Examining Factors That Influence Weight in Children from the Queen Mothers Hospital, Glasgow

Group 1

March 2025

1 Introduction

Steadily increasing weight is an indicator of a healthy child. Weight change differs between children, so it is of interest to know what factors influence a child's weight. A study conducted at the Queen Mothers hospital in Glasgow sampled 127 new born babies gathered the following variables:

`Wt24` - The child's weight at 24 months (numerical variable, continuous).

`Wt1` - The child's weight at 1 month (numerical variable, continuous).

`Solids` - The age at which the child was first introduced to solid food (numerical variable, discrete).

`Sex` - The child's sex (categorical variable, with categories Male or Female).

The goal of this report is to examine which variables influence a child's weight at 24 months first by assessing how well `Wt24` can be predicted using a linear model with `Wt1` as an explanatory variable. We then see if the model if improved by adding `Sex` or `Solids` to the model. Finally, we see if there is any evidence that the difference in weight (`Wt24 - Wt1`) is different for male and female children.

Section 2 consists of exploratory analysis explores the potential relationships between variables by use of data visualisation and numerical summaries. Section 3 uses linear regression modelling to answer the questions of interest. Section 4 summarises the findings of the report as well as discuss some limitations of the study.

2 Exploratory Analysis

Figure 1 produces a scatterplot of `Wt24` against `Wt1` (left) and `Wt24` aginst `Solids` to assess a potential relationship between the response and explanatory variables. From the scatterplot of `Wt24` against `Wt1`it appears that there is a positive, moderate to strong linear relationship between `Wt24` and `Wt1`. Moreover, with the exception of a few points there appears to be a distinction by `Sex` with female children having lower weight and male children having a higher weight. The scatterplot of `Wt24` against `Solids` examines the relationship by having separate regression lines for males and females. The plot suggests a moderate negative linear relationship for females, indicating that introducing solids later may be associated with lower weight at 24 months. In contrast, the regression line for males remains relatively flat, suggesting a weak relationship between `Wt24` and `Solids` for male children.

Figure 2 presents boxplots illustrating the distribution of `Wt24` and `Wt1` categorised by `Sex`. The boxplot of `Wt1` grouped by `Sex` suggests that male children tend to have a slightly higher

Figure 1: Scatterplots of Wt24 against Wt1 coloured by sex (left) and Wt24 against Solids coloured by sex (right)

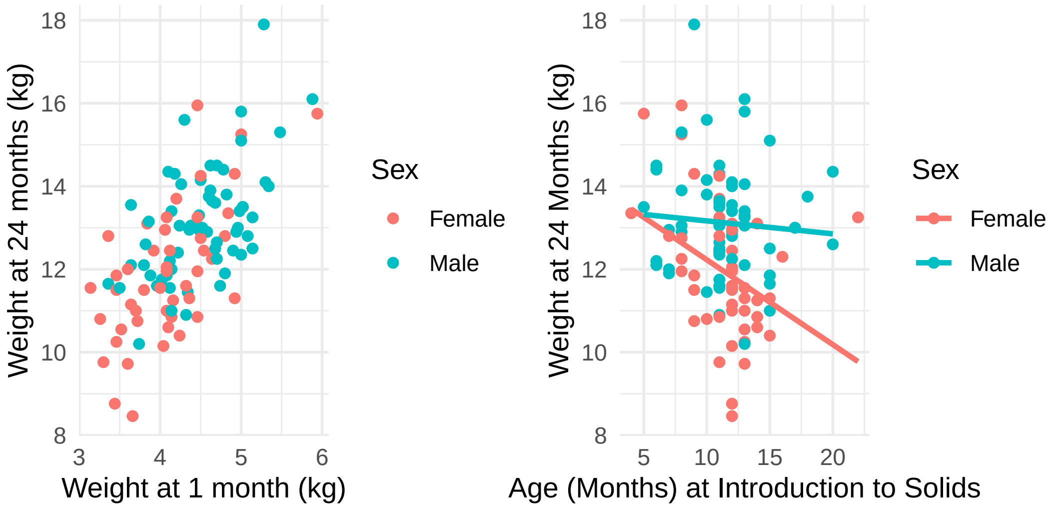


Figure 2: Boxplots of weight at 24 months by sex (left) and weight at 1 month by sex (right)

