Statistical learning
Assignment 1

Jakub Skalski

March 20, 2024

1. Problem 1

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

1.1 Is A symmetric?

A is symmetric since $A = A^T$.

1.2 Perform the spectral decomposition of A

By solving $det(A - \lambda I) = 0$, we obtain $\lambda_1 = 2, \lambda_2 = 4$ and their corresponding eigenvectors $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. And so the decomposition is such that $P = \frac{1}{\sqrt{2}} \begin{bmatrix} v_1 & v_2 \end{bmatrix}$ and $\Lambda = diag(\lambda_1, \lambda_2)$.

1.3 Alternative form of spectral decomposition of A

$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}$

1.4 Finding $\sqrt{A}$ through spectral decomposition

Using the explicit formula $\sqrt{A} = P\Lambda P^T$ we obtain $\sqrt{A} = \frac{1}{2} \begin{bmatrix} \sqrt{2}+2 & \sqrt{2}-2 \\ \sqrt{2}-2 & \sqrt{2}+2 \end{bmatrix}$.

2. Problem 2

2.1 The inverse of P is P^T

Since $e_i^T e_i = 1$ and $e_i^T e_j = 0$ for $i \neq j$

$$P^T P = \begin{bmatrix} e_1^T \\ - \\ ... \\ - \\ e_n^T \end{bmatrix} \begin{bmatrix} e_1 & | & ... & | & e_n \end{bmatrix} = \begin{bmatrix} e_1^T e_1 & ... & e_1^T e_n \\ ... & ... & ... \\ e_n^T e_1 & ... & e_n^T e_n \end{bmatrix} = I$$

2.2 Determinant of Λ is equal to the product of the terms on the diagonal

We can prove this by induction. Assume $det(\lambda_n) = \prod_i^n \lambda_i$ and show that it holds for $n+1$ by $n+1$ matrix too. To that end we perform laplace expansion on the n+1-th row:

$$det(\Lambda_{n+1}) = det(\Lambda_n) * \lambda_{n+1} + 0 + ... + 0$$

Therefore, by our inductive hypothesis, $det(\Lambda_{n+1}) = \prod_i^{n+1} \lambda_i$.

2.3 Dederminant of A is the same as that of Λ

Using the fact 2. and $det(P) = det(P^T) = \pm 1$ we have:

$$det(A) = det(P \Lambda P^T) = det(P) det(\Lambda) det(P^T) = det(\Lambda)$$

2.4 Finding inverse of Λ

Let $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_n)$ then $\Lambda^{-1} = diag(\lambda_1^{-1}, \lambda_2^{-1}, ..., \lambda_n^{-1})$.

2.5 Finding A^{-1} through spectral decomposition

It suffices to show that $AA^{-1} = I$:

$$AA^{-1} = (P \Lambda P^T)(P \Lambda^{-1} P^T) = P \Lambda \Lambda^{-1} P^T = PP^T = I$$

3. Problem 3

3.1 Distribution parameters

Distribution is parameterized with a vector mean $\boldsymbol{\mu} = \begin{bmatrix} 3343 \\ 49.8 \end{bmatrix}$ and a covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} 528^2 & 990 \\ 990 & 6.25 \end{bmatrix}$.

3.2 Density function

The joint probability density function of p normal distributions with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is given by:

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

3.3 Eigenvalues and eigenvectors of Σ.

$$P = \begin{bmatrix} -0.999993695 & 0.003551149 \\ -0.003551149 & -0.999993695 \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 2.787875e+05 \\ 2.734341e+00 \end{bmatrix}$$

