Estimation of Resting Metabolic Rate Using Height and Simple Linear Regression*

Subtitle

Joseph Burks

October 29, 2025

Resting metabolic rate (RMR) is a measure of the energy the body requires to maintain basic life functions. People can use RMR to better tailor diet and nutrition plans, but obtaining an accurate measurement requires laboratory-based tests. In this report, we developed a predictive equation using data from 95 observations, ranging from 18 to 65 years old. We used simple linear regression with height as the predictor variable: RMR = -2082.267 + 22.075(height) with ^2 = 0.469. We found height is a moderately accurate predictor of RMR. However, our model explains less of the variation of RMR than more well-established formulas.

1 Introduction

Resting metabolic rate (RMR) is the amount of energy the body requires to maintain basic life functions while at rest. In other words, it is the amount of calories one burns at rest. RMR accounts for approximately 60-70% of the total calories burnt daily (Plaza-Florido and Alcantara 2023). Knowing one's RMR can help tailor nutrition and diet plans to an individual's needs.

Accurately measuring RMR requires laboratory testing using gas analysis. By measuring the oxygen consumed and the carbon dioxide expelled while at rest, one can obtain an accurate estimate of RMR. However, due to cost and accessibility, several formulas have been developed to estimate RMR. The most widely used one is the Mifflin-St Jeor Equation, which uses four predictors: weight, height, age, and sex. The Mifflin-St Jeor equation was obtained using multiple linear regression. Although considered the most accurate formula, it leaves approximately 30% of the variance in RMR unexplained (Mifflin et al. 1990).

*Project repository available at: https://github.com/GrumioEstCoquus/MATH261A-project1.

The report investigates whether height alone is a good predictor of resting metabolic rate. We fit a simple linear regression model, with height as the predictor and resting metabolic rate as the response variable. The fitted model is $\hat{Y} = -2082.267\ kcal/day + 22.075cm \cdot X$. On average, RMR increases by 22.075 kcal/day for each additional centimeter of height.

The remainder of this report is structured as follows: Section 2 discusses the data, Section 3 discusses the simple linear regression model, Section 4 discusses the results, and Section 5 discusses the findings, weaknesses, and further questions.

2 Data

The data we used in this report comes from the Harvard Dataverse and was used in the study *Gender-agnostic Inclusive Estimation for Resting Metabolic Rate.* (Navalta 2025). The dataset contains 95 observations of healthy individuals. Participants' ages range from 18 to 65 years. Sex was not recorded, so it is unknown whether male or female participants were oversampled, but besides that, there are no concerns about the representativeness of the sample. Participants' RMR was measured using indirect calorimetry. Indirect calorimetry is a non-invasive method of measuring the oxygen consumption and carbon dioxide production of an individual. The variables of interest are height and resting metabolic rate. Height is measured in centimeters and has a mean of 168.399 cm and a standard deviation of 10.185 cm. RMR is measured in kcal per day with a mean of 1635.097 kcal/day and a standard deviation of 328.428 kcal/day.

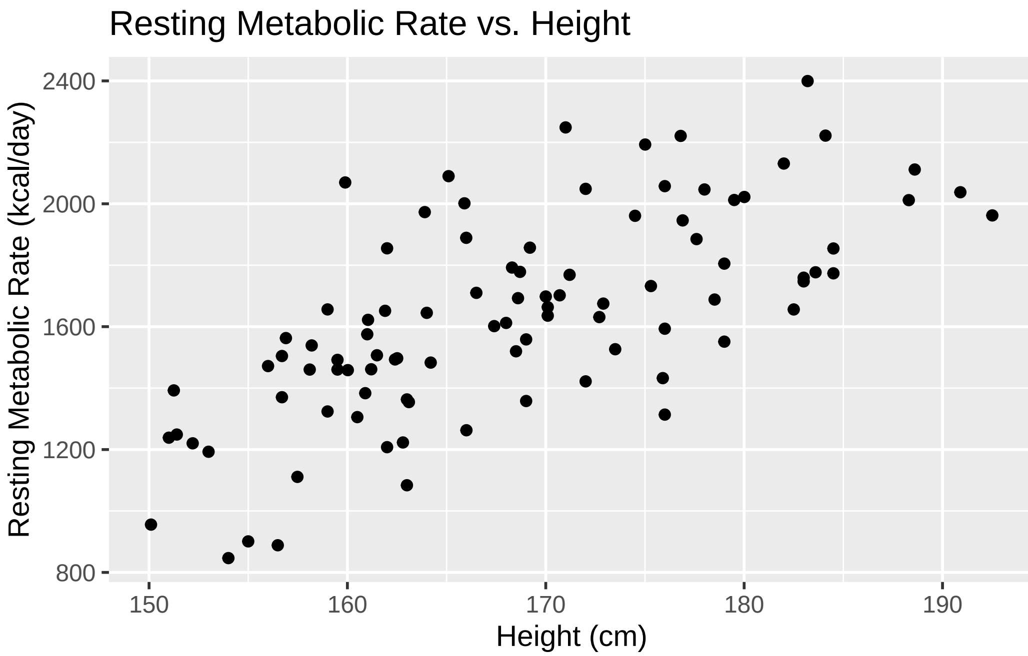

Figure 1: Scatter plot of Resting Metabolic Rate (kcal/day) vs. Height (cm)

A preliminary scatterplot Figure 1 indicated a strong positive linear relationship between height and RMR, suggesting that linear regression would likely be an appropriate model.