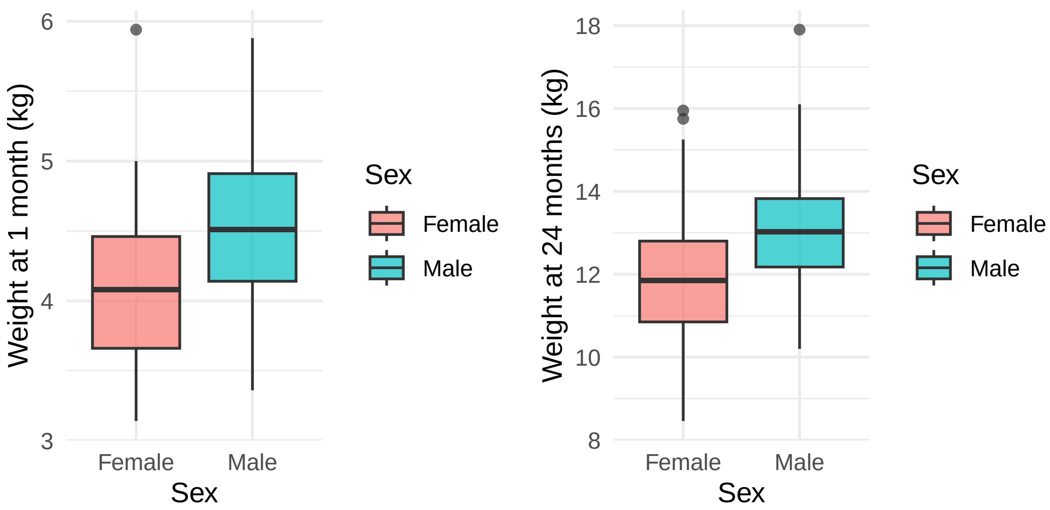


Table 1: Correlations between numerical variables by sex

term	Wt1	Wt24	Solids
Wt1	1.00	0.56	−0.13
Wt24	0.56	1.00	−0.07
Solids	−0.13	−0.07	1.00

term	Wt1	Wt24	Solids
Wt1	1.00	0.64	−0.23
Wt24	0.64	1.00	−0.37
Solids	−0.23	−0.37	1.00

median weight compared to females, with a slightly wider range of values. The presence of outliers, particularly among female infants, indicates some variation in early weight distribution. The boxplot of `Wt24` grouped by `Sex` highlights that male children exhibit a higher median weight than female children, and the overall distribution of weight is more spread out for males. The increased variation, particularly in the upper range, suggests that male children may experience a broader range of growth trajectories compared to females.

Table 1 shows the correlation between numerical variables for male and female children. For male children **?@tbl-corrs-1** highlights that for male children `Wt1` and `Wt24` are moderately, positively correlated, whereas `Solids` against `Wt24` and `Wt1` are weakly negatively correlated. This is slightly different for female children as from **?@tbl-corrs-2** there appears to be a higher positive correlation between `Wt24` and `Wt1`. Additionally, there appears to be a moderate negative correlation between `Solids` and `Wt24` for female children and a weak-moderate correlation between `Solids` and `Wt1`.

Table 4: Mean, median and standard deviation (sd) Wt24, Wt1 and Solids by sex.

	Wt24			Wt1			Solids		
Sex	Mean	Median	Std.Dev	Mean	Median	Std.Dev	Mean	Median	Std.Dev
Male	13.12	13.03	1.35	4.51	4.51	0.52	11.50	11.00	3.18
Female	11.94	11.85	1.59	4.11	4.08	0.55	11.43	12.00	2.91

Table 4 shows the mean, median and standard deviation of `Wt24`, `Wt1` and `Solids` by `Sex`. On average, there appears to be a slight difference in `Wt24` and `Wt1`by `Sex`. There does not appear to be a difference in `Solids` by `Sex` on average.

3 Formal Analysis

3.1 Primary Model

To investigate how well weight at 24 months can be predicted using a linear model with weight at 1 month as the explanatory variable, we begin by fitting the following regression model, written as:

$$Wt24_i = \alpha + \beta \; Wt1_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \tag{1}$$

Where, α is the intercept, β is the slope, $Wt24_i$ and $Wt1_i$ is `Wt24` and `Wt1` for the ith child. Note that, whilst the intercept is biologically meaningless this parameter is necessary.

Table 5: Estimates of the regression model coefficients.

term	estimate	p_value
intercept	4.82	0.00
Wt1	1.80	0.00

Table 5 shows the parameter estimates and p-values for Equation 1. From this we get the following fitted model:

$$\widehat{Wt24}_i = 4.82 + 1.80 Wt1_i \tag{2}$$

The p-value for `Wt1` is close to zero, thus `Wt1` is statistically significant in predicting `Wt24` . The fitted model can be interpreted as follows, for a 1kg increase in `Wt1`, we would expect `Wt24` to increase by 1.80kg, and, holding `Wt1` constant we would expect `Wt24` to be 4.82kg.

Figure 3: Relationship between Weight at 1 Month and Weight at 24 months with regression line superimposed.

