**Orthogonal polynomial contrasts Addendum Sep. 10, 2022**

Consider modeling longitudinal data with 4 equally spaced time points (e.g., Ramus data). We model time as a class variable. We can test for polynomial trends (up to cubic) using orthogonal polynomial contrasts.

**L**=(-3, -1, 1, 3) linear **Q**=(1, -1, -1, 1) quadratic

**C**=(-1, 3, -3, 1) cubic.

These coefficients will determine the strength of the respective polynomial trends. For example, **Lτ** will have greater magnitude when **τ** has a stronger linear pattern, and closer to 0 when it doesn’t. Similar for **Qτ** and **Cτ**.

For 3 equally spaced contrasts, we can test up to quadratic trends: **L**=(-1, 0, 1) linear,

**Q**=(1, -2, 1) quadratic.

Note that the coefficients for the linear trend suggest that the middle time point has no impact on the slope. It’s true! Try performing a simple linear regression on 3 points where x values are equally spaced, and y could be anything. If you change the middle point, the slope will not change; the line will just move up or down.

Back to PCA and the Ramus data

PC’s roughly were

PC1≈ 0.5H1 + 0.5H2 + 0.5H3 + 0.5H4

PC2≈ –0.6H1 – 0.4H2 + 0.3H3 + 0.6H4

PC3≈ 0.6H1 – 0.5H2 – 0.4H3 + 0.4H4

PC4≈ –0.3H1 + 0.6H2 – 0.7H3 + 0.4H4

If you compare these to the previous orthogonal contrast coefficients (for r=4), they are not too different (after scaling).

As we discussed, orthogonal polynomial contrasts are intuitive for variables that have an intrinsic ordering such as those collected over time. But if we have 4 correlated variables not collected over time, we may not end up with PC’s that express polynomial trends.

Consider the COPDGene study, where 4 types of measures were taken on subjects with moderate COPD at baseline: FEV1 (higher=stronger lung), adjusted lung density (lower values indicate more emphysema), 6-minute walk distance (higher=better exercise capacity) and a symptom score (lower=better health).

The PRINCOMP Procedure

|  |  |
| --- | --- |
| **Observations** | 1788 |

| **Simple Statistics** | | | | |
| --- | --- | --- | --- | --- |
|  | **FEV1\_utah** | **adj\_density\_mesa** | **distwalked** | **SGRQ\_scoreTotal** |
| **Mean** | 1.886233781 | 78.84545793 | 1315.024609 | 32.48958613 |
| **StD** | 0.509255161 | 23.05728164 | 362.602371 | 21.35909880 |

| **Correlation Matrix** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | **FEV1\_utah** | **adj\_density\_mesa** | **distwalked** | **SGRQ\_scoreTotal** |
| **FEV1\_utah** | FEV1, post-Utah | 1.0000 | 0.0253 | 0.3103 | -.1675 |
| **adj\_density\_mesa** | Lung density, CT sponge model adjusted | 0.0253 | 1.0000 | -.0088 | 0.0259 |
| **distwalked** | Distance walked, ft | 0.3103 | -.0088 | 1.0000 | -.4355 |
| **SGRQ\_scoreTotal** | SGRQ score: Total | -.1675 | 0.0259 | -.4355 | 1.0000 |

| **Eigenvalues of the Correlation Matrix** | | | | |
| --- | --- | --- | --- | --- |
|  | **Eigenvalue** | **Difference** | **Proportion** | **Cumulative** |
| **1** | 1.62210443 | 0.61423268 | 0.4055 | 0.4055 |
| **2** | 1.00787175 | 0.17220585 | 0.2520 | 0.6575 |
| **3** | 0.83566590 | 0.30130799 | 0.2089 | 0.8664 |
| **4** | 0.53435791 |  | 0.1336 | 1.0000 |

| **Eigenvectors** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | **Prin1** | **Prin2** | **Prin3** | **Prin4** |
| **FEV1\_utah** | FEV1, post-Utah | **0.482** | 0.184 | **0.808** | -.284 |
| **adj\_density\_mesa** | Lung density, CT sponge model adjusted | -0.014 | **0.976** | -.215 | -.004 |
| **distwalked** | Distance walked, ft | **0.651** | -.0147 | -.122 | **0.749** |
| **SGRQ\_scoreTotal** | SGRQ score: Total | **-0.586** | 0.112 | **0.534** | **0.599** |

PC1 appears to be a ‘health variable’ for the clinical outcomes; PC2 essentially captures the emphysema variable and reinforces the fact that it is not very correlated with the clinical outcomes; PC3 and PC4 are sort of oddballs since key variables have positive coefficients but where ‘better health’ scales are opposite. BUT…usually it’s the secondary PC’s that are the most novel/interesting.