**Homework 8 Ref: Chapter 8 Two-Sample Inference**

Last Name: Zuo First Name: Lian

* Read Activities 8 very carefully before answering the homework problems.
* Data for this hw can be copied and pasted into SAS along with necessary data steps.

**Homework 8.1: Paired T-test**

A study is intended to evaluate effect of certain diet in reducing the cholesterol of patients. The data below is a sample of 12 patients with the measurements of their cholesterol before and after the diet.

| **id** | **before** | **after** |
| --- | --- | --- |
| 3082 | 168 | 145 |
| 3354 | 159 | 146 |
| 3373 | 197 | 169 |
| 3399 | 250 | 202 |
| 3415 | 195 | 146 |
| 3423 | 197 | 178 |
| 3513 | 244 | 208 |
| 3630 | 192 | 184 |
| 3725 | 236 | 215 |
| 3746 | 205 | 178 |
| 3920 | 145 | 155 |
| 3965 | 178 | 137 |

Let and refers to the mean cholesterol of patients in the population before and after the diet, measured in milligrams (mg) of cholesterol per deciliter (dL) of blood (mg/dL). We wish to test the null hypothesis vs at 5% level of significance.

1. What test statistic should you use?
2. Find the value of the test statistic and p-value for the test.
3. What conclusion can you reach regarding the test?

Solution：

**/\*Ref: Act 8 SAS solution BioSTAT class;\*/**

**data** hw8pt1;

input id before after;

diff=before-after;

cards;

3082 168 145

3354 159 146

3373 197 169

3399 250 202

3415 195 146

3423 197 178

3513 244 208

3630 192 184

3725 236 215

3746 205 178

3920 145 155

3965 178 137

;

**run**;

\*ods select Statistics ConfLimits TTests;

ods trace on;

**proc** **ttest** data= hw8pt1 alpha=**0.05**;

paired before\*after;

title "Results with paired t for hw 8.1";

**run**;

ods trace off;

**proc** **print** data=hw8pt1 (drop=id diff) noobs;**run**;

ods trace on;

\*ods select Statistics ConfLimits TTests;

**proc** **ttest** data=hw8pt1 alpha=**0.05**;

var diff;

title "Results with univariate ttest for hw 8.1";

**run**;

ods trace off;

**Homework 8.2: Equal variance t-test**

One basis of two independent samples of sizes n1=45, n2=35, the following sample measurements are available:

Sample 1: mean=49.25, sd=12.5

Sample 2: mean=48.21, sd=11.3

Assume that the variances of the two populations samples come from are equal and unknown.

1. Find an estimate of the pooled standard deviation .
2. Find the margin of error for the 95% CI estimate of .
3. Find a 95% confidence interval for the difference of the two population means.
4. We wish to test vs and at 5% level of significance.
5. What test should you use?
6. Find the value of the test statistic.
7. Find the p-value for the test.
8. What is your conclusion about the test?

Solution:

**data** hw8pt2;

input n1 n2 x1bar x2bar s1 s2;

df=n1+n2-**2**;

sp=sqrt(((n1-**1**)\*s1\*\***2**+(n2-**1**)\*s2\*\***2**)/df);

se=sp\*sqrt(**1**/n1+**1**/n2);

me=quantile("t",**0.975**,df)\*se;

ciL=(x1bar-x2bar)-me;

ciU=(x1bar-x2bar)+me;

t=(x1bar-x2bar)/se;

p=**2**\*cdf("t",-abs(t),n1+n2-**2**);

keep df sp me ciL ciU t p;

cards;

45 35 49.25 48.21 12.5 11.3

;

**run**;

**proc** **print**;

title "Results for hw 8.2";**run**;

**Homework 8.3: Unequal variance t-test**

In reference to the activity 12.2, assume that the variances of the two populations samples come from are unequal and unknown.

1. Find a 95% confidence interval for the mean difference in the mean heart rate between Caucasian and African American newborns under the new assumption.
2. We wish to test vs and at 5% level of significance.
3. What test should you use?
4. Find the value of the test statistic.
5. Find the p-value for the test.
6. What is your conclusion about the test?

Solution:

**data** hw8pt3;

input n1 n2 x1bar x2bar s1 s2;

num=(s1\*\***2**/n1+s2\*\***2**/n2)\*\***2**;

den1=(s1\*\***2**/n1)\*\***2**/(n1-**1**);

den2=(s2\*\***2**/n2)\*\***2**/(n2-**1**);

den=den1+den2;

dfa=round(num/den);

se=sqrt(s1\*\***2**/n1+s2\*\***2**/n2);

me=quantile("t",**0.975**,dfa)\*se;

t=(x1bar-x2bar)/se;

ciL=(x1bar-x2bar)-me;

ciU=(x1bar-x2bar)+me;

p=**2**\*cdf("t",-abs(t),dfa);

keep dfa se me ciL ciU t p;

cards;

218 156 125 133 11 12

;

**run**;

**proc** **print**;

title "Results for hw 8.3";**run**;

**Homework 8.4: F-test for equality of variances**

In reference to the Activity 12.2, we wish to test the null hypothesis vs at 5% level of significance.

1. What test statistic should you use?
2. Find the value of the test statistic and p-value for the test.
3. What conclusion can you reach regarding the test?