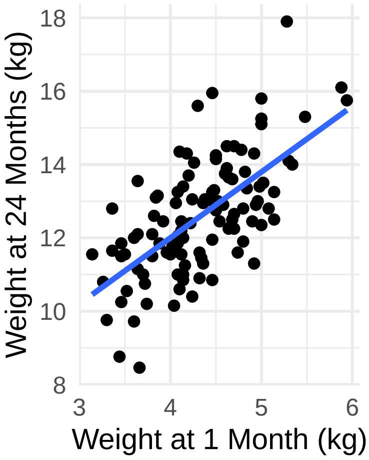


Figure 3 shows a plot of `Wt24` against `Wt1` with the regression model superimposed. The model appears to fit the data well, although we do see some outliers they do not appear to be influential.

Figure 4: Residuals against fitted values (left) and histogram of residuals (right)

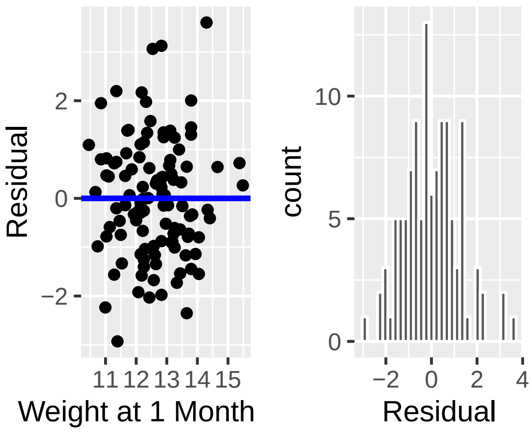


Figure 4 displays the residuals against fitted values (left) and a histogram of the residuals (right). From the plot of residuals against fitted values, it is reasonable to assume that the residuals have mean zero and constant variance as the points are scattered evenly above and below the line, and there does not appear to be any patterns such as fanning. The histogram of residuals allows us to check that the residuals are normally distributed. Although there appears to be some extremes far from the center of the data, the residuals appear to be normally distributed.

3.2 Secondary Models

To determine if the model in Equation 1 can be improved by adding the child's sex and/or age at introduction to solids as further explanatory variables we fit the following model:

$$Wt24_i = \alpha + \beta_1 Wt1_i + \beta_2 Solids_i + \beta_3 \mathbb{I}_{male} + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2)$$

Where, α is the intercept, $Wt24_i$, $Wt1_i$ and $Solids_i$, is `Wt24`, `Wt1` and `Solids` for the ith child. β_1 and β_2 are the slope terms for `Wt1` and `Solids` respectively. β_3 represents the change in `Wt24` by `Sex` with $\mathbb{I}_{male}$ being an indicator variable that takes the following value 1 if the child is male, and 0 if the child is female.

Table 6: Estimates of the regression model coefficients.

term	estimate	p_value
intercept	6.02	0.00
Wt1	1.58	0.00
Solids	−0.05	0.17
Sex: Male	0.56	0.02

From Table 6 we see that the p-value for Solids is 0.17 not statistically significant at a significance level of $\alpha = 0.05$. Therefore, we drop Solids from the model and refit with only Wt1 and Sex as predictors.

Table 7: Estimates of the regression model coefficients.

term	estimate	p_value
intercept	5.23	0.00
Wt1	1.63	0.00
Sex: Male	0.53	0.03

All coefficients in Table 7 appear to be statistically significant due to the low p-values. Therefore, we get the following regression equation:

$$\widehat{Wt24}_i = 5.23 + 1.63 Wt1_i + 0.53 \mathbb{I}_{male} \tag{3}$$

Figure 5 shows as scatterplot of `Wt24` against `Wt1` coloured by `Sex`, with separate regression lines for male and female children. Although we do see some outliers towards the top of Figure 5 for males and towards the bottom for females, these do not appear to be influential.

Figure 6 displays the residuals versus fitted values split by sex (left) and a histogram of residuals (right). From the residuals against fitted values plot it is reasonable to assume that the residuals have mean zero and constant variance as the points appear to be evenly scattered above and below the line and there does not seem to be any pattern among the residuals. From the histogram of residuals we can see that although it is not a perfect bell

Figure 5: Relationship between Wt24 and Wt1 by Sex with the regression lines superimposed.