3 Methods

We used a simple linear regression model in this report. The general model is as follows: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon$. Y_i are observed responses. In this case, the observed RMRs. β_0 is the intercept. It represents the average response value when the predictor is 0. In this case, it contains little meaning, since the predictor variable is strictly positive. β_1 is the slope of the line of best fit. It represents the average increase in resting metabolic rates for a one-centimeter increase in height. X_i are the observed predictor values. In this case, these are the heights. The $\varepsilon_i$are the error terms and represent the variability in Y not explained by a linear function of X.

There are several assumptions for simple linear regression. One is that the relationship between the response and predictor must be linear. Meaning, $E[Y_i] = \beta_0 + \beta_1 X_i$ is correct. The others relate to the ε_i. The ε_i must be independently and identically distributed normal with constant variance ($\varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2$. If the above assumptions are true, we can use hypothesis testing to test the model estimates. The null hypotheses are $\beta_0 = 0$ and $\beta_1 = 0$. We fit the simple linear regression model using the lm() function in R (R Core Team 2024).

4 Results

The fitted model is: $\hat{Y}_i$ = -2082.267+22.075 X_i

The estimated intercept parameter is b_0 = -2082.267. This represents the average RMR for a person with a height of 0cm. Since height is strictly positive, the intercept does not contain much meaning.

The estimated slope parameter is b_1 = 22.075. This means that for each one centimeter increase in height, the resting metabolic rate increases on average by 22.075.

This simple linear regression has an R-squared of 0.469. This means the model explains around 46.9% of the variation of RMR.

To test the assumptions of linearity, constant variance, and independence, we plot the residuals $(\hat{Y}_i - Y_i)$ vs the fitted values (Figure 2).

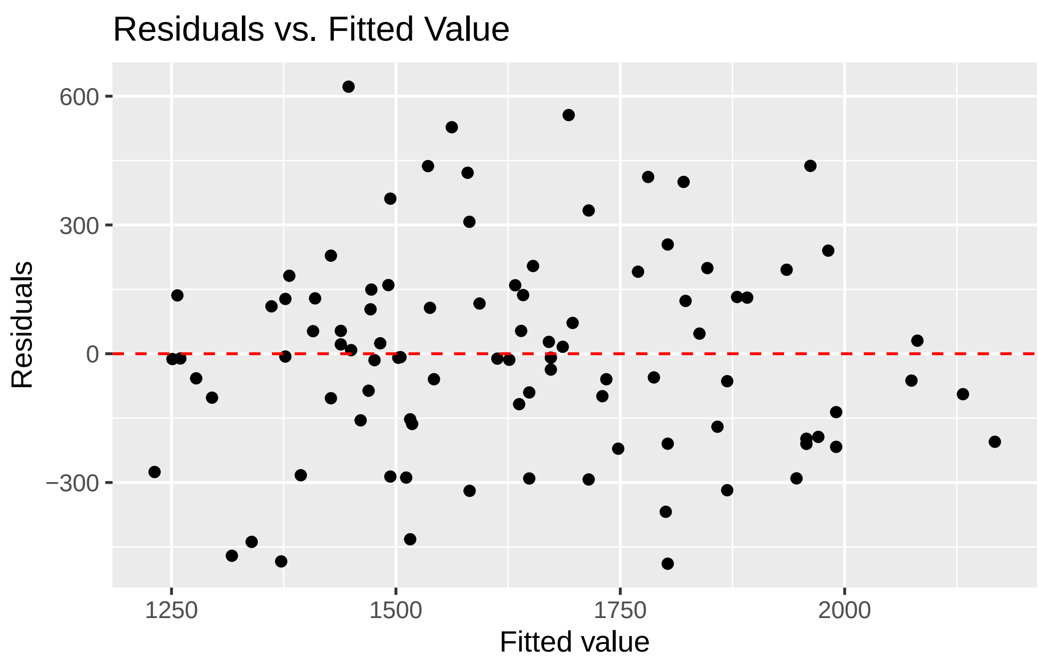

Figure 2: Residuals (y-axis) vs fitted values (x-axis) for a simple linear regression model with resting metabolic rate (kcal/day) as the response and height (cm) as the predictor

Figure 2 shows the residuals evenly distributed vertically, with no clear patterns and no systematic change in variance. While there are a few oddities with extreme points, these are likely due to a lack of data rather than assumption violations. Based on Figure 2, the assumptions of independence, constant variance, and linearity seem satisfied.

