

Data Analysis: Report II

Jane Bloggs

9 January 2023

Descriptive statistics

In this data set, we have measured the weight (in kg) and height (in cm) of 6068 participants (4082 males, 1986 females). In Table 1, we show the mean and separately for males and females.

Statistical model

We will model the relationship between weight and height as a varying intercepts normal linear model as follows.

For each $i \in 1 \dots n$,

$$y_i \sim N(\mu_i, \sigma^2),$$
$$\mu_i = \beta_0 + \beta_1 x_i + \beta_2 z_i,$$

where y_i , x_i , z_i are the weight, height, and gender of participant i .

In R, this analysis can be easily performed as follows.

```
model <- lm(weight ~ height + gender, data = weight_df)
```

The R^2 for this model is 0.45, $F(2, 6065) = 2495.83$, $p < 0.01$.

More information about varying-intercept models can be found in Gelman and Hill (2006).

References

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Table 1: Descriptive summary

gender	weight_avg	height_avg	weight_stdev	height_stdev
female	67.76	162.85	10.98	6.42
male	85.52	175.62	14.22	6.86