University of New Mexico

Hypothesis Testing-6 (Fall 2016)

PH 538: Public Health Biostatistical Methods I (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:

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

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) <u>The decision rule (about H_0) is:</u>

(d) <u>Conduct the test in STATA:</u>
sdtesti n $\bar{x}$ **s** σ_0

```
sdtesti 25 126 17 15
sdtesti 25 . 17 15
```

(e) <u>Get the p-value from STATA's output:</u>

(f) <u>Decision:</u>

(g) <u>Conclusion:</u>

(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/diabetesfall16.dta) to test whether the true population standard deviation of systolic blood pressure is higher than 15 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 STATA:

```
univar bp_1s
sdtest bp_1s==20
```

(e) Get the p-value from STATA's output:

(f) Decision:

(g) Conclusion:

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

```
hist bp_1s, normal
swilk bp_1s
```

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/diabetesfall16.dta) 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 STATA:

```
sort gender
sdtest bp_1s, by(gender)
```

(e) Get the p-value from STATA's output:

(f) Decision:

(g) Conclusion:

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

```
hist bp_1s if gender=="female", normal
swilk bp_1s if gender=="female"

hist bp_1s if gender=="male", normal
swilk bp_1s if gender=="male"
```

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

```
robvar bp_1s, by(gender)
```

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 STATA:
sdtesti n_1 $\bar{x}_1$ s_1 n_2 $\bar{x}_2$ s_2

```
sdtesti 30 28 3 45 21 4
sdtesti 30 . 3 45 . 4
```

(e) Get the p-value from STATA's output:

(f) Decision:

(g) Conclusion:

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

REMARK 1: For paired variances one could use either *sdpair* or *sdtest*. For example, to test if the population variance of systolic blood pressure is the same as that of diastolic blood pressure one could use either of the followings:

```
findit sdpair
sdpair bp_1s bp_1d
sdtest bp_1s==bp_1d
```

What is the conclusion from this test?

REMARK 2: To get the $(1-\alpha)\times 100$ confidence interval about σ, one could use the command $cisd1$. For example, to find the 95% confidence interval about the true population standard deviation of systolic blood pressure one could use the followings: [1]

```
findit cisd1
univar bp_1s
cisd1 bp_1s
```

Please give an interpretation for the found confidence interval?

[1]To get the $(1-\alpha)\times 100$ confidence interval about σ^2, firstly find the C.I. for σ by using the command $cisd1$ and secondly square the found results.