My favourite awesome statistics lesson
Exploring statistical measures by estimating the ages of famous people

What are the key statistical concepts or ideas that underpin this lesson?

- How can we use data in more than one way to create meaning? *Depending on what we want to find out, we need to think about what and how the features of the data will help us.*

- How can we measure and compare variation? *Students get the relationship between "closeness" or "consistency" and the size of the measure used.*

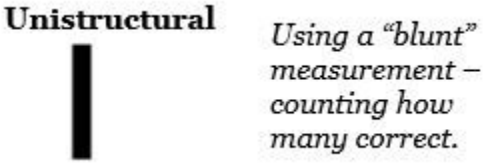
- Why do we need to think beyond the data in front of us? *We could consider the influence of other variables (e.g. the age of the estimator) in creating a measure. Students need to be aware of multivariate relationships.*

What are the key learning activities of the lesson?

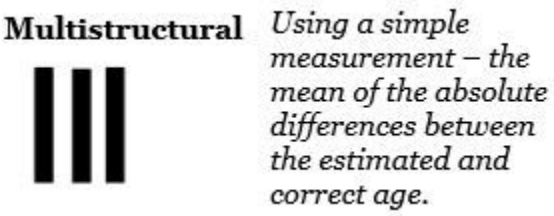
1. Explain to the class that they are going to do an activity requiring them to estimate the ages of different people, and that we want to know who is the best at estimating ages. Then show the students the photos of 10 "famous" people and ask them to write down each person's name and an estimate for their age. Go through the photos again and tell the students the correct ages of the people, getting them to write down the correct ages beside their estimates. [Hint: I set up a spreadsheet with each person's birthday so I always had their current age no matter what year I did the activity.]

Person	Birthday	Current age
Adele	5/05/1988	27
John Key	9/08/1961	53
Sonny Bill Williams	3/08/1985	30
Beyonce	4/09/1981	33
Justin Bieber	1/03/1994	21
Miley Cyrus	23/11/1992	22
Barack Obama	4/08/1961	53
Matt Damon	8/10/1970	44
Jennifer Lopez	24/07/1969	46
Usher	14/10/1978	36

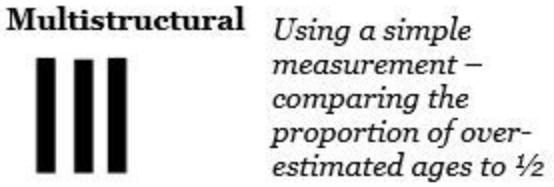
2. For the first measure of "estimating skill", use a count of how many ages each student got correct i.e. their individual score out of 10. This is a really blunt "hit or miss" measure. Depending on how much time you have, you could collect the scores and plot them using a suitable display. You should at least get each student to state their score so the class can get a feeling for the measure. Decide on the "winner" (or winners) using this measure, checking students understand that for this measure, the higher the value, the better in terms of "estimating skill", and that there are only 11 possible values for the measure. Ask the students if they think this is a good way to measure estimating skill - I bet they'll tell you no, and suggest another measure that takes into account how close your estimate was to the actual age of each person.



3. For the second measure of "estimating skill", get students to calculate "how many years off" they were from the actual age for each person. Use the absolute difference and show the students examples of how to calculate this e.g. real age 85 estimated age 88 → 3 years off, real age 24 estimated age 20 → 4 years off. Get two students to share their absolute differences and use these differences to construct parallel dot plots with the same axis/scale. Discuss with the class how to get an overall measure for "estimating skill" using the absolute differences - guide them towards using the mean of these differences (do this with visual representations - not just by calculation). Reinforce this with language e.g. Ernie was off with his estimates by around 3.2 years while Bert was off with his estimates by around 1.1 years. Decide on the "winner" using this measure after comparing results within the class. You should discuss positive and negative aspects to this measure as the students identify them e.g. what about a student who gets one estimate way wrong, and so their mean "years" off is "unfairly" high.

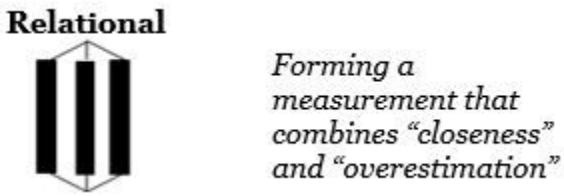


4. For the third measure of "estimating skill", get the students to think about whether it is better to over or underestimate someone's age and whether this matters in terms of the skill. You could inject a little bit of humour at this stage e.g. by getting the class to estimate your age to make your point (do your students overestimate your age too?). Ask the class how we can use the data to determine if someone is typically an over or under estimator. They'll need to go back and add positive and negative signs to the "number of years off" values. Discuss what you will do with estimates that were correct (kind of like a sign test). Guide students towards ignoring the estimates that were correct (the zero differences), and calculating the proportion of incorrect estimates that were overestimates e.g. ⅝. This proportion could be interpreted in comparison to ½ to decide if someone is generally overestimating or not. Evaluate this measure - what does it tell you, what doesn't it tell you? What if a student was correct for most of the ages, is it fair to judge overestimation based on the few they got wrong?

