

Standard Setting for the OSCE

DR HELEN HYNES, MEDICAL EDUCATION UNIT, UCC

This presentation, produced by the Medical Education Unit in UCC, explains how to carry out a borderline regression analysis when setting a standard for Objective Structured Clinical Examinations (OSCEs).

We have also produced presentations on setting the standard in MCQ examination and an overview of the theory behind standard setting.



Why not just use 50%?

Using a pre-fixed pass mark was the traditional approach.

Many other departments still use this.

Why can't we?



Traditionally universities just used a set pass mark for examinations. In many faculties this has been 40%. In medical education, the traditional pass mark has been 50%. So why can't we just keep on doing what we did in the past?

Checklist category / subject (a box to be ticked for each item)	Inadequate	Adequate	Good
introduction: Candidates introduces self, establishes rapport - (Adequate= first name, surname and student)	0	2	
Establishes patient's name and age	0	2	
Presenting Complaint:			
Appropriate use of open question	0	2	
History of Presenting Complaint:			
Pain history: location / triggering factors / character / radiation / exacerbating factor / relieving factors / severity / periodicity / duration of headache (G=8 or 9 of the above, A= 6 or 7 of these)	0	3	5
Timeline: Duration of symptoms and recent change in pattern. (G=both clearly defined)	0	2	4
Enquires about aura and altered vision	0	3	
Enquires about nausea / vomitting	0	2	
Asks re other neurological symptoms – visual disturbance, dizziness, paraesthesia, loss of power, tremor, balance. A=4	0	3	
Past Medical History - Asks about Past medical history	0	2	
Drug history - A=asks re meds. G=mentions COCP	0	2	4
Family History - Enquires about family history (G= asks headaches/migraine)	0	2	3
Social History	0	2	
Asks re impact of symptoms on life and work situation	0	3	
Summaries history back to patient	0	2	
Summarises to Examiner (G= comprehensive, systematic summary)	0	3	- 5
QUESTION: What is the most likely diagnosis? - Migraine/tension headache (G= links OCP)	0	2	3
Appropriate verbal and non verbal behaviour, appears empathic	0	1	3
Global rating - Fall / Borderline / Clear Pass / Good / Outstanding			

Before we explore the rationale for standard setting, let's just look at a typical OSCE mark sheet.

This is a checklist style answer sheet from a neurology history OSCE station used in a previous Third Med examination.

Examiners tick off various steps in the history taking process as either Good, Adequate or Inadequate. The examiners are not aware of the exact amount of marks each component receives.

On completion of the station they are also asked to rate the student's overall performance as one of the following – Fail / Borderline / Clear Pass / Good / Outstanding.

In this particular OSCE – the maximum number of marks a student can get is 50. In parallel with that we also have the examiner's rating of the overall standard of the student's performance.



Data from past OSCEs

- Med 3 / GEM 2 OSCE.
- 7 stations in the OSCE 3 history taking, 3 examination 1 procedural skills.
- 198 candidates.

No Standard Setting

20XX	
1H	184
2H	14
P	0
F	0

With Standard Setting

20XX	
1H	34
2H	145
Р	28
F	1

Now let's look at some data from past OSCEs. What difference does standard setting make?

This data is from a previous Med 3 / GEM 2 OSCE. There were 198 candidates, and 7 different stations in the OSCE.

I have anonymsed the year because this is real data.

In the first table we see the breakdown of overall OSCE marks if no standard setting is

• 184 first class honours grades, and 14 second class honours grades, no passes, no fails.

The second table shows the same cohort of students after standard setting was carried out using borderline regression analysis. This time we can see there are:

• 34 first class honours, 145 second class honours, 28 passes and 1 fail.

Very different outcomes, but how do we know that this is fair? Maybe they were a really good class and all deserved their honours grades?

