5. Placenta Mean Shapes

Given a sample of 1113 placenta configurations, I computed the extrinsic mean shape and normal coordinates. To do this, I first perform a Principal Component Analysis (PCA) on the cloud of points. Using these results, we can construct one and two sample tests. This facilitates the process of comparing the data to the sample mean estimate. As shown in Figure 8, the principal components explain more than 95% of the variance. This suggests that the principal directions are a good representation of the shape space, which means that the landmarks are highly connected.

6. Relation between placenta shapes and FPR

An objective of my project was to study placenta shapes and use them to predict disease outcomes. I performed a Principal Component Analysis (PCA) on the cloud of points. Table 1 shows the proportion variation and cumulative principal variation explained by the first 18 principal components. The first two principal components explain almost 1/3 of the total variance and the first 10 components explain more than 95% of the variance. This suggests that the principal directions are a good representation of the shape space, which means that the landmarks are highly connected.

What does the distribution of points along the principal directions tell us about placenta shape? How are placenta shapes related to the disease? In this section, we study the relation between placenta shape and Foetal Placental Ratio (FPR). Figure 9 illustrates the change in the 41 intrinsic mean shape caused by perturbation along the first principal direction. The projections are computed for times 0, 0.5, 1, and 2, where 0 is the standard deviation for the component.
is any interaction between $x_5$, $x_a$ and test whether the model has any non-zero coefficient other than $\hat{\beta}_0$, i.e. if there is any interaction between $y$ and $x$. Table 2 shows the results of my analysis. Column 1 shows the shape components used in the model explaining FPR as a function of shape. Column 2 lists the estimates of the coefficients in the model. Column 3 is the $t$-value of the estimated coefficient, column 4 is the $p$-value for the t-test carried out to test for interaction between $y$ and $x$. If $p$-value is less than 0.05, I accept the hypothesis that the model has some non-zero coefficient other than $\hat{\beta}_0$, and hence is a good model.

### Table 2: Regression of FPR on shape($x_a$)

<table>
<thead>
<tr>
<th>Shape Component</th>
<th>$\hat{\beta}_j$</th>
<th>$t$-value</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_5$</td>
<td>0.0089</td>
<td>2.87</td>
<td>0.0040</td>
</tr>
<tr>
<td>$x_a$</td>
<td>0.00119</td>
<td>0.58</td>
<td>0.562</td>
</tr>
</tbody>
</table>

Note that FPR seems to depend on the first principal component ($p$-value for $F$-ratio = 0.0040), meaning adding the new principal component does not add new information and just increases the dimension thereby making the model inefficient.

### Figure 9: Histogram of shape distance from mean shape

![Figure 9: Histogram of shape distance from mean shape](image)

### Figure 10: Scatter plot of FPR against $x_5$, along with best fitting quadratic model

![Figure 10: Scatter plot of FPR against $x_5$, along with best fitting quadratic model](image)

Here I use nonparametric density estimation on embarrassingly, to estimate the posterior distribution of FPR (i.e. given the placenta shape). To do that, I divide the FPR values into 5 equal frequency bins. Based on the FPR values from each bin, I estimate the conditional density, and then estimate the probability that containing the placenta shape alone, the placenta will fall into a particular FPR class.

To get the partition points dividing the FPR classes, I maximise the weighted sum of squares (of the distances of the sample points). I then get the partition points in (5) by the sample empirical distribution $P_2$. For this specific sample, $\hat{\gamma}_1 = 0.0119$, $\hat{\gamma}_2 = 0.02$, $\hat{\gamma}_3 = 0.0338$, $\hat{\gamma}_4 = 0.1$, $\hat{\gamma}_5 = 0.0089$. I then maximize the weighted sum of squared residuals $R^2 = \sum_{i<j} \hat{\gamma}_j |X_i - X_j|^2$ so as to maximize $R^2$.

### Figure 11: Histogram of FPR values classified into 5 classes

![Figure 11: Histogram of FPR values classified into 5 classes](image)

### Figure 12: Scatter plot of FPR against $x_5$, along with best fitting quadratic model

![Figure 12: Scatter plot of FPR against $x_5$, along with best fitting quadratic model](image)

Here $(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4, \hat{\gamma}_5)$ represents the conditional shape density for the class $y = \hat{\gamma}_k$. We estimate that by the kernel density estimate,

$$f_n(y; \hat{\gamma}_k) = \frac{1}{n \hat{\gamma}_k} \sum_{i=1}^{n} K\left(\frac{y - X_i}{\hat{\gamma}_k}\right)$$

for appropriately chosen $K$. Here $x_5$ is the $i$-th shape sample and $y$ is the bin number of shapes in the class $y = \hat{\gamma}_k$. Then we estimate the posterior probability $P(y = \hat{\gamma}_k)$ by

$$P(y = \hat{\gamma}_k) = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{y - X_i}{\hat{\gamma}_k}\right)$$

where $\hat{\gamma}_k$ is the proportion of $y = \hat{\gamma}_k$. For our sample, they are $\hat{\gamma}_1 = 0.0119$, $\hat{\gamma}_2 = 0.02$, $\hat{\gamma}_3 = 0.0338$, $\hat{\gamma}_4 = 0.1$, $\hat{\gamma}_5 = 0.0089$. Table 3 shows the posterior probabilities of $y = \hat{\gamma}_k$ for a few shape samples.

### Table 3: Posterior probabilities ($P(y = \hat{\gamma}_k)$) for a few placenta

<table>
<thead>
<tr>
<th>Shape Component</th>
<th>Posterior Probability ($P(y = \hat{\gamma}_k)$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_5$</td>
<td>0.0119</td>
</tr>
<tr>
<td>$x_a$</td>
<td>0.02</td>
</tr>
<tr>
<td>$x_6$</td>
<td>0.0338</td>
</tr>
</tbody>
</table>

### Figure 13: Bar plot of FPR values along with the chosen 5 classes

![Figure 13: Bar plot of FPR values along with the chosen 5 classes](image)

### Figure 14: Scatter plot of FPR against $x_5$, along with best fitting quadratic model

![Figure 14: Scatter plot of FPR against $x_5$, along with best fitting quadratic model](image)

Here the first 5 placenta’s in the table on the left are closest to the (intrinsic) mean shape. The next 10 are the ones with the largest $d$-distance in the middle, and the last 10 have shapes furthest from the mean. Note how the FPR distribution changes for the 3 shape groups. The first group placenta shapes seem to have the most homogeneous conditional FPR distribution, while for the last 10, the distribution seems to be a mixture. Finally, to measure placenta shape more accurately, I may have to estimate the proportion of variation in FPR and their posterior probabilities given a shape. This is what I do using the model.

$\pi_1 = \sum_{i=1}^{n} \pi_i(x)$

where $\pi_i(x)$ is the probability 1 which does not seem to be consistent with Figure 14, so as to maximize $R^2$.

### Figure 15: Histogram of FPR values along with the chosen 5 classes

![Figure 15: Histogram of FPR values along with the chosen 5 classes](image)

Here is the fundamental problem in putting in appropriate classification. These features can tell us a lot about the shape. Finally, to measure placenta shape more accurately, I may have to increase the number of appropriate features.

### Figure 16: Scatter plot of FPR against $x_5$, along with best fitting quadratic model

![Figure 16: Scatter plot of FPR against $x_5$, along with best fitting quadratic model](image)

Here $(\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3, \hat{\gamma}_4, \hat{\gamma}_5)$ represents the conditional shape density for the class $y = \hat{\gamma}_k$. We estimate that by the kernel density estimate,