Mouth size data

Measurements were made on the size of mouths of 27 children at four ages: 8, 10, 12, and 14. There are 11 girls (Sex=1) and 16 boys (Sex=0).

<table>
<thead>
<tr>
<th>Obs $i$</th>
<th>Age8</th>
<th>Age10</th>
<th>Age12</th>
<th>Age14</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21.0</td>
<td>20.0</td>
<td>21.5</td>
<td>23.0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>21.0</td>
<td>21.5</td>
<td>24.0</td>
<td>25.5</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>20.5</td>
<td>24.0</td>
<td>24.5</td>
<td>26.0</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>22.5</td>
<td>25.5</td>
<td>25.5</td>
<td>26.0</td>
<td>0</td>
</tr>
<tr>
<td>26</td>
<td>23.0</td>
<td>24.5</td>
<td>26.0</td>
<td>30.0</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>22.0</td>
<td>21.5</td>
<td>23.5</td>
<td>25.0</td>
<td>0</td>
</tr>
</tbody>
</table>
\[ Y = x \beta z' + R \]

\[ = \begin{pmatrix} 1_{11} & 1_{11} \\ 1_{16} & 0_{16} \end{pmatrix} \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \\ \delta_0 & \delta_1 & \delta_2 & \delta_3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{pmatrix} + R \]
Under the hypothesis that the quadratic and cubic terms are zero:

\[ Y = x \left( \begin{array}{cccc} \beta_0 & \beta_1 & 0 & 0 \\ \delta_0 & \delta_1 & 0 & 0 \end{array} \right) z' + R. \]

Put the \( z \) on the other side:

\[ Y^z = Y(z')^{-1} = x \left( \begin{array}{cccc} \beta_0 & \beta_1 & 0 & 0 \\ \delta_0 & \delta_1 & 0 & 0 \end{array} \right) + R^z. \]

Now the last two columns (quadratic and cubic) of \( Y^z \) have zero means, so they can be used as covariates.
\[ Y^z = \begin{pmatrix}
\text{Constant} & \text{Linear} & \text{Quadratic} & \text{Cubic} \\
21.375 & 0.375 & 0.625 & -0.125 \\
23.000 & 0.800 & 0.250 & -0.150 \\
23.750 & 0.850 & -0.500 & 0.200 \\
24.875 & 0.475 & 0.125 & 0.075 \\
\vdots \\
24.250 & 1.950 & -1.000 & 0.400 \\
24.875 & 0.525 & -0.625 & 0.175 \\
25.875 & 1.125 & 0.625 & 0.125 \\
23.000 & 0.550 & 0.500 & -0.150 \\
\end{pmatrix} \]
The original model:

\[ Y^z = (Y^z_{01}, Y^z_{23}) = x \left( \begin{array}{cccc} \beta_0 & \beta_1 & 0 & 0 \\ \delta_0 & \delta_1 & 0 & 0 \end{array} \right) + R^z. \]

Now shift the quadratic and cubic terms to the other side (condition on them), to get the covariate-adjusted model:

\[ Y^z_{01} = (x, Y^z_{23}) \left( \begin{array}{cc} \beta_0 & \beta_1 \\ \delta_0 & \delta_1 \\ \gamma_{02} & \gamma_{03} \\ \gamma_{12} & \gamma_{13} \end{array} \right) + R^*. \]
The estimate of the beta parameters from the original model on $Y^z$ ($\nu = 27 - 2 = 25$):

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>$t$</th>
<th>Linear</th>
<th>$t$</th>
<th>Quadratic</th>
<th>Cubic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>24.969</td>
<td>0.784</td>
<td>9.12</td>
<td>0.203</td>
<td>-0.05</td>
<td></td>
</tr>
<tr>
<td>Girls - Boys</td>
<td>-2.321</td>
<td>-3.05</td>
<td>-0.305</td>
<td>-2.26</td>
<td>-0.214</td>
<td>0.073</td>
</tr>
</tbody>
</table>

And from the covariate-adjusted model on $Y^z_{01}$ ($\nu = 27 - 4 = 23$):

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>$t$</th>
<th>Linear</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>24.937</td>
<td>0.827</td>
<td>9.13</td>
<td></td>
</tr>
<tr>
<td>Girls - Boys</td>
<td>-2.272</td>
<td>-2.86</td>
<td>-0.350</td>
<td>-2.54</td>
</tr>
<tr>
<td>Quadratic</td>
<td>-0.189</td>
<td>-0.191</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cubic</td>
<td>-1.245</td>
<td>0.063</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

p-values for the Girls–Boys linear effect: 0.033 and 0.018.
Likelihood & BIC

Consider submodels. \( M_{p^*l^*} \): the \( p^* \) indicates the \( x \)-part of the model, where \( p^* = 1 \) means just use the constant, and \( p^* = 2 \) uses the constant and the sex indicator; the \( l^* \) indicates the degree of polynomial represented in the \( z \) matrix, where \( l^* = degree + 1 \). (\( l^* = 1 \) means just the constant term, \( l^* = 2 \) means the constant+linear terms, etc.)

