

Report 1

Kamil Zaborniak

Contents

Vector and Matrix Algebra	2
Properties of the spectral decomposition	3
The join distribution of length and weight of newborn children	4
Operations on a multivariate normal distribution	6
Analysis of weight and length data of newborn children	6
Analysis of weight and length data of newborn children and their parents' heights	9

Vector and Matrix Algebra

Let's consider the following symmetric matrix:

$$A = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}.$$

Let's perform the spectral decomposition of **A**. Let's start by determining the roots of the characteristic polynomial, i.e. the eigenvalues of the above matrix.

$$\begin{vmatrix} 3-\lambda & -1 \\ -1 & 3-\lambda \end{vmatrix} = (3-\lambda)^2 - 1 = (2-\lambda)(4-\lambda).$$

So we got eigenvalues: $\lambda_1 = 2$, $\lambda_2 = 4$. To determine the eigenvectors of the matrix: $e_1 = [x_1, y_1]^T$, $e_2 = [x_2, y_2]^T$, let us solve the systems of equations:

$$\mathbf{A}e_1 = 2e_1$$

$$\mathbf{A}e_2 = 4e_2$$

$$\begin{cases} 3x_1 - y_1 = 2x_1 \\ -x_1 + 3y_1 = 2y_1 \end{cases}$$

$$\begin{cases} x_1 = \frac{1}{\sqrt{2}} \\ y_1 = \frac{1}{\sqrt{2}} \end{cases}$$

$$\begin{cases} 3x_2 - y_2 = 4x_2 \\ -x_2 + 3y_2 = 4y_2 \end{cases}$$

$$\begin{cases} x_2 = \frac{1}{\sqrt{2}} \\ y_2 = -\frac{1}{\sqrt{2}} \end{cases}$$

$$e_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

$$e_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

The spectral decomposition of A has form:

$$A = P\Lambda P^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

We see that the columns of matrix P are subsequent eigenvectors of matrix A: $P = [e_1, e_2]$, what is more $\Lambda = \text{diag}(\lambda_1, \lambda_2)$.

We can also write matrix **A** as:

$$\begin{aligned} \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T &= \frac{1}{2} \left(2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} + 4 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \right) \\ &= \frac{1}{2} \left(\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix} \right) \\ &= \frac{1}{2} \left(\begin{bmatrix} 6 & -2 \\ -2 & 6 \end{bmatrix} \right) \\ &= \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}. \end{aligned}$$

To find \sqrt{A} , we use the spectral decomposition of A :

$$\begin{aligned}\sqrt{A} &= P\sqrt{\Lambda}P^T \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1+\sqrt{2}}{2} & \frac{\sqrt{2}-2}{2} \\ \frac{\sqrt{2}-2}{2} & \frac{1+\sqrt{2}}{2} \end{bmatrix}.\end{aligned}$$

Using spectral decomposition we can prove that $A = \sqrt{A}\sqrt{A}$:

$$\sqrt{A}\sqrt{A} = P\sqrt{\Lambda}P^T P\sqrt{\Lambda}P^T = P\sqrt{\Lambda}\mathbb{I}\sqrt{\Lambda}P^T = P\sqrt{\Lambda}\sqrt{\Lambda}P^T = P\Lambda P^T = A.$$

Properties of the spectral decomposition

Let's consider the spectral decomposition of a positive definite matrix as given in Lecture 1:

$$A = P\Lambda P^T,$$

where the columns of P are made of eigenvectors e_i , $i = 1, \dots, n$ ($P = [e_1, \dots, e_n]$, $e_i = [e_{1,i}, \dots, e_{n,i}]^T$) and they are orthonormalized, i.e. their lengths are one and they are orthogonal (perpendicular) one to another. The diagonal matrix Λ has the corresponding (positive) eigenvalues on the diagonal ($\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$).

Properties:

1. $P^T = P^{-1}$, because:

$$\begin{aligned}P^T &= P^{-1} & | \cdot_R P \\ P^T P &= \mathbb{I} \\ \begin{bmatrix} e_{1,1} & \cdots & e_{n,1} \\ \vdots & \ddots & \vdots \\ e_{1,n} & \cdots & e_{n,n} \end{bmatrix} \begin{bmatrix} e_{1,1} & \cdots & e_{1,n} \\ \vdots & \ddots & \vdots \\ e_{n,1} & \cdots & e_{n,n} \end{bmatrix} &= \mathbb{I} \\ \begin{bmatrix} e_1^T e_1 & e_1^T e_2 & \cdots & e_1^T e_n \\ e_2^T e_1 & e_2^T e_2 & \cdots & e_2^T e_n \\ \vdots & \vdots & \ddots & \vdots \\ e_n^T e_1 & e_n^T e_2 & \cdots & e_n^T e_n \end{bmatrix} &= \mathbb{I},\end{aligned}$$

where $\forall_{i \in \{1, \dots, n\}} e_i^T e_i = 1$ and $\forall_{i \in \{1, \dots, n\}} \forall_{j: j \neq i} e_i^T e_j = 0$.

