

Project #2: Human Height and Weight

In this project you will revisit the data sets BBhtwt.dat and Bigheightwt.dat, this time looking at the weight data as well as the height data. First, calculate the mean and the standard deviations for the weight data in both BBhtwt.dat and Bigheightwt.dat. Now we will look for correlations between height and weight. Plot the weight vs. height for the two data sets (be sure to make the plot show points!). What about the plots indicates that height and weight are correlated? Calculate the correlation σ_{hw} (h is height and w is weight) and the Pearson's correlation coefficient for each data set. Finally, we will look at the Body Mass Index (BMI), given by the formula $BMI = 703 \left(\frac{w}{h^2} \right)$, where the weight w is in pounds, the height h is in inches, and the factor 703 converts from english to metric units. The BMI is used as an improved measure of health than just weight since it accounts for the fact that taller people should weigh more. We can calculate the average value of the BMI of a data set just by using the average values of height and weight. We would like to use error propagation to find the standard deviation in BMI in a data set given the standard deviations in height and weight. However, this is complicated by the fact that height and weight are correlated. In this case the formula becomes (for two correlated variables)

$$\sigma_q = \sqrt{\left(\frac{df}{dX} \right)^2 \sigma_x^2 + \left(\frac{df}{dY} \right)^2 \sigma_y^2 + 2 \left(\frac{df}{dX} \right) \left(\frac{df}{dY} \right) \sigma_{xy} + \dots} \quad (1)$$

where the derivatives are all evaluated at the average values of X and Y . Calculate a numerical value for the standard deviation of the BMI for both data sets both with and without σ_{hw} in order to gauge its effect. Use python or matlab to make a list of BMI values for each data set and calculate the means and standard deviations directly.

Discussion: Compare the Pearson's coefficients for the two data sets and discuss why baseball players' heights and weights might be more or less correlated than adolescents'. Discuss what effect correlation has on the standard deviation of quantities that combine correlated values. When does it increase the standard deviation, and why? When does it decrease the standard deviation, and why? Discuss how well your direct calculations of the BMI average and standard deviation match the results of your error propagation calculation. Finally, compare the means and the standard deviations of BMI for the two data sets and discuss why they might be different. Now, the argument for using BMI instead of weight as an indicator of health is that your weight depends not just on your fitness but also how tall you are. Thus some of the variability in weight is just due to variation in height. This would suggest that BMI should have less variability than weight, since it supposedly accounts for a person's height. Now, since BMI and weight have different average values, we need to consider relative variability, i.e. the ratio of the standard deviation and the average. This tells us by what percentage the quantity varies; for instance, if $\sigma = 10\text{lbs}$ in a weight sample and the average weight is 100lbs then we would say that there is about a 10% variation in weight in the sample. Compare the relative variability of weight and BMI for the samples. Does this comparison suggest that BMI is a better indicator of fitness than weight?