Power, sample size + Paired t test tutorials

Initial intro ~ 5 minutes. Re-explain ORAIP and dialysis very briefly.

ORAIP levels - control versus ESRD patients

The file `Control.dat` contains data on the ORAIP levels of the control group. The file `ESRD.dat` contains data on the ORAIP levels for the ESRD patients. The first columns is a patient ID. The second column is ORAIP concentration before dialysis and the third column is ORAIP concentration after dialysis of the same patient.

1. Perform a Student t test to check whether the ORAIP levels in ESRD patients before dialysis are the same as those in the control group. Use a 95% confidence level and compute the p-value. Assume equal variance. Load the data files. Note the skip header part. Do without, see the error, then fix it by adding it. Then, ask them to do the test. Allocate 15 minutes

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

control = np.genfromtxt("control.dat",skip_header =1);
ESRD_data = np.genfromtxt("ESRD.dat",skip_header =1);

ESRD_before = ESRD_data[:,1];#Second column

#Standard t test between control and ESRD in terms of ORAIP conc
num_control = len(control);
num_ESRD = len(ESRD_before);

mean_ESRD = np.mean(ESRD_before);
mean_control = np.mean(control);
#std calculations, manual way with for loops also OK.
std_ESRD = np.std(ESRD_before,ddof=1);
std_control = np.std(control,ddof=1);

#calculate pooled variance
s_p=np.sqrt(((num_control-1)*std_control**2 + \
(num_ESRD-1)*std_ESRD**2)/(num_control+num_ESRD-2));

#Correct value is 22.84 for t stat
t_stat = (mean_ESRD-mean_control)/  \
(s_p*np.sqrt(((1.0)/num_ESRD)+((1.0)/num_control)));
#Correct value for t_crit is 1.984 << 22.4
t_crit = stats.t.ppf(0.975,num_control+num_ESRD-2);
#Note this p_value is zero as too small for Python
p_value = 2*(1-stats.t.cdf(abs(t_stat),num_control+num_ESRD-2));
```

2. Compute the power of the test to detect a difference of $\delta = 5ng/ml$ (assume $\sigma = s_p$) Code the power calculations

```
#Power calculation
delta=5;
sigma = s_p;
D = delta/(sigma*(np.sqrt(1/num_control + 1/num_ESRD)));
t_star = t_crit-D;#t* in the slides
power = 1-stats.t.cdf(t_star,num_control+num_ESRD-2);#0.366
```

3. Obtain a plot of power versus δ for a range between 2 and 30 in increments of 0.1 (assume $\sigma = s_p$) Code the vector of deltas and powers. Ask them to do the for loop

```
#loop for power calculation, increasing delta
all_deltas = np.arange(2,30,0.1);
powers = np.zeros(len(all_deltas));
for i in range(0,len(all_deltas)):
    D = all_deltas[i]/(sigma*(np.sqrt(1/num_control+1/num_ESRD)));
    t_star = t_crit-D;
    powers[i] = 1-stats.t.cdf(t_star,num_control+num_ESRD-2);

plt.figure (1);
plt.plot(all_deltas,powers,'k-');
```

4. Obtain a plot of power versus σ for a range between 2 and 50 in increments of 0.1 (assume $n_{control} = n_{ESRD} = 10$) and $\delta = 10ng/ml$. Code the last plot

```
#loop for power calculation, increasing sigma, n=10 for each
all_sigmas = np.arange(2.0,50,0.1);
powers_sigmas = np.zeros(len(all_sigmas));
n=10;
t_crit = stats.t.ppf(0.975,n+n-2);
for i in range(0,len(all_sigmas)):
    sigma=all_sigmas[i];
    D = 10.0/(sigma*(np.sqrt(1/n+1/n)));
    t_star = t_crit - D;
    powers_sigmas[i] = 1-stats.t.cdf(t_star,n+n-2);

plt.figure(2);
plt.plot (all_sigmas,powers_sigmas,'r-');
```

Paired t test

Determine whether the dialysis treatment affect ORAIP levels (use a 99% confidence level) by performing a t test between data before and after the dialysis session of ESRD patients. Paired t test is done as

```
#Paired t test between ESRD before and after
ESRD_after = ESRD_data[:,2];
diff_vector =ESRD_before - ESRD_after;
mean_diff = np.mean(diff_vector);
```

```
diff_s_sq = 0;
for i in range(0,num_ESRD):
    diff_s_sq = diff_s_sq + (diff_vector[i]-mean_diff)**2;

diff_std = np.sqrt(diff_s_sq/(num_ESRD-1));
t_stat_paired = mean_diff/(diff_std/np.sqrt(num_ESRD));
eta_paired = num_ESRD-1;
t_crit_paired = stats.t.ppf(1-0.01/2,eta_paired);
if (np.fabs(t_stat_paired) > t_crit_paired):
    #Reject
    print("Data support a difference in ORAIP levels between befoire and af
    #We observe that, on average, ORAIP levels increased.
    #This increase is statistically significant at the 99% level
else:
    #Unable to reject
    print("Data failed to support any difference in ORAIP levels \
          between befoire and after dialysis")
#We already know p<0.01
p_value_paired = 2*(1-stats.t.cdf(abs(t_stat_paired),num_ESRD-1));#0.0034
```