University of New Mexico

Hypothesis Testing-6 (Spring 2018)

BIOM 505: Biostatistical Methods II (by Fares Qeadan)

HYPOTHESIS TESTS FOR ONE OR TWO POPULATIONS' VARIANCES OR STANDARD DEVIATIONS

Hypothesis Testing about a Single Population Variance (or Standard Deviation)[for sample statistics]: (Chi-Square test for the variance (or standard deviation))

The assumptions of this test are:

- The sample data is random.
- The observations are independent.
- In the population, the data follow the normal distribution.

The null and alternative hypotheses are:

$H_0 : \sigma^2 = \sigma_0^2.$
$H_1 :$ Either $\sigma^2 < \sigma_0^2$ or $\sigma^2 > \sigma_0^2$ or $\sigma^2 \neq \sigma_0^2$

REMARK: Sometimes, the variance (or standard deviation) may be more important than the mean. For example, health practitioners are not only interested in the average systolic or diastolic blood pressure of patients, but how the measurements vary. In particular, recent studies have shown that fluctuating blood pressure levels could be a better warning signal that a patient is t risk of a stroke than high average readings.

(1) **Problem 1:** In a pilot study of 25 patients, the sample mean of systolic blood pressure was found to be 126 mm HG with standard deviation of 17 mm HG. Please verify the claim, at the 5% significance level, that the population standard deviation of systolic blood pressures is higher than 15 mm HG? Assume that systolic blood pressures follow the normal distribution.

(a) The significance level α is:

(b) The null and alternative hypotheses are:

(c) The decision rule (about H_0) is:

(d) Conduct the test in SAS:

```
%MACRO std1(n,xbar,s,sigma0,alpha);
data _null_;
file print;
n = &n.; xbar = &xbar.; s = &s.; sigma0 = &sigma0.; alpha= &alpha.;
s2=s**2;
df =n - 1;
testchi = df*s2/(sigma0**2);
LeftSidedPvalue=CDF('CHISQ',testchi,DF);
RightSidedPvalue=SDF('CHISQ',testchi,DF);
TwoSidedPvalue=2*min(LeftSidedPvalue,RightSidedPvalue);
chilower = QUANTILE('CHISQ',alpha/2,df);
chiupper = QUANTILE('CHISQ',1-alpha/2,df);
SigmaLL=sqrt(df * s2/chiupper);
SigmaUL=sqrt(df * s2/chilower);
Sigma2LL=df * s2/chiupper;
Sigma2UL=df * s2/chilower;
sigmaCI=cat("(",SigmaLL,"-",SigmaUL,")");
sigma2CI=cat("(",Sigma2LL,"-",Sigma2UL,")");
put '===== Test for one sample standard deviation =====';
put 'n = ' n ' xbar =' xbar ' s =' s ;
put 'sigma0 = ' sigma0;
put 'Chi square test statistic = ' testchi;
put 'Pr > |chi|: ' TwoSidedPvalue pvalue.;
put 'Pr > chi: ' RightSidedPvalue pvalue.;
put 'Pr < chi: ' LeftSidedPvalue pvalue.;
put 'C.I about sigma'  sigmaCI;
put 'C.I about sigma2'  sigma2CI;
%MEND std1;

%std1(25,126,17,15,0.05);

===== Test for one sample standard deviation =====
n = 25  xbar =126  s =17
sigma0 = 15
Chi square test statistic = 30.826666667
Pr > |chi|: 0.3177
Pr > chi: 0.1588
Pr < chi: 0.8412
C.I about sigma(13.274082294-23.649586848)
C.I about sigma2(176.20126074-559.30295806)
```

(e) Get the p-value from the SAS output:

(f) Decision:

(g) Conclusion:

(h) Were the assumptions of this test satisfied? Explain.

Hypothesis Testing about a Single Population Variance (or Standard Deviation)[for sample data]: (Chi-Square test for the variance (or standard deviation))

(2) **problem 2:** Consider the Diabetes and obesity, cardiovascular risk factors data set we have used in Lab 1 (link: `http://www.mathalpha.com/lab1/diabetesfall17.sas7bdat`)) to test whether the true population standard deviation of systolic blood pressure is higher than 21 mm HG?

