**BMI has an expectations management problem**

Body Mass Index (BMI) is an attempt to give a quick back-of-envelope answer to the question “if someone weighs *W* kg, is that a lot or not very much?” Clearly the answer to that question has to take into account at a minimum the person’s height; in general, whatever may constitute a healthy weight, it has to be higher for tall people. So BMI – the weight (in kilograms) divided by the height (in metres) squared – gives a scaled version of weight that aims to be rougly comparable for people of different heights. It is actually extremely successful in this regard; certainly it tells you a *lot* more about someone than either the weight or height individually.

However, BMI is one of those single dimensional metrics that gets widespread criticism on account of being a single dimensional metric. Criticisms of these types of measures typically come in three varieties:

1. reality is much more complex than any single measure M could represent;
2. …but if you tweaked it (by method Y) it would somehow be much better;
3. …anyway, I think Z is really important. M is a really bad measure of Z.

Gross Domestic Product wins the award for attracting the biggest collection of criticisms in this pattern, but BMI would be high up in the list of also-rans.

The first of the criticisms listed above is always correct, but uninteresting given we need to somehow model the world into a simpler form to grasp it. A one-to-one scale map is no use for navigation.

The third criticism is often (not always) a red herring. In particular, it can introduce a straw man fallacy, implicitly or explicitly accusing the metric of not being something it never claimed to be. This can divert discussion into technical areas when the real disagreement is about something more substantive.

The second criticism (“if you tweaked it, it would be better”) can be interesting and fruitful, but in my experience is often standing as a proxy for one of the other two criticisms, which can cause more confusion and talking-past-one-another.

BMI attracts criticisms in all three categories. Clearly, a person’s state of health is much more complex than can be represented by a single number and “BMI doesn’t tell the whole story” (criticism one with a smattering of three). Equally clearly, sometimes – perhaps always – whether someone is obese, normal weight or underweight is not the most important thing to know about them (criticism three). BMI is an inaccurate measure of body fat content (well, yes, of course) and a ‘terrible measure of health’ (still the straw man of criticism three).

On average and controlling for other factors, obesity and high BMI is associated with poor health outcomes, but in some unusual circumstances obesity seems to be a protective factor (criticism three – again, not clear why this is seen as a critique of BMI as a *measure* rather than an interesting fact about how that measure relates to outcomes). Some very well-muscled and clearly healthy individuals have high BMIs that would put them in the ‘overweight’ category by WHO standards (I think this is a variant of criticism one with a dusting of three). Or BMI is just calculated the wrong way – it should have height to the factor of 2.5, not 2 (criticism two).

I want to pass on most of this debate, not because it’s not important and interesting (it is clearly both) but because I don’t have enough time and space just now. Relatedly, I have wondered (as I’m sure many many people have before me), why divide weight by height to the power of 2? Why 2? Why not 3, given that if people of different heights had the same proportions we would expect weight to be proportionate to height cubed? It turns out that this is because people’s horizontal dimensions don’t increase as much as by their height; but by how much less?

The body-mass index that you (and the National Health Service) count on to assess obesity is a bizarre measure. We live in a three-dimensional world, yet the BMI is defined as weight divided by height squared. It was invented in the 1840s, before calculators, when a formula had to be very simple to be usable. As a consequence of this ill-founded definition, millions of short people think they are thinner than they are, and millions of tall people think they are fatter.

New formula: BMI = 1.3 \* weight(kg) / height(m) ^ 2.5

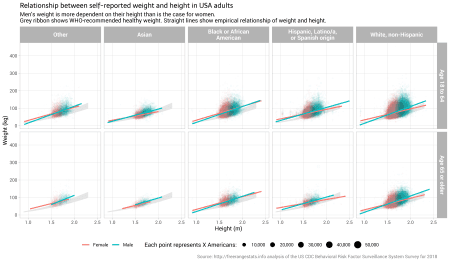
The scaling (which is designed to make the new BMI identical to the old for someone who is 1.69 metres tall) seems an unnecessary complication. After all, BMI is a completely artificial measure and we can put the boundaries for underweight, healthy, overweight etc wherever we wish; if we want a “new BMI” measure there is no reason for it to appear similar in magnitude to the old one (and in fact some good arguments against).