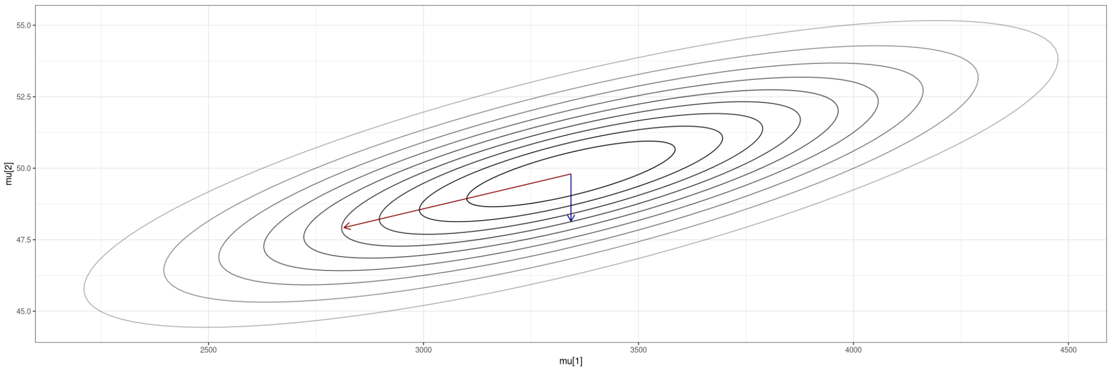

Figure 1: Joint distribution density

The drawn ellipses are characterized by the previous eigen decomposition. Their axes point in the direction of eigenvectors which are scaled by their eigenvalues. Variance of the marginal distributions dictates their stretch along the respective axis. The strong skew of the ellipses indicates a high correlation coefficient.

3.4 Parameter count

Bivariate normal distribution is characterized by $\frac{p(p+3)}{2}$ parameters.

3.5 Marginal distribution

$L \sim N(49.8, 6.25)$

3.6 Best child's length guess

Mean (49.8cm) would be the best guess for the child's length. With 68% chance the error of this prediction is within one standard deviation (2.5cm).

4. Problem 4

4.1 Conditional distribution

The conditional distribution of L given W is the following:

$$L|W = w \sim \mathcal{N}(\mu_L + \rho\frac{\sigma_L}{\sigma_W}(w - \mu_W), \sigma_L^2(1 - \rho^2))$$

Assuming $w = 4025$ we calculate $L|W = w \sim \mathcal{N}(52.22187, 2.734375)$.

4.2 Improving on the previous guess

Mean (52.22187cm) is still the best guess for the child's length and with 68% confidence the error of this prediction is within one standard deviation of 1.65359456941537. We observe that the standard deviation is now smaller compared to the previous estimate, meaning the guess is now more accurate.

5. Problem 5

5.1 Definining variables

$X_i \sim \mathcal{N}(\mu_{(i)}, \Sigma_{(i)})$
$V_i \sim \mathcal{N}(\mu_i, \Sigma_i)$

$$(V_1, V_2) \sim \mathcal{N}(\mu, \Sigma) = \mathcal{N}(\begin{bmatrix}\mu_1 \\ \mu_2\end{bmatrix}, \begin{bmatrix}\Sigma_1 & \Sigma_{12} \\ \Sigma_{21} & \Sigma_2\end{bmatrix})$$

5.2 Solution

$\mu_1 = \frac{1}{4}\mu_{(1)} - \frac{1}{2}\mu_{(2)} + \frac{1}{4}\mu_{(3)}$
$\mu_2 = \frac{1}{4}\mu_{(1)} - \frac{1}{2}\mu_{(2)} - \frac{1}{4}\mu_{(3)}$
$\Sigma_1 = \frac{1}{4^2}\Sigma_{(1)} - \frac{1}{2^2}\Sigma_{(2)} + \frac{1}{4^2}\Sigma_{(3)}$
$\Sigma_2 = \frac{1}{4^2}\Sigma_{(1)} - \frac{1}{2^2}\Sigma_{(2)} - \frac{1}{4^2}\Sigma_{(3)}$
$\Sigma_{12} = \Sigma_{21}^T = E[V_1^T V_2] - E[V_1^T]E[V_2] = \frac{1}{4^2}\Sigma_{(1)} - \frac{1}{2^2}\Sigma_{(2)} - \frac{1}{4^2}\Sigma_{(3)}$

