Adv Data Programming with R - HW 1

Sathvik Srinath Gaurav - 24213364

Introduction to Dataset

- The **GaltonFamilies** dataset comes from a 19th-century heredity study by *Sir Francis Galton*.
- It includes height data from 204 families of 934 individuals, with father, mother, and multiple children in each.
- Each child is identified by gender and birth order, and their height is recorded.
- This dataset helps us explore how parent height influences child height.

> **Galton's Approach**
>
> - Galton's goal was to investigate how parental characteristics influence their offspring, particularly focusing on height as a measurable inherited trait.
> - Galton created a "mid-parent height" by adjusting and averaging the parent's heights to study inheritance patterns.
>
> $midparentHeight = \frac{(father + 1.08 * mother)}{2.08}$

Summary Statistics

Using the **summary()** function on the GaltonFamilies dataset, it gives a overview of the key statistics, such as the minimum, maximum, mean, and quartiles for each variable.

```
    family        father          mother      midparentHeight    children
185    : 15   Min.   :62.0   Min.   :58.00   Min.   :64.40   Min.   : 1.000
066    : 11   1st Qu.:68.0   1st Qu.:63.00   1st Qu.:68.14   1st Qu.: 4.000
120    : 11   Median :69.0   Median :64.00   Median :69.25   Median : 6.000
130    : 11   Mean   :69.2   Mean   :64.09   Mean   :69.21   Mean   : 6.171
166    : 11   3rd Qu.:71.0   3rd Qu.:65.88   3rd Qu.:70.14   3rd Qu.: 8.000
097    : 10   Max.   :78.5   Max.   :70.50   Max.   :75.43   Max.   :15.000
(Other):865
   childNum         gender     childHeight
Min.   : 1.000   female:453   Min.   :56.00
1st Qu.: 2.000   male  :481   1st Qu.:64.00
Median : 3.000                Median :66.50
Mean   : 3.586                Mean   :66.75
3rd Qu.: 5.000                3rd Qu.:69.70
Max.   :15.000                Max.   :79.00
```

- Most fathers are between 68–71 inches tall.
- Most children are around 66–70 inches tall.
- Almost equal numbers of male and female children.
- On average, each family had about 6 children, with some families having up to 15.

Mid-parent height is calculated and summary statistics is provided, which includes the means and standard deviations of both child and parent heights.

Table 1: Summary Statistics of Parent and Child Heights

mean_child	mean_parent	sd_child	sd_parent
66.75	66.54	3.58	1.73

- On average, the children in the dataset are about 66.75 inches tall.
- The average mid-parent height (weighted average of father's and mother's height) is 66.54 inches, which is comparatively similar to average child height.
- More variation in child height compared to parents.

Analysis - Mid-Parent vs Child Height

- For exploring, we analyze the relationship between mid-parent height (a weighted average of the father's and mother's heights) and the child's height.
- The mid-parent height places an important key as it accounts for differences in average height between men and women, making it a balanced estimate of expected genetic influence. By plotting child height against this mid-parent value, we can visually inspect the strength and direction of inheritance.

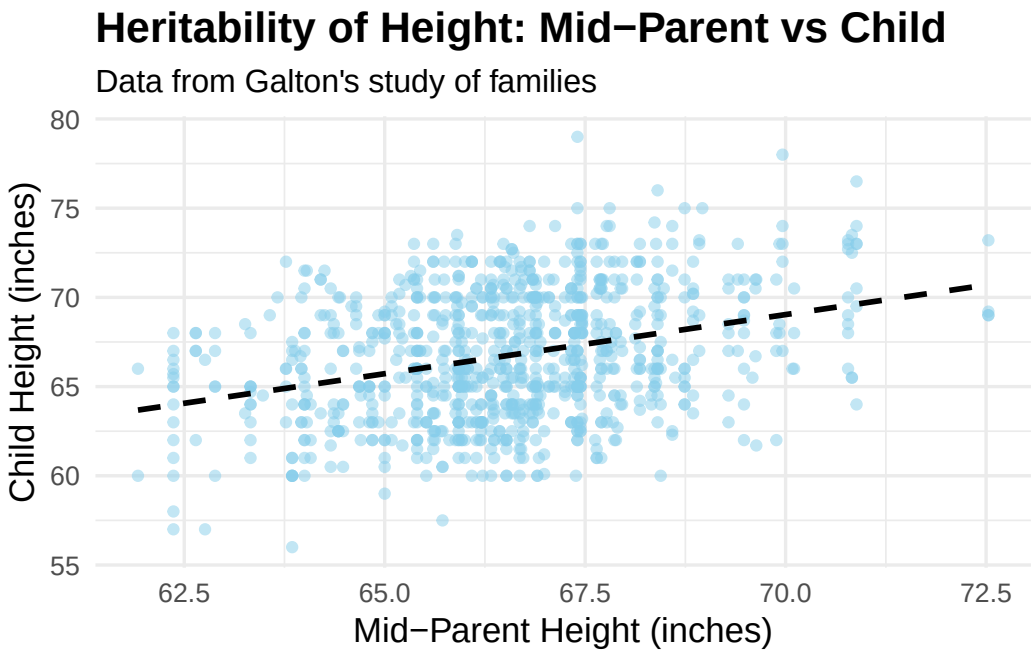

Figure 1: Relationship between mid-parent height (weighted average of father and mother's height) and child height, based on Galton's 19th-century heredity study

- The positive slope of the trend line suggests a **direct correlation**, as mid-parent height increases, child height increases.
- Some variability is observed around the line, indicating parental height is influential, other factors also affect child height.
- The plot illustrates Galton's insight, where parents with extremely tall or short height often have children whose heights are closer to the average.

Conclusion

In summary, this plot illustrates that children's heights generally relate to their parent's average height, reflecting genetic influence, while also highlighting the natural variation present in human growth.