BMI, nearly always uncritically, for example in the subheading “BMI exaggerates thinness in short people and fatness in tall people” the result comes out closer to height ^ 2.5 than height ^ 2”. But despite the widespread acceptance of Trefethen’s critique everywhere from the popular press to Reddit sites for tall people, I can’t see any actual evidence for the use of 2.5.

**How do height and weight relate in a real population?**

Very well. Let us look at some data. The US Centers for Disease Control and Prevention (CDC) Behavioral Risk Factor Surveillance System (BRFSS) is an extraordinary on-going survey of around 400,000 adult interviews each year. It claims to be the largest continuously conducted health survey system in the world, and I imagine it is one of the largest continuous surveys of any purpose. It includes questions on (self-reported) height and weight as well as standard demographics.

Let’s start by looking at the relationship between weight and height, without any calculated summary of the two:



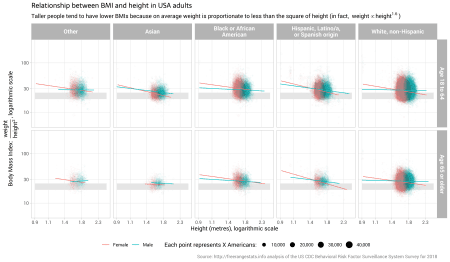
There’s a few interesting things here. It is clear that the relationship between weight and height varies with sex, ethnicity and age. In particular, the relationship between height and weight is stronger for men (as in a steeper line) in all races and age groups. Also, the average values of these things change. Women are shorter and weigh less than men; people with an Asian ethnic background than the general population; etc.

I’ve applied as a layer of grey the healthy weights for a person of given height implied by the World Health Organization’s BMI guidelines. This makes it clear that most US Americans weigh more than is recommended for their height, which of course is well known.

**What does this mean for BMI?**

My next step is to convert those weight and height measures into BMI. If BMI is a “fair” measure of how far an individual is from the average weight at their height, we would expect to see no relationship between height and BMI. That is, BMI should capture all the structural information about the norm for weight available in height. The essence of Trefethen’s argument is that this is not the case, and that traditional BMI is systematically higher for taller people because the impact of height is shrunk by choosing a power of only 2 rather than 2.5.

Let us see:



In fact, we see that there is a very marked *negative* relationship between height and BMI, particularly for women, but also present for men. This is the precise opposite of Trefethen’s claim, which is that tall people get exaggeratedly high BMIs in the current standard calculation. Based on sheer empiricism, we could make the case that BMI *understates* the scaled weight of tall people.

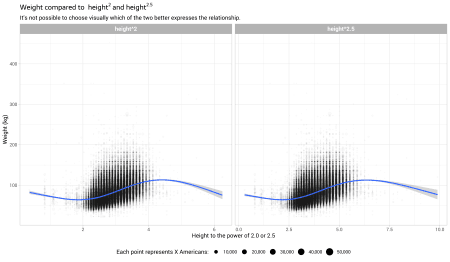
(There’s an important nuance here, which is that I’m only looking at a ‘positive’ relationship. We can’t deduce a normative statement from this. But I think that’s a question for another day. I’m also disregarding for now measurement error, a survey in one time and place, and other restrictions on making general inferences from this sample. After all, this is only a blog.)

Here’s an interesting point for those interested in gender relations and the development of science and measurement. The slopes in the chart above for men are nearly horizontal, certainly more so than for women; and most horizontal for people described as ‘white, non-Hispanic’. One interpretation for this would be that:

“the current BMI calculation is a fair representation of the relationship between height and weight for white, non-Hispanic men, but underplays the weight of taller people (and exaggerates the weight of shorter) of other racial backgrounds and of women in general.”

**What power for height is weight proportional to?**

I provided a second hand quote of Trevethen earlier in which he argued that “if you plot typical weights of people against their heights, the result comes out closer to height ^ 2.5 than height ^ 2”. Is this the case? Actually, there’s nothing visual to tell at all, in my opinion:



We need to calculate. Luckily, if weight is expected to be proportionate to height to the power of something, then the logarithms of the two will have a linear relationship and the height coefficient of a regression of log(weight) on log(height) will tell us what the power is.

An interesting statistical issue arises here – should we do this regression for the full population of Americans, or should we let that coefficient vary by race, sex and age? From a modelling perspective, it *definitely* is a different value for different race, sex and age.

