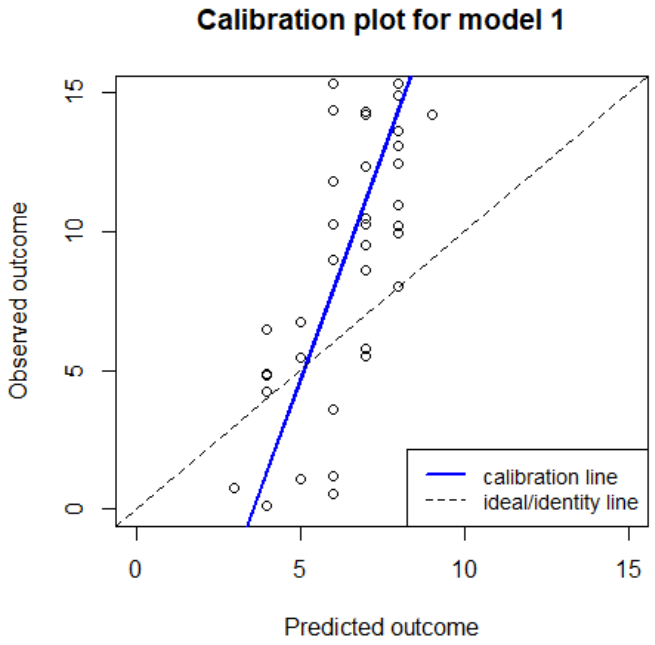
Key concept:

- You will need to plot a xy-graph and scatter **points** having the **predicted outcome as x-coordinate** and the **observed outcome as y-coordinate** for each of the 10 individuals. Therefore, you will need to plot 10 points on your xy-graph. Careful with the scaling of your axes.
- Then, try to sketch a line of best fit for these points, which will be your **calibration line**, showing how much observed and predicted outcomes agree.

TASK 2 -Analysing calibration plots

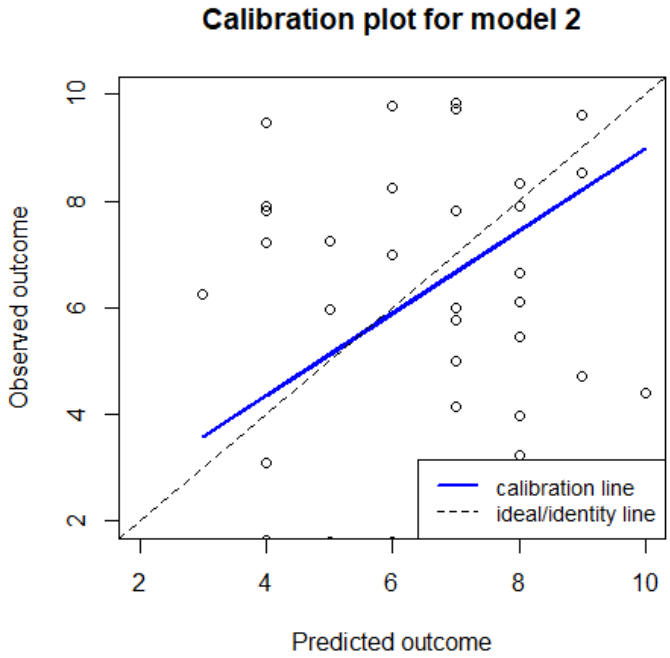
Look at the calibration plots for different prediction models with continuous outcomes applied on new data. For each plot, answer the following questions:

1) is the calibration slope negative, positive or null?
2) if positive, is the calibration slope larger or lower than 1?