6. Project

6.1 Part One

6.1.1 Using the date estimate the mean and the covariance for the length and the weight of children

Parameter	Value
μ_W	3233.545
μ_L	49.23764
σ_{WL}	915.2955

Table 1: Estimated distribution parameters

6.1.2 Verify graphically normal distribution of the data

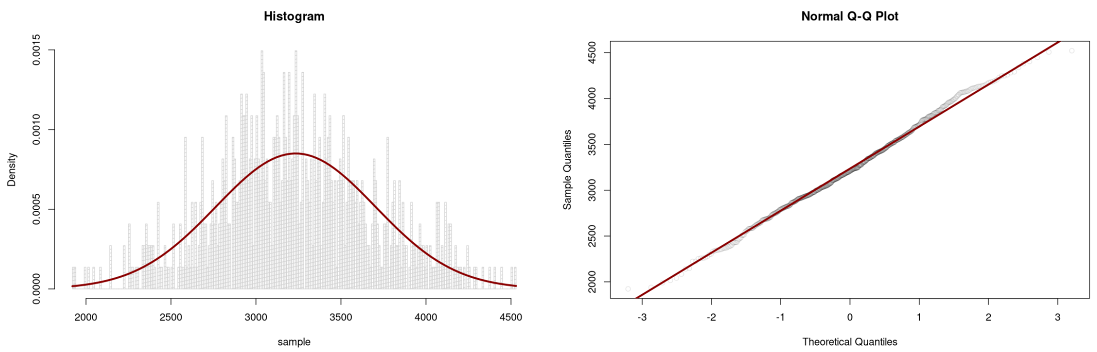

Figure 2: Distribution of weight

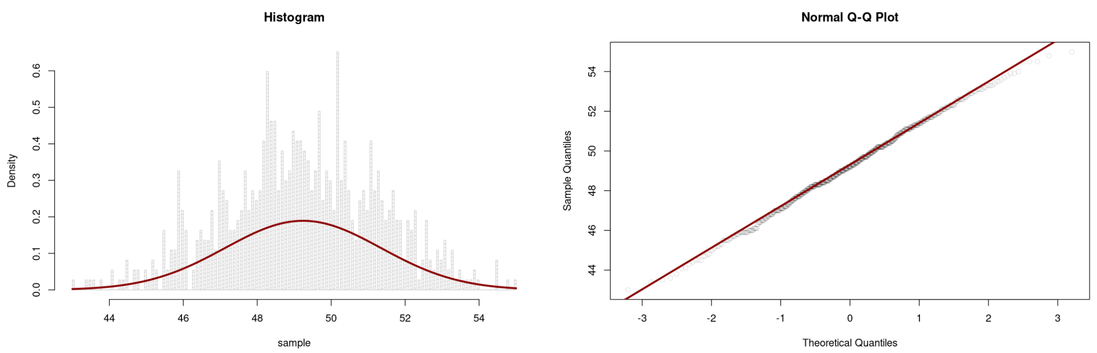

Figure 3: Distribution of height

6.1.3 Find classification regions

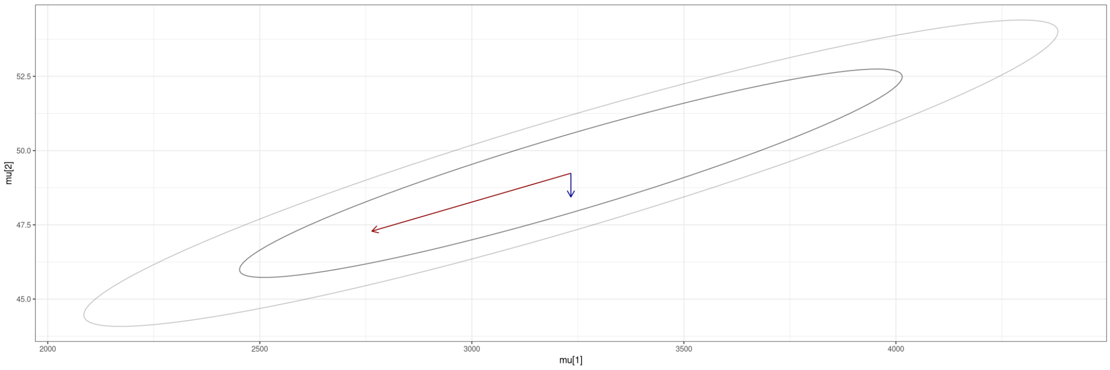

Figure 4: Classification regions

6.1.4 Scores and classification plot

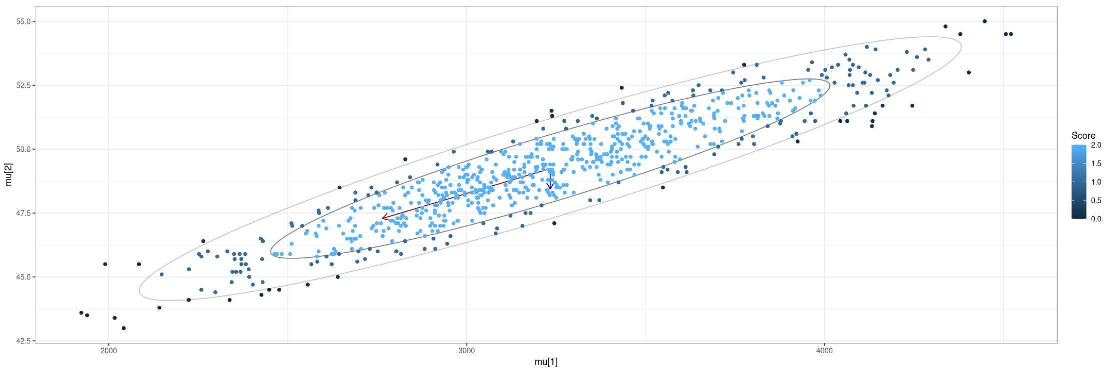

Figure 5: Classification plot

Score	Number of children
0	38
1	157
2	541

Table 2: Scores

6.1.5 Spectral decomposition of estimated covariance matrix

$$P = \begin{bmatrix} -0.9999914 & -0.004155185 \\ 0.004155185 & -0.9999914 \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 220280.5 \\ 0.6400477 \end{bmatrix}$$