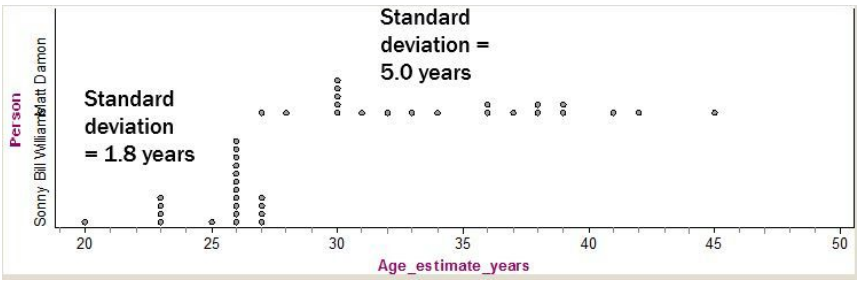


5. For the fourth measure, each student needs to use their own data to create a scatterplot with the actual ages on the vertical axis and the estimated ages on the horizontal axis. Ask students to add the y = x line to represent estimates that were correct - discuss why it makes sense to add this line to the plot. Ask students to draw a vertical line from each estimate down or up to the y = x line. Discuss with students what each vertical line represents. Then ask your students to create a measure that summarises two things: how close they were overall to the correct age, and whether they tend to overestimate or underestimate. For this measure, they can use what they have already done for the second and third measures. What we want is to see if they can combine these together to give one new measure e.g. if they are mostly underestimating and their mean absolute difference is 2.3, the

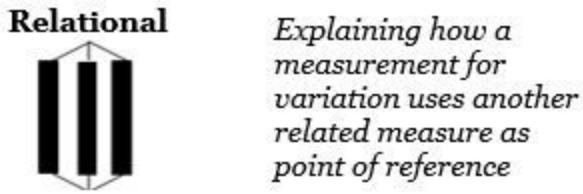
measure would be -2.3



6. For the fifth measure, the focus is on building understanding of the standard deviation which requires moving away from using the true age as a point of reference. There should be two people within the set of 10 for which the class as a whole were more and less consistent with their estimates. Choose two such people and create two parallel dot plots with the class estimates. The question now is - which of these two people was the class more consistent at estimating their age? We are not considering the true age of the person or accuracy. Visually using the dot plots, students should be able to explain which person got more consistent estimates from the class. Ask the students how we could put a number on this amount of consistency - hopefully your students will suggest using something to do with the mean number of years off, similar to an earlier measure. The key difference for this measure is that we need a point of reference - the mean - so we can compare variation across the two different people. I don't make students calculate the standard deviation using a formula, but provide them with the value for the standard deviation and discuss whether it makes sense as a number to match what they can see in the plots.

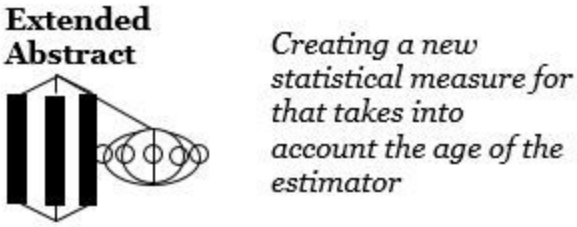


Ask the students to describe (including drawing on the plots) how they think the value of the standard deviation was obtained. Guide them, if necessary, towards using the mean (estimated visually from the plot) as the point of reference for measuring how different the estimates are from each other. This activity is just an introduction to the standard deviation measure, and there are many other great activities you can use to teach standard deviation visually and conceptually. Discuss positive and negative aspects of this measure e.g. that one student in every class that estimates a person's age as a ridiculous number :-)



7. The fifth measure should lead you to the last activity - which is to consider the influence of the age of the person making the estimates. Ask the class - could it be that you are better at estimating ages of people closer in age to you? What could we do to take this into account if we wanted to compare the age estimation skills of two people with quite different ages? [You could also open up discussion about other factors that would influence age estimation, and how some jobs it is important to have very good age estimation skills]. To finish off this lesson, challenge the class to come up with a way of comparing two differently aged people who had completed the same age guessing activity as them e.g. one estimator aged 16 and one estimator aged 32. Provide the class with two different sets of data - the age estimates for each of these estimators (you'll need to create these sets of data in

advance of the lesson). Ask your students to work in groups to create a measure to compare the two estimators and get the groups to report back to the class who is “the best at estimating ages” by explaining the statistical measure they have created.



It should be noted that through this lesson we should be cautious about making inferences about overall age estimation skill of an individual. You could use words like “For these 10 people….” or “For this activity….” to limit the scope. The focus of the learning is on the creation of statistical measures, not the use of these measures to make inferences.

What are some ideas for extending this lesson?

I would use how-old.net or something similar.

For my photo, the gender is correct, but the age is the wrong decade :-) I would run the same photos used for the activity through how-old.net and compare the estimates from the software to either the student/class estimates and the actual age of each person. A question could be "Is the software or a human better at estimating someone's age?"

Related links or resources

Upload photos of people and see if the technology can determine the correct gender and age how-old.net

Evaluating and creating a statistical measure was assessed in an AP Statistics exam in 2009. Have a read of the commentary and be inspired to try out this kind of activity with your students! 2009 AP Statistics exam Q6: commentary document