T-test and Comparing Two Popoulation Mean

Two independent populations

When comparing two populations, we consider $\mu_1 - \mu_2$. When we have two **independent** randomly-chosen large samples, we will use two sample means and two sample std.

- mean of $Xbar - Ybar$ = $\mu_1 - \mu_2$
- std of $Xbar - Ybar$ = $\sigma_X - Y = \sqrt{\frac{(\sigma_1)^2}{m} + \frac{(\sigma_2)^2}{n}}$

CI for two populations

$$(x - y) \pm Z_{\alpha/2} * \sqrt{\frac{(s_1)^2}{m} + \frac{(s_2)^2}{n}}$$

CLT

If both m and n are larger then 40,then sample variance can be used as point estimates for population variance.

two sample t test

$$t = \frac{(\bar{x} - \bar{y}) - \Delta_0}{\sqrt{\frac{(s_1)^2}{m} + \frac{(s_2)^2}{n}}}$$

Code

Randomly divide to two groups

use PROC RANK with option GROUPS=2 to divide the subject to 0 or 1, and create a new data set containing the treatments treat

```
Proc rank data=data_name groups=2;
out=treat;
var var_name;
run;
```

Use PROC FORMAT to change from 0-1 to A-B

```
proc format;
value zerone 0='A' 1='B'
Run;

proc sort data=treat;
BY Name;
Run;
```

run ttest

```
proc ttest data=Milk;
class diet;
var yeild;
run;
```

Wilcoxon Rank-sum test

```
proc npar1way data=dataset wilcoxon;
class location;
var cont;
exact wilcoxon;
run;
```

Two Dependent population (paired data)

scenarios

a) samples draw from same population

b) sample draw from different population but highly related

Definitions

$\mu_D = E(X - Y) = \mu_1 - \mu_2$'

Code

Paired T-test using proc mean

```
# Calculate difference before datalines;
Diff = before-after

# use proc mean with T and PRT(p-val)
Proc means data=data N Mean T PTR
```

Note SAS default is a two sided test. To get the correct one-sided p-value, divide SAS p-value by 2

Paired T-test using proc ttest

```
# calcualted diff before datalines;
PROC TTEST DATA = PAIRED SIDES=U;
VAR DIFF;
RUN;
```

```
# no need to calcualte diff
ROC TTEST DATA = PAIRED SIDES=U;
PAIRED BEFORE*AFTER;
RUN;
```