For example:

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^{-1} = \frac{1}{-1} \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T.$$

2. $\det(\Lambda) = \prod_{i=1}^n \lambda_i$:

$$\det \left(\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \right) = \lambda_1 \det \left(\begin{bmatrix} \lambda_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \right) = \lambda_1 \lambda_2 \det \left(\begin{bmatrix} \lambda_3 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \right) = \cdots = \lambda_1 \lambda_2 \cdots \lambda_n$$

For example:

$$\det \left(\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \right) = 2 \cdot 4 - 0 \cdot 0 = 8.$$

3. $\det(A) = \det(\Lambda)$:

$$\det(A) = \det(P\Lambda P^T) = \det(P) \det(\Lambda) \det(P^T) = \det(\Lambda) \det(PP^T) = \det(\Lambda \mathbb{I}) = \det(\Lambda).$$

For example:

$$\begin{aligned} \det\left(\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}\right) &= \det\left(\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}\right) \det\left(\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\right) \det\left(\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}\right) \\ &= \det\left(\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}\right) \det\left(\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\right) \\ &= \det(\mathbb{I}) \det\left(\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}\right) \end{aligned}$$

4. Inverse matrix to Λ :

$$\Lambda^{-1} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{\lambda_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_n} \end{bmatrix}$$

For example:

$$\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{bmatrix}.$$

5. A simple way to determine the inverse of a matrix A from its spectral decomposition is through $A^{-1} = P\Lambda^{-1}P^T$:

$$\begin{aligned} A^{-1} &= P\Lambda^{-1}P^T \quad | \cdot_L A \\ AA^{-1} &= P\Lambda P^T P\Lambda^{-1}P^T \\ \mathbb{I} &= P\Lambda\mathbb{I}\Lambda^{-1}P^T \\ \mathbb{I} &= PP^T \\ \mathbb{I} &= \mathbb{I}. \end{aligned}$$

For example:

$$\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{3}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{8} \end{bmatrix}.$$

Then:

$$\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} \frac{3}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{8} \end{bmatrix} = \mathbb{I}.$$

The join distribution of length and weight of newborn children

In a medical study, length L and weight W of newborn children is considered. It was assumed that (L, W) will be modeled through a bivariate normal distribution. The following information has been known: the mean weight is 3343[g], with the standard deviation of 528[g], while the mean length is 49.8[cm], with the standard deviation of 2.5[cm]. Additionally the correlation between the length and the weight has been established and equal to 0.75.

The join distribution of (W, L) is bivariate normal, i.e.

$$\begin{bmatrix} W \\ L \end{bmatrix} \sim N\left(\begin{bmatrix} 3343 \\ 49.8 \end{bmatrix}, \begin{bmatrix} 278784 & 990 \\ 990 & 6.25 \end{bmatrix}\right).$$

The density of the above vector has form:

$$\begin{aligned} f(X) &= \frac{1}{2\pi} \det \left(\begin{bmatrix} 278784 & 990 \\ 990 & 6.25 \end{bmatrix} \right)^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \left(X - \begin{bmatrix} 3343 \\ 49.8 \end{bmatrix} \right)^T \begin{bmatrix} 278784 & 990 \\ 990 & 6.25 \end{bmatrix}^{-1} \left(X - \begin{bmatrix} 3343 \\ 49.8 \end{bmatrix} \right) \right) \\ &= \frac{\sqrt{7}}{4620\pi} \exp \left(-\frac{1}{2} \left(X - \begin{bmatrix} 3343 \\ 49.8 \end{bmatrix} \right)^T \begin{bmatrix} \frac{1}{121968} & \frac{-1}{770} \\ \frac{-1}{770} & \frac{1}{175} \end{bmatrix} \left(X - \begin{bmatrix} 3343 \\ 49.8 \end{bmatrix} \right) \right) \end{aligned}$$

Let's find the eigenvalues and eigenvectors of the covariance matrix Σ :

$$\det \left(\begin{bmatrix} 278784 - \lambda & 990 \\ 990 & 6.25 - \lambda \end{bmatrix} \right) = \lambda^2 - 278790.25\lambda + 762300,$$

$\lambda^2 - 278790.25\lambda + 762300 = 0$ for $\lambda_1 = 2.734$ and $\lambda_2 = 278787.516$. The eigenvectors are $e_1 = [0.004, -1]^T$ and $e_2 = [-1, -0.004]^T$.