Solution:

**data** hw8pt4;

input n1 n2 s1 s2 ;

f=s1\*\***2**/s2\*\***2**;

pL=cdf("f",f,n1-**1**,n2-**1**);

pR=**1**-pL;

pT=**2**\*min(pL,pR);

f2=s2\*\***2**/s1\*\***2**;

pL2=cdf("f",f2,n2-**1**,n1-**1**);

pR2=**1**-pL;

pT2=**2**\*min(pL,pR);

keep f pT f2 pT2;

cards;

218 156 11 12

**run**;

**proc** **print**;

title "Results for hw 8.4";

**run**;

**Homework 8.5: Estimation and Test Observed Data**

The desirable levels of an HDL (high-density lipoprotein), known as good cholesterol, for healthy men 20 or over is 40 mg/dL or higher, while the desirable levels for women 20 or over is 50mg/dL or higher. Two random samples from male and female patients treated in a health facility are considered to compare their HDL levels. Data from the two samples are reported at the end of questions below:

1. Find a 95% CI estimates for the mean difference in the levels of cholesterol between male and female population over 20 or more. Assume that the two population variances are unknown and equal.
2. Find a 95% CI estimates for the mean difference in the levels of cholesterol between male and female population over 20 or more. Assume that the two population variances are unknown and unequal.
3. Test vs at 5% level of significance assuming that the two population variances are equal.
4. Test vs at 5% level of significance.
5. Test vs at 5% level of significance assuming that the two population variances are unknown and unequal.
6. Test vs at 10% level of significance assuming that the two population variances are unknown and unequal.

| **id** | **hdl** | **gender** |
| --- | --- | --- |
| 5532 | 57 | Male |
| 5425 | 60 | Male |
| 5907 | 62 | Male |
| 5276 | 62 | Male |
| 5887 | 60 | Male |
| 5818 | 62 | Male |
| 5602 | 62 | Male |
| 7384 | 78 | Female |
| 7708 | 70 | Female |
| 7769 | 71 | Female |
| 7378 | 73 | Female |
| 7797 | 63 | Female |
| 7757 | 66 | Female |
| 7702 | 69 | Female |
| 7387 | 72 | Female |

Solution(a,b,c):

data hw8pt5;  
input n1 n2 x1bar x2bar s1 s2;  
df=n1+n2-2;  
sp=sqrt(((n1-1)\*s1\*\*2+(n2-1)\*s2\*\*2)/df);  
se\_a=sp\*sqrt(1/n1+1/n2);  
me\_a=quantile("t",0.975,df)\*se\_a;  
ciL\_a=(x1bar-x2bar)-me\_a;  
ciU\_a=(x1bar-x2bar)+me\_a;  
t=(x1bar-x2bar)/se\_a;  
p\_a=2\*cdf("t",-abs(t),n1+n2-2);  
  
num=(s1\*\*2/n1+s2\*\*2/n2)\*\*2;  
den1=(s1\*\*2/n1)\*\*2/(n1-1);  
den2=(s2\*\*2/n2)\*\*2/(n2-1);  
den=den1+den2;  
dfa=round(num/den);  
se\_b=sqrt(s1\*\*2/n1+s2\*\*2/n2);  
me\_b=quantile("t1",0.975,dfa)\*se\_b;  
t1=(x1bar-x2bar)/se\_b;  
  
keep me\_a me\_b t p\_a;  
cards;  
7 8 60.71 70.25 1.89 4.53  
;  
run;  
proc print;  
title "Results for hw 8.5";run;

(e.f)

data hw8pt5;  
input n1 n2 x1bar x2bar s1 s2;  
df=n1+n2-2;  
sp=sqrt(((n1-1)\*s1\*\*2+(n2-1)\*s2\*\*2)/df);  
se\_a=sp\*sqrt(1/n1+1/n2);  
me\_a=quantile("t",0.975,df)\*se\_a;  
ciL\_a=(x1bar-x2bar-3)-me\_a;  
ciU\_a=(x1bar-x2bar-3)+me\_a;  
t2=(x1bar-x2bar-3)/se\_a;  
p\_a=2\*cdf("t2",-abs(t),n1+n2-2);  
  
keep me\_a me\_b t3 p\_a;  
cards;  
7 8 60.71 70.25 1.89 4.53  
;  
run;  
proc print;  
title "Results for hw 8.5";run;

solution: in R

[a,c]

Male <- c(57,60,60,62,62,62,62,62)

Female <- c(63,66,69,70,71,72,73,78)

t.test(Male,Female, var.equal=T)

Two Sample t-test

data: Male and Female

t = -5.1728, df = 13, p-value = 0.0001794

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-13.518214 -5.553215

sample estimates:

mean of x mean of y

60.71429 70.25000

[b]

t.test(Male,Female, var.equal=F)

Welch Two Sample t-test

data: Male and Female

t = -5.4399, df = 9.62, p-value = 0.0003261

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-13.462460 -5.608969

sample estimates:

mean of x mean of y

60.71429 70.25000

[d]

var.test(Male,Female)

F test to compare two variances

data: male and female

F = 0.17422, num df = 6, denom df = 7, p-value = 0.04908

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.0340359 0.9922422

sample estimates:

ratio of variances

0.174216

p-value<0.05,so reject

[e]

t.test(Male,Female,mu=3,conf.level=0.95)

Welch Two Sample t-test

data: Male and Female

t = -7.1514, df = 9.62, p-value = 3.798e-05

alternative hypothesis: true difference in means is not equal to 3

95 percent confidence interval:

-13.462460 -5.608969

sample estimates:

mean of x mean of y

60.71429 70.25000

[f]

t.test(Male,Female,mu=3,conf.level=0.9)

Welch Two Sample t-test

data: Male and Female

t = -7.1514, df = 9.62, p-value = 3.798e-05

alternative hypothesis: true difference in means is not equal to 3

90 percent confidence interval:

-12.725582 -6.345846

sample estimates:

mean of x mean of y

60.71429 70.25000