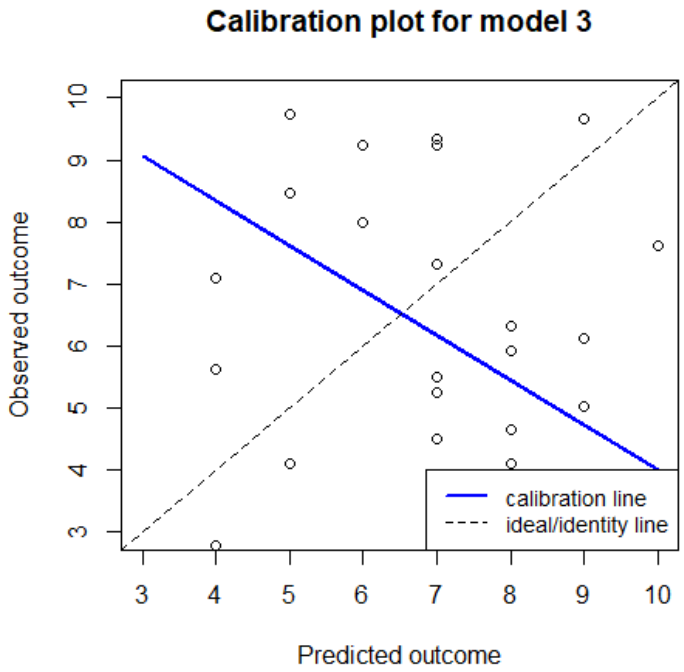
1) _____

2) _____



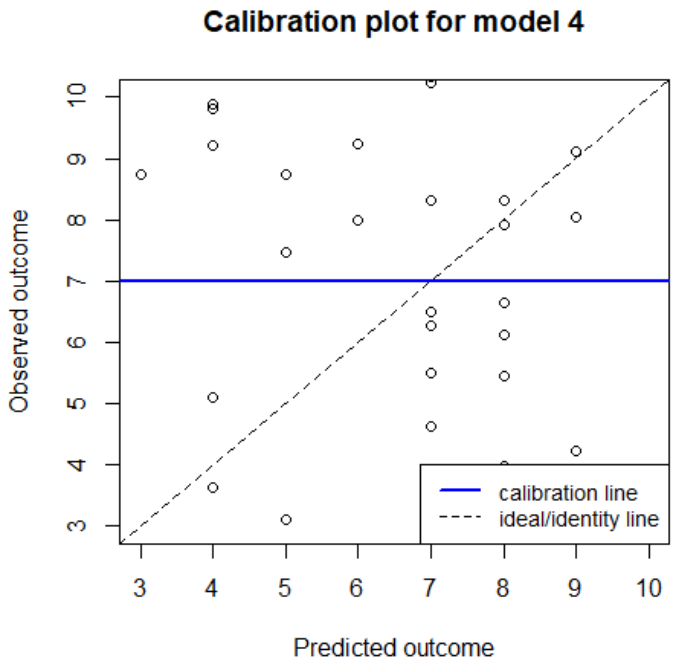
1)_____

2)_____



1)_____

2)_____



1)_____

2)_____

Key concept:

- The **calibration slope** is the **gradient** of the calibration line. The gradient of a straight line indicates the steepness of the line. If you consider many straight lines all crossing the origin of the xy-graph, the steepest line will have the highest gradient.
- Recall that the equation of a generic straight line is $y = \boldsymbol{m}x + c$, where $\boldsymbol{m}$ is the gradient and c is the y-coordinate of the point in which the line crosses the y-axis. Therefore, for example, a calibration line with calibration slope equal to 1.3 will have an equation of the form $y = 1.3x + c$, where y is the observed outcome and x is the predicted outcome.

TASK 3 - Assessing calibration performance

Now, I will give you the equations for the calibration line for each model calibration plot in Task 2:

Model 1 calibration line equation: $observed = 3 \times predicted - 10$
Model 2 calibration line equation: $observed = 0.75 \times predicted + 1.7$
Model 3 calibration line equation: $observed = -0.6 \times predicted + 11$
Model 4 calibration line equation: $observed = 7$

1) In pairs, comment on each model's calibration performance on new data by specifying if the model is overfitting or underfitting the new data or returning inverse predictions or having null calibration (i.e., the observed outcome completely unrelated to predicted outcome).
2) Is there any model with good calibration performance?
3) If you replied 'yes', say which one and explain why and if you replied no, would you be able to correct miscalibration in any of the models?