To test the assumption of normality, we graph a Quantile-Quantile plot, the sample quantiles vs the theoretical normal quantiles (Figure 3). If the data is normal, the points on the Quantile-Quantile plot will lie on a straight line.

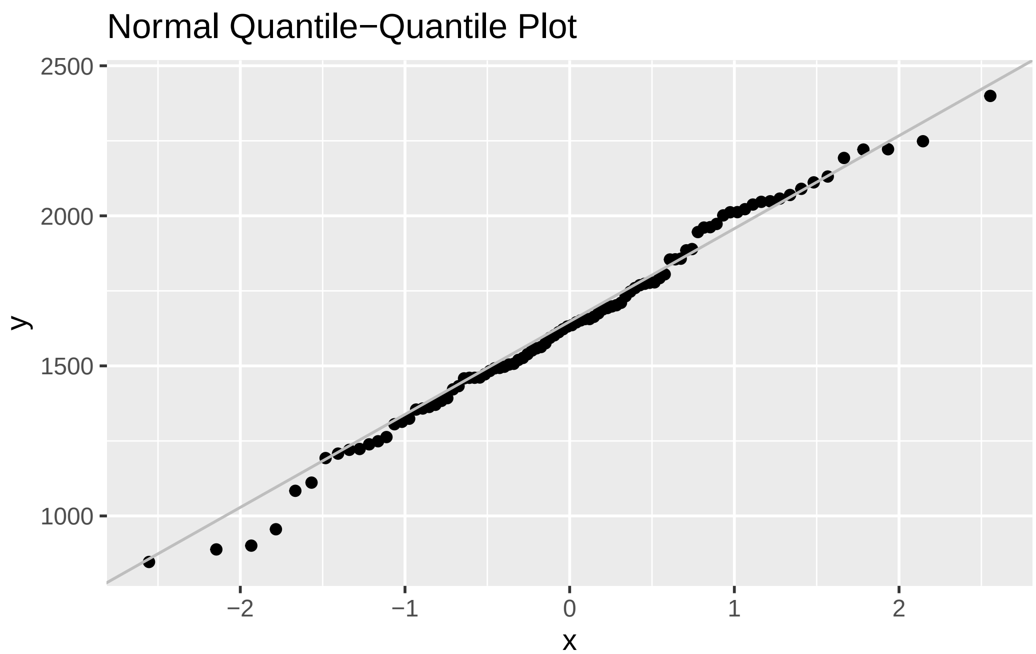

Figure 3: Plot of quantiles of residuals vs the theoritical normal distribution quantiles

Figure 3 shows very little deviations from the straight line, indicating that the assumption of normality of the error terms is also satisfied.

Because the data satisfy the assumptions, we can use hypothesis testing to check the model fit. We conducted a two-sided t-test with null hypothesis $\beta_1 = 0$ and alternative hypothesis $\beta_1 \neq 0$. The associated p-value for β_1 is $2.8057667 \times 10^{-14}$ which is much smaller than the classical significance of $\alpha = 0.05$. Hence, we reject the null hypothesis. There is evidence to suggest that $\beta_1 \neq 0$.

5 Discussion

The simple linear regression model indicates that height is a moderately strong predictor of RMR, explaining nearly half of its variation $R^2 = 0.469$. Moreover, the result of the hypothesis test indicates that the estimated slope is statistically significant. However, while height alone is a decent predictor, it leaves a lot of variation unexplained. The Mifflin-St Jeor Equation, which includes height, weight, age, and gender, is more accurate and likely better.

Height may also be a proxy for weight. Taller people will weigh more than shorter people, on average. Future research could test whether weight provides a stronger relationship or use BMI as a way to standardize height and weight.

A valuable next step would be fitting a multiple linear regression using additional body composition variables as predictors, and could explore if it is necessary for men and women to have separate equations, as in the Mifflin-St Jeor Equation.

References

Mifflin, MD, ST St Jeor, LA Hill, BJ Scott, SA Daugherty, and YO Koh. 1990. "A New Predictive Equation for Resting Energy Expenditure in Healthy Individuals." *The American Journal of Clinical Nutrition* 51 (2): 241–47. https://doi.org/https://doi.org/10.1093/ajcn/51.2.241.

Navalta, James. 2025. "Data for: A Gender-agnostic Inclusive Estimation for Resting Metabolic Rate." Harvard Dataverse. https://doi.org/10.7910/DVN/MIOESN.

Plaza-Florido, Abel, and Juan M. Alcantara. 2023. "Resting Metabolic Rate of Individuals." *Metabolites* 13 (8): 926. https://doi.org/10.3390/metabo13080926.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.