	Position in class	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6	Station 7	Overall grade before standard setting	Overall grade after standard setting
Student 1	First	Good	Good	Clear Pass	Outstanding	Outstanding	Good	Good	1H	1H
Student 2	Median	Good	Clear Pass	Good	Good	Clear Pass	Clear Pass	Clear Pass	1H	2н
Student 3	10 th percentile	Clear Pass	Clear Pass	Clear Pass	Good	Clear Pass	Clear Pass	Good	1H	P
Student 4	Lowest	Clear Pass	Borderline	Clear Pass	Borderline	Fail	Clear Pass	Clear Pass	214	F

We can check that because we also have data of how the examiners rated each student overall on each OSCE station.

So here we see the data for 4 specific students from this exam:

- the student who came top overall;
- the median student;
- · the student at the 10th centile; and
- the student with the lowest score on the exam.

We can see the overall rating that the examiners gave these students on each of the 7 stations in the OSCE and we can see what their end overall grade was both before and after standard setting.

Looking at the top scoring student, their ratings were Good, Good, Clear pass, Outstanding, Outstanding, good and good. The overall grade both before and after standard setting was a 1H and I think we can agree that this is an appropriate grade when you take the examiners' ratings into consideration.

Look now at student 2. Their ratings were good, clear pass, good, clear pass, clear

pass and clear pass. Before the standard setting, their overall grade was a first class honour but taking the examiner ratings into consideration, this is not the profile of a first class honours student. The overall grade after standard setting was a second class honour, which seems more appropriate given the examiner ratings.

Looking at students 3 and 4 we see the same pattern. Student 3 was rated Clear Pass by 5 examiners and good by 2. This came out as a first class honour before standard setting, but I think you'll agree that these ratings do not describe a first class honours performance.

And for student 4, who was the lowest scoring student in the class, their ratings contained 1 fail, 2 borderlines, and 4 passes. This is not the pattern of a second class honours student, which is the grade they would have received before standard setting.



How do we set standards in an OSCE?

Possible methods are:

- Borderline Regression ¹
- Borderline Group ¹
- Modified Angoff 1,2
- Hofstee Method ^{1,3}

Borderline regression and Borderline Group methods are describes as "Examinee Centred"-

- Standard is translated to a passing score based on the review of examinees' performance on tasks.
- Judges determine whether the performance they review depicts someone possessing the knowledge and skills needed to meet the standard.

What methods of standard setting can we use for the OCCE?

The AMEE Guide 85 describes a number of methods for carrying out standards setting on an OSCE.

- Borderline Regression ¹
- Borderline Group ¹
- Modified Angoff ^{1,2}
- Hofstee Method^{1,3}

Borderline regression and Borderline Group methods are described as "Examinee Centred"-

- The standard is translated to a passing score based on the review of examinees' performance on tasks.
- Judges determine whether the performance they review depicts someone possessing the knowledge and skills needed to meet the standard.

Modified Angoff and Hofstee Methods are described in detail in the MEU's presentation on Setting the Standard in the MCQ Exam and are not explored further in this presentation.



The Borderline Candidate

Borderline regression and Borderline group methods both require the examiners to identify the borderline candidate.

This is the candidate who is neither qualified nor unqualified to meet the standard.

These methods depend on examiner training in advance of each exam so that the examiner has a clear picture of the standard expected.

The examiner must be familiar with the material being examined and must also have a clear picture of the standard required of a candidate at this stage of their training.

Borderline regression and Borderline group methods both require the examiners to identify the borderline candidate.

This is the candidate who is neither qualified nor unqualified to meet the standard.

These methods depend on examiner training in advance of each exam so that the examiner has a clear picture of the standard expected.

The examiner must be familiar with the material being examined and must also have a clear picture of the standard required of a candidate at this stage of their training.



Borderline Group Method ¹

Examiners:

- Mark each student's performance on a checklist = Score
- Rate each student's performance as Fail, Borderline, Pass, Good, Outstanding = Rating

Descriptors vary from test to test but the important group is the borderline.