Key concepts

- **Overfitting:** a prediction model overfits the data when it overestimates high outcome values and underestimates low outcome values. In other words: high predictions will be too high and low predictions will be too low. It means that the model will perform very well on the development data but will have poor performance on new data.
- **Underfitting:** a prediction mode underfits the data when it underestimates high outcome values and overestimates low outcome values. In other words: high predictions will be too low and low predictions will be too high.
- When the calibration slope is negative, it means that the model predictions are the opposite sign compared to the observed outcome values: we have **inverse predictions**. This kind of calibration is very poor and cannot be corrected.
- **Null calibration slope**: there is no association between observed and predicted outcomes. A model with null calibration slope cannot be recalibrated.

Final assignment

The final assignment consists of two parts: an essay and a problem set, which will have the same weigh on your final grade. The success criteria for the essay are the same as the ones for the baseline assignment: the exposition should be clear and use appropriate language. The problem set answers should mirror the homework tasks of Tutorials 3 and 4 and the tasks of Tutorial 5.

Essay: What is and how do we develop and assess a prediction model?

Write a 1000-word essay introducing prediction models for treatment outcome and explaining the assessment of their performance, considering what you have learnt during the Uni Pathways course.

By structuring the essay with an introduction, a main body (that can be paragraphed using subheadings) and a conclusion, you will need to include and show your understanding of:

- What prediction models for treatment outcome are and why we develop them.
- The type of study needed to unbiasedly measure treatment outcomes and why.
- The steps to develop a general prediction model.
- What prediction model performance is in terms of discrimination and calibration.
- The measures of prediction model performance for continuous outcomes.
- The process of validation of prediction model performance and why this is necessary, focusing on external validation.

If you plan to include graphs in the essay, they should be numbered, labelled, and captioned. You should be able to use the correct statistical language throughout the text, enriching the essay with some personal reading of at least one article in the list of useful references I will send you via email to help you with your essay. When you include any information from your additional readings, please reference the text correctly.

Problem set: Assess the performance of a prediction model for treatment outcome

A randomised controlled trial was run to predict severity of depression on 20 people with depression offered either a psychological treatment called electroconvulsive therapy (ECT) or treatment as usual. The model had this formula:

$$y = -3.82t - 0.61y_0 + 20.29$$

Where t is the binary variable indicating **treatment randomisation** (1 for ECT and 0 for usual treatment), y_0 measures the **baseline depression severity scores** for depression (outcome measured before randomisation) and the outcome y measures the **depression severity score after treatment**.

