

How General Health Index (GHI) performs better to predict health levels*

A comparative analysis of GHI and BMI against body measurements

Aamishi Avarsekar

November 30, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

BMI is a common indicator for general health but it has its limitations. It does not take into consideration the actual body composition of an individual and other important factors such as sex and age. Through this paper, I would like to propose a new way categorising people based on other body measurements such as the waist-to-hip ratio and wrist and ankle circumferences measurements as they are more indicative of a person's visceral fat, this proving a more accurate representation of general health. I would like to propose a newer formula for General Health Index (GHI) through the BodyM dataset. I plan on using the Isaac Kuzmar dataset that has accurate measurements of body fat to create a Confusion Matrix to see how my model compares in providing more accurate body fat estimations through easlily measurable data such as waist and hip circumference.

- this paper uses 2 datasets to capture the difference in fat assessment for popular methods. BMI is a popular way of estimating health issues in an individual but it fails to capture details like body composition. In its initial design, BMI was designed using only male [] subjects and thus cannot be applied to females. However, due to excess popularity, it is used as a self-assessment to estimate your body composition and health risks. However, due to its inability to capture body composition, a lot of false positives and false negatives are produced. These indivuals often identify health risks [] later in the onset. My aim with this paper is to introduce simple measurements that can be recorded at home and help indivuals recognise when they should seek mdeical attention.

*Code and data are available at: <https://github.com/aamishi/ImprovedGeneralHealthIndex/>

1.1 BMI Classification:

```
bmi_canada_gov <- data.frame(  
  Category = c("Underweight", "Normal weight", "Overweight", "Obese - Class I", "Obese - Class II", "Obese - Class III"),  
  `BMI Range` = c("Below 18.5", "18.5-24.9", "25.0-29.9", "30.0-34.9", "35.0-39.9", "40.0 and above")  
)  
  
# Create a table using knitr::kable  
knitr::kable(  
  bmi_canada_gov,  
  col.names = c("Category", "BMI Range"),  
  caption = "BMI Categories as per the Government of Canada Guidelines"  
)
```

Table 1: BMI Categories as per the Government of Canada Guidelines

Category	BMI Range
Underweight	Below 18.5
Normal weight	18.5–24.9
Overweight	25.0–29.9
Obese - Class I	30.0–34.9
Obese - Class II	35.0–39.9
Obese - Class III	40.0 and above

1.1.1 The difference between an underweight and normal person:

Let us take a closer look at the BMI levels below 25 for underweight and normal weight adults.

```
body_mass_data_raw <- read_parquet("../data/raw_data/body_mass_data.parquet")  
body_mass_clean <- body_mass_data_raw %>%  
  clean_names() %>%  
  select(gender_m1f2, age, height_cm, w_kg, bmi, fat_mass_kg, fm, ffm_kg, bone_mass_kg, muscle_mass_kg)  
  mutate(  
    bmi_category = case_when(  
      bmi < 18.5 ~ "Underweight",  
      bmi >= 18.5 & bmi <= 24.9 ~ "Normal weight",  
      bmi >= 25.0 & bmi <= 29.9 ~ "Overweight",  
      bmi >= 30.0 & bmi <= 34.9 ~ "Obese - Class I",  
      bmi >= 35.0 & bmi <= 39.9 ~ "Obese - Class II",  
      bmi >= 40.0 ~ "Obese - Class III"  
    )  
  )
```

