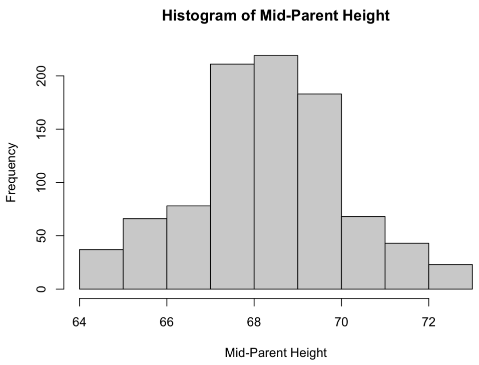
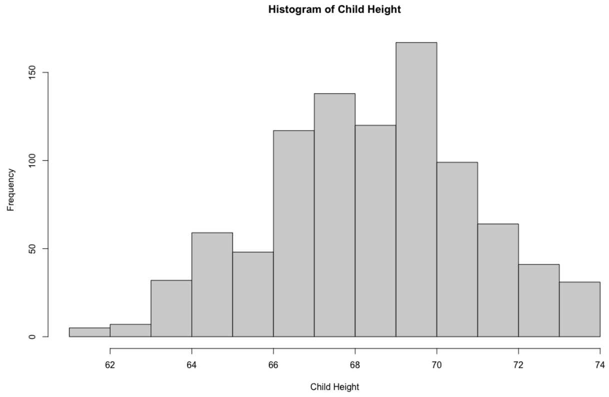
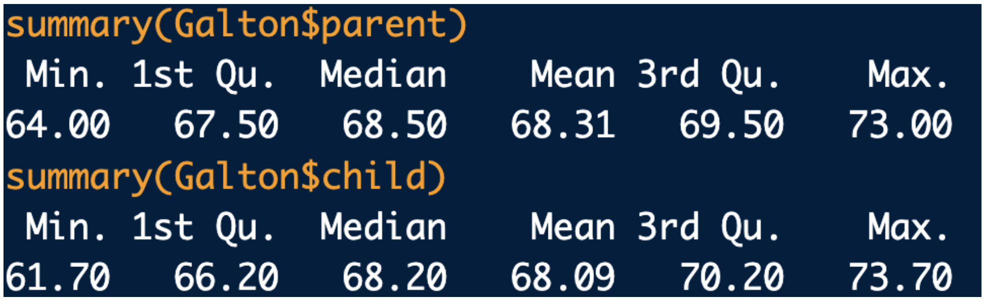
Derin Gezgin | Johnny Andreasen | Sababa Ahmed
Group Project #1

1. **Install the package and get it ready for this session.**
   R-Code: install.packages("UsingR")
2. **Open help file for the data and use it to understand variables.**
   R-Code: ?Galton
   There are two variables in the data:
   - **Parent:** A numeric vector: height of the mid-parent (average of father and mother) in inches.
   - **Child:** A numeric vector: height of the child in inches.
3. **What is the goal of this study? State the response and predictor variables.**
   The study aims to see if the mid-parents' height in inches is directly related to the child's height in inches.
   - **Predictor Variable (X):** Mid-parent's height in inches.
   - **Response Variable (Y):** Child's height in inches.
4. **Check the data summaries and make a plot for each variable.**





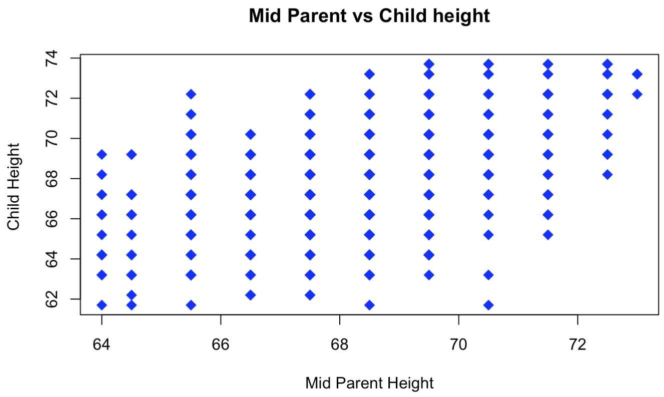


```
summary(Galton$parent)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
64.00   67.50   68.50   68.31   69.50   73.00
summary(Galton$child)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
61.70   66.20   68.20   68.09   70.20   73.70
```