I decided fairly arbitrarily that I wanted a single coefficient to put height to the power to, but that I wanted that coefficient calculated after controlling for other obviously important characteristics. Fitting a regression of the type

log(weight) ~ log(height) + sex + race + age

gets me a slope for log(height) of 1.55. In other words, after controlling for sex, race and age, American’s weight varies in proportion to their height to the power of 1.55.

Alternatively, if we treat the population as one and fit the regression:

log(weight) ~ log(height)

We get 1.87.

I imagine I would get different values again with different populations or times.

So what do I conclude? Frankly, that “2” is good enough. I think the Trefethen critique of BMI is at the higher level an unnecessary distraction, and at the level of detail he almost certainly gets the direction of misrepresentation wrong (it is short people who are made to look fatter than they are by the standard calculation, not tall people).

Overall, having poked and prodded this measure a bit, I think that it meets many of the criteria of a useful measure. It’s easy to calculate, easy to explain, lends itself to simple benchmarking, doesn’t need specialist skills for data collection, and is clearly closely related to major outcomes of interest.

I doubt BMI is the best single number to represent someone’s health, but I don’t think this has often been claimed for it. It probably isn’t even the best single number regarding the cluster of issues relating to weight, shape, size and body fat content, but it’s not a bad starting point. It might well be the most cost-effective such measure in many situations. So I’m happy to say that as a metric, BMI isn’t a bad one so long as you don’t expect it to do more than is reasonable.

**Population versus individuals**

*I want to make a final point about one point in the BMI debate. We often see BMI critiqued (or defended) from the angle that “BMI is meant to be a population measure about averages, not something that is meaningful for individuals”. As sometimes expressed, this seems to me completely incoherent. If the average BMI tells us anything about a population, it is because it is meaningful for individuals; the population effect is the sum of a bunch of small individual effects. I think this argument seems to be a misguided extrapolation from the wise caution that one shouldn’t try to diagnose the health of an individual from just their BMI. That seems absolutely fair. But this doesn’t mean BMI tells us nothing about the individual, just that the BMI alone is not determinative of outcomes.*

**Code**

Here’s all the R code for the above in a single chunk. It’s not terribly exciting, but there are some interesting visualisation challenges associated with:

* the lumping of reported height and weight (I jitter points to make them more distinguishable)
* taking care to convert measurements which can be either in metric or medieval measurements to a common unit
* unequal survey weights
* large sample sizes (with nearly half a million points, how to represent them on the screen? I use a high degree of transparency to get around this).

When it comes to the actual modelling that treats the survey weights and the sampling structure appropriately, Thomas Lumley’s survey package rules supreme in terms of easy of use.

library(tidyverse)

library(scales)

library(patchwork)

library(MASS)

library(rio)

library(survey)

library(Cairo)

#--------------CDC annual Behavioral Risk Factor Surveillance System (BRFSS) survey-

# 69.5 MB download, so only uncomment this and run it once:

# download.file("<https://www.cdc.gov/brfss/annual_data/2018/files/LLCP2018XPT.zip>",

# destfile = "LLCP2018XPT.zip", mode = "wb")

#

# unzip("LLCP2018XPT.zip")

llcp <- import("LLCP2018.XPT")

dim(llcp) # 437,436 respondents and 275 questions!

# Survey weights. These are very unbalanced so take care:

summary(llcp$"\_LLCPWT")

# Codebook:

# <https://www.cdc.gov/brfss/annual_data/2018/pdf/codebook18_llcp-v2-508.pdf>

# Function to clean and tidy up variables from the original

code\_vars <- function(d){

d %>%

# height in metres:

mutate(height = case\_when(

HEIGHT3 >= 9000 & HEIGHT3 <= 9998 ~ as.numeric(str\_sub(HEIGHT3, 2, 4)) / 100,

HEIGHT3 < 800 ~ (as.numeric(str\_sub(HEIGHT3, 1, 1)) \* 12 +

as.numeric(str\_sub(HEIGHT3, 2, 3))) \* 2.54 / 100,

TRUE ~ NA\_real\_

)) %>%

# weight in kg:

mutate(weight = case\_when(

WEIGHT2 < 1000 ~ WEIGHT2 \* 0.453592,

WEIGHT2 >= 9000 & WEIGHT2 <= 9998 ~ as.numeric(str\_sub(WEIGHT2, 2, 4)),

TRUE ~ NA\_real\_

),

bmi = weight / height ^ 2,

trefethen = 1.3 \* weight / height ^ 2.5,

# BMI cutoff points from <https://www.who.int/gho/ncd/risk_factors/bmi_text/en/>

who\_cat = cut(bmi, c(0, 18.5, 25, 30, Inf), labels = c("Underweight", "Healthy",