After the exam, the mean score of all candidates rated as borderline is calculated and this becomes the cut score or pass mark.

This method only works with a large cohort of candidates. With smaller cohorts there may be very few candidates ranked as borderline and then the pass mark is not very reliable.

We can overcome this by using Borderline Regression Analysis which uses data from all the candidates not just those who were ranked as borderline.

When carrying out standard setting using the Borderline Group Method:

- Examiners mark each student's performance on a checklist: This is called the Score
- They also rate student performance as Fail, Borderline, Pass, Good, Outstanding: This is called the Rating
- Descriptors (fail, borderline, pass and so on) vary from test to test but the important group is the borderline group.
- After the exam, the mean score of all candidates rated as borderline is calculated and this becomes the cut score or pass mark.

This method only works with a large cohort of candidates. With smaller cohorts there may be very few candidates ranked as borderline and then the pass mark is unreliable.

We can overcome this by using Borderline Regression Analysis which uses data from all the candidates, not just those who were ranked as borderline.



Borderline Regression Analysis

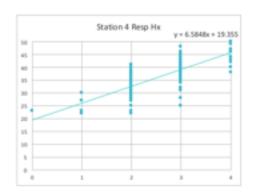
Evaminare

- Mark each student's performance on a checklist = Score
- Rate each student's performance as Fail, Borderline, Pass, Good, Outstanding = Rating

All candidates results are plotted on Excel (equivalent software).

The candidates' score on the checklist is the dependent variable and these are plotted on the Y axis. Their ratings are the independent variable and are plotted on the X axis.

We can then draw a scatter plot of all the data allowing for a regression line to be drawn.



Borderline regression analysis overcomes the problem with the borderline group method by considering data from all of the candidates in the examination and not just those ranked as borderline.

In the same manner as already described for the Borderline Group Method, in Borderline Regression Analysis examiners:

- Mark each student's performance on a checklist: this is the Score
- Rate each student's performance as Fail, Borderline, Pass, Good, Outstanding: this is the rating Rating

All candidates' results are then plotted on Excel (or equivalent software).

The candidates' score on the checklist is the dependent variable and these are plotted on the Y axis. Their examiner ratings are the independent variable and are plotted on the X axis.

We can then draw a scatter plot of all the data allowing for a regression line to be drawn.



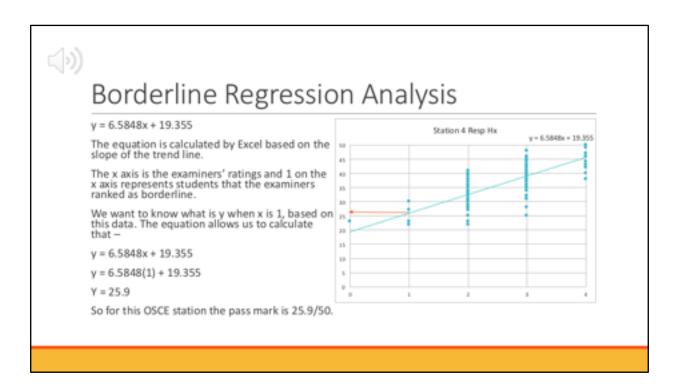
Let's look at that chart in more detail. This is Data from a past OSCE. It represents students' performance on a respiratory history station. 198 students are represented in total.

The Y axis depicts the score obtained out of 50 by each candidate using the checklist. The X axis depicts examiners' rating of each candidate.

- 0=Fail
- 1=Borderline
- 2=Clear Pass
- 3 = Good
- 4= Outstanding

Using Excel, we draw a scatter plot. We then instruct Excel to insert a trend line and an equation.

The equation allows us to calculate the borderline score which becomes the pass mark.



The equation in this case is y = 6.5848x + 19.355. This is calculated by Excel based on the slope of the trend line.

The x axis depicts the examiners' ratings of each student. 1 on the x axis represents students that the examiners ranked as borderline.