(a) The significance level α is:

(b) The null and alternative hypotheses are:

(c) The decision rule (about H_0) is:

(d) Conduct the test in SAS:

```
proc summary data=biom505.diabetesfall17 print;
var bp_1s;
run;
```

The SUMMARY Procedure

Analysis Variable : bp_1s bp_1s				
N	Mean	Std Dev	Minimum	Maximum
398	136.9045226	22.7410332	90.0000000	250.0000000

```
%std1(398, 136.9045226, 22.7410332,21,0.05)

===== Test for one sample standard deviation =====
n = 398  xbar =136.9045226  s =22.7410332
sigma0 = 21
Chi square test statistic = 465.55640052
Pr > |chi|: 0.0199
Pr > chi: 0.0099
```

```
Pr < chi: 0.9901
C.I about sigma(21.263320109-24.441165759)
C.I about sigma2(452.12878208-597.37058363)
```

(e) Get the p-value from the SAS output:

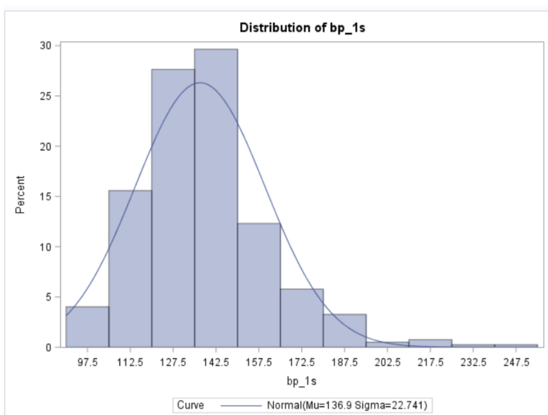
(f) Decision:

(g) Conclusion:

(h) Were the assumptions of this test satisfied? Explain.

```
proc capability data=biom505.diabetesfall17 graphics normaltest;
var bp_1s;
histogram bp_1s/normal;
run;
```

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.937664	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.129946	Pr > D	<0.0100
Cramer-von Mises	W-Sq	0.882882	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	5.222324	Pr > A-Sq	<0.0050



Hypothesis Testing about Two Population Variances (or Standard Deviations)[for sample data]: (F-test for Equality of Two Variances (or homogeneity of variances))

The assumptions of this test are:

- The two samples are random.
- The observations are independent.
- In each population, the data follow the normal distribution.

The null and alternative hypotheses are:

$$H_0 : \sigma_1^2 = \sigma_2^2.$$
$$H_1 : \text{Either } \sigma_1^2 < \sigma_2^2 \qquad \text{or} \qquad \sigma_1^2 > \sigma_2^2 \qquad \text{or} \qquad \sigma_1^2 \neq \sigma_2^2$$

OR

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1.$$
$$H_1 : \text{Either } \frac{\sigma_1^2}{\sigma_2^2} < 1 \qquad \text{or} \qquad \frac{\sigma_1^2}{\sigma_2^2} > 1 \qquad \text{or} \qquad \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

(3) **problem 3:** Consider the Diabetes and obesity, cardiovascular risk factors data set we have used in Lab 1 (link: `http://www.mathalpha.com/lab1/diabetesfall17.sas7bdat`) to test whether the true population standard deviation of systolic blood pressure among females is higher than that of males?

(a) The significance level α is:

(b) The null and alternative hypotheses are:

(c) The decision rule (about H_0) is:

(d) Conduct the test in SAS:

```
proc ttest data=biom505.diabetesfall17;
class gender;
var bp_1s;
run;
```

```
proc glm data=biom505.diabetesfall17;
class gender;
model bp_1s = gender;
MEANS gender/ HOVTEST=BARTLETT welch;
MEANS gender/ HOVTEST=BF welch;
MEANS gender/ HOVTEST=LEVENE(TYPE=ABS)welch;
run;
```

The TTEST Procedure

Variable: bp_1s (bp_1s)

gender	N	Mean	Std Dev	Std Err	Minimum	Maximum
female	231	136.3	24.9280	1.6401	90.0000	250.0
male	167	137.7	19.3557	1.4978	100.0	199.0
Diff (1-2)		-1.4232	22.7588	2.3117		

gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
female		136.3	133.1	139.5	24.9280	22.8432	27.4348
male		137.7	134.8	140.7	19.3557	17.4787	21.6879
Diff (1-2)	Pooled	-1.4232	-5.9679	3.1215	22.7588	21.2782	24.4626
Diff (1-2)	Satterthwaite	-1.4232	-5.7899	2.9436			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	396	-0.62	0.5385
Satterthwaite	Unequal	393.96	-0.64	0.5221

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	230	166	1.66	0.0006

The GLM Procedure

Brown and Forsythe's Test for Homogeneity of bp_1s Variance ANOVA of Absolute Deviations from Group Medians					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
gender	1	1316.6	1316.6	5.65	0.0179
Error	396	92299.0	233.1		

Bartlett's Test for Homogeneity of bp_1s Variance			
Source	DF	Chi-Square	Pr > ChiSq
gender	1	11.8641	0.0006

