STAT:2100/4200

Lab 7. paired t-test and t-test for 2 independent samples

1 Paired-sample problems (one-sample t test)

To carry out the hypothesis test of interest, we apply one-sample procedures to the *differences* between values measured on members of each pair.

Example: To study the effect of cigarette smoking on platelet aggregation, researchers drew blood samples from 11 individuals before and after they smoked a cigarette and measured the percentage of blood platelet aggregation. This study can be found in Levine, P. H. (1973). An acute effect of cigarette smoking on platelet function, Circulation, 48, 619-623. The data is available in ICON/modules/lab worksheets/smoking.dat.

We test the null hypothesis that the means before and after are the same. Use $\alpha = 0.05$.

We will do a two-sided test, because we are not sure in advance whether to expect μ_1 (mean percentage of blood platelet aggregation before smoking) to be higher or lower than μ_2 (mean percentage of blood platelet aggregation after smoking).

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

or equivalently:

$$H_0 : \mu_2 - \mu_1 = 0$$
$$H_a : \mu_2 - \mu_1 \neq 0$$

or equivalently:

$$H_0 : \delta = 0$$
$$H_a : \delta \neq 0$$

where δ denotes $\mu_2 - \mu_1$.

We will use the *observed differences* between the percentage after and before smoking observed on each subject as our data to carry out the hypothesis test regarding δ at the .05 significance level.

We will carry out a t-test. In theory, the assumptions for a t-test to be valid are SRS, and approximate normality of the population distribution (here the distribution of differences). In practice, use the rules of thumb regarding one-sample t procedures (see chap18-19-extra notes, page 3). Are the rules satisfied here?

Note that by default, `proc univariate` tests the null hypothesis that $\mu = 0$ ($\delta = 0$ if using the symbols above), so in this case we don't have to specify a value for `mu0` in the command line.

```
data smoking ;
input before after ;
diff = after - before ;
datalines ;
* note: copy and paste data in here ;
;
run ;

proc univariate plot data = smoking ;
var diff ;
run ;

proc means data = smoking n mean stddev stderr clm alpha = .05 ;
var diff ;
run ;
```

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	4.271609	Pr > |t|	0.0016
Sign	M	4.5	Pr >= |M|	0.0117
Signed Rank	S	32	Pr >= |S|	0.0020

The SAS System

The MEANS Procedure

Analysis Variable : diff					
N	Mean	Std Dev	Std Error	Lower 95% CL for Mean	Upper 95% CL for Mean
11	10.2727273	7.9761007	2.4048848	4.9143099	15.6311446

An alternative way to do paired t test in SAS is to use proc ttest:

```
proc ttest data = smoking ;
paired after*before ;
run ;
```

2 Two-independent-sample problems (2-sample t-test) using proc ttest - Review

We wish to test whether the population mean of rainfall produced per cloud is the same for unseeded clouds as for seeded clouds. The *cloud.dat* is a data set containing measurements of rainfall in acrefeet from 52 clouds, 26 of which were chosen at random and seeded with silver nitrate.

We will log-transform the rainfall amounts to get more symmetrical distributions of sample values (in order to reduce the skewness in the data set). Denote the 52 logged rainfall values by $x_{u,1}, x_{u,2}, \cdots, x_{u,26}$ and $x_{s,1}, x_{s,2}, \cdots, x_{s,26}$.

Denote
μ_u = the population mean of log rain fall produced per unseeded cloud, and
μ_s = the population mean of log rain fall produced per seeded cloud.

We will test the following hypotheses at significance level $\alpha = .05$:

$$H_0 : \mu_u = \mu_s$$
$$H_a : \mu_u \neq \mu_s$$

Are the assumptions for doing a 2-sample t-test satisfied here? (Use the rules of thumb regarding one-sample t procedures (see chap18-19-extra notes))

```
data cloud ;
input rainfall seeded $ ;
lograin = log(rainfall) ;
datalines;
1202.6  U
...
;
run;

proc sort data=cloud;
by seeded;
run;

proc univariate plot data=cloud;
var rainfall lograin;
by seeded;
run;

proc ttest ;
class seeded ;
var  lograin ;
run;
```