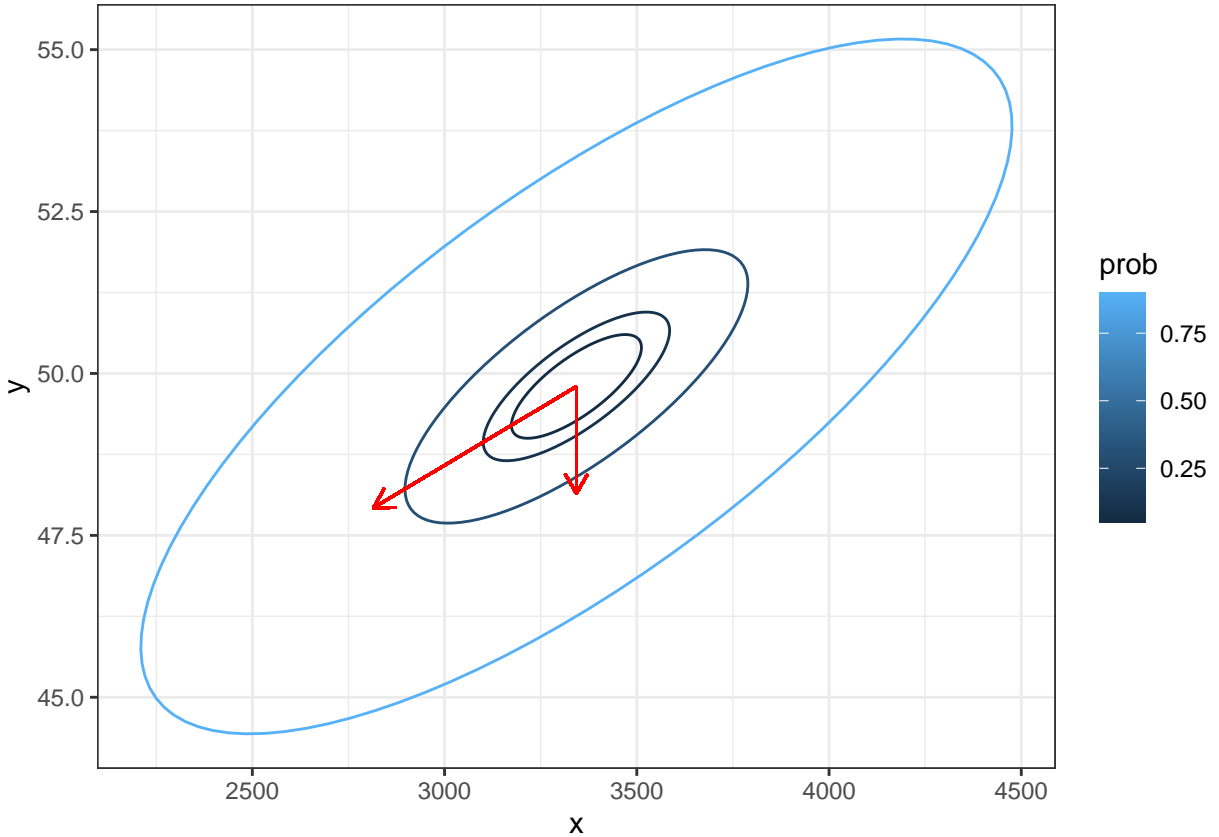


Figure 1: Ellipses corresponding to the constant density contours of the join distribution W and L

The vectors in the above graph are perpendicular, but the differences in the scales of the x and y axes mean that we cannot see this.

The bivariate normal distribution is characterized by 5 parameters (2 means, 2 standard deviations and covariance between variables). The p -dimensional normal distribution is characterized by $p + p(p+1)/2$ (p means, p variances, $p(p-1)/2$ covariances).

Variable of length: L, is normal distributed: $L \sim N(49.8, 2.5)$.

The 3- σ rule tells us that the variable L belongs to the interval $[49.8 - 3 \cdot 2.5, 49.8 + 3 \cdot 2.5] = [42.3, 57.3]$ with a probability of 99.73%.

When the child's weight is known ($W = 4025$), the variable L has a conditional normal distribution.

$$\begin{aligned} L|W &\sim N\left(49.8 + 2.5 \frac{4025 - 3343}{528} \cdot 0.75, 6.25\left(1 - \frac{9}{16}\right)\right) \\ &\sim N\left(\frac{16711}{320}, \frac{175}{64}\right) \end{aligned}$$

Now the 99.73% CI for L is $[44.019, 60.425]$. Additional information about the child's weight should have narrowed the CI, but it did not do so in this case. The reason for this is that $L|W$ has a larger variance.

Operations on a multivariate normal distribution

Let X_1 , X_2 , and X_3 be independent $N(\mu, \Sigma)$ random vectors of a dimension p . We have two vectors:

$$V_1 = \frac{1}{4}X_1 - \frac{1}{2}X_2 + \frac{1}{4}X_3,$$

$$V_2 = \frac{1}{4}X_1 - \frac{1}{2}X_2 - \frac{1}{4}X_3.$$

These vectors has distributions:

$$V_1 \sim N\left(\frac{1}{4}\mu - \frac{1}{2}\mu + \frac{1}{4}\mu, \frac{1}{16}\Sigma + \frac{1}{4}\Sigma + \frac{1}{16}\Sigma\right) =: N\left(0, \frac{3}{8}\Sigma\right),$$

$$V_2 \sim N\left(\frac{1}{4}\mu - \frac{1}{2}\mu - \frac{1}{4}\mu, \frac{1}{16}\Sigma + \frac{1}{4}\Sigma + \frac{1}{16}\Sigma\right) =: N\left(-\frac{1}{2}\mu, \frac{3}{8}\Sigma\right).$$

The join distribution of $V = [V_1, V_2]^T$ is $2p$ dimensional multivariate normal distribution $N(\mu_V, \Sigma_V)$. We see that:

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}.$$

So:

$$\begin{aligned} \mu_V &= \begin{bmatrix} \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{2}\mu \end{bmatrix} \\ \Sigma_V &= \begin{bmatrix} \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{2} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma & 0 \\ 0 & 0 & \Sigma \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{3}{8}\Sigma & \frac{1}{4}\Sigma \\ \frac{1}{4}\Sigma & \frac{3}{8}\Sigma \end{bmatrix}. \end{aligned}$$

Analysis of weight and length data of newborn children

Healthcare providers and insurance companies aim to determine the necessary medical tests for newborns. They've devised a scoring system to identify whether a newborn requires special attention or if they're healthy and don't need extra care. This scoring system primarily relies on the newborn's weight and length, as these are reliable indicators of their health status.

Here's how the scoring works: If a newborn's weight or length is exceptionally high or low, falling outside the top or bottom 5% of typical values for newborns, they're given a score of zero. This suggests that they might need additional medical tests. If their measurements fall between the top 5% and top 25%, they receive a score of one, indicating they might need some extra attention. Otherwise, if their weight and length are within the range of what's typical for newborns (the bottom 75%), they're assigned a score of two, indicating good health.

To analyze this, we've gathered data from 736 recently born children in a specific region. While their records contain various information, we're particularly interested in their weight and length, which we'll use to determine how many newborns may require extra medical attention and how many are healthy without further tests. This data is extracted from the file *WeightLength.txt*.

