LAsBEST@USC: Final Poster Project Instructions (SUMMER 2021)

This final poster project will evaluate your skills with data management/exploration, developing and implementing a data analysis plan, and the summarization/critical interpretation of your results. This will be a group project, allowing you the opportunity to further develop skills for teamwork. At the end of the project, each group will be expected to produce a poster that summarizes the setup of the study from which the dataset comes from, the methods used in analyzing the data, the main findings from the analysis and conclusions drawn from the main findings.

Dataset Description:

The dataset will be drawn from the landmark and ongoing Southern California Children's Health Study, hereafter referred to as CHS, that has been collecting rich information annually for several years on more than 11,700 children from several communities in Southern California. The CHS study has five different cohorts of children that were recruited in 1993 (Cohorts A, B and C), in 1996 (Cohort D) and in 2002 (Cohort E). The primary objective of the study was to examine the long-term effects of environmental factors (especially air pollution exposure) on children's respiratory health.

The dataset you will be analyzing will be based on the baseline information (i.e., data collected at the beginning of the study) on a subset of 1,000 children from Cohort E, residing in 8 Southern California communities. Note that this dataset includes only a non-representative sample from the actual study cohort. Hence, it is solely intended for instructional purposes and should not be used to draw definitive (and substantively meaningful) conclusions about actual study objectives.

The dataset containing the participant-level variables on health outcomes, socio-demographic and other characteristics is provided in the file "CHS_cohortE_final_subset.csv". A separate dataset will be provided containing community-level variables on environmental exposure variables.

Variable listing and description for file "chs_ cohortE.csv" on children's health and socio-demography:

Variable	Variable Name	Description
Subject ID	id	Unique participant identifier
Community	townabbr	Community abbreviation of where child's residence is located • AN = Anaheim • GL = Glendora • LB = Long Beach • ML = Mira Loma • RV = Riverside • SA = Santa Barbara • SD = San Dimas • UP = Upland
Age	age	Age of child at lung function testing (years)
Sex	male	Binary indicator for sex • 0 = Female • 1 = Male
Race	race	Multi-category variable for self-reported race • Asian • African American • Caucasian • Mixed • Others • Unknown or Missing

Hispanic	hisp	Binary indicator for self-reported Hispanic ethnicity: • Hispanic • Non-Hispanic • Unknown or Missing
Asthma	active_asthma	Binary indicator for asthma activity in past 12 months
Height	height	Height of child (cm)
Weight	weight	Weight of child (kg)
BMI	bmi	Body mass index (kg/m^2)
Parental Education	educ	Multi-category variable for parental education • 1 = < 12^{th} Grade • 2 = Grade 12 • 3 = Some post high-school • 4 = 4 years of college • 5 = Some post-graduate
Home age	HomeBuilt	Period in which the child's home was built
Gas Stove	BaseGasstove	Binary indicator for presence of gas stove in the residence
Pets	BasePets	Binary indicator for presence of any pets in the residence
Second Hand Smoke	ETS_base	Binary indicator for exposure of child to second hand smoke at home
Wheeze	wheeze	Binary indicator for child's wheezing status
FEV_1	fev1	Forced expiratory volume in 1 second (ml)
FVC	fvc	Forced vital capacity (ml)
PM2.5	pm25	Particulate matter concentration with aerodynamic diameter less than 2.5 micrometers (ug/m^3)
Sulfate	sulfate	Sulfate particle concentration (ug/m^3)
Nitrate	nitrate	Nitrate particle concentration (ug/m^3)
Elemental Carbon	ec	Elemental carbon particle concentration (ug/m^3)
Dust	dust	Dust particle concentration (ug/m^3)
Longitude	longitude	Longitude coordinate of child's residence
Latitude	latitude	Latitude coordinate of child's residence

Expected Outcomes:

The final product from this project will be a group poster on findings from each of the following four topics:

Group 1: Risk factors (with and/or without air pollution) for Respiratory Health in CHS children: Overall and by sex. Main outcomes will be continuous [FEV_1], and binary [Asthma].

Group 2: Risk factors (with and/or without air pollution) for Respiratory Health in CHS children: Overall and by sex. Main outcomes will be continuous [FVC], and binary [Asthma].

Group 3: Risk factors (with and/or without air pollution) for Respiratory Health in CHS children: Overall and by ethnicity (Hispanic vs. NHW). Main outcomes will be continuous [FEV_1], and binary [Asthma].

Group 4: Risk factors (with and/or without air pollution) for childhood obesity in CHS children: Overall and by sex. Main outcomes will be continuous [BMI] and binary [obesity indicator].

Group 5: Risk factors (with and/or without air pollution) for childhood obesity in CHS children: Overall and by ethnicity (Hispanic vs. NHW). Main outcomes will be continuous [BMI] and binary [obesity indicator].

Analysis Objectives:

You are required to analyze the data with the following major steps in mind:

1) To provide descriptive statistics of the data through summary statistics [**All Groups**]
2) To provide descriptive statistics of the data through visualization techniques [**All Groups**]
3) To fit appropriate regression models within each group as follows:
    a. To fit linear (for continuous) and logistic (for binary) regression models to assess effects of risk factors on respiratory health outcomes (i.e., FEV, asthma) on the entire data set and by sex **[Group 1]**
    b. To fit linear (for continuous) and logistic (for binary) regression models to assess effects of risk factors on respiratory health outcomes (i.e., FVC, asthma) on the entire data set and by sex **[Group 2]**
    c. To fit linear (for continuous) and logistic (for binary) regression models to assess effects of risk factors on respiratory health outcomes (i.e., FVC, asthma) on the entire data set and by ethnicity (focusing on Hispanic White and Non-Hispanic White groups only due to sample size limitations) **[Group 3]**
    d. To fit linear (for continuous) and logistic (for binary) regression models to assess effects of risk factors on obesity related outcomes (BMI, obesity indicator) on the entire data set and by sex **[Group 4]**
    e. To fit linear (for continuous) and logistic (for binary) regression models to assess effects of risk factors on obesity related outcomes (BMI, obesity indicator) on the entire data set and by ethnicity (focusing on Hispanic White and Non-Hispanic White groups only due to sample size limitations) [**Group 5]**
4) To summarize findings from the models using both tabular and graphical displays [**All Groups**]
5) To provide interpretation for the main findings of the final models [**All Groups**]
6) Prepare a poster summarizing your goals, methods, and main findings [**All Groups**]
7) Make a group poster presentation during the last day of the summer program [**All Groups**]

Poster Preparation Instructions:

Each group will produce one poster, on which all group members are expected to contribute. Your poster should include:

1. A descriptive title and a listing of all members of each group along with affiliations - in alphabetical order.
2. A summary of the overall setup, methods, data description, main findings, and conclusions from each group project.
3. Acknowledgements of funding sources, and individuals/groups that have helped with the conduct of the project and/or preparation of the poster.
4. A disclaimer as follows: "*Note that the results on this poster are based on a dataset that includes only a non-representative sample from the actual study cohort. Hence, it is solely intended for instructional purposes and should not be used to draw definitive (and substantively meaningful) conclusions about actual study objectives.*"

Poster dimensions:
Poster boards are **portrait format (tall and narrow)**. Maximum poster dimensions are 66 cm wide x 100 cm tall (26 inches wide x 39 inches tall)