Verify analytically that the Residual standard error ($\hat{\sigma} = \sqrt{\frac{SSE}{n-p}}$) is identical to the sample standard deviation of the heights in the simple mean model.

The residual standard error ist by definition the standard deviation σ in the normal distribution for the errors ($\varepsilon_i \sim N(0, \sigma)$) in the model. We don't know σ but estimate it from the residual sum of squares of errors (SSE). Residuals are defined as the difference between the actual value and the model-predicted values:

$r_i = height_i - \widehat{height}_i = height_i - \overline{height_i}$

The standard deviation of the residuals is:

$SD(r_i)^2 = \frac{1}{n-1} \sum_{i=1}^{n} (r_i - \bar{r}_i)^2$

$\overline{r_i} = \frac{1}{n} \sum_{i=1}^{n} (height_i - \overline{height_i}) = \frac{1}{n} \sum_{i=1}^{n} (height_i) - \frac{n \cdot \overline{height_i}}{n} = \overline{height_i} - \overline{height_i} = 0$

Hence,

$SD(r_i) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} r_i^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (height_i - \overline{height_i})^2} = SD(height_i) = \sqrt{\frac{SSE}{n-1}} = \hat{\sigma}$

which is just the definition of the residual standard error and an estimate for σ

In general, this is not true since the estimator for $\hat{\sigma} = \sqrt{\frac{SSE}{n-p}}$, where p is the number of parameters. We only had 1 parameter (mean) to estimate, hence $p = 1$.

The residual standard error $\hat{\sigma}$ is an estimation for σ which tells us, how widely spread the residuals are around the model-estimation. In our example, we can expect to find approximately two thirds of heights within $\overline{height_i} \pm 1 \cdot 7.742$ and approximately 95% of observations within $\overline{height_i} \pm 2 \cdot 7.742$.