Inference, *z*-scores, and determining when "average" results are appropriate for an individual

One very important factor to consider when we are performing and interpreting the results of regressions (or calculating other point estimates—e.g., calculating means or proportions), is that these statistical methods are extremely limited in terms of what they can tell us about ***individual cases***. These numbers are useful for largescale population decisions (e.g., telling us what policies will work best ***on average*** if we just use them across the whole population), but they are NOT necessarily useful for saying much of anything about what is best for an ***individual***, who may vary quite a lot from the most common individual used in the research studies. For "atypical" individuals, it is inappropriate to apply inferences from general population means.

One way that we can tell how "typical" an individual is (and therefore how relevant the results of a particular study may be for them) is to use *z*-scores as a measure.

Let's look at some examples based on height:

Ales was assigned a female sex at birth, and is unusually tall compared to others who were assigned female at birth in the US population. Let's consider how this might impact some different medical decisions that might be made about them.

Alex is 190cm tall (or just under 6 ft 3 in).

The mean height of adult women in the US is 162.1cm, and the standard deviation is 7.11cm.

Calculating a *z*-score for this height tells us that Alex is 3.92 standard deviations above the mean in terms of height, which is pretty unusual. Anyone used to using *z*-scores immediately knows how unusual this is, but if you are not used to thinking in terms of the number of standard deviations away from the mean that a particular number is, you can start by looking up this *z*-score (again, a *z*-score is just the number of standard deviations above or below the mean an observation is) on a chart of the standard normal distribution to get a sense of what percentage of observations will be above or below that particular score. So, how unusual exactly is it to be 3.92 standard deviations above the mean?

Looking at a standard normal distribution (there are many tables and calculators online where you can do this—here is one example: https://www.inchcalculator.com/z-table/), we see that 99.996% of all adults in the US with an assigned sex of female are shorter than Alex (and so only 0.004% of adults in the US with the same assigned sex are taller).

That number can also be hard to get our heads around, so we can also convert this to a set number of people: for example, 1 out of how many people with this assigned sex in the US are taller? Some simple math (0.004/100 is equal to 1/n, and we just solve for n) tells us that roughly 1 out of 25,000 adults in the US with a female assigned sex are going to be taller than Alex.

What consequences might this have for Alex's medical care?

For some things, likely no impact at all. For example, there are many measures that are not expected to vary significantly by height (if we restrict ourselves to just measuring adults), such as someone's blood pressure, their vitamin D blood levels, etc.

However, many other medical tests may vary significantly by height. Let's look at a few:

Aortic root diameter

The aortic root diameter is the measure of how wide someone's aorta is. This might be measured via ultrasound if someone is at risk for certain diseases. This can be really important, because if someone's aorta is dilated (i.e., wider than expected), this can mean that they are at higher risk for aortic dissection, or a medical emergency in which the aorta bursts--when this happens there is an extremely high chance that someone will die. So, doctors want to know in advance if this is a risk so that they can intervene. On the other hand, labeling someone has having a dilated aorta when they are not actually at higher risk for aortic dissection also has its costs—it can turn someone into a "cardiac cripple", or someone who is told to severely restrict otherwise healthy activities or important life activities (e.g., severely limit many kinds of weight-bearing exercise; stop lifting their kids). In addition, it can expose the person with this label to unnecessary medical interventions (e.g., blood pressure lowering medications, which have their own side effects) and to very elevated levels of stress related to worry about their diagnosis, which has its own negative consequences for health.

A doctor has sent Alex to get an ultrasound to measure the aortic root (at the sinuses of the Valsalva, measuring from the leading edge to the leading edge, if you are interested in the medical terminology :)). Alex's measurement was 35.0 mm. The ultrasound report listed the reference range for those with an assigned female sex as 22.6-34.4mm, and thus, the doctor flagged Alex as having a dilated aorta and began a string of interventions.

This reference range, as is typical in much of medicine, is based on measurements taken using a sample which is supposed to be representative of the population, in which typical "healthy" adults have some value measured, those values are recorded and used to generate a dataset, a normal distribution is assumed, and then the reference range (or "normal" range) is just the range of values that occur in the middle 95% of the population (or from -1.96 to 1.96 standard deviations around the mean). This reference range comes from a population study in which for adults with an assigned female sex, the mean was 28.5 and the standard deviation was 3.0.

However, we know that the aortic root is bigger in taller people (also in older people and people with a larger BSA). So, is this value actually abnormal for Alex, given their size?

One thing to consider is that unless research studies have 25,000 people in them, someone who is 3.92 standard deviations above the mean is unlikely to have even one person as tall or taller than them in the study. Almost all of the people in the study are much shorter than them. So considering a population mean may not be a very appropriate thing to do when treating an individual whom we know is very far from that population in terms of certain key parameters that are correlated with the mean of interest.