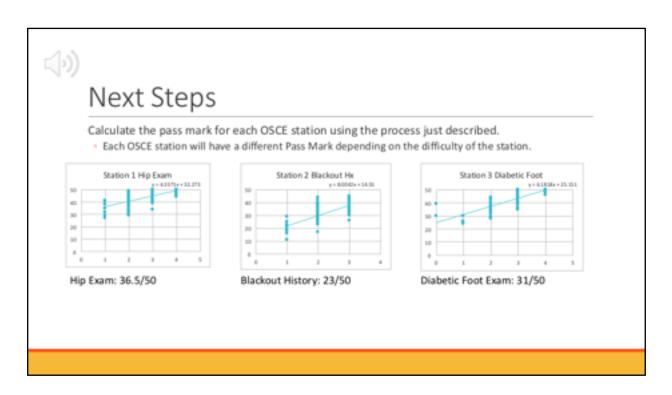
We want to know what is y when x is 1, based on this data. The equation allows us to calculate that –

y = 6.5848x + 19.355

y = 6.5848(1) + 19.355

Y = 25.9

So for **this** OSCE station the pass mark is 25.9/50.



The next step is to calculate the pass mark for each OSCE station using the process just described.

Each OSCE station will have a different Pass Mark depending on the difficulty of the station and the slope of the trendline.

(a) Editing the students' mark to reflect UCC Marks and Standards

As UCC uses 50% as the Pass Mark for examinations in the medical degree programmes, the students' actual marks are amended taking into account the cut score.

This is the formula used:

Amended mark = (actual mark X old pass mark)/ new pass mark.

Old pass mark = 50%

For example, if the student' actual score is 60/100 and the new pass mark (cut-score) is 55%, then the student's amended mark is (30 X 50)/55= 54.5%

In the case of OSCE's we usually mark out of 50.

Therefore the old pass mark = 25/50 (50%).

Amended mark = (actual mark X 25)/new pass mark.

As UCC uses 50% as the Pass Mark for examinations in the medical degree programmes, the students' actual marks are amended taking into account the cut score.

This is the formula used:

Amended mark = (actual mark X old pass mark)/ new pass mark.

For UCC Medical exams the "old pass mark" = 50%

For example, if the student's actual score is 60/100 and the new pass mark (cut-score) is 55%, then the student's amended mark is (60 X 50)/55=54.5%



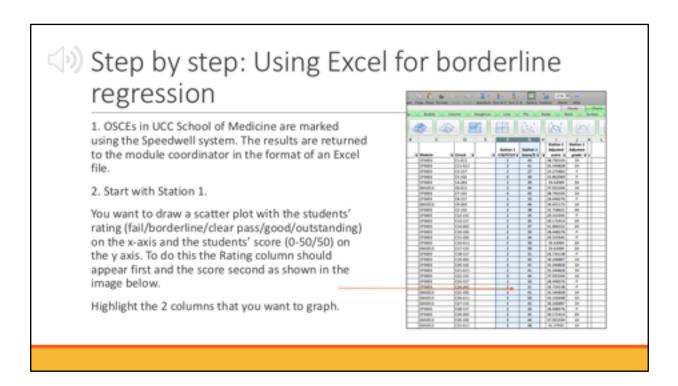
Guidelines for choosing a method

The method must: 4

- · Produce standards consistent with the purpose of the test
- · Rely on informed expert judgement
- Demonstrate due diligence
- · Be easy to explain and implement
- · Be supported by a body of research

Whichever method you choose, the method must:

- Produce standards consistent with the purpose of the test
- · Rely on informed expert judgement
- · Demonstrate due diligence
- Be easy to explain and implement
- Be supported by a body of research