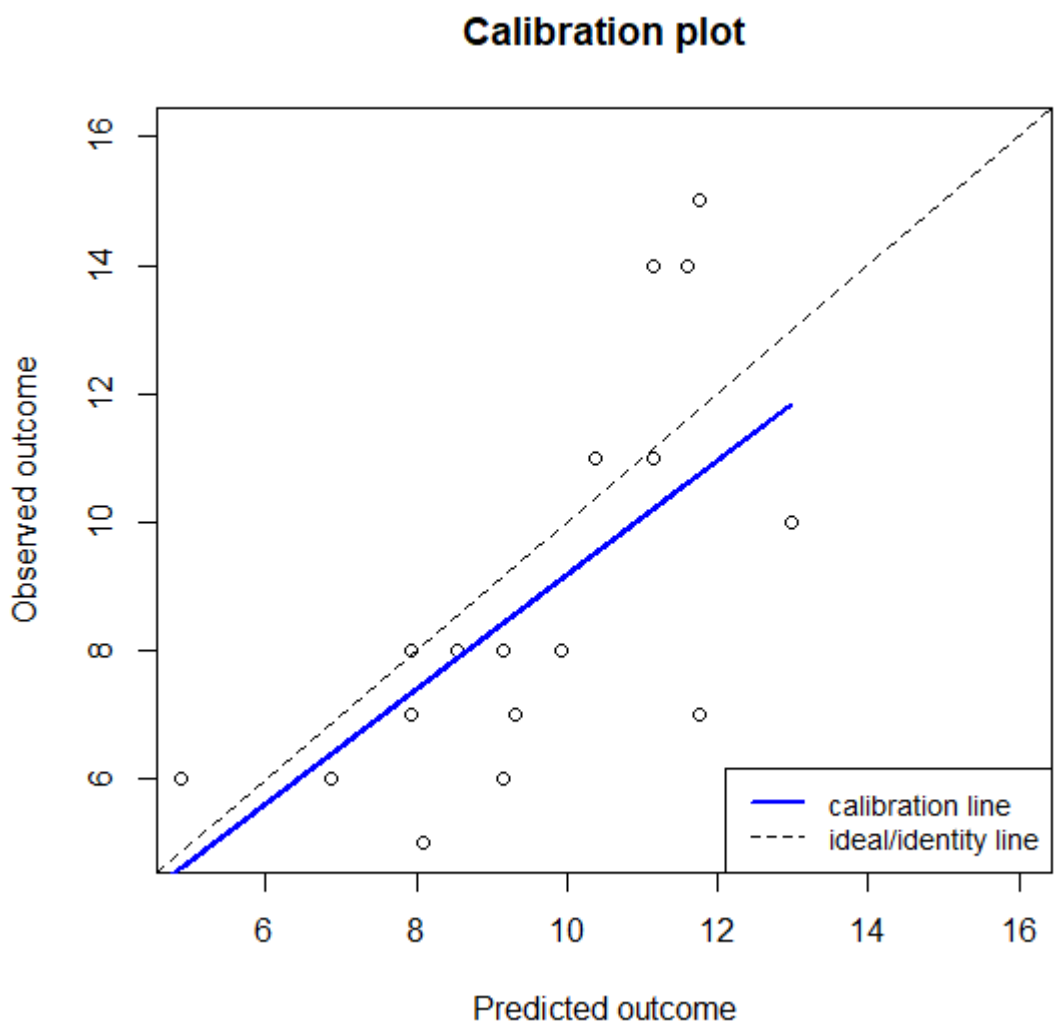
The model was then tested on 20 different patients with depression to validate its performance. The data for the validation sample are in the table below.

1) By completing the table, work out the validated MSE and R-squared, knowing that the population variance of the depression severity scores after treatment for the 20 new patients is 9.75.

2) Assess the validated discriminative ability of the model.

New Individuals $n = 20$	Treatment allocation (t)	Baseline cognitive ability score (y_0)	Cognitive ability score post treat. (y)	Predicted cognitive ability score $(\hat{y})$	Error $y - \hat{y}$	Squared Error $(y - \hat{y})^2$
1	0	12	10			
2	0	15	14			
3	0	12	10			
4	0	14	7			
5	0	20	5			
6	0	17	8			
7	0	22	6			
8	0	14	15			
9	0	15	11			
10	0	18	7			
11	1	19	3			
12	1	13	8			
13	1	19	6			
14	1	10	11			
15	1	12	6			
16	1	12	8			
17	1	14	7			
18	1	8	14			
19	1	20	6			
20	1	14	8			

3) Assess the validated calibration performance of the model on this new data by interpreting the calibration plot. Now, consider both the model validated discrimination performance (answer to question 2) and the validated calibration performance: would this model be good for clinical practice?



In the problem set, you will need to show me that you are able to:

- Calculate and interpret a validated measure of discrimination for a given prediction model for continuous outcomes.
- Interpret calibration graphs to describe the validated calibrative performance of the prediction model for continuous outcomes.

Tutorial 6 – Feedback tutorial

What is the Purpose of Tutorial 6?