In the case of height and aortic root diameter, it turns out that there actually is research that could help us to answer this question:

Other research has generated linear regression equations that estimate the mean aortic diameter by including: height or BSA, assigned sex, and age, and these produce different reference ranges based on these combinations of variables. Plugging Alex's height, age, and assigned sex into these equations gives us the following information about Alex's aortic root measurement:

Based on a 2012 study, Alex's aortic root measurement is 0.77 standard deviations above the mean for adults of their same height, age, and assigned sex.

Based on a 2017 study, Alex's aortic root measurement is 0.82 standard deviations above the mean for adults of their same height and assigned sex (age was not included as a covariate in this study).

a. *Is Alex's aortic root dilated or not? Which criteria should we use for Alex in determining whether their aorta's diameter is unusually large? Why?*
b. *When doctors use "normal" ranges for aortic root diameter that have been derived from the general population (without adjusting these using regression that includes other factors such as height to determine what is "normal" for people of different heights), what are the potential risks for Alex?*

Alex has a friend, Pat, who instead of being very tall compared to others in the US, is quite short. Pat is 134.2 cm tall, or 3.92 standard deviations *below* the mean height for those assigned a female sex at birth. Pat's aortic diameter is 34mm. Recall that the reference range for the overall population is 22.6-34.4mm. Thus, according to this range, Pat's aorta is not dilated, and they are not at elevated risk for aortic dissection, nor is it suggested that they receive any additional monitoring or interventions to reduce their risk.

Based on a 2012 study, Pat's aortic root measurement is 2.25 standard deviations above the mean for adults of their same height, age, and assigned sex.

c. *Is Pat's aortic root dilated or not? Which criteria should we use for Pat in determining whether their aorta's diameter is unusually large? Why?*
d. *When doctors use "normal" ranges for aortic root diameter that have been derived from the general population (without adjusting these using regression that includes other factors such as height to determine what is "normal" for people of different heights), what are the potential risks for Pat?*

High Birthweight

People are pregnant with babies who have a higher birthweight are often flagged as "high risk" for several reasons: 1) studies with the general pregnant population have shown that on average as birthweight goes up, the risk of birth complications goes up; and 2) higher birthweight babies are on average more likely to have been born by mothers with gestational diabetes, which comes with its own health risks.

Many studies have shown that the risk of adverse birth outcomes from a larger baby goes up as the height of the mother goes down.

Babies are often flagged as high birthweight if they weigh more than 4000g, and extremely high birthweight if they weigh more than 4500g. These numbers are not adjusted for the height of the mother—the same numbers are applied to flag all pregnancies, regardless of maternal height.

One study in the general US population suggested that the mean birthweight is 3,510 grams with a standard deviation of 385 grams.

Another study suggested that a babies' birthweight is directly proportional to the height of the mother, with each additional inch in maternal height increasing the baby's birthweight by about 100g.

Recall that the mean height of adult women in the US is 162.1cm, and the standard deviation is 7.11cm.

e. *How many standard deviations above the mean would a baby have to be in order to have a high birthweight of 4000g?*
f. *How many standard deviations above the mean would a baby have to be in order to have an extremely high birthweight of 4500g?*
g. *Discuss the relationship between the standard deviation of a baby's birthweight and the standard deviation of a mother's height (just based on the simplified information provided here—in real life, the relationship is slightly more complex).*
h. *Based on the above information from various studies, what birthweight would we expect for a baby for Alex (recall that their height is 190cm tall (or just under 6 ft 3 in)?*
i. *Based on the above information from various studies, what birthweight would we expect for a baby for Pat (recall that their height is 134.2 cm tall (or just under 4 ft 5 in)?*
j. *Will Alex's baby be flagged as having a high or extremely high birthweight? Is this reasonable? Why or why not?*
k. *Will Pat's baby be flagged as having a high or extremely high birthweight? Is this reasonable? Why or why not?*
l. *Suppose that Alex's baby were born at a birthweight of 3500g. This is close to the mean birthweight in the general US population. What might a doctor who considers only the mean population values think about this birthweight? Knowing what we know about the relationship between birthweight and maternal height, what are some concerns we might have about this birthweight in this case?*
m. *Use your answers to questions e through l to describe i) how z-scores allow us to compare different measures in a way that is useful or important; ii) why population means become less and less appropriate for making inferences about subjects that are farther and father away from the mean on key variables.*
n. *Think about the examples we have explored so far, and now consider how this might be relevant to education research. Generate at least two examples of cases in which a conclusion arrived at by a study of a larger sample or population may be inappropriate to apply to students or learners who fall far away from the mean on some key variables of interest.*