6.1.6 Plot the transformed data

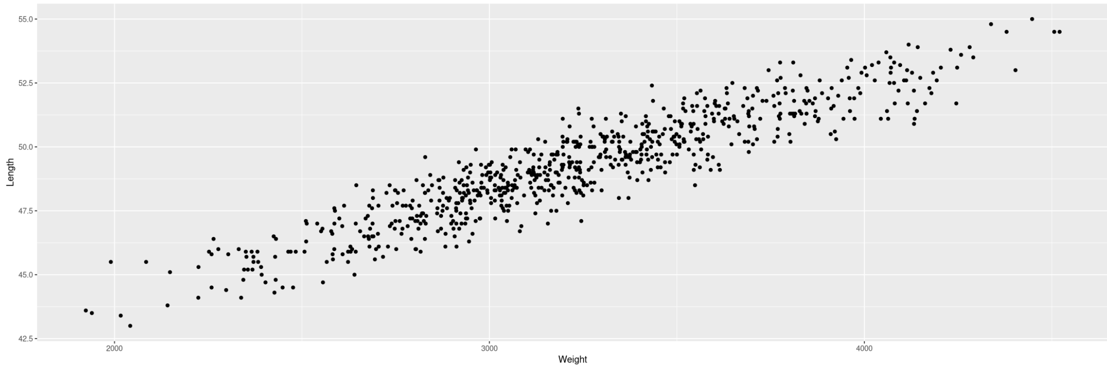

Figure 6: Original data

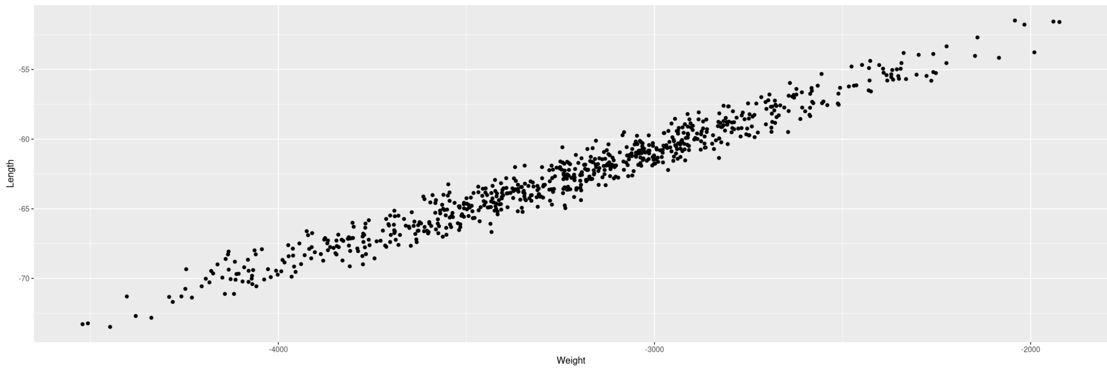

Figure 7: Transformed data

It looks like the transformation reduced the variance of one of the marginal distributions. The vector mean was also not preserved.

6.2 Part Two

6.2.1 Using the date estimate the mean and the covariance for the length and the weight of children

Variable	Mean
Father height	177.416
Mother height	166.9196
Weight	3233.545
Length	49.23764

Table 3: Vector mean

Variable	Father height	Mother height	Weight	Length
Father height	12.6121	0.631	931.859	3.2895
Mother height	0.631	9.7721	827.2878	2.8521
Weight	931.859	827.2878	220276.7	915.2955
Length	3.2895	2.8521	915.2955	4.4433