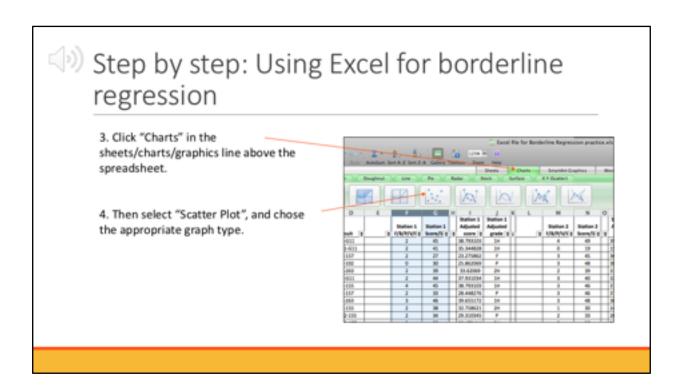


- 1. The remaining slides in this presentation are a step by step guide to using Excel to carry out Borderline Regression analysis for an OSCE.
- 2. OSCEs in UCC School of Medicine are marked using the Speedwell system. The results are returned to the module coordinator in the format of an Excel file.

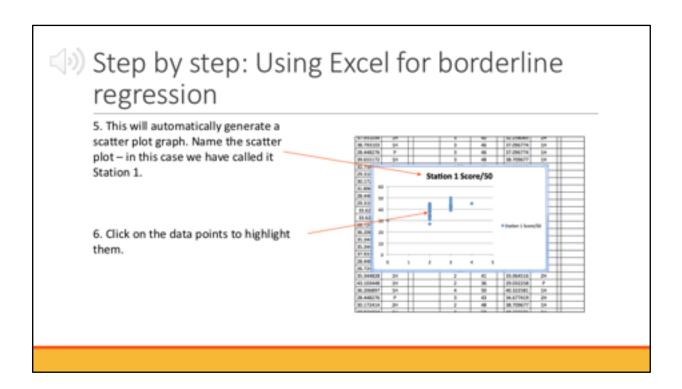
2. Start with Station 1.

You want to draw a scatter plot with the students' rating (fail/borderline/clear pass/good/outstanding) on the x-axis and the students' score (0-50/50) on the y axis. To do this, the Rating column should appear first and the Score second as shown in the image below.

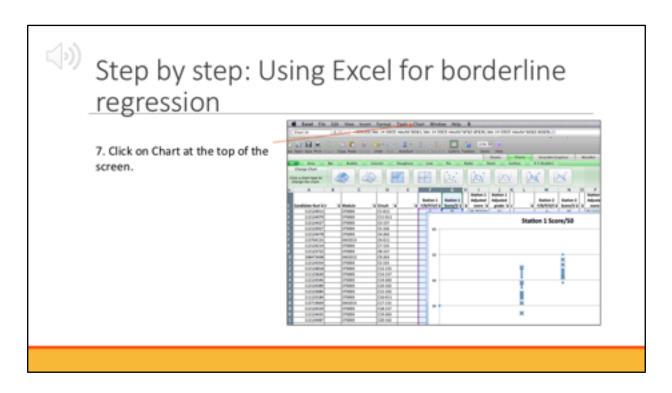
Highlight the 2 columns that you want to graph.



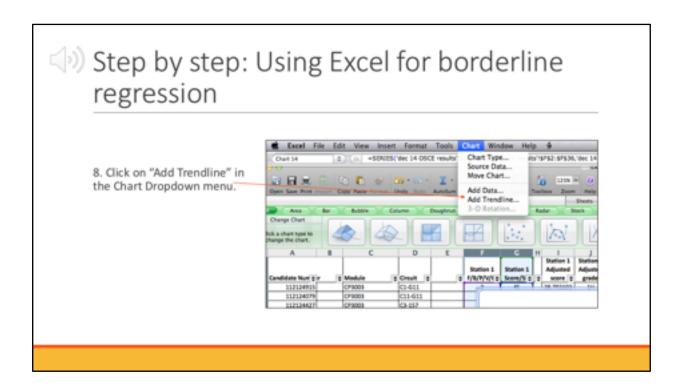
- 3. Click "Charts" in the sheets/charts/graphics line above the spreadsheet.
- 4. Then select "Scatter Plot", and chose the appropriate graph type.



