This is the readme file for the R script "JRSS C codes cleft data", which contains R codes related to Section 5, Analyses of the Cleft Lip Smile Data, in the paper. In this section, cleft subjects are compared with the control subjects.

The codes have been divided into 8 chunks with indices ranged from 0 to 7.

**Code chunk 0** contains codes for reading in the data. The data for 13 cleft subjects at the three frames (first, middle and last frames) are stored in a list with name "cleft\_data", where the corresponding element for each subject is also a list containing the landmark coordinates at the three frames. The same format is used for the control subjects and the 12 control subjects are stored in a list with name "control\_data".

**Code chunk 1** contains codes for computing the signed elementary feature vector

defined in equations (2.1) - (2.3) in Section 2. The results are stored for the cleft and control subjects separately at the three frames. The signed elementary feature vectors for cleft subjects at the three frames are saved with names:

1. d\_cle\_fir;

2. d\_cle\_mid;

3. d\_cle\_la.

The signed elementary feature vectors for control subjects at the three frames are saved under names:

1. d\_con\_fir;

2. d\_con\_mid;

3. d\_con\_la.

**Code chunk 2** involves codes for computing the three composite asymmetry scores defined in Section 3.1.

This chunk has been divided into 4 sub-chunks: 2.1, 2.2, 2.3 and 2.4:

(a). 2.1 contains codes for computing the asymmetry score \phi^\*\_{L\_1} defined in equation (3.9), where the results for cleft and control subjects are saved in two different matrices with names "phi\_star\_L1\_cle" and "phi\_star\_L1\_con" respectively, where rows are for subjects and columns for frames.

(b). 2.2 contains codes for computing the asymmetry score \phi\_{L\_1} defined in equation (3.6), where the results for cleft and control subjects are saved in two different matrices with names "phi\_L1\_cle" and "phi\_L1\_con" respectively, where rows are for subjects and columns for frames.

(c). 2.3 contains codes for computing the asymmetry score \phi\_{L\_2} defined in equation (3.7), where the results for cleft and control subjects are saved in two different matrices with names "phi\_L2\_cle" and "phi\_L2\_con" respectively, where rows are for subjects and columns for frames.

(d). 2.4 contains codes for computing the weighted \phi\_{L\_1} and \phi\_{L\_2}. We first compute the mean shapes at the first, middle and last frames and store them in "mu\_sym\_fir", "mu\_sym\_mid" and "mu\_sym\_la" respectively. Then the weights are computed and stored in "weight\_fir", "weight\_mid" and "weight\_la" respectively. Then the weighted \phi\_{L\_1} are computed and saved in two different matrices with names "phi\_L1\_cle\_scale" and "phi\_L1\_con\_scale" respectively, where rows are for subjects and columns for frames. The weighted \phi\_{L\_2} is computed and saved in a similar manner with names "phi\_L2\_cle\_scale" and "phi\_L2\_con\_scale" for cleft and control respectively.

**Code chunk 3** contains codes for performing the univariate tests - t-test and Mann-Whitney U test. This chunk has been divided into 2 sub-chunks: 3.1 and 3.2:

(a). 3.1 contains codes for carrying out traditional t-test with equal variance assumption (chunk 3.1.1) and Welch's t-test (chunk 3.1.2). The tests have been performed at the three frames separately. \phi^\*\_{L\_1} is used here.

(b). 3.2 contains 4 sub-chunks. The first 2 sub-chunks involve the codes for performing the Mann-Whitney U tests on \phi\_{L\_1} (chunk 3.2.1) and \phi\_{L\_2} (chunk 3.2.2). The next 2 sub-chunks contain codes for performing the Mann-Whitney U tests on weighted \phi\_{L\_1} (chunk 3.2.3) and \phi\_{L\_2} (chunk 3.2.4). These tests have been performed at the three frames separately.

**Code chunk 4** contains codes for carrying out feature selection method given in Section 3.2 in the paper. The separate test statistics v\_j (see equation (3.15)) are computed in chunk 4.1. The cleft and control data at each frame are combined in chunk 4.2. The bootstrap procedures are performed in chunk 4.3. Codes for Mann-Whitney U tests on \phi\_{L\_1} and \phi\_{L\_2} computed on various subsets of landmarks are in 4.4.

**Code chunk 5** contains codes for Fisher's (chunk 5.1) and Pearson's methods (chunk 5.2). The separate p-values are saved in variables:

1. p\_wilcoxon\_fir;

2. p\_wilcoxon\_mid;

3. p\_wilcoxon\_la

respectively.

**Code chunk 6** contains the codes for Hotelling's T^2 test. The three frames are used separately in chunk 6.1, 6.2 and 6.3.

**Code chunk 7** contains codes for making the dot plots given in Figure 5 and 6 in Section 5.3 in the paper. Codes for producing Figure 5 are given in chunk 7.1, whereas the codes for producing Figure 6 are given in chunk 7.2.