PDAT 617 Zoom Discussion 6

Statistics in Python:

- **SciPy**: is a Python module that provides an extended range of random number generators, probability distributions and statistical tests.
- **Statsmodels**: is a Python module that provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration.

Scipy can compute the PDF, CDF and inverse CDF, fitting parameters using MLE, and compute various moments.

- **dist.rvs**
- **dist.pdf**
- **dist.cdf** ,
- **dist.ppf** : Inverse CDF evaluation
- **dist.median/dist.mean** .
- **dist.moment** .
- **dist.varr/ dist.std** .

Hypothesis test in Scipy/Statsmodels

- import scipy.stats as stats; import statsmodels.api as sm
- **One sample t-tes** : stats.ttest-1samp(data, .11) ; d1 = sm.stats.DescrStatsW(data) d1.ttest-mean(value=0.11, alternative=two-sided) .
- **One sample Proportion Z test** :sm.stats.proportions-ztest(x, n, p0, alternative = larger,prop-var= 0.5)
- **2- sample t-test (Independent)** : stats.ttest-ind(carpted, uncarpted); sm.stats.ttest-ind(carpted,uncarpted,alternative=larger)
- **2 sample t-test (Paired)** :stats.ttest-rel(father, son); d2.ttest-mean(value=0, alternative=smaller)
- **2 sample proportion test** : sm.stats.proportions-ztest(x1, n1,alternative = smaller)

- **Analysis of Variance Test (ANOVA):stats.f-oneway()** :

- **Pearson's Correlation Coefficient** : stats.pearsonr(data1, data2)

Linear Regression in Python :

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x + ... + \beta_n x$$

Our task: estimate $\beta_i' s$ based on the available data

- from sklearn.linear-model import LinearRegression
- lm = LinearRegression().fit(X,Y)
- model = lm.fit(X-train, Y-train)