The SAS System

The TTEST Procedure

Variable: lograin

seeded	N	Mean	Std Dev	Std Err	Minimum	Maximum
S	26	5.1342	1.5995	0.3137	1.4110	7.9178
U	26	3.9904	1.6418	0.3220	0	7.0922
Diff (1-2)		1.1438	1.6208	0.4495		

seeded	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
S		5.1342	4.4881	5.7802	1.5995	1.2544	2.2080
U		3.9904	3.3272	4.6536	1.6418	1.2876	2.2664
Diff (1-2)	Pooled	1.1438	0.2409	2.0467	1.6208	1.3562	2.0148
Diff (1-2)	Satterthwaite	1.1438	0.2408	2.0467			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	50	2.54	0.0141
Satterthwaite	Unequal	49.966	2.54	0.0141

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	25	25	1.05	0.8971

We perform a 2 sample t-test for the hypotheses. Specifically, we use the Satterthwaite method, which does not assume equal population variance. The resulting p-value is **.0141**. Since, this is smaller than the pre-specified significance level $\alpha = .05$. We say that the data contains strong evidence that the mean of (the log of) the amount of rain of seeded clouds and unseeded clouds are different, where the seeded clouds produced more rain.

3 More practice, using the body temperature example

A random sample of 130 healthy people is taken and each person's temperature, gender, and heart rate are recorded in "normtemp.dat".

3.1 Recall we did a t-test in lab 9 for a one sample problem (proc univariate)

It is widely believed that the average body temperature for healthy humans is 98.6 F. We think that might not be true, so we decide to do a two-sided significance test at significance level $\alpha = .05$:

$$H_0 : \mu_{\text{temp}} = 98.6$$
$$H_a : \mu_{\text{temp}} \neq 98.6$$

We reformat the variable gender in the data step.

```
proc format ;
value sexfmt 1 = 'M' 2 = 'F' ;
run ;

data normtemp ;
input temp gender heart ;
format gender sexfmt. ;
datalines;
* note: copy and paste data in here ;
;
run ;
```

We performed a t test using **proc univariate**.
(Remember to check the rule of thumb before accepting/interpreting the results.)

```
proc univariate mu0 = 98.6 data = normtemp ;
var temp;
run ;
                         The UNIVARIATE Procedure
                              Variable:  temp

                       Tests for Location: Mu0=98.6
             Test           -Statistic-    -----p Value------

             Student's t    t  -5.45482    Pr > |t|  <.0001
             Sign           M       -21    Pr >= |M| 0.0002
             Signed Rank    S     -1963    Pr >= |S| <.0001
```

Conclusion:

3.2 Two-sample t-test for a two-independent-sample problem (proc ttest)

We might wish to test whether the population mean body temperature is the same in men and in women (at significance level $\alpha = .05$):

$$H_0: \quad \mu_M - \mu_F = 0$$
$$H_A: \quad \mu_M - \mu_F \neq 0$$

We can view the (fe)males in our study as a random sample from the (fe)male population, then the data contains two independent samples from the two populations of interest.

We can perform a two-sample t test using **proc ttest**. (Check the rule of thumb first.)

```
proc ttest data = normtemp ;
class gender ;
var temp ;
run ;
```

The SAS System

The TTEST Procedure

Variable: temp

gender	N	Mean	Std Dev	Std Err	Minimum	Maximum
M	65	98.1046	0.6988	0.0867	96.3000	99.5000
F	65	98.3938	0.7435	0.0922	96.4000	100.8
Diff (1-2)		-0.2892	0.7215	0.1266		

gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
M		98.1046	97.9315	98.2778	0.6988	0.5959	0.8449
F		98.3938	98.2096	98.5781	0.7435	0.6340	0.8990
Diff (1-2)	Pooled	-0.2892	-0.5396	-0.0388	0.7215	0.6429	0.8221
Diff (1-2)	Satterthwaite	-0.2892	-0.5396	-0.0388			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	128	-2.29	0.0239
Satterthwaite	Unequal	127.51	-2.29	0.0239

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	64	64	1.13	0.6211

Conclusion: The p-value of the 2-sample t-test (the Satterthwaite test) is .0239, smaller than the significance level $\alpha = .05$. Hence the data suggests that the mean body temperature is different in the male and the female population. Specifically, the male has lower body temperatures. We are 95% confident that the mean male body temperature is between .5396 to .0388 degrees lower than the mean female body temperature.