Table 4: Covariance matrix

6.2.2 Verify graphically normal distribution of the data

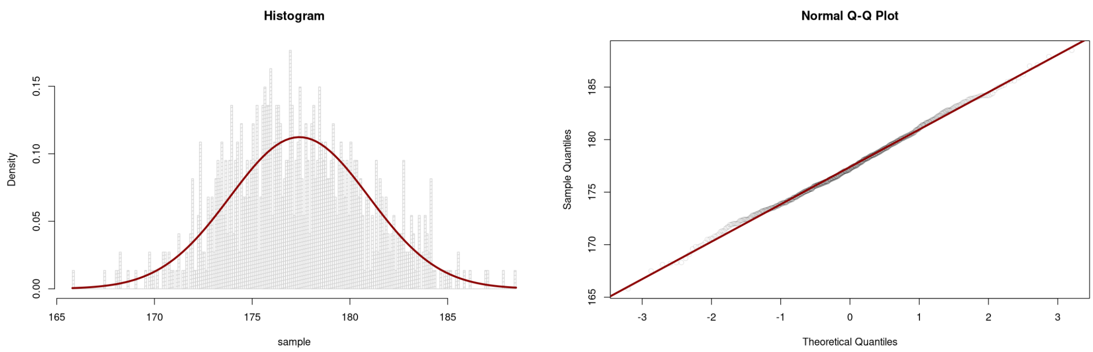

Figure 8: Distribution of father height

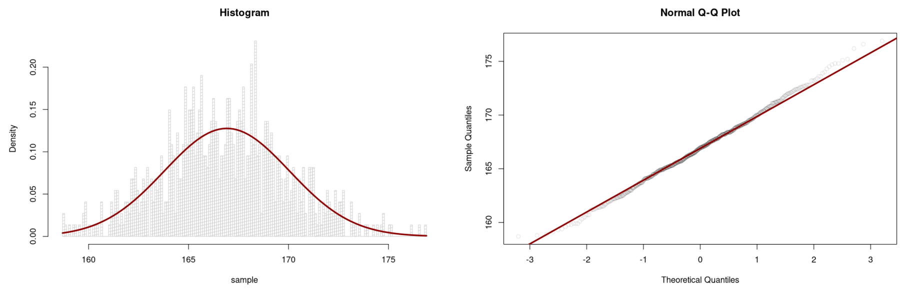

Figure 9: Distribution of mother height

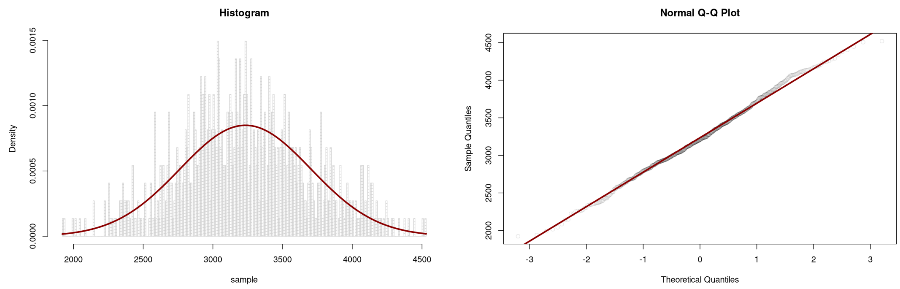

Figure 10: Distribution of child weight

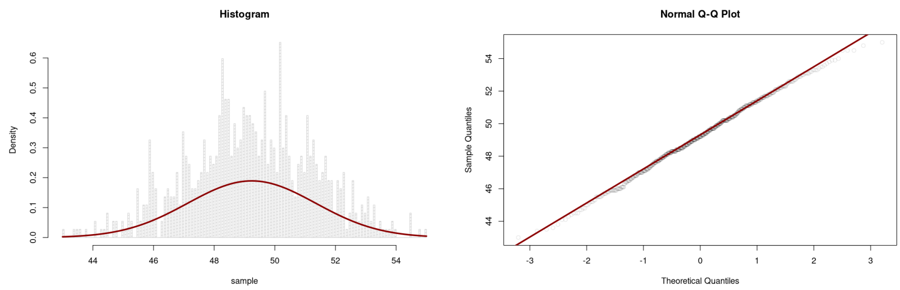

Figure 11: Distribution of child height

6.2.3 Identify the conditional distribution of the weight and length of a child given the heights of parents. Find an estimate of the covariance matrix of the conditional distribution and compare it with the original unconditional covariance.

The conditional distribution of L,W given F,M is as follows:

$$L, W|F = f, M = m \sim \mathcal{N}(\mu_{WL} + \Sigma_{FMWL}\Sigma_{FM}^{-1}(\begin{bmatrix} f \\ m \end{bmatrix} - \mu_{FM}), \Sigma_{WL} - \Sigma_{FMWL}\Sigma_{FM}^{-1}\Sigma_{WLFM})$$

$$\Sigma_{WL} = \begin{bmatrix} 220276.6577 & 915.295511 \\ 915.2955 & 4.443303 \end{bmatrix} \quad \Sigma^* = \begin{bmatrix} 88857.9955 & 456.846078 \\ 456.8461 & 2.843767 \end{bmatrix}$$