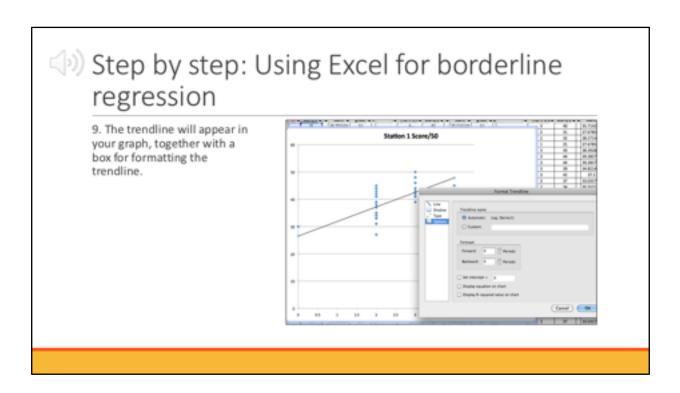
- 5. This will automatically generate a scatter plot graph. Name the scatter plot in this case we have called it Station 1 Score / 50.
- 6. Click on the data points to highlight them.



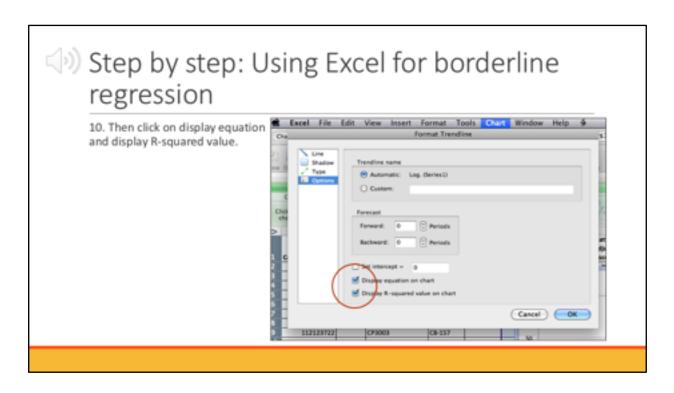
7. Click on Chart at the top of the screen.



8. Click on "Add Trendline" in the Chart Dropdown menu.



9. The trendline will appear in your graph, together with a box for formatting the trendline.



10. Then click on "Display equation" and "Display R-squared value".



Step by step: Using Excel for borderline regression

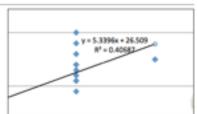
The equation will appear in your graph. Use this to calculate the pass mark.

1 on the x-axis corresponds to "borderline" so use x=1, to calculate y - the pass mark for that station.

In the example we have been using the equation is: y=5.3396x + 26.509.

If we use an x of 1, then $y = (5.3396 \times 1) + 26.509 = 31.83$.

Therefore the pass mark is 31.83 for that particular station.



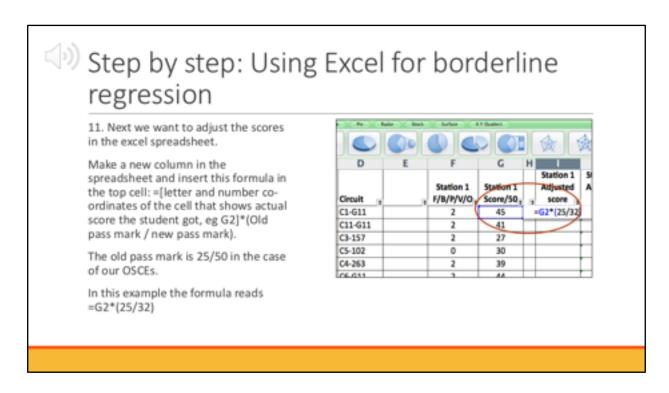
10.a The equation will appear in your graph. Use this to calculate the pass mark.