The GLM Procedure

Levene's Test for Homogeneity of bp_1s Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
gender	1	1345.3	1345.3	5.83	0.0162
Error	396	91418.6	230.9		

(e) Get the p-value from the SAS output:

(f) Decision:

(g) Conclusion:

(h) Were the assumptions of this test satisfied? Explain.

```
 proc sort data=biom505.diabetesfall17;
 by gender;
 run;
proc capability data=biom505.diabetesfall17 graphics normaltest;
BY gender;
var bp_1s;
histogram bp_1s/normal;
run;
```

The CAPABILITY Procedure
Fitted Normal Distribution for bp_1s (bp_1s)

gender=female

Parameters for Normal Distribution		
Parameter	Symbol	Estimate
Mean	Mu	136.3074
Std Dev	Sigma	24.928

Goodness-of-Fit Tests for Normal Distribution					
Test	Statistic		DF	p Value	
Kolmogorov-Smirnov	D	0.141284		Pr > D	<0.010
Cramer-von Mises	W-Sq	0.661989		Pr > W-Sq	<0.005
Anderson-Darling	A-Sq	4.119306		Pr > A-Sq	<0.005
Chi-Square	Chi-Sq	337.053865	6	Pr > Chi-Sq	<0.001

(i) Levene (1960) proposed a test statistic for equality of variances that was found to be robust under non-normality. Please use Levene's test to resolve the violation of the assumptions of the F-test.

Hypothesis Testing about Two Population Variances (or Standard Deviations)[for sample statistics]: (F-test for Equality of Two Variances (or homogeneity of variances))

(4) **problem 4:** In a study from 2002 in the US, the sample BMI mean was found to be 28 kg/m^2 with a standard deviation of 3 kg/m^2 among 30 adults of the age 30 years or more while it was found to be 21 kg/m^2 with a standard deviation of 4 kg/m^2 among 45 adults of the ages between 18 and 29 years old inclusive. Assuming that the data is normally distributed in the underlying two populations, test the claim that the true population standard deviation of BMI among adults of the age 30 years or more is less than that of adults of the ages between 18 and 29 years old inclusive

(a) The significance level α is:

(b) The null and alternative hypotheses are:

(c) The decision rule (about H_0) is:

(d) Conduct the test in SAS:

```
%MACRO std2(n1,xbar1,s1,n2,xbar2,s2,alpha);
data _null_;
file print;
n1 = &n1.; xbar1 = &xbar1.; s1 = &s1.; n2 = &n2.; xbar2 = &xbar2.; s2 = &s2.; alpha= &alpha.;
v1=s1**2; v2=s2**2;
df1 =n1 - 1; df2 =n2 - 1;
STATISTIC= v1/ v2;
PVAL =CDF('F',STATISTIC,df1, df2);
LeftSidedPvalue=PVAL;
RightSidedPvalue=1 - PVAL;
TwoSidedPvalue=2 * min(PVAL, 1 - PVAL);
upper=STATISTIC / QUANTILE('F',alpha/2,df1,df2);
lower=STATISTIC / QUANTILE('F',1-alpha/2,df1,df2);
sqlower=sqrt(lower);
squpper=sqrt(upper);
sigmaRatioCI=cat("(",sqlower,"-",squpper,")");
varRatioCI=cat("(",lower,"-",upper,")");
```

```
put '===== Test for two samples standard deviation =====';
put 'n1 = ' n1 ' xbar1 =' xbar1 ' s1 =' s1 ;
put 'n2 = ' n2 ' xbar2 =' xbar2 ' s2 =' s2 ;
put 'F test statistic = ' STATISTIC;
put 'Pr > |F|: ' TwoSidedPvalue pvalue.;
put 'Pr > F: ' RightSidedPvalue pvalue.;
put 'Pr < F: ' LeftSidedPvalue pvalue.;
put 'C.I about sigma1/sigma2     '  sigmaRatioCI;
put 'C.I about var1/var2    '  varRatioCI;
%MEND std2;

%std2(30,28,3,45,21,4,0.05);

===== Test for two samples standard deviation =====
n1 = 30  xbar1 =28  s1 =3
n2 = 45  xbar2 =21  s2 =4
F test statistic = 0.5625
Pr > |F|: 0.1045
Pr > F: 0.9477
Pr < F: 0.0523
C.I about sigma1/sigma2     (0.5417013041-1.0631217342)
C.I about var1/var2    (0.2934403029-1.1302278217)
```

(e) Get the p-value from the SAS output:

(f) Decision:

(g) Conclusion:

(h) Were the assumptions of this test satisfied? Explain.