6.3 How the ellipsoids based on the conditional distribution will look like?

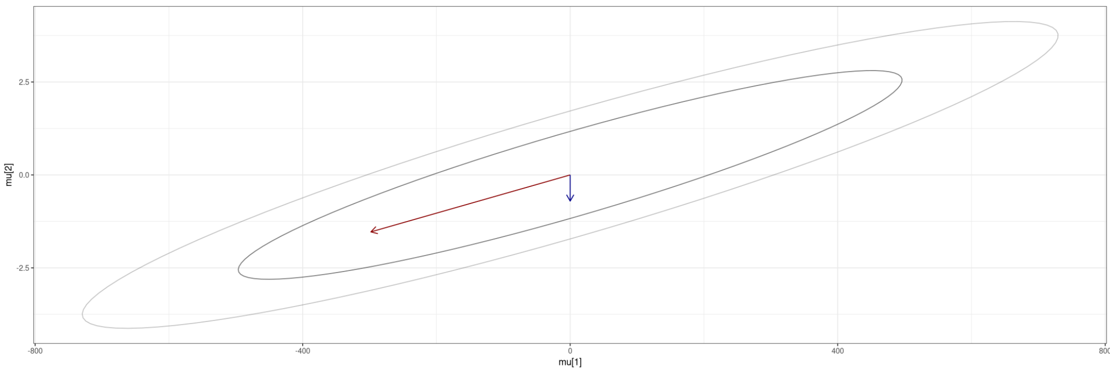

Figure 12: Shape of a conditional ellipsoid

6.3.1 Scores and classification plot

The plot below shows the conditional classification with an uncoditional classification zones superimposed.