"Overweight", "Obese"))) %>%

mutate(sex = case\_when(

SEX1 == 1 ~ "Male",

SEX1 == 2 ~ "Female",

TRUE ~ NA\_character\_

)) %>%

mutate(race = case\_when(

`\_HISPANC` == 1 ~ "Hispanic, Latino/a, or Spanish origin",

`\_PRACE1` == 1 ~ "White, non-Hispanic",

`\_PRACE1` == 2 ~ "Black or African American",

`\_PRACE1` == 3 ~ "American Indian or Alaskan Native",

`\_PRACE1` == 4 ~ "Asian",

`\_PRACE1` == 5 ~ "Pacific Islander",

`\_PRACE1` == 6 ~ "Other",

TRUE ~ NA\_character\_

),

race = fct\_relevel(race, "White, non-Hispanic")) %>%

mutate(race2 = fct\_collapse(race,

"Other" = c("Pacific Islander", "Other",

"American Indian or Alaskan Native")),

race2 = str\_wrap(race2, 20),

race2 = fct\_reorder(race2, `\_LLCPWT`, .fun = sum)) %>%

mutate(age = case\_when(

`\_AGE65YR` == 1 ~ "Age 18 to 64",

`\_AGE65YR` == 2 ~ "Age 65 or older",

NA ~ NA\_character\_

)) %>%

mutate(race3 = str\_wrap(race2, 20),

race3 = fct\_reorder(race3, `\_LLCPWT`, .fun = sum)) %>%

mutate(hhinc = case\_when(

INCOME2 == 1 ~ "Less than $10k",

INCOME2 == 2 ~ "$10k to $15k",

INCOME2 == 3 ~ "$15k to $20k",

INCOME2 == 4 ~ "$20k to $25k",

INCOME2 == 5 ~ "$25k to $35k",

INCOME2 == 6 ~ "$35k to $50k",

INCOME2 == 7 ~ "$50k to $75k",

INCOME2 == 8 ~ "More than $75k",

TRUE ~ NA\_character\_

),

hhinc = fct\_reorder(hhinc, INCOME2)) %>%

mutate(education = case\_when(

EDUCA == 1 ~ "Never attended school or only kindergarten",

EDUCA == 2 ~ "Grades 1 through 8",

EDUCA == 3 ~ "Grades 9 through 11",

EDUCA == 4 ~ "Grade 12 GED",

EDUCA == 5 ~ "College 1 to 3 years",

EDUCA == 6 ~ "College 4 years or more",

TRUE ~ NA\_character\_

),

education = fct\_reorder(education, EDUCA)) %>%

# remove people taller than the world record height:

filter(height < 2.72) %>%

rename(psu = `\_PSU`,

survey\_weight = `\_LLCPWT`) %>%

dplyr::select(height, weight, sex, race, race2, race3, age, hhinc, education, bmi, psu, survey\_weight)

}

the\_caption = "Source: <http://freerangestats.info> analysis of the US CDC Behavioral Risk Factor Surveillance System Survey for 2018"

# small subsampled version of the data used during dev, not referred to in blog:

set.seed(123)

llcp\_small <- llcp %>%

sample\_n(20000, weight = `\_LLCPWT`, replace = TRUE) %>%

code\_vars()

nrow(distinct(llcp\_small)) # 17,370 - remembering higher weight observations likely to be resampled twice, and dropping NAs

# full data, with tidied variables

llcp\_all <- llcp %>%

code\_vars()

# Make a data frame for shading of healthy BMI, to get around problems with rectangles and log axes. See

healthy <- tibble(

xmin = 0.9,

xmax = 2.6,

ymin = 18.5,

ymax = 25,

race3 = unique(llcp\_small$race3)

) %>%

drop\_na()