1 on the x-axis corresponds to "borderline" so use x=1, to calculate y - the pass mark for that station.

In the example we have been using the equation is: y=5.3396x + 26.509.

If we use an x of 1, then y becomes 5.3396+26.509=31.83.

Therefore the pass mark is 31.83 for that particular station.



11. Next we want to adjust the scores in the excel spreadsheet.

Make a new column in the spreadsheet and insert this formula in the top cell: =[letter and number co-ordinates of the cell that shows actual score the student got, eg G2]*(Old pass mark / new pass mark).

The old pass mark is 25/50 in the case of our OSCEs.

In this example the formula reads" =G2*(25/32)"



Step by step: Using Excel for borderline regression

12. Now make this apply to the whole column by clicking on the small rectangle ____ at the bottom right hand corner of the cell you have been working on, and then drag the mouse to the bottom of that column. Now the formula applies to the whole column.

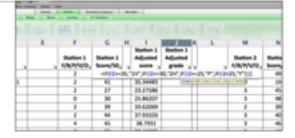
Circuit	Station 1 F/B/P/V/O	Station 1 Score/90	Station 1 Adjusted	Grade 1 Adjusted
C4-263	2	45	38.793003	214
06-611	2	41		
C3-157	2	27		
C26-G11	0	30		
C1-611	2	39		
C7-155	2	-64		
C21-G11	4	45		
C17-155	2	33		
04-263	2	46		
C26-G11	3	38		
C9-263	2	34		

12. Now make this apply to the whole column by clicking on the small rectangle at the bottom right hand corner of the cell you have been working on, and then drag the mouse to the bottom of that column. Now the formula applies to the whole column.



Step by step: Using Excel for borderline regression

13. Next we need to input a grade for the marks in that station. We are now using the adjusted score that the student was given (the score in the new adjusted column we have just made.) In this case the top box of this column is labelled "12"



Based on the adjusted score, use 50% as the pass mark, 60% as 2H and 70% as 1H, then we would just programme in the code =IF(I2>=35,"1H",IF(I2>=30,"2H",IF(I2>=25,"P",IF(I2<25,"F"))) (presuming that I2 is the box where the student's adjusted score on that station is listed).

13. Next we need to input a grade for the marks in that station. We are now using the adjusted score that the student was given (the score in the new adjusted column we have just made.) In this case the top box of this column is labelled "I2"

Based on the adjusted score, we use 50% as the pass mark, 60% as 2H and 70% as 1H. Then we would programme in the code:

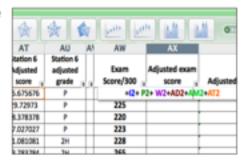
=IF(I2>=35,"1H",IF(I2>=30,"2H",IF(I2>=25,"P",IF(I2<25,"F")))) (presuming that I2 is the box where the student's adjusted score on that station is listed).



Step by step: Using Excel for borderline regression

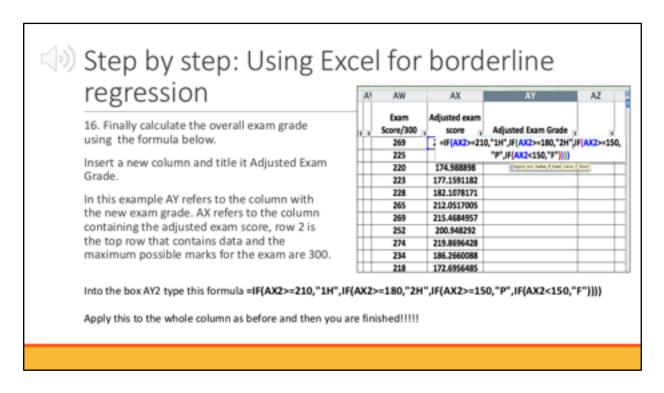
Do this for all the OSCE stations. Then add a new column and title it Adjusted Exam Score. In the example we have here, the new columns with the adjusted scores for the stations are I, P, W, AD, AM and AT and the top row of data is row 2.