Let's assume vector $X = [W, L]^T$, where the variable W determines the weight [g] and L - the length [cm] of the child.. We know that vector X comes from a two-dimensional normal distribution $N(\mu, \Sigma)$. The estimators of the parameters of this distribution are:

$$\hat{\mu} = [3233.545, 49.237]^T,$$

$$\hat{\Sigma} = \begin{bmatrix} 220276.658 & 915.296 \\ 915.296 & 4.443 \end{bmatrix}.$$

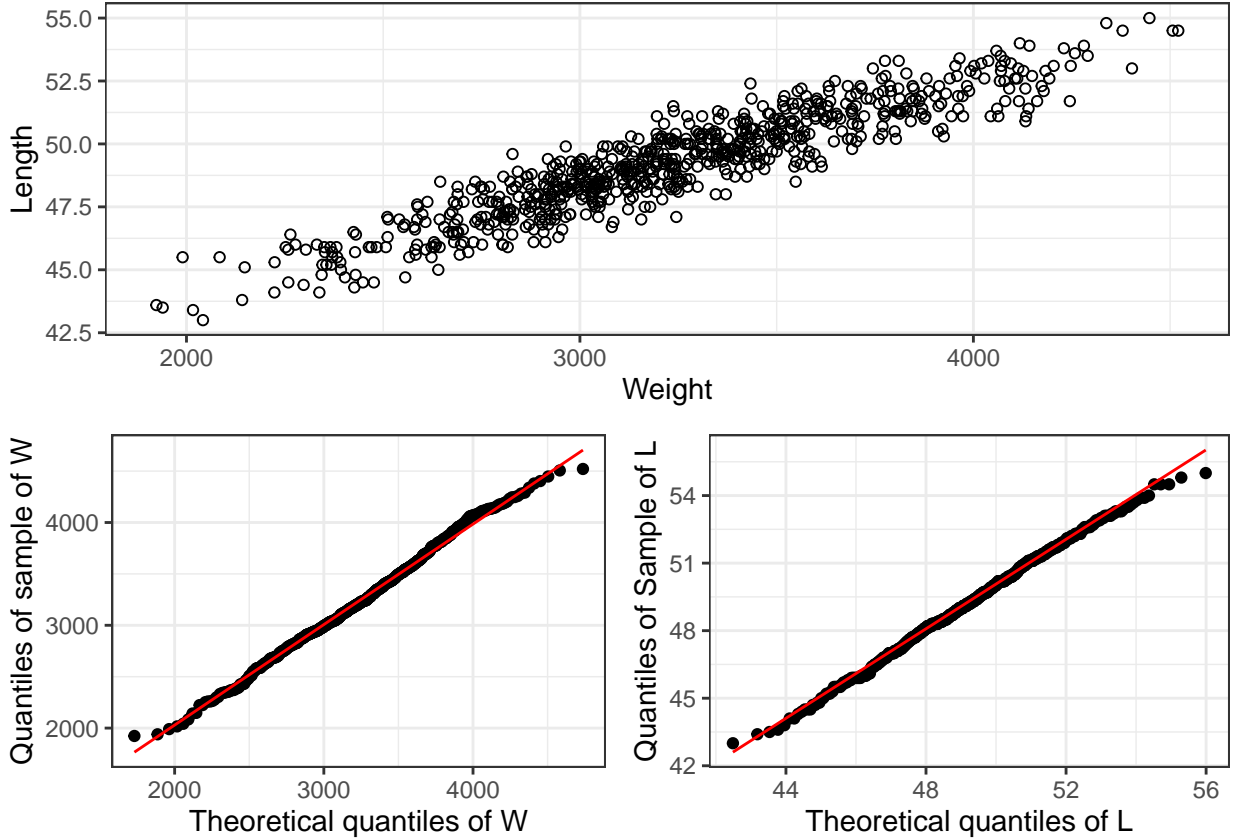


Figure 2: Scatterplot and qq-plots for the marginal distributions of data.

In the scatterplot we see that the two-dimensional variables $X = [W, L]^T$ have a multivariate normal distribution and are positively correlated with each other. Quantile-quantile charts prove that marginal distributions are normal distributions: $W \sim N(3233.545, 220276.658)$, $L \sim N(49.237, 4.443)$.

In order to assign points to children, we will use the construction . The k th child will receive 2 points if the corresponding vector $X_k = [Weight_k, Length_k]^T$ satisfies the inequality:

$$(X_k - \hat{\mu})^T \hat{\Sigma}^{-1} (X_k - \hat{\mu}) \leq \chi_2^2(0.75).$$

Similarly, we will award point 1:

$$\chi_2^2(0.75) \leq (X_k - \hat{\mu})^T \hat{\Sigma}^{-1} (X_k - \hat{\mu}) \leq \chi_2^2(0.95).$$

From the above it follows that the number of children who received 2 points is 541, 1 point: 157, 0 points: 38.