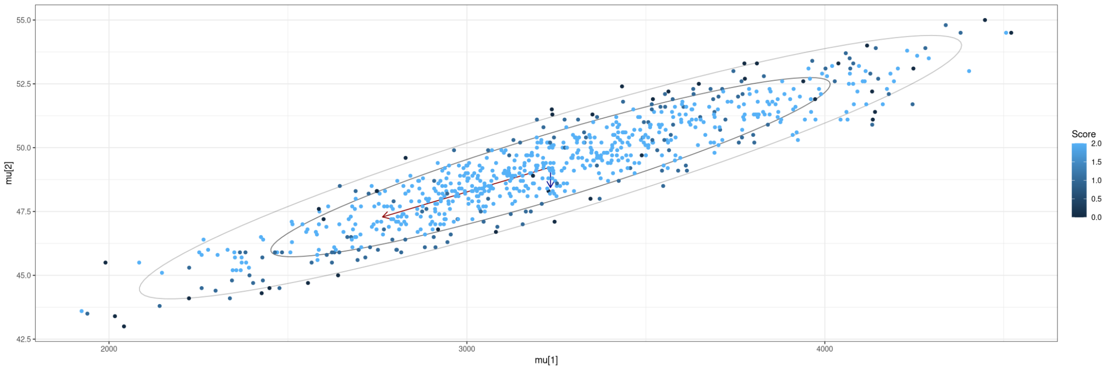

Figure 13: Classification plot

Score	Number of children
0	40
1	137
2	559

Table 5: Scores

6.3.2 Plot the classification ellipsoids for a child given the heights of its parents

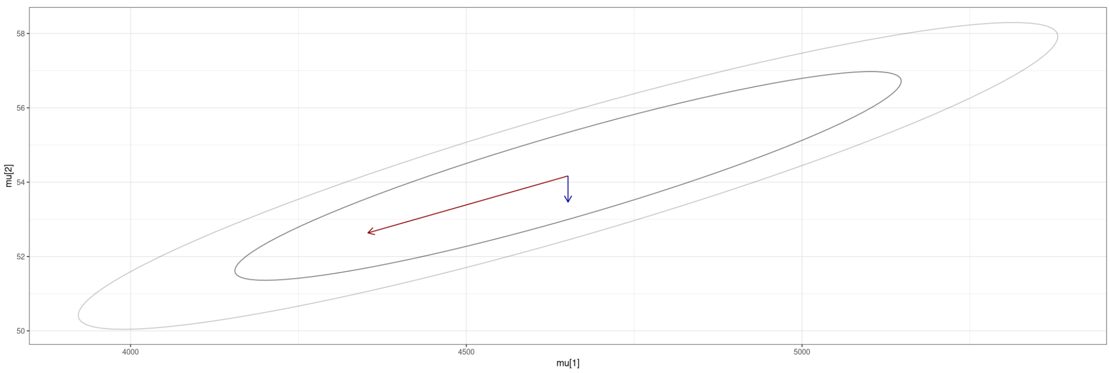

Figure 14: Classification zones for a specific child

6.3.3 Spectral decomposition of covariance matrix for conditional distribution

