**Python (version ≥3.6) for R Users: Stat Modules II**

**CMU MSP 36601, December 6, 2017, Howard Seltman**

1. **random module**

Example: generate a random dataset

import random

import pandas as pd

N = 30

rxVals = ('placebo', 'txA', 'txB')

id = ["P" + str(i + 1) for i in range(N)]

rx = pd.Series([random.choice(rxVals) for i in range(N)])

male = pd.Series([int(random.random() < 0.4) for i in range(N)])

age = pd.Series([round(random.normalvariate(mu=35, sigma=7))

for i in range(N)])

dtf = pd.DataFrame([(r, m, a) for (r, m, a) in

zip(rx, male, age)], index=id,

columns=['rx', 'male', 'age'])

print(dtf)

print(dtf.shape)

1. **scipy.stats module**

import numpy as np

dir(stats)

help(stats.binom)

stats.binom.pmf(range(5), n=5, p=0.5)

stats.norm.cdf(np.arange(6.0, 14.1, 2.0), loc=10, scale=2)

stats.poisson.rvs(size=5, mu=6)

1 - stats.chi2.cdf(15, df=9)

a = stats.norm.rvs(size=20, loc=7, scale=2)

b = stats.norm.rvs(size=20, loc=8.5, scale=2)

stats.ttest\_ind(a, b)

1. **statsmodels module** (<http://patsy.readthedocs.io/en/latest/>)

import statsmodels.formula.api as smf

m = smf.ols(formula="age ~ male + C(rx)", data=dtf).fit()

m.summary()

dir(m)

m.resid

1. Many statistical models with a machine learning approach: scikit-learn