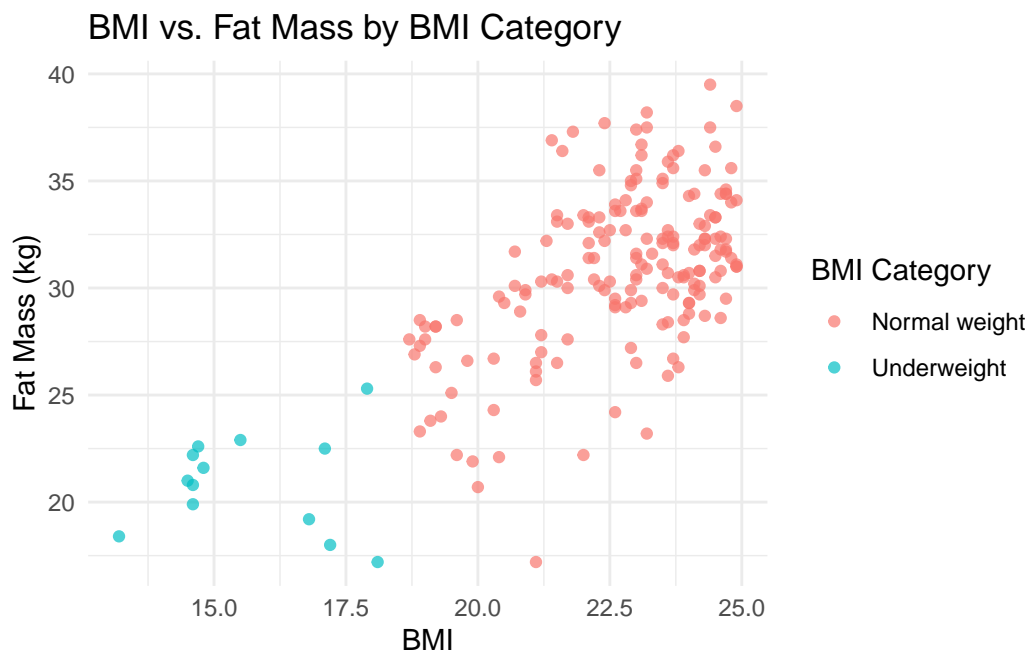
```

    bmi >= 35.0 & bmi <= 39.9 ~ "Obese - Class II",
    bmi >= 40.0 ~ "Obese - Class III"
  ))

# Data Section: Part 1: What is the difference between an underweight / athletic person / normal weight person / overweight person / obese person?
bmi_under_25 <- body_mass_clean %>%
  filter(bmi < 25)

ggplot(bmi_under_25, aes(x = bmi, y = fm)) +
  geom_point(aes(color = bmi_category), alpha = 0.7) +
  labs(
    title = "BMI vs. Fat Mass by BMI Category",
    x = "BMI",
    y = "Fat Mass (kg)",
    color = "BMI Category"
  ) +
  theme_minimal()

```



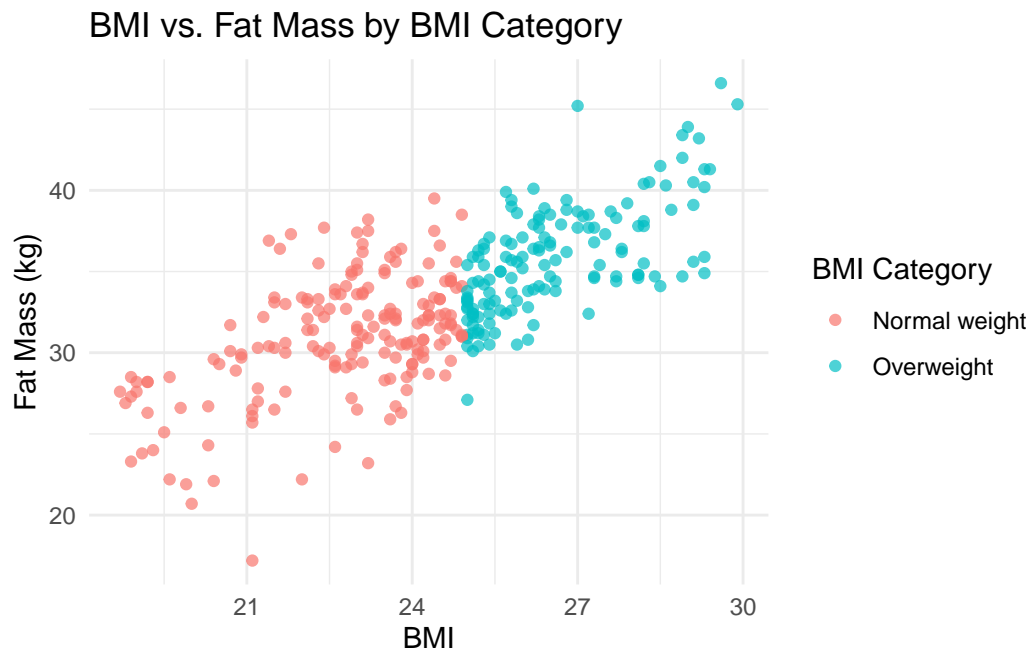
1.1.2 The difference between a normal and overweight person:

```

bmi_normal_over <- body_mass_clean %>%
  filter(18.5 < bmi & bmi < 30)