Find the \( BIC(M_{p^*l^*} ; y^z) \)’s.
$M_{22}$ is the model we just fit: Constant and linear terms, including the Boys vs. Girls effect —

$$
Y(z) = (Y_{01}^{(z)}, Y_{23}^{(z)}) = x \left( \begin{array}{cccc}
\beta_0 & \beta_1 & 0 & 0 \\
\delta_0 & \delta_1 & 0 & 0
\end{array} \right) + R^z.
$$

$$
\hat{\Sigma}_{z,aa \cdot b} = \frac{1}{27} \left( y_{01}^{(z)} - (x y_{23}^{(z)}) \left( \begin{array}{c}
\hat{\beta} \\
\hat{\gamma}
\end{array} \right) \right)' \left( y_{01}^{(z)} - (x y_{23}^{(z)}) \left( \begin{array}{c}
\hat{\beta} \\
\hat{\gamma}
\end{array} \right) \right)
$$

$$
= \frac{1}{27} y_{01}^{(z)'} Q_{(x,y_{23}^{(z)})} y_{01}^{(z)} = \left( \begin{array}{cc}
3.313 & 0.065 \\
0.065 & 0.100
\end{array} \right)
$$

and

$$
\hat{\Sigma}_{z,bb} = \frac{1}{27} y_{23}^{(z)'} y_{23}^{(z)}
$$

$$
= \left( \begin{array}{cc}
0.266 & -0.012 \\
-0.012 & 0.118
\end{array} \right).
$$
Deviances

Model $M_{22}$:

\[
\text{deviance}(M_{22}; y^{(z)}) = n \log(|\hat{\Sigma}_{z,aa}b|) + n \log(|\hat{\Sigma}_{z,bb}|) + nq
\]
\[
= 27(-1.116 - 3.463 + 4) = -15.643.
\]

And $d_{22} = 4 + 10 = 14$: 4 nonzero $\beta_{ij}$'s and 10 elements in $\Sigma_z$.

Big model, $M_{24}$:

\[
\hat{\Sigma}_z = \frac{1}{27} y^{(z)'} Q_x y^{(z)}.
\]

\[
\text{deviance}(M_{24}; y^{(z)}) = n \log(|\hat{\Sigma}_z|) + nq = -18.611,
\]

And $d_{24} = 8 + 10 = 18$: 8 nonzero $\beta_{ij}$'s and 10 elements in $\Sigma_z$. 
Likelihood ratio test

Consider testing the two models:

\[ H_0 : M_{22} \text{ versus } H_A : M_{24}. \]

The likelihood ratio statistic is

\[ \text{deviance}(M_{22}) - \text{deviance}(M_{24}) = -15.643 + 18.611 = 2.968. \]

That value is obviously not significant, but formally the chi-square test would compare the statistic to the cutoff from a \( \chi^2_{df} \) where \( df = d_{24} - d_{22} = 18 - 14 = 4. \) (Or just note we set four parameters to 0.)
$$\log(27) \approx 3.2958:$$

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_{22}$</td>
<td>$-15.643 + 2(14) = 12.357$</td>
<td>$-15.643 + \log(27)(14) = 30.499$</td>
</tr>
<tr>
<td>$M_{24}$</td>
<td>$-18.611 + 2(18) = 17.389$</td>
<td>$-18.611 + \log(27)(18) = 40.714$</td>
</tr>
</tbody>
</table>

$$\hat{P}^{BIC}[M_{22} | \mathbf{y}^{(z)}] = \frac{e^{-\frac{1}{2}BIC(M_{22})}}{e^{-\frac{1}{2}BIC(M_{22})} + e^{-\frac{1}{2}BIC(M_{24})}} = 0.994.$$ 

Between these two models, the smaller one has an estimated probability of 99.4%, quite high.
All the models

<table>
<thead>
<tr>
<th>(p^*)</th>
<th>(l^*)</th>
<th>Deviance</th>
<th>(d_{p^<em>l^</em>} = p^<em>l^</em> + 10)</th>
<th>BIC</th>
<th>(\hat{P}^{BIC})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>36.322</td>
<td>11</td>
<td>72.576</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>-3.412</td>
<td>12</td>
<td>36.138</td>
<td>0.049</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>-4.757</td>
<td>13</td>
<td>38.089</td>
<td>0.018</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>-4.922</td>
<td>14</td>
<td>41.220</td>
<td>0.004</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>30.767</td>
<td>12</td>
<td>70.317</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-15.643</td>
<td>14</td>
<td>30.499</td>
<td>0.818</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>-18.156</td>
<td>16</td>
<td>34.577</td>
<td>0.106</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-18.611</td>
<td>18</td>
<td>40.714</td>
<td>0.005</td>
</tr>
</tbody>
</table>

\(\hat{P}[M_{22}] = 81.8\%\).