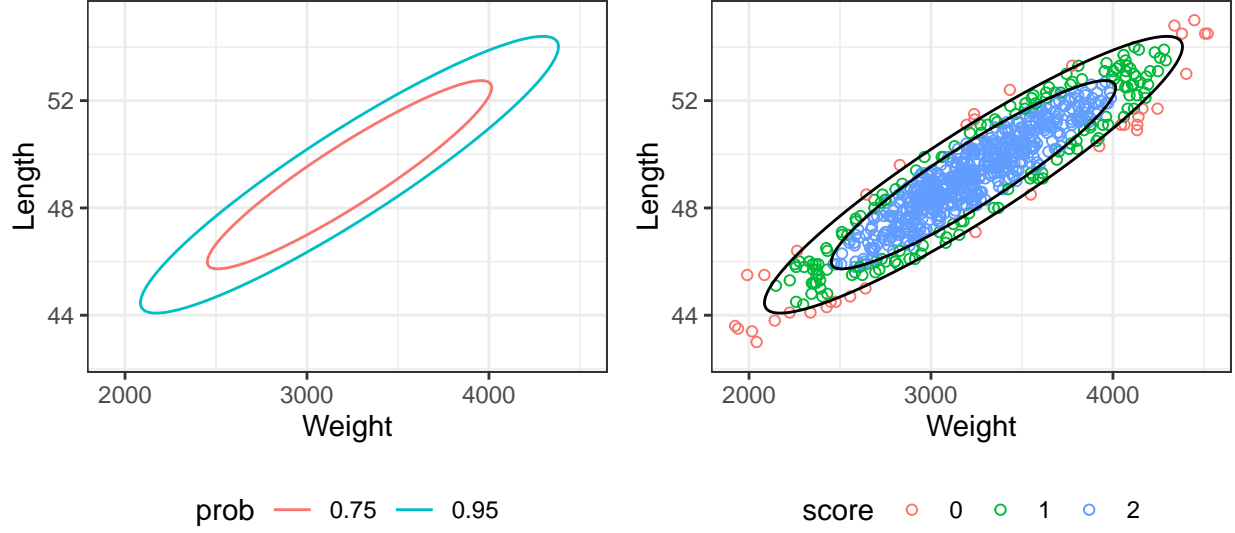


Figure 3: Plot of confidence ellipsoids and scatter plot with classification region for Weight and Length of newborn children.

Let's find the spectral decomposition of the $\hat{\Sigma}$. The eigenvalues of the matrix $\hat{\Sigma}$ are $\lambda_1 = 0.64$ $\lambda_2 = 220280.5$ and their corresponding eigenvectors, respectively: $v_1 = [0.004, -1]^T$, $v_2 = [-1, -0.004]^T$. The spectral decomposition of $\hat{\Sigma}$ has form:

$$\hat{\Sigma} = P\Lambda P^T = \begin{bmatrix} 0.004 & -1 \\ -1 & -0.004 \end{bmatrix} \begin{bmatrix} 0.64 & 0 \\ 0 & 220280.5 \end{bmatrix} \begin{bmatrix} 0.004 & -1 \\ -1 & -0.004 \end{bmatrix}$$

Let's transform our data by left-multiplying each vector $[W, L]$ by P^T .

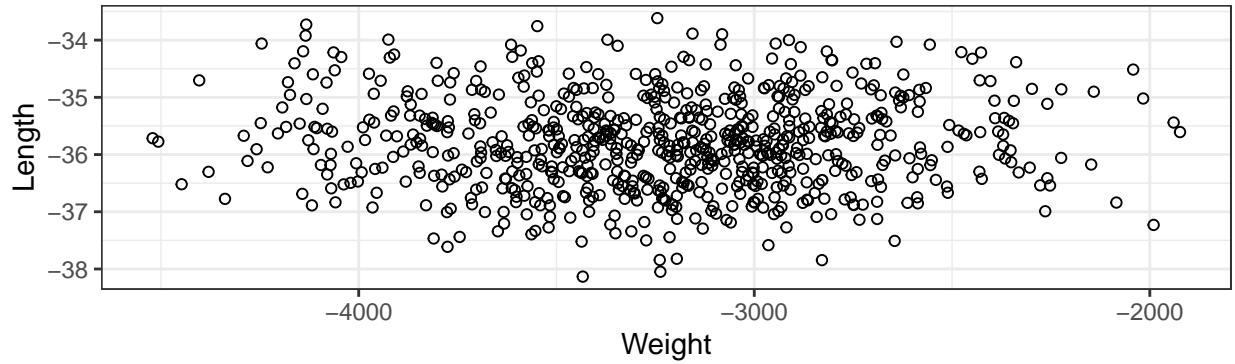


Figure 4: Scatterplot of transformed data.

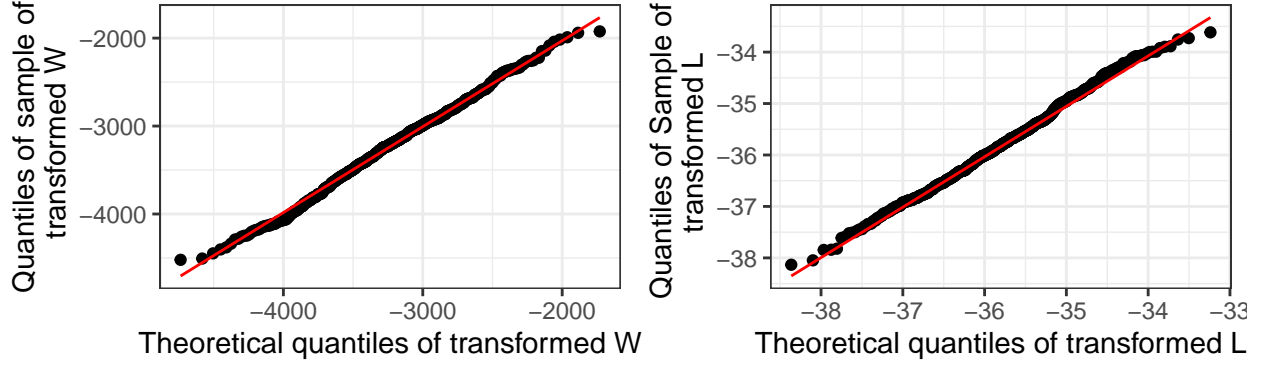


Figure 5: Qq-plots for the marginal distributions of transformed data.

We see that the data comes from a bivariate normal distribution (more precisely $N\left(P^T\hat{\mu}, P^T\hat{\Sigma}P\right)$). The transformation reduced the covariances between the variable W and L , we can see in the scatterplot that the points do not accumulate along any straight line. Moreover, the transformation changed the values and the sign of the data.

Analysis of weight and length data of newborn children and their parents' heights

Let's analyze the children's previous data with additional data including the heights of their mothers and fathers (from the file *ParentsWeightLength.txt*). Let's also check what impact additional data has on children's scores.