- To receive feedback on your final assignment
- To respond to the feedback from your Uni Pathways teacher
- To write targets for improvement on your final assignment

Final assignment feedback from your Uni Pathways Teacher

(Remember to look at the mark scheme to help you understand what you have done well so far, and how you can do even better in your final assignment)

Here are three things that my Uni Pathways Teacher thought I did well in my draft assignment
•   •   •

Here are three things that my Uni Pathways Teacher thinks that I could do to get a higher mark in my final assignment
•

•   •

Tasks from my Uni Pathways Teacher to do during the feedback tutorial to help me improve
•   •   •
My response:

Actions I will take to improve my final assignment after this tutorial…
•   •   •

Hand in date for my final assignment:	

Tutorial 7 – Final tutorial

What is the Purpose of Tutorial 7?

- To receive feedback and a grade on your final assignment.
- To reflect on the programme including what you enjoyed and what was challenging.
- To ask any questions you may have about university.

Final assignment feedback from my Uni Pathways Teacher

Final mark:	University style grade:
Feedback: Here are three things that my Uni Pathways teacher thought I did well in my final assignment • • •	

Here are three things that my Uni Pathways teacher thinks I should remember for when I am doing this kind of study in the future
• • •

University

What questions do you still have about University after taking part in Uni Pathways?
•   •   •

Reflecting on Uni Pathways

What did you most enjoy about Uni Pathways?
•   •   •

What did you find challenging about the programme?	How did you overcome these challenges?
• • •	• • •

Appendix 1 – Referencing correctly

When you get to university, you will need to include references in the assignments that you write, so we would like you to start getting into the habit of referencing in your Brilliant Club assignment. This is really important, because it will help you to avoid plagiarism. Plagiarism is when you take someone else's work or ideas and pass them off as your own. Whether plagiarism is deliberate or accidental, the consequences can be severe. In order to avoid losing marks in your final assignment, or even failing, you must be careful to reference your sources correctly.

What is a reference?

A reference is just a note in your assignment which says if you have referred to or been influenced by another source such as book, website or article. For example, if you use the internet to research a particular subject, and you want to include a specific piece of information from this website, you will need to reference it.

Why should I reference?

Referencing is important in your work for the following reasons:

- It gives credit to the authors of any sources you have referred to or been influenced by.
- It supports the arguments you make in your assignments.
- It demonstrates the variety of sources you have used.
- It helps to prevent you losing marks, or failing, due to plagiarism.

When should you use a reference?

You should use a reference when you:

- Quote directly from another source.
- Summarise or rephrase another piece of work.
- Include a specific statistic or fact from a source.

How do I reference?

There are a number of different ways of referencing, and these often vary depending on what subject you are studying. The most important to thing is to be consistent. This means that you need to stick to the same system throughout your whole assignment. Here is a basic system of referencing that you can use, which consists of the following two parts:

1. **A marker in your assignment**: After you have used a reference in your assignment (you have read something and included it in your work as a quote, or re-written it your own words) you should mark this is in your text with a number, e.g. [1]. The next time you use a reference you should use the next number, e.g. [2].
2. **Bibliography**: This is just a list of the references you have used in your assignment. In the bibliography, you list your references by the numbers you have used, and include as much information as you have about the reference. The list below gives what should be included for different sources.

    a. **Websites** – Author (if possible), title of the web page, website address, [date you accessed it, in square brackets].

    E.g. Dan Snow, 'How did so many soldiers survive the trenches?', http://www.bbc.co.uk/guides/z3kgjxs#zg2dtfr [11 July 2014].

    b. **Books** – Author, date published, title of book (in italics), pages where the information came from.

    E.g. S. Dubner and S. Levitt, (2006) *Freakonomics*, 7-9.

    c. **Articles** – Author, 'title of the article' (with quotation marks), *where the article comes from* (newspaper, journal etc.), date of the article.

    E.g. Maev Kennedy, 'The lights to go out across the UK to mark First World War's centenary', Guardian, 10 July 2014.

Notes

UNI PATHWAYS

UNI PATHWAYS

thebrilliantclub.org