Politecnico di Milano
STATISTICS (079086), 2009/2010, Prof. A.Barchielli
Problem set n. 9

4th June 2010

Exercise 1 The quantity of sugar included in a pear juice is normally distributed. A random sample of $n = 10$ pear juices has sample variance $s^2 = 23.04\ mg^2$.

1. Verify $H_0 : \sigma^2 = 18\ mg.^2$ vs $H_1 : \sigma^2 \neq 18\ mg.^2$ at level 5%. *[We know that $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$. Thus $P_{H_0}\big(\chi^2_{\frac{\alpha}{2},n-1} \leq (n-1)\frac{S^2}{\sigma_0^2} \leq \chi^2_{1-\frac{\alpha}{2},n-1}\big) = 1-\alpha$. Hence, we reject H_0 if $s^2 < \chi^2_{\frac{\alpha}{2},n-1}\ \frac{\sigma_0^2}{n-1}$ or if $s^2 > \chi^2_{1-\frac{\alpha}{2},n-1}\ \frac{\sigma_0^2}{n-1}$. With the data above, we do not reject H_0 at level 5% ($\chi^2_{0.025,9} = 2.7$, $\chi^2_{0.975,9} = 19.023$)]*

2. Compute the bilateral confidence interval at level 95%. *[We know that $1-\alpha = P\big(\chi^2_{\frac{\alpha}{2},n-1} \leq (n-1)\frac{S^2}{\sigma^2} \leq \chi^2_{1-\frac{\alpha}{2},n-1}\big) = P\big((n-1)S^2\frac{1}{\chi^2_{1-\frac{\alpha}{2},n-1}} \leq \sigma^2 \leq (n-1)S^2\frac{1}{\chi^2_{\frac{\alpha}{2},n-1}}\big)$; thus, the confidence interval is given by $\big((n-1)s^2\frac{1}{\chi^2_{1-\frac{\alpha}{2},n-1}}; (n-1)s^2\frac{1}{\chi^2_{\frac{\alpha}{2},n-1}}\big) = (10.9; 76.8)$]*

Exercise 2 [Mood, Graybill, Boes, n. 25 page 476][1]

A cigarette farm sent two samples of cigarettes to two different laboratories. Each laboratory observed the quantity of nicotine (in milligrams) contained in 5 cigarettes. The first laboratory found: (*i*) 24, 27, 26, 21, 24, the second: (*ii*) 27, 28, 23, 31, 26. Assume normality of the distribution of the quantity of nicotine.

1. Check that the variances of the populations can be considered as equal, at level 10%. [$s_x^2 = 5.3$, $s_y^2 = 8.5$. Under H_0, $\frac{S_y^2}{S_x^2} \sim F_{n_y,n_x}$. Since $\frac{s_y^2}{s_x^2} = 1.6$ and $F_{4,4,0.95} = 6.39$ and $F_{4,4,0.05} = 0.16$, we do not reject H_0. (P-value=0.6584)]

2. Do the data allow you to say that the two samples contain the same tobacco, at level 5%? [yes, t-pooled test; $\bar{x} = 24.4$, $\bar{y} = 27$, $t_0 = 1.565$, $t_{8,0.05} = 1.86$, p-value$\in (0.1, 0.2)$]

Exercise 3 In the production process of the semiconductors, in order to remove the silicon from the back of the wafers before the metallization, it is often used a technique known as "chemical and humid attack". In this process, the attack velocity is an essential characteristic and it is known to be normally distributed. Two different attack solutions have been compared. Precisely, the attack velocities (in ml/min) on a sample of 10 wafers exposed to solution 1 have been measured:

$$\text{Solution 1}: 9.9 \quad 9.4 \quad 9.3 \quad 9.6 \quad 10.2 \quad 10.6 \quad 10.3 \quad 10.00 \quad 10.3 \quad 10.1;$$

and the attack velocities (in ml/min) on a sample of 10 wafers exposed to solution 2 have been measured:

$$\text{Solution 2}: 10.2 \quad 10.6 \quad 10.7 \quad 10.4 \quad 10.5 \quad 10.00 \quad 10.2 \quad 10.7 \quad 10.4 \quad 10.3.$$

1. Assuming normality, check that the variances of the populations can be considered as equal, at level 5%. [$s_x^2 = 0.178$, $s_y^2 = 0.053$. Under H_0, $\frac{S_x^2}{S_y^2} \sim F_{n_x,n_y}$. Since $\frac{s_x^2}{s_y^2} = 3.335$ and $F_{9,9,0.975} = 4.03$, we do not reject H_0.]