Let the vector $X'_i = [W_i, L_i, F_i, M_i]^T$ correspond to the i -th record from the data set, where the variables W_i and L_i correspond, as before, to the weight and length of the i -th child, and F_i and M_i correspond to the heights of the father and mother (respectively) of the i th child.

Vector $X'_i \sim N\left(\hat{\mu}', \hat{\Sigma}'\right)$, where:

$$\hat{\mu}' = \begin{bmatrix} 3233.545 \\ 49.238 \\ 177.416 \\ 166.920 \end{bmatrix},$$

$$\hat{\Sigma}' = \begin{bmatrix} 220276.658 & 915.296 & 931.859 & 827.288 \\ 915.296 & 4.443 & 3.290 & 2.852 \\ 931.859 & 3.290 & 12.612 & 0.631 \\ 827.288 & 2.852 & 0.631 & 9.7721065 \end{bmatrix}.$$

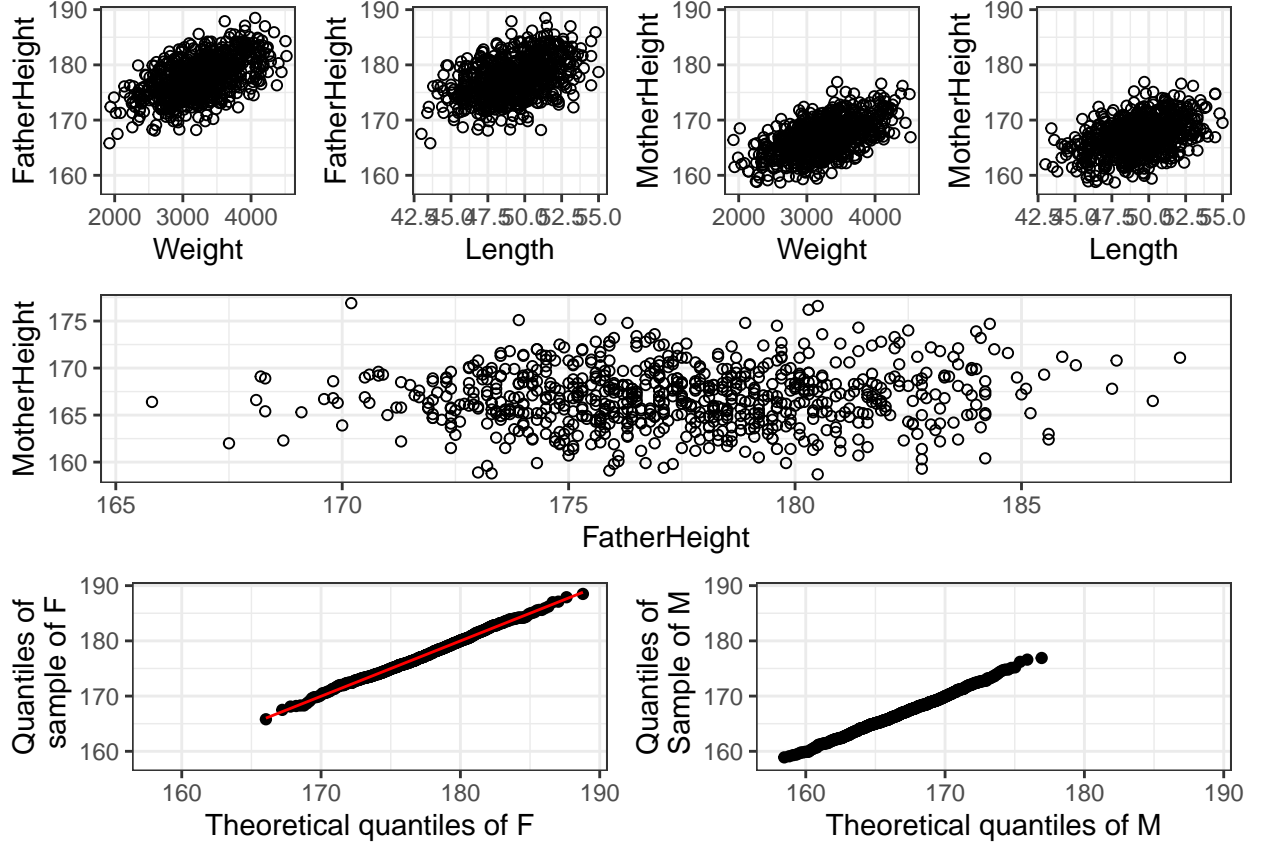


Figure 6: Scatterplot of new data and qq-plots of Heights of parents.

The scatterplots above confirm that the data have a four-dimensional normal distribution. The first 4 graphs show that there is a relationship between the child's metrics and the parent's height. The fifth chart shows that there is no relationship between the heights of both parents. We also see that qq-plots confirm that adult height is also normally distributed.

The conditional distribution of the weight and length of a child given the heights of parents has form $X|Y = y \sim N(\hat{\mu}^*, \hat{\Sigma}^*)$, where $Y_i = [F_i, M_i]^T$,

$$\begin{aligned}
 \hat{\mu}^* &= \begin{bmatrix} 3233.545 \\ 49.238 \end{bmatrix} + \begin{bmatrix} 931.859 & 827.288 \\ 3.290 & 2.852 \end{bmatrix} \begin{bmatrix} 0.0795 & -0.005 \\ -0.005 & 0.103 \end{bmatrix} \left(y - \begin{bmatrix} 177.416 \\ 166.920 \end{bmatrix} \right) \\
 &= \begin{bmatrix} 3233.545 \\ 49.238 \end{bmatrix} + \begin{bmatrix} 69.876 & 80.146 \\ 0.247 & 0.276 \end{bmatrix} \left(y - \begin{bmatrix} 177.416 \\ 166.920 \end{bmatrix} \right) \\
 \hat{\Sigma}^* &= \begin{bmatrix} 220276.658 & 915.296 \\ 915.296 & 4.443 \end{bmatrix} - \begin{bmatrix} 931.859 & 827.288 \\ 3.290 & 2.852 \end{bmatrix} \begin{bmatrix} 0.0795 & -0.005 \\ -0.005 & 0.103 \end{bmatrix} \begin{bmatrix} 931.859 & 3.290 \\ 827.288 & 2.852 \end{bmatrix} \\
 &= \begin{bmatrix} 88857.996 & 456.846 \\ 456.846 & 2.844 \end{bmatrix}.
 \end{aligned}$$