15. Finally we come up with a total score for the OSCE by adding the adjusted scores for all 6 stations. Programme this in the excel using this formula- =12+ P2+ W2+AD2+AM2+AT2 (where the co-ordinates I2, P2 etc refer to the actual cells in the which the adjusted scores are in that row.) Apply to the whole column by clicking on the small rectangle at the bottom right hand corner of the cell as before and dragging the mouse to the bottom of that column.



14. Do this for all the OSCE stations. Then add a new column and title it Adjusted Exam Score. In the example we have here, the new columns with the adjusted scores for the stations are I, P, W, AD, AM and AT and the top row of data is row 2.

15. Finally we come up with a total score for the OSCE by adding the adjusted scores for all stations. Programme this in the spreadsheet using this formula- =12+ P2+ W2+AD2+AM2+AT2 (where the co-ordinates I2, P2 etc refer to the actual cells in the which the adjusted scores are in that row.) Apply to the whole column by clicking on the small rectangle at the bottom right hand corner of the cell as before and dragging the mouse to the bottom of that column.



16. Finally calculate the overall exam grade using the formula below. Insert a new column and title it Adjusted Exam Grade.

In this example:

- AY refers to the column with the adjusted exam grade;
- AX refers to the column containing the adjusted exam score;
- Row 2 is the top row that contains data; and
- The maximum possible marks for the exam are 300.

Using UCC's marks and standards, >=70% (210/300) is a 1H, >=60% (180/300) is a 2H, >=50% (150/300) is a Pass and <50% (150/300) is a fail.

Into the box AY2 type this formula =IF(AX2>=210,"1H",IF(AX2>=180,"2H",IF(AX2>=150,"P",IF(AX2<150,"F"))))

Apply this to the whole column as before and then you are finished!



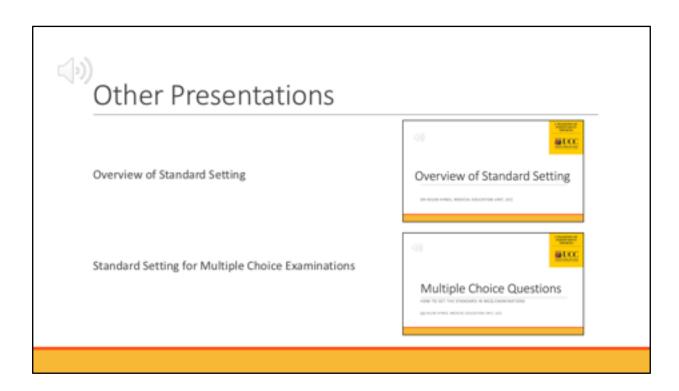
If you have used Excel in the past at a high level, then doing borderline regression analysis using Excel is quite easy.

If you are less familiar with Excel, then the first time you carry out a borderline regression analysis it is challenging. I would advise you to have a version of this presentation open so that you can refer back to the steps and formulas as needed.

You should also save a practice version of your results so that you can have a trial run.

With experience it is possible to standard set a 6 or 7 station OSCE is less than an half an hour.

Contact me on H.hynes@ucc.ie for help if needed!



The Medical Education Unit has also produced presentations on an Overview of Standard Setting and Standard Setting for MCQ examinations.

References

- McKinley DW, Norcini JJ. How to set standards on performance-based examinations: AMEE Guide No. 85. Medical Teacher 2014; 36(2):97-110
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington DC: American Council on Education
- Hofstee WKB. The case for compromise in educational selection and grading. In: Anderson SB, Helminck JS, editors. On Educational Testing. San Francisco: Jossey-Bass; 1983. pp. 109–27.
- Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V, Roberts T. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. Medical Teacher 2010; 33:3