5. **How many missing values do we have? Do you see any outliers? Anything unusual?**
   - We have no missing values in the Galton data set.
   - There are also no outliers in the data set. We can also see this from the scatterplot in the next question.
   - **R-Code**: sum(is.na(Galton)): Returns 0 which means no missing values

6. **Using correlation coefficient and scatter plot, comment on the relationship.**
   - **Multiple R^2:** 0.459

Derin Gezgin | Johnny Andreasen | Sababa Ahmed
Group Project #1



- The correlation coefficient is average with some positive correlation. The scatterplot shows some relation between the two variables but it's far from being a strong relationship.

7. Fit a linear regression model and show the fitted model.

The fitted model is $\hat{Y}$ **= 23.942 + 0.646x.**
x = **Mid Parent's Height**
$\hat{Y}$**= Estimated Child Height**

8. Interpret the regression coefficients (slope and intercept) of the model?

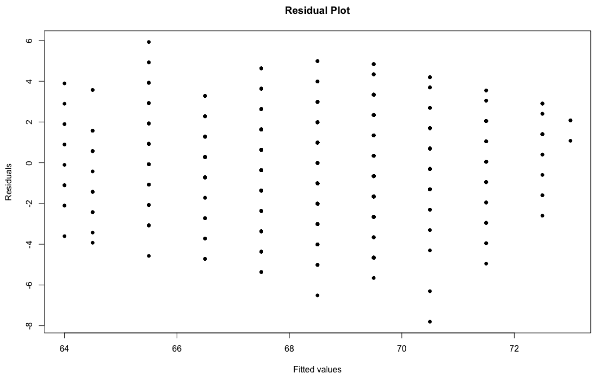
The slope estimate is ***0.646***. Which is the expected difference in the estimated "child height" for each 1 inch increase in the mid parent's height
The intercept estimate is ***23.942***. Which is the expected child height when the mid-parent's height is 0.

9. Check goodness of fit of this model?

a. Coefficient Determination
For this factor, we can see that the Multiple R-squared values is *0.2105* which means that the 21% of the child height can be explained by mid-parent height. It's the variability in Y explained by X.

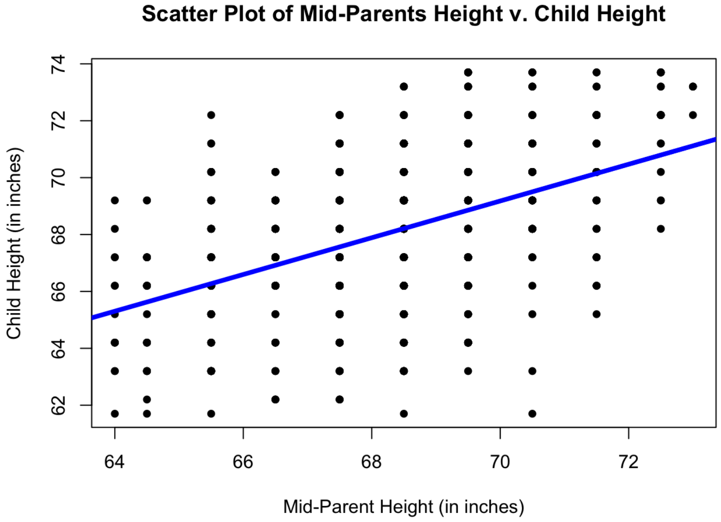
b. Variability in Errors
The average distance between the observed values of Child height and fitted values of it is ± 2.239.

c. Residual Plot



Derin Gezgin | Johnny Andreasen | Sababa Ahmed
Group Project #1

10. Show fitted values in the scatterplot.



11. Are regression coefficients statistically significant at 5% significance level?

Step 1: H_0: B_1=0 and H_a: B_a≠0
Step 2: let ***a*** = 0.05
Step 3: **t** value = 8.517
Step 4: **p** value = 2e-16 which is approximately 0. So reject H_0
Step 5: At 5% significance level, mid-parent height is a significant predictor of child height.

12.Report 99% Cis for regression parameters.

We're %99 confident that the actual difference in the estimated "child height" for each 1-inch increase in the mid parent's height is between 0.540 and 0.752.

We're %99 confident that the actual child height when the mid-parent's height is 0.
is between 16.686 and 31.197