The diameters of the ellipses will be determined by the eigenvectors of the matrix $\hat{\Sigma}^*$, and the center of the ellipses will be the vector $\hat{\mu}^*$. The ellipsoids based on the conditional distribution they are narrower and smaller, i.e. they have a smaller area compared to the previous ones. The new ellipse has also been rotated.

This makes sense because new data gives us more information about the distribution of given metrics.

Let's assume $y = \hat{\mu}^*$, then the classification areas are shown by first below plot:

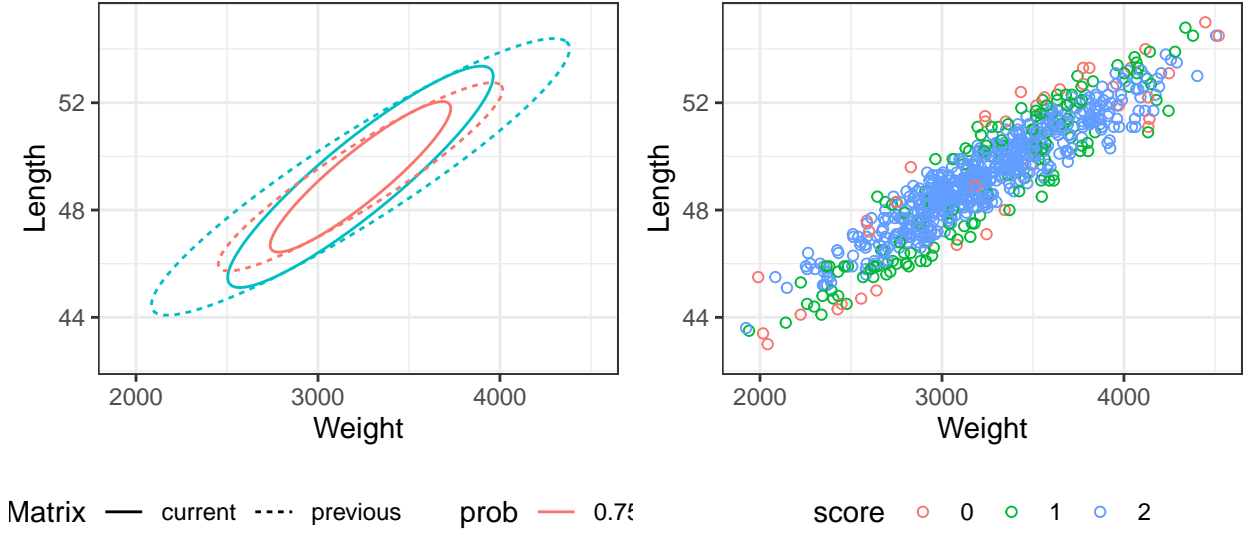


Figure 7: Plot of confidence ellipsoids and scatter plot with classification region for Weight and Length of newborn children under conditional distribution.

We can see that the areas have decreased significantly, but the scoring system worked better. In this case we award the k th child 2 points if:

$$(X_k - \hat{\mu}^*)^T \hat{\Sigma}^{*-1} (X_k - \hat{\mu}^*) \leq \chi_2^2(0.75),$$

1 point if:

$$\chi_2^2(0.75) \leq (X_k - \hat{\mu}^*)^T \hat{\Sigma}^{*-1} (X_k - \hat{\mu}^*) \leq \chi_2^2(0.95),$$

else 0. In this case, there are 40 children with a score of 0, 137 with a score of 1, and 559 with a score of 2. Confidence ellipses are not shown in a scatter plot showing classified points because for each point the pair of ellipses has a different center.

Below plot shows classification ellipsoids when $y = [185, 178]^T$.

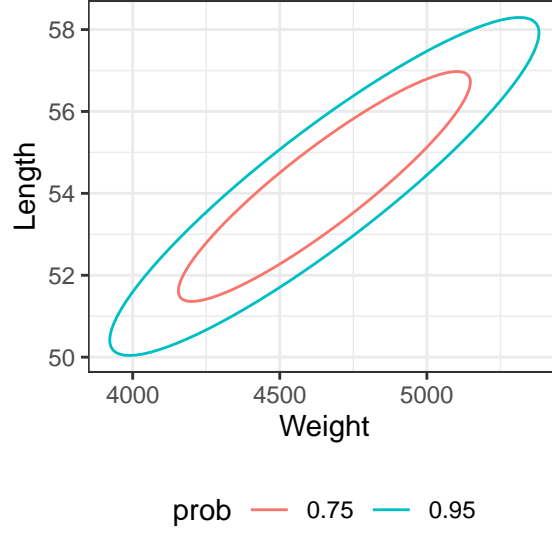


Figure 8: Classification region for Weight and Length of newborn children, when father of a child is 185[cm] tall and mother is 178[cm] tall.