$$P = \begin{bmatrix} -0.9999868 & -0.005141266 \\ 0.005141266 & -0.9999868 \end{bmatrix} \qquad \Lambda = \begin{bmatrix} 88860.34 \\ 0.494969 \end{bmatrix}$$

6.3.4 Plot the transformed data

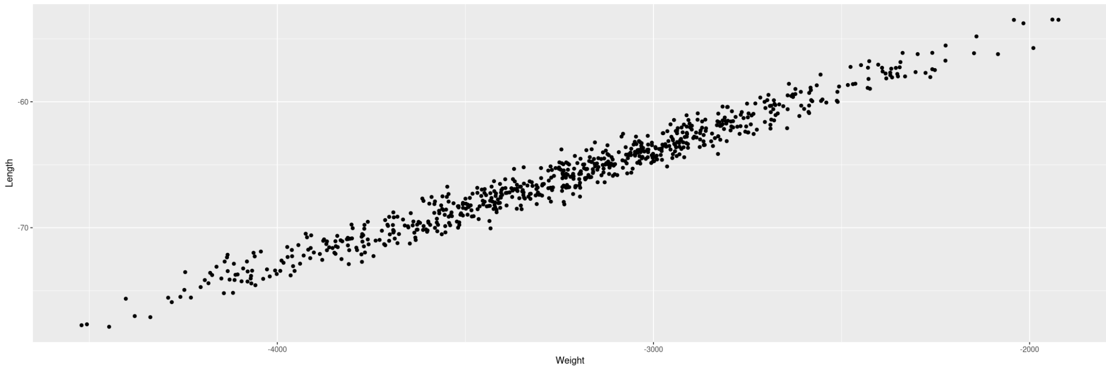

Figure 15: Original data

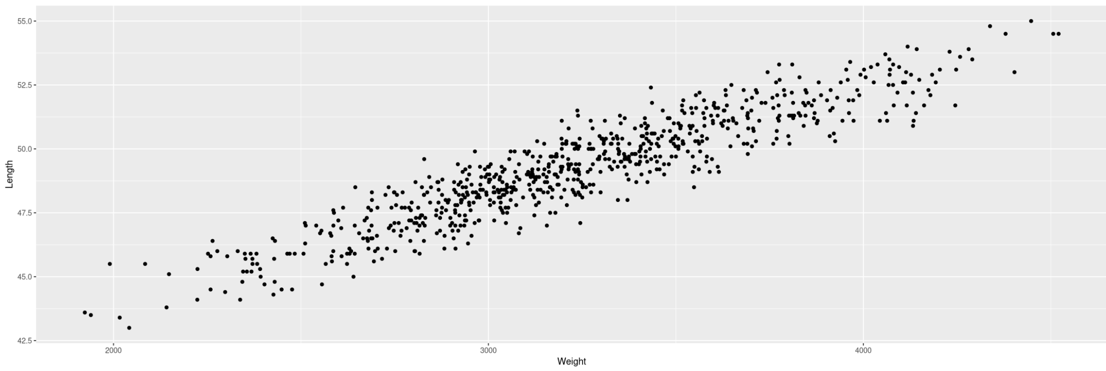

Figure 16: Transformed data