```

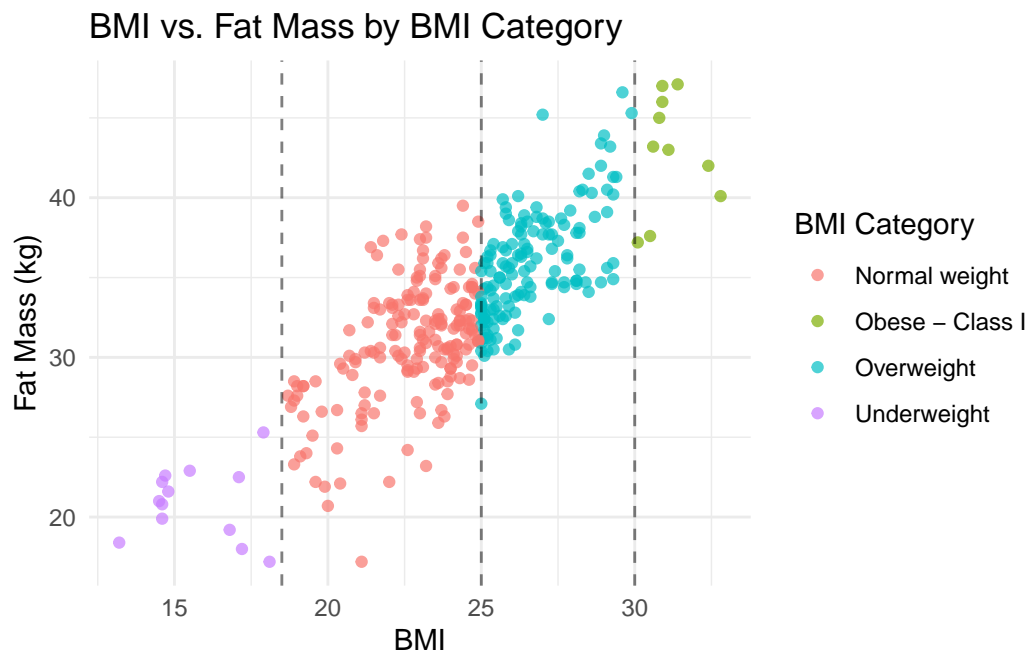
```
# this is what i chose
ggplot(bmi_normal_over, aes(x = bmi, y = fm)) +
  geom_point(aes(color = bmi_category), alpha = 0.7) +
  labs(
    title = "BMI vs. Fat Mass by BMI Category",
    x = "BMI",
    y = "Fat Mass (kg)",
    color = "BMI Category"
  ) +
  theme_minimal()
```



1.1.3 All categories

```
# this is what i chose
ggplot(body_mass_clean, aes(x = bmi, y = fm)) +
  geom_point(aes(color = bmi_category), alpha = 0.7) +
  geom_vline(xintercept = c(18.5, 25, 30), linetype = "dashed", color = "black", linewidth = 1) +
  labs(
    title = "BMI vs. Fat Mass by BMI Category",
    x = "BMI",
    y = "Fat Mass (kg)",
  )
```

```
color = "BMI Category"
) +
theme_minimal()
```



A new way of estimating general body health is to use WHR. This ratio aims to capture the excess fat deposit in a person's abdominal area which is where most people store adipose fat. A high WHR indicates high levels of fat. However, this is not enough. This is simply the most effective way of identifying the most obvious and dangerous place of fat accumulation.

2 Data

2.1 Overview

2.1.1 Body Composition Data - Isaac Kuzmar Et Al.

2.1.1.1 talk about this data

2.1.1.2 how was it collected

2.1.1.3 what is collected

2.1.1.4 graph what i will be using in my model

2.1.2 Measurements Data - BodyM - AWS.

2.1.2.1 talk about this data

2.1.2.2 how was it collected

2.1.2.3 what is collected

2.1.2.4 graph what i will be using in my model

2.2 How are these data sets important are how they relate to each other

2.2.0.1 desitination

2.2.0.2 well as any relationships between the variables.

Overview text

2.3 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

Talk more about it.

Talk way more about it.

2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (1)$$

$$\mu_i = \alpha + \beta_i + \gamma_i \quad (2)$$

$$\alpha \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta \sim \text{Normal}(0, 2.5) \quad (4)$$

$$\gamma \sim \text{Normal}(0, 2.5) \quad (5)$$

$$\sigma \sim \text{Exponential}(1) \quad (6)$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

- I wanted to provide a categorisation of fat levels in an individual based on simple measurements. Using measurements such as height and weight that are used in the formula and WHR and ankle and wrist measurements, I wanted to capture a person's body composition as much as possible. I expected that there is a positive relationship between a person's excess body fat levels and their WHR. In addition to WHR, the height if a person is also important to capture the distribution of fat, a short person with a high WHR would be at a higher risk of obesity related diseases.
- men are more likely to collect fat in their abdomen area and women in their hip area. To capture nature body shapes and leanness, I also considered....

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ .

Table 2: Explanatory models of flight time based on wing width and wing length

4 Results

Our results are summarized in Table 2.

5 Discussion

5.1 First discussion point

- height is not all, personality is more important ## Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.2 Third discussion point

5.3 Weaknesses and next steps

- the BodyM dataset does not include age. An older person has lower metabolism [] and a higher chance of collecting fat on their abdomen.

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

we compare the posterior with the prior. This shows...

References

- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.