Agenda 04/20:

- Concatenated front and back torso curves:
    - Rescaled curve coordinates to align the deepest point of each curve with the corresponding "Depth" measurement in our dataset
    - Used predicted coefficients, along with the depth and anterior-posterior-length measurements, to produce plots of the front and back curves, and combined them to create a complete curve

Training Back R-Square:	[0.24	0.41	0.52	0.43]
Testing Back R-Square:	[0.05	0.17	0.31	0.34]
Training Front R-Square:	[0.58	0.4	0.63	0.41]
Testing Front R-Square:	[0.38	0.22	0.33	0.33]

    - Found inconsistencies between the anterior-posterior-length measurements in the dataset and the values calculated from the rescaled plot (to be discussed during next meeting)
- PCA and K-means for African American and White group;
- Heatmap across different races display;
- Logistic Regression and Random Forest race prediction;
- Report progress check.

Notes 04/20:

- To-do:
    - Create a list of the inconsistencies between the anterior-posterior length in the original dataset and in the predicted values (percentage errors)
        - These inconsistencies may not affect the analysis much, but it's still important to find their source
    - What is the heatmap telling us?
        - Dendrograms tell us the similarities between features
        - The first four features share characteristics, and so do the last eight features
        - Cluster is not telling us the features have good predictive ability
    - Produce same results on the imbalanced data set
        - Consider plotting a Precision Recall (PR) Curve
    - Logistic regression does not allow for interactions – could explain why it is not performing as well as Random Forest and XG Boost methods
    - Repeat analysis for White vs. As. American
    - Leave races with small sample sizes out – leave a note in "Limitations" section on report
    - Comment from Dr. Basu: As long as you reproduce some of the analyses (e.g. the logistic and RF fit) on White-Asian, we should be good on the analysis fronts. Some of the caveats such as spline fits, discrepancy in anterior-posterior distance, could be listed as potential limitations but not necessarily addressed in this report.

Agenda 03/30:

- PCA and K-means
- Logistic regression race prediction
- Random Forest race prediction
- Heatmap across African American and White group

Notes 03/30:

- Use R^2 values for predicted curve performance in addition to MAPE (sensitive to small values)
- Don't be alarmed by large MAPE value – could be a good thing; may be additional information embedded in the curves, predicted coefficients don’t necessarily need to equal the true coefficients
- Remove race as a variable from heatmaps
- Normalize variables properly before creating heatmap (don’t use min/max values)
- Presentation/Report
    - Include boxplots in presentation
    - Share presentation with Drs. Basu and Baytar
    - Final presentation will be in a recorded format

Agenda 03/23:

- Curve Fitting
    - Discuss:
        - Literature Reviews;
        - Prediction of coefficients;
            - Predictors:
                - BMI, Height, Max_Hip, Anterior_posterior_Length, Depth, Crotch_curve_length_at_back_waist

	Coef1	Coef2	Coef3	Coef4
MAPE_train	0.1399	0.2387	0.1204	1.4367
MAPE_test	0.1573	0.2973	0.1517	0.3978

            -
        - Graphs of the predicted curve
            - MPS_Curve/Analysis/Curve_Fitting/03_23_23_Curve_Fitting/Plot/Predict_Curve
    - Ask questions:
        - Assure the fitting curves going below or above actual curves at the bottom part?

Notes 03/23:

- Use K-means / theoretical clustering to group subjects based on features (selected 4 features, all features, etc.)
- See if there is any overlap in clustering for different ethnic groups (Race)
- Look into using heat maps for more than 3 predictors
- Logistic regression, random trees – consider these as well.
- Use principal component analysis as well.
- Area under ROC curve (R package: AUC)

Agenda 03/16:

- Curve Fitting
    - Discuss:
        - Literature Reviews that we have done so far.
    - Ask questions:
        - Are there any viable smoothing techniques other than conreg() that Dr. Basu would recommend?
        - How to avoid the fitting curves going below actual curves at the bottom part?

Notes 03/16:

- Literature reviews
    - For each paper, consider the question the researchers are asking!
    - Dr. Baytar will (hopefully) provide more guidance about what to look for in each paper
    - Show the client you are interested/invested in his/her domain of expertise.
    - For each paper, consider:
        1. What problem are you solving?
        2. Why is this problem important/significant to solve?
        3. If it is SO important, why has not people solved it yet? In other words, what is hard/challenging about this problem?
        4. What is your solution?
        5. How did you know it worked (simulated data: matched ground truth? real data: did better than existing alternatives?)
    - Prepare a bullet-point summary of each.
    - (Second paper had descriptions on metrics)
    - Consolidate all summaries into one file!

Agenda 03/09:

- Curve Fitting
    - Discuss:
        - Removal of non-convexity from curves and whether the curves look acceptable to Dr. Baytar
        - How the fitted curves change for different model parameters
        - How to find the relationship between the coefficients and waist features
    - Ask questions:
        - Conreg function in R was giving us trouble – are there alternative methods for removing non-convexity?
        - How should we assess the goodness of fit of our curves? RMSE?