\(\hat{P}[\text{Boys} \neq \text{Girls}] = \hat{P}[M_{21}] + \hat{P}[M_{22}] + \hat{P}[M_{23}] + \hat{P}[M_{24}] = 92.9\%\).
MLE with real covariates: Leprosy data

(Exercise 9.6.13) The model is

\[(Y^{(b)}, Y^{(a)}) = x\beta + R = \begin{bmatrix} (1 & 1 & 1) \otimes 1_{10} \end{bmatrix} \begin{pmatrix} \mu_b & \mu_a & 0 & 0 & \alpha_a & \beta_a \end{pmatrix} + R.\]

where \(Y^{(b)}\) has the before measurements, \(Y^{(a)}\) has the after measurements, \(\alpha_a\) is the after Drug vs. Placebo effect, and \(\beta_a\) is the after Drug A vs. Drug D effect.
Condition on $Y^{(b)}$ to get the conditional model for $Y^{(a)}$:

$$Y^{(a)} = (x \ y^{(b)}) \begin{pmatrix} \mu^* \\ \alpha_a \\ \beta_a \\ \gamma \end{pmatrix} + R^*.$$ 

The marginal model for $Y^{(b)}$ is

$$Y^{(b)} = \left[ \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -2 & 0 \end{pmatrix} \otimes 1_{10} \right] \begin{pmatrix} \mu_b \\ 0 \\ 0 \end{pmatrix} + R^{(b)} = 1_{30} \mu_b + R^{(b)}.$$ 

Both models are regular multivariate regressions (actually, in this case just univariate regressions), so finding the MLE’s, and likelihoods, is straight forward.
Conditional

Using R:

```R
> bothsidesmodel(cbind(x, leprosy[,1]), leprosy[,2], 1)
```

\[
\begin{pmatrix}
\mu^* \\
\alpha_a \\
\beta_a \\
\gamma
\end{pmatrix} = \left[ (x y^{(b)})'(x y^{(b)}) \right]^{-1} (x y^{(b)})' y^{(a)} = \begin{pmatrix}
-2.696 \\
-1.131 \\
-0.054 \\
0.987
\end{pmatrix},
\]

\[
\widehat{\sigma^2}_{aa,b} = \frac{1}{30} y^{(a)'} Q_{(x,y^{(b)})} y^{(a)} = \frac{26}{30} 16.04625 = 13.90675.
\]

(The "16.04625" is from the bothsidesmodel function, which gives the unbiased estimate with \( \nu = 30 - 4 \), not the MLE.)
Here the design matrix is just $\mathbf{1}_{30}$, and $\mathbf{Q}_{130} = \mathbf{H}_{30}$. Using R:

```r
> bothsidesmodel(rep(1,30),leprosy[,1],1)
```

$$\hat{\sigma}_{bb}^2 = \frac{1}{30} \mathbf{y}^{(b)'} \mathbf{H}_{30} \mathbf{y}^{(b)} = \frac{29}{30} 22.96092 = 22.19556.$$
Deviance

The observed likelihood multiplies the conditional and marginal likelihoods.

\[ L = \left[ \left| \hat{\sigma}_{aa}^2 b \right|^{-n/2} e^{-\frac{1}{2} nq_a} \right] \times \left[ \left| \hat{\sigma}_{bb}^2 \right|^{-n/2} e^{-\frac{1}{2} nq_b} \right]. \]

So the deviance is

\[
-2 \log(L) = n \log(\hat{\sigma}_{aa}^2) + n \log(\hat{\sigma}_{bb}^2) + nq
= 30 \log(13.90675) + 30 \log(22.19566) + 30 \cdot 2
= 231.9681.
\]

The number of parameters: 4 for the coefficients (2 \( \mu \)'s and one each of \( \alpha_a \) and \( \beta_a \)) and 3 for the \( 2 \times 2 \) covariance, \( \Rightarrow d = 7 \).

\[ BIC = 231.9681 + 7 \log(30) = 255.7765. \]
Exercise 9.6.14: Other models, which set $\alpha_a$ and/or $\beta_a$ to zero. All that changes is the $x$ for the conditional model: leave out the columns corresponding to the parameters set to zero. Have to adjust $d$ and $\nu$. 