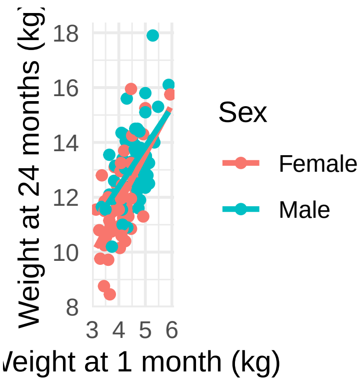
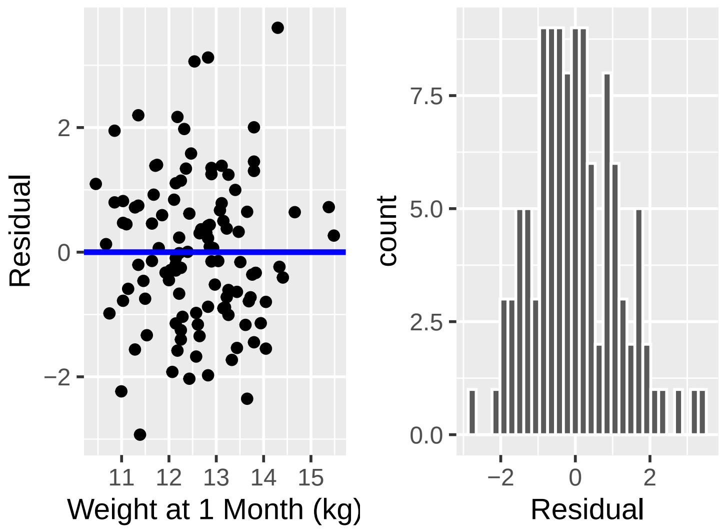


Figure 6: Residuals versus fitted values by sex (left) and Histogram of residuals (right)



curve due to a few outliers, the residuals approximatley follow a normal distribution. We conclude that it is reasonable to fit this linear regression model. However, we compare the models given in Equation 1 and Equation 3 using AIC.

Table 8: AIC values of the 2 modles

AIC Comparison Table

Model	AIC
Model with Wt1 and Sex	40.50
Model with Wt1 Only	43.48

From Table 8 we see that the more favorable model is the model that includes both `Wt1` and `Sex` as predictor variables of `Wt24` due to having a lower AIC.

To determine if the change in weight from 1 month to 24 months is different for male and female children, define `WtDiff` = `Wt24 - Wt1` to represent the change in weight between 1 month and 24 months for the ith child and fit the following one-way ANOVA model:

$$WtDiff_i = \alpha + \beta \mathbb{I}_{male} + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma^2)$$

Where $WtDiff_i$ is the difference in weight for the ith child, α is the intercept, $\mathbb{I}_{male}$ is an indicator variable that takes on the value 1 if the child is male and 0 if the child is female and β represents the shift in weight difference for male compared to female children.

Table 9: Estimates of the regression model coefficients.

term	estimate	p_value
intercept	7.82	0.00
Sex: Male	0.79	0.00

Since the coefficients displayed in Table 9 are all very close to 0, we conclude that they are all statistically significant and should be kept in the model.

Figure 7: Residuals versus fitted values by sex (left) and Histogram of residuals (right)

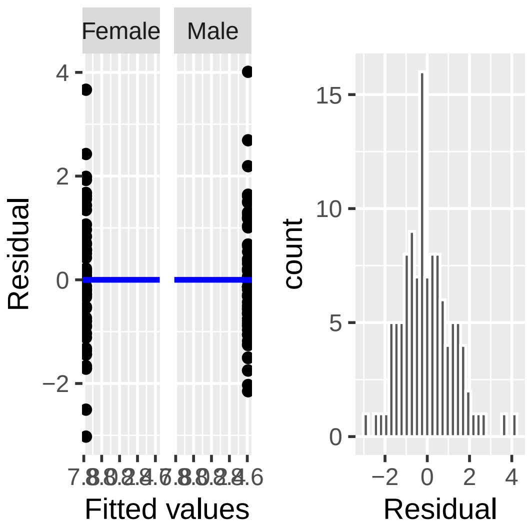


Figure 7 displays the fitted values versus the residuals split by sex as well as a histogram of residuals. From the residuals versus fitted values plot it is reasonable to assume that residuals have mean zero and constant variance as there appears to be the same number of observations both above and below and there does not appear to be any patterns such as fanning. Although we do see some outliers for both sex, these do not appear to be influential. From the histogram of the residuals we can see that although there appears to be a few extremes, the residuals appear to follow a normal distribution.

Since the model assumptions are satisfied it is reasonable to assume that there is statistically significant evidence that change in weight from 1 month to 24 months is different for male and female babies.

4 Conclusion

In conclusions, based on our explanatory analysis in *** and formal analysis in Section 3 we believe that....

We also conclude that this model can be improved by adding the child's sex however it is not improved by adding their age at introduction to solids. These comparisons and conclusions were done using Akaike Information Criterion as per Table 8. Additionally, we conclude that the change in weight from 1 month to 24 months is different for male and female babies due to the statistical significance of the coefficients seen in Table 9.

The limitations of the analysis carried about above is

5 References

- https://www.nhs.uk/conditions/baby/babys-development/height-weight-and-reviews/baby-height-and-weight/
- https://www.babycenter.com/baby/baby-development/average-weight-and-growth-chart-for-babies-toddlers-and-beyo_10357633