Notes 03/09:

- Highlights from papers recently uploaded to Box:
    - Focus on paper written by Xu
- Paper summary:
    - 1 to 2 slides summarized high level what those papers wrote about (Bullet points)
    - May include some screenshots from the papers
- Curve fitting:
    - Trade-off between Complexity and error
    - 90% fit is ok

Next step:

- Literature Review
- Sent some Conreg function sample codes to Dr. Basu

Agenda 03/02:

- Curve Fitting
    - fixed the model so that the model can fit the starting (top) point well
    - reduced the number of parameters to 5
    - uploaded the code and all the plots to the box under Analysis/03_02_23_Curve_Fitting folder

Notes on meeting 03/02:

- Generate spreadsheet with four curve parameters for each individual
- Don’t use Google Docs in Box anymore; instead use Docx (Word)
- Provide the ID number in each graph of the lower torso curve
- Try to fit some curves to the torso plots exactly (or almost exactly)
- Members other than Fanjun: use conreg() function to modify curves to remove non-convexity
- Communicate with Fatma; ask her to answer two questions (reminder email)
- Metrics for assessing the goodness of fit of the curve (RMSE)

Agenda 02/23:

- Review univariate summary statistics and pairwise correlations between major explanatory variables
- Present histograms for each variable and for the residuals in the linear models
- Show pairwise correlation plots and a correlation table for numerical explanatory variables
- Discuss new feature for the difference between the length of the front and back crotch and our spline model
- Talk about new findings after adding the weight variable and compare the BMIs of different races

Notes on meeting 02/23:

- Put the notes and agenda on the same file
  - Makes it more continuous
- Recap what we discussed last meeting first
- Residual plot to check error (systematic or randomness)
- QQ plot for better diagnostic for normality
- If highly correlated, we can't trust that variable even it is statistically significant
- 11 pieces is too complex
- tradeoff between interpretation and accuracy
- Starting point (Top) match!
- Try to use four numbers of polynomials to represent the curves, and to see if the numbers change across different BMI, Race, etc.

- Clean data (check for NA values) once and for all
- Knitting is a large advantage - if possible, go through knitted file rather than the Rmd file itself
- Start writing up report
- Everything should be uploaded to Box!

Agenda 02/16:

Discuss tasks:

- Prepare univariate summary statistics and pairwise correlations.
    - Prepare histograms of each variable
    - Look at the residuals – is there systematic error or randomness?
- Create a feature with front/back crotch (difference between them, indicator function) - model the asymmetry
- Incorporate both **height and weight** in model
- (Split curve at midpoint and rotate by 90 degrees, fit polynomial function to each part)

Ask questions:

- How should the splines be constructed?
    - How many knots in each region?

Agenda 02/09:

- Present on progress made
  - Try to find the relationships between numerical dependent variables (find the association between BMI, Height and columns D-I)
  - Import curve files into R
  - Do paired t test to see BMI difference in different races
- Next plans
  - Extract features from the curve
  - Create some new features from the curve (Angle, slope, length on part of the curve, convex/concave)
  - create other variables like weight?
  - Divide the curve into 3 parts: front, middle, back
- Ask Dr. Baytar
  - How were the quantitative variables collected? (G-K columns)
  - How is the curve generated? (by side? Front? Back?)
  - What does the curve represent in the real world?
  - How does the .txt file connect to the .csv file?
    - Do the numbers stand for the subject ID? Some are missing.
  - What's the unit for each variable?

Notes on meeting 02/09:

- Dr. Baytar had a personal conflict
- Presented on linear regression
    - Crotch curve length:
        - BMI, height had positive coefficients (and significant) - makes sense
        - R^2 is around 0.5 (reasonably strong fit)
    - Depth:
        - Similar patterns – positive association with BMI, max hip
        - Significant effects from racial groups (Asian, Other)
        - Even adjusting for other vars, we see some difference in race
    - Anterior/posterior length:
        - RaceAsian and RaceWhite are significant; why is RaceWhite negative here
        - High R^2 (around 0.88)
    - Depth and Anterior-posterior length have positive correlation.
    - There is still significant variability in the body types of two people that have the same BMI and height
    - Good practice for multivariate analysis: prepare univariate summary statistics and pairwise correlations.
        - Prepare histograms of each variable
        - Look at the residuals – is there systematic error or randomness?
    - Create a feature with front/back crotch (difference between them, indicator function) - model the asymmetry
    - Incorporate both **height and weight** in model
- Questions for / answers from Dr. Baytar:
    - Brief description of dependent variables:
        - Anterior/posterior length: straight-line distance between front and back where LH side is front, RH is back
        - Crotch curve length / Trochanter: length of entire torso curve (from front to back)
            - Sum of front crotch and back crotch
        - Max_hip: hip circumference at the widest point of lower body
        - Depth: horizontal line to lowest point
        - Midpoint: point that splits front and back crotch (?)
    - What to look for in a curve:
        - Flatness – length of flat region vs. entire length
        - Abdomen part is easily squished
    - Units for independent variables:
        - Height in mm
        - Weight in kg

