Steps to Run a Complete FDA with Categorical Data

Carlos Gonzalez, Berkley Shands, Elena Deych, William Shannon

January 31, 2020

Below we will define the steps required to complete an FDA with our data

1. Load the `Actigraphy` package:

```
>         library(Actigraphy)
```

2. Read in the activity and covariate data and store them as the datasets covariate and activity:

```
>         data(clinic_29pt_ahi)
>         data(act_29pt)
>         covariate <- clinic_29pt_ahi
>         activity <- act_29pt
```

3. Change the continious variable AHI in the covariate file into a categorical variable:

```
>         covariate$ahicat <- as.factor(
+                 ifelse(covariate$AHI >= 0 & covariate$AHI <= 5, 1,
+                 ifelse(covariate$AHI > 5 & covariate$AHI <= 15, 2,
+                 ifelse(covariate$AHI > 15 & covariate$AHI <= 30, 3,
+                 ifelse(covariate$AHI > 30, 4, 0)))))
```

4. Match the data in the activity and covariate datasets and since the covariate is a factor, apply user-defined category labels:

```
>         matchid <- fda.matchid(activity, covariate[,-2], "factor", c("normal", "mild", "
```

5. Smooth the data and plot it:

- L is the number of entries in the activity file

```
>         L <- nrow(activity)
>         FDinterest <- fda.smoothdata(matchid)
>         ts.plot(predict(FDinterest$fd$fd, 1:L), main="Smoothed Activity Data")
```

6. Apply FLM to the data with the `flm_cate` function:

```
>         geftinterest <- flm_cate(FDinterest)
```

7. Set up the plot parameters for the next step:

- ylim is a vector of length 2 that contains the Y-axis boundries based on the results on the `flm_cate` function
- lb is a vector of labels for the X-axis
- xat is a vector of positions for the labels (lb) on the X-axis

```
>         ypred <- as.vector(geftinterest$freg$yhatfdobj$y)
>         ylim <- c(0, max(ypred) + 100)
>         lb <- c("Midnight", "6AM", "Noon", "6PM", "Midnight")
>         xat <- c(0, L/4, L/2, 3*L/4, L)
```

8. Plot results from FLM results and F-test if indicated (with the categorical flm plotting function):

```
>         cat.flm.results <- cat_flm_plot(FDinterest, matchid, geftinterest, TRUE, 5, lb,
```

Notes:

- Columns in Activity file data MUST represent subjects
- Covariate file MUST only contain 2 columns; subject identifier (id) and one covariate (numeric or factor). If user is interested in 2 or more covariates, they should be put into separate datasets and analyzed separately.
- Subject identifier should be the same in both Activity (first column) and Covariate (row names) files