2. Do the data allow you to say that the mean attack velocity is the same for both solutions, at level 5%? [no, t-pooled test; $\bar{x} = 9.97$, $\bar{y} = 10.4$, $t_0 = 2.8278$]

[1] Mood, A. M., Graybill, F. A., Boes, D. C. (1988) *Introduzione alla statistica* Milano: McGraw-Hill libri Italia.

3. Calculate the p-value of the test in 2. $[0.01 < p - value < 0.025]$
4. Construct a 95% confidence interval for the difference of the mean attack velocities. $[(-0.7495, -0.1105)]$

Exercise 4 A building firm needs to buy special metallic bars. Two suppliers, Supplier 1 and Supplier 2, produce this type of bars and they both declare that the length of the bars is 127 mm. Two samples, each of size 10, of the lengths of the bars produced by the two suppliers are taken: the sampling standard deviation of the lengths of the bars produced by Supplier 1 is $s_1 = 0.13$ mm., while the sampling standard deviation of the lengths of the bars produced by Supplier 2 is $s_2 = 0.17$ mm.

1. Assuming that the lengths are normally distributed, can you say at significance level $\alpha = 0.1$ that Supplier 2 is better than Supplier 1? (in the sense that the lengths of the bars produced by 2 are less variable than the lengths of the bars produced by 1?) [no]

Exercise 5 We have recorded the logarithmic repairing times of 7 faults occurring to photocopiers of type A:

$$3.466 \quad 4.431 \quad 3.611 \quad 3.738 \quad 4.357 \quad 4.127 \quad 4.078$$

We can assume that these data are a random sample from a normal distribution with unknown mean μ_A and unknown variance σ_A^2.

1. Construct a confidence interval of level 95% for σ_A^2.

We now also have the logarithmic repairing times of 6 faults occurring to photocopiers of a different type, that we shall call type B:

$$3.664 \quad 4.710 \quad 4.007 \quad 4.663 \quad 4.500 \quad 4.466$$

We can assume that these data are a random sample from a normal distribution with unknown mean μ_B and unknown variance σ_B^2.

2. Verify the null hypothesis that the variances σ_A^2 and σ_B^2 can be considered equal, at significance level 10%.
3. On the basis of the result to point 2, verify the null hypothesis that the means μ_A and μ_B can be considered equal, at level 10%.

Solution.

1. We have
$$s_A^2 = \frac{\sum_{i=1}^7 x_i^2 - 7\bar{x}^2}{6} = \frac{111.305 - 7 \times 3.973^2}{6} \simeq 0.139.$$
The confidence interval is given by $\left((n_A - 1)s_A^2/\chi^2_{0.975,6}; (n-1)s_A^2/\chi^2_{0.025,6}\right)$, where $\chi^2_{0.975,6} = 14.449$ and $\chi^2_{0.975,6} = 1.237$. We thus get the confidence interval (0.058,0.068).

2. We have
$$\bar{x}_A = 3.973, \ \bar{x}_B = 4.335, \ s_b^2 = \frac{\sum_{i=1}^6 y_i^2 - 6\bar{y}^2}{6} = \frac{113.604 - 6 \times 4.335^2}{5} \simeq 0.170.$$
We want to verify $H_0 : \sigma_A^2 = \sigma_B^2$ versus the alternative $H_1 : \sigma_A^2 \neq \sigma_B^2$. We know that under H_0, $\frac{S_B^2}{S_A^2} \sim F_{n_B, n_A}$. Since $\frac{s_B^2}{s_A^2} = 1.222$ and $F_{0.95,5,6} = 4.387$, we do not reject the null hypothesis at significance level 10%.

3. Given the result to the previous point, we can assume $\sigma_A^2 = \sigma_B^2 = \sigma^2$, and we can estimate the common variance by the pooled sample variance
$$s_p^2 = \frac{6\, s_A^2 + 5\, s_B^2}{7 + 6 - 2} = \frac{0.139 \times 6 + 0.170 \times 5}{6 + 5} \simeq 0.153.$$
We want to verify $H_0 : \mu_A = \mu_B$ versus the alternative $H_1 : \mu_A \neq \mu_B$. We know that under H_0,
$$\frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2 \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$
has a t-student distribution with $n_A + n_B - 2 = 11$ degrees of freedom. Since $\left|(\bar{x}_A - \bar{x}_B) / \sqrt{s_p^2 \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}\right| = 1.66 < t_{0.95,11} = 1.796$, we do not reject the null hypothesis at significance level 10%.