Task list:

- Prepare univariate summary statistics and pairwise correlations.
    - Prepare histograms of each variable
    - Look at the residuals – is there systematic error or randomness?
- Create a feature with front/back crotch (difference between them, indicator function) - model the asymmetry
- Incorporate both **height and weight** in model
- (Split curve at midpoint and rotate by 90 degrees, fit polynomial function to each part)

Agenda 02/02:

1. Review the five tasks we completed and discuss improvements to be made
    a. Starting draft report – Aristotle
    b. Reviewing and summarizing data in spreadsheets – Diane
    c. Script in R that appends spreadsheets for different body types – Zongjie, Wenxuan
    d. Script in R Markdown that provides summary data – Aristotle
    e. Summarizing articles – Aristotle (Gu), Fanjun (McKinney)
2. Discuss next steps on project
    a. Begin data analysis
        i. Plot out data
        ii. Determine which techniques we want to use – regression (linear, polynomial), machine learning (lasso, ridge, etc.), curve estimation
3. Ask questions
    a. How to analyze crotch curves in R or using other software

Notes from Meeting 2/2:

1. **Presentation**
    a. Starting draft report – Aristotle
    b. Reviewing and summarizing data in spreadsheets – Diane
    c. Script in R that appends spreadsheets for different body types – Zongjie, Wenxuan
    d. Script in R Markdown that provides summary data – Aristotle
    e. Summarizing articles – Aristotle (Gu), Fanjun (McKinney)

2. **Some Findings from Data:**
    a. Features of the curve are different from different races;
    b. Higher BMI implies wider curve;
    c. The higher the depth, the longer the vertical length of the curve;

3. **TO DO LIST:**
    a. Ask for Dr. Fatma Baytar next week:
        i. How do we get those numbers for each variable? (D-I columns)
        ii. How does the curve come from? (by side?Front?Back?)
        iii. How does the txt file connect to the csv file? (probably the number stands for the subject id? Some are missing though)
    b. Tasks:
        i. First focus on at White and AfricanAmerican's data;
        ii. Try to find the relationships between numerical dependent variables; (find the association between BMI, Height and columns D-I)
        iii. Extract features from the curve; (coding activities)
        iv. Create some new features from the curve; (Angle, slope, length on part of the curve, convex/concave)
        v. create other variables like weight?
        vi. Spline might be helpful. (divide the curve into 3 parts: front, middle, back)
        vii. Try to do pair t test or other statistical tests to see if there is a significant difference in BMI

Agenda 01/26:

- Listen to and take notes on client presentation
    - Think about which programming language we will use
    - Get access to data
    - Sign NDAs if any
- Discuss frequency with which we will meet with advisor and clients

Our group:

- Meet weekly on Zoom Tuesday at 5:30 pm
- Meet biweekly in person Tuesday at 5:30 pm
- Location: Olin or Uris Library

Notes from Meeting 01/26:

Dr. Fatma Baytar: Prof. in Dept. of Fiber Science and Apparel Design (Human Ecology)

1. **Presentation**
    - Body shape analysis to inform customized pattern-making for women's pants particularly
    - Pants are the most problematic garments to fit (either because of tightness, looseness, mismatch with body type)
    - One pattern shape doesn't fit all
        - The current fashion system assumes bodies are standard and makes people feel bad about themselves
    - Pant components:
        - Two tubular shapes ((front + back) + (front + back))
        - Crotch curves should be close to person's body shape
    - In general, researchers take many measurements to draft pant patterns:
        - Circumference measurements:
            1. Waist arc
            2. Hip arc
            3. Upper thigh
            4. Mid thigh
            5. Knee
            6. Calf
            7. Ankle
            8. Foot entry
    - When drafting pants patterns, crotch curve is randomly drawn.
    - Dr. Baytar's friend who runs a company has reported having a lot of trouble with rise shapes, esp. with non-flat/curvy body types

- In this case, researchers collected data using CAESAR 3D body scan database
  - Women between 18-45

- **Questions**:
  1. How can we extract features of the curves?
  2. Are there any associations among height, weight, basic measurements (e.g., hip) and curve features that may help us group the curve shapes based on a few inputs and suggest using certain shapes when drafting patterns?

2. Scheduling and Logistics
   a. Programming language: R
   b. Zoom meetings
      i. Meet with Dr. Baytar once every two weeks Thursday 4:15-5:15 pm
      ii. Meet with Dr. Basu once every week Thursday 4:15-5:15 pm
      iii. Dr. Basu will attend biweekly meetings and is willing to meet more frequently than that if necessary
3. Next steps
   a. Send calendar invite(Aristotle)
   b. Create folder structure (zongjie)