The eigenvalues of the matrix $\hat{\Sigma}'$ are $\lambda'_1 = 0.478$, $\lambda'_2 = 4.788$, $\lambda'_3 = 10.708$, $\lambda'_4 = 22.029$ and their corresponding eigenvectors, respectively: $v'_1 = [0.005, -0.981, -0.122, -0.149]^T$, $v'_2 = [0.005, 0.193, -0.564, -0.803]^T$, $v'_3 = [0.001, 0.014, -0.816, 0.577]^T$, $v'_4 = [1, 0.004, 0.004, 0.004]^T$. The spectral decomposition of $\hat{\Sigma}$ has form:

$$\begin{aligned} \hat{\Sigma} &= P' \Lambda' P'^T \\ &= \begin{bmatrix} 0.005 & 0.005 & 0.001 & 1 \\ -0.981 & 0.193 & 0.0137 & 0.004 \\ -0.122 & -0.564 & -0.816 & 0.004 \\ -0.149 & -0.803 & 0.577 & 0.004 \end{bmatrix} \begin{bmatrix} 0.478 & 0 & 0 & 0 \\ 0 & 4.788 & 0 & 0 \\ 0 & 0 & 10.708 & 0 \\ 0 & 0 & 0 & 22.029 \end{bmatrix} \begin{bmatrix} 0.005 & -0.981 & -0.122 & -0.149 \\ 0.005 & 0.193 & -0.564 & -0.803 \\ 0.001 & 0.014 & -0.816 & 0.577 \\ 1 & 0.004 & 0.004 & 0.004 \end{bmatrix} \\ &= \begin{bmatrix} 220276.658 & 915.296 & 931.859 & 827.288 \\ 915.296 & 4.443 & 3.290 & 2.852 \\ 931.859 & 3.290 & 12.612 & 0.631 \\ 827.288 & 2.852 & 0.631 & 9.772 \end{bmatrix}. \end{aligned}$$

Let's transform the data as before, multiplying the data matrix on the left by P'^T .

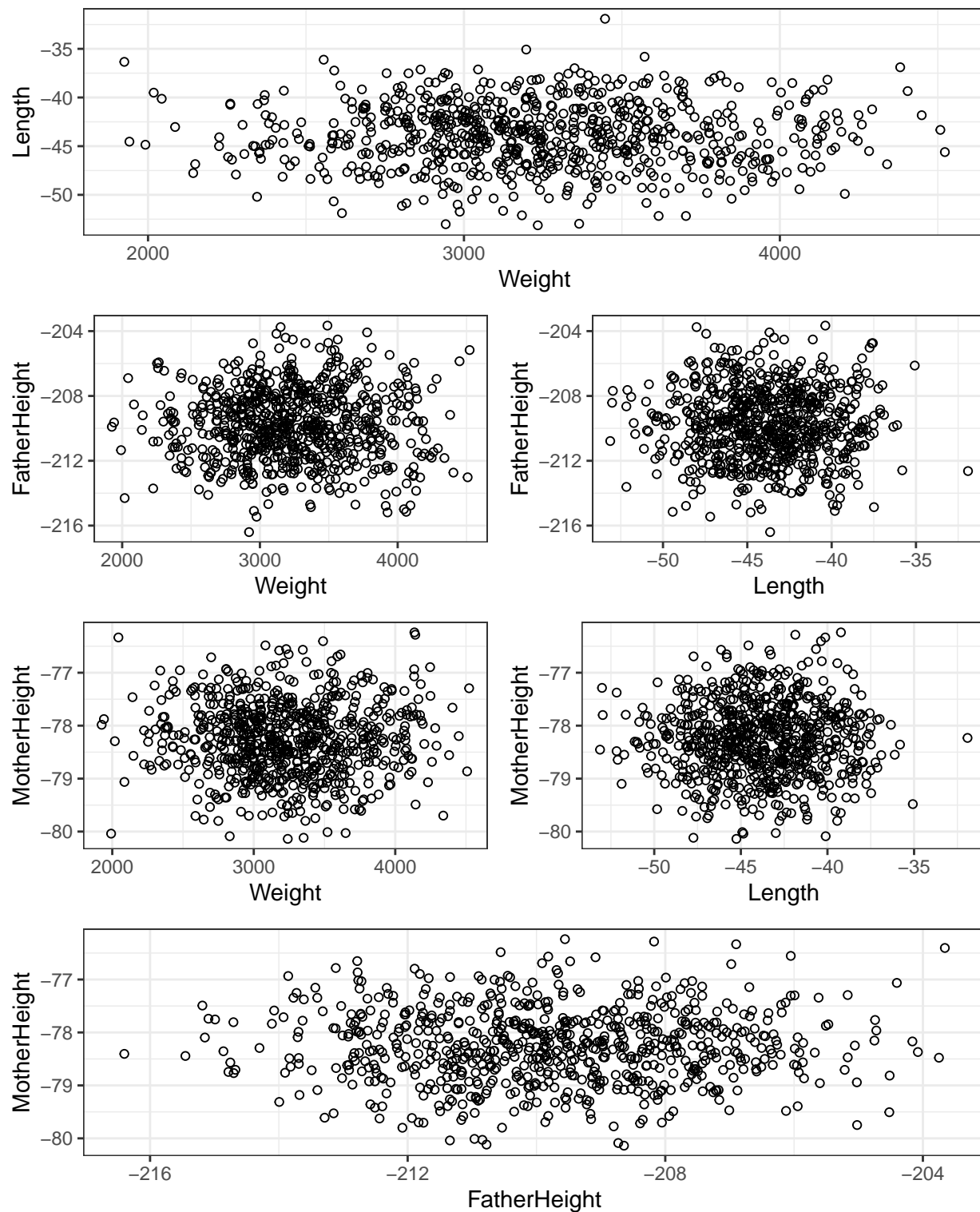


Figure 9: Scatterplot of transformed data (data of children and parents).

As before, the transformation reduced the dependency between the variables. In each of the above graphs, the points are arranged in groups that are not clustered around any straight lines.