# data frame to use for shading of healthy weight implied by BMI

healthy\_by\_height <- healthy %>%

dplyr::select(ymin, ymax, race3) %>%

gather(variable, bmi, -race3) %>%

mutate(link = 1) %>%

full\_join(tibble(height = seq(from = 0.9, to = 2.3, length.out = 100), link = 1), by = "link") %>%

mutate(weight = bmi \* height ^ 2) %>%

dplyr::select(-link, -bmi) %>%

spread(variable, weight)

p1b <- llcp\_all %>%

dplyr::select(height, weight, sex, age, race3, survey\_weight) %>%

drop\_na() %>%

ggplot(aes(x = height)) +

facet\_grid(age ~ race3) +

geom\_ribbon(data = healthy\_by\_height, aes(ymin = ymin, ymax = ymax),

colour = NA, fill = "grey90") +

geom\_jitter(alpha = 0.02, aes(y = weight, colour = sex, size = survey\_weight)) +

geom\_smooth(method = "rlm", se = FALSE, aes(y = weight, colour = sex)) +

scale\_size\_area(label = comma) +

theme(panel.grid.minor = element\_blank()) +

labs(title = "Relationship between self-reported weight and height in USA adults",

subtitle = "Men's weight is more dependent on their height than is the case for women.

Grey ribbon shows WHO-recommended healthy weight. Straight lines show empirical relationship of weight and height.",

x = "Height (m)",

y = "Weight (kg)",

caption = the\_caption,

colour = "",

size = "Each point represents X Americans:") +

guides(size = guide\_legend(override.aes = list(alpha = 1)))

print(p1b)

p3 <- llcp\_all %>%

drop\_na() %>%

ggplot(aes(x = height, y = bmi, colour = sex)) +

facet\_grid(age ~ race3) +

geom\_rect(data = healthy, aes(xmin = xmin, xmax = xmax, ymin = ymin, ymax = ymax),

colour = NA, fill = "grey90", inherit.aes = FALSE) +

geom\_jitter(alpha = 0.02, aes(size = survey\_weight)) +

geom\_smooth(method = "rlm", se = FALSE, size = 0.5) +

scale\_size\_area(label = comma) +

scale\_x\_log10(breaks = c(0.9, 1.1, 1.4, 1.8, 2.3)) +

scale\_y\_log10() +

theme(panel.grid.minor = element\_blank()) +

labs(title = "Relationship between BMI and height in USA adults",

subtitle = bquote("Taller people tend to have lower BMIs because on average weight is proportionate to less than the square of height (in fact, "~weight%prop%height^1.6~")"),

x = "Height (metres), logarithmic scale",

y = bquote("Body Mass Index: "~frac(weight, height ^ 2)~", logarithmic scale"),

caption = the\_caption,

colour = "",

size = "Each point represents X Americans:") +

guides(size = guide\_legend(override.aes = list(alpha = 1)))

print(p3)

#------------

p4 <- llcp\_all %>%

mutate(h2 = height ^ 2,

h2.5 = height ^ 2.5) %>%

dplyr::select(h2, h2.5, weight, survey\_weight) %>%

drop\_na() %>%

gather(variable, value, -weight, -survey\_weight) %>%

mutate(variable = gsub("h2", expression(height^2), variable)) %>%

ggplot(aes(x = value, y = weight)) +

facet\_wrap(~variable, scales = "free\_x") +

geom\_jitter(alpha = 0.02, aes(size = survey\_weight)) +

geom\_smooth(method = 'gam', formula = y ~ s(x, k = 5)) +

scale\_size\_area(label = comma) +

labs(title = bquote("Weight compared to "~height^2~"and"~height^2.5),

subtitle = "It's not possible to choose visually which of the two better expresses the relationship.",

y = "Weight (kg)",

x = "Height to the power of 2.0 or 2.5",

size = "Each point represents X Americans:") +

guides(size = guide\_legend(override.aes = list(alpha = 1)))

print(p4)

#-------Models---------------

llcp\_svy <- svydesign(~psu, weights = ~survey\_weight, data = llcp\_all)

model0 <- svyglm(log(weight) ~ log(height), design = llcp\_svy)

model1 <- svyglm(log(weight) ~ log(height) + sex + race + age, design = llcp\_svy)

anova(model1)

summary(model1)

# best exponent for height is 1.55, not 2 (this is the raw coefficient for height)

# men of the same race, age and height weigh 6% more than women (exp(0.056))

# Asians weight 89% of white non-hispanic of same sex and age:

exp(coef(model1))