Politecnico di Milano
STATISTICS (079086), 2009/2010, Prof. A.Barchielli
Problem set n. 6

6th May 2010

Exercise 1 The production manager of a firm that produces pasta has to decide wether to revision its production plant. To take the decision, he verifies the reliability of the weight declared on the spaghetti cardboard boxes (500g). A random sample of 60 cardboard boxes gives a sample mean of 496g. The production manager supposes that the quantity of spaghetti contained in one cardboard box is normally distributed with mean μ and variance 200g^2 (known from previous surveys).

(a) What decision will the production manager take? Use tests with significance levels: 1%, 5% and 10%. *[We have $X \sim N(\mu, \sigma^2)$, with σ known; given the sample $x_1, \dots, x_n$, with sample mean $\bar{x}$, we want to test $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. The test lead us to reject H_0 at significance level α if $|\bar{x} - \mu_0| > z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, where $z_{\frac{\alpha}{2}}$ is such that $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$, where Z is a standard gaussian random variable. With the data above ($\sigma^2 = 200$, $n = 60$, $\bar{x} = 496$, $\mu_0 = 500$), we do not reject H_0 at level 1% ($z_{0.005} = 2.58$), but we reject it at levels 5% ($z_{0.025} = 1.96$) and 10% ($z_{0.05} = 1.64$).]*

(b) Compute the minimum significance level that would lead the production manager to revision the production plant. *[We reject H_0 if $\frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} > z_{\frac{\alpha}{2}}$. With the data above, the smallest $z_{\frac{\alpha}{2}}$ that leads to rejection of H_0 is $z_{\frac{\alpha}{2}} = 2.19$, i.e. α=2.8%]*

(c) For what value of $\mu \in \{490, 493.5, 496, 497.1\}$ the probability of type II error is greater? *[The type II error is made when we do not reject H_0 even if H_0 is false. Its probability is greater when the true value of μ is close to μ_0. Formally, we need to verify for what value of μ, if $X \sim N(\mu, \sigma^2)$ and thus $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, $P_\mu(|\bar{X} - \mu_0| \leq z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = P(\frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}} - z_{\frac{\alpha}{2}} \leq Z \leq \frac{\mu_0-\mu}{\frac{\sigma}{\sqrt{n}}} + z_{\frac{\alpha}{2}})$ is greater. We obtain $\mu = 497.1$]*

Now suppose that the same data are obtained in a survey carried out by a consumer association.

(d) Can the consumer association conclude, at level 1%, that the cardboard boxes contain less than what is declared? *[In this case the alternative is: $H_1 : \mu < \mu_0$. The threshold of the critical region is $z_{0.01} = 2.33$. Hence, it is not possible to reject H_0 at level 1%.]*

Exercise 2 In order to control the precision of a multimeter, an engineer measures 5 times the potential difference between the two poles of a generator of a known potential μ_0: $\mu_0 = 10^{-1}$V. We assume the results $X_1, X_2, \dots, X_5$ of these measurements be a gaussian random sample with known mean μ_0 and unknown variance σ^2. The actual results of the 5 measurements are $x_1 = 0.997055 \times 10^{-1}$, $x_2 = 0.996954 \times 10^{-1}$, $x_3 = 0.999521 \times 10^{-1}$, $x_4 = 1.00664 \times 10^{-1}$ and $x_5 = 1.00261 \times 10^{-1}$.

The engineer declares the multimeter to be "precise" if $\sigma \leq \sigma_0 = 0.5 \times 10^{-2}$V, otherwise the multimeter is not precise.

1. Determine an unilateral confidence interval of the form $(c, +\infty)$ (that is a confidence lower bound) for the standard deviation σ, with confidence level 95%.
2. Verify the null hypothesis H_0: "the multimeter is precise", versus H_1: "the multimeter is not precise" with a test of significance level 5%.
3. Find the expression of the power function of the test at point 2.
4. Find approximately the probability of second type error in $\sigma = 0.6 \times 10^{-2}$ for the test in point 2, or, eventually a numerical interval for this probability.

Solution.

1. We have $\frac{1}{\sigma^2}\sum_{i=1}^5 (X_i-\mu_0)^2 \sim \chi^2(5)$. It follows that the CI of level 95% for σ^2 is given by

$$\left(\frac{\sum_{i=1}^5 (x_i-\mu_0)^2}{\chi^2_{.95}(5)}\,,+\infty\right),$$

and the numerical value of the lower bound is $\frac{\sum_{i=1}^5 (X_i-\mu_0)^2}{\chi^2_{.95}(5)} = \frac{0.690825\times 10^{-8}}{11.07} = 0.062405\times 10^{-8}$. Then, the confidence interval for σ is $(\sqrt{0.062405\times 10^{-8}}\,,+\infty) = (0.24981\times 10^{-4}\,,+\infty)$.

2. The two hypotheses are: $H_0:\sigma\le\sigma_0 = 0.5\times 10^{-2}$ versus $H_1:\sigma>\sigma_0 = 0.5\times 10^{-2}$.
The critical region is given by CR $=\left\{\sum_{i=1}^5 (X_i-\mu_0)^2 > \sigma_0^2\chi^2_{.95}(5)\right\}$. In other terms, we reject the null hypothesis if $\sigma_0 < \sqrt{\frac{\sum_{i=1}^5 (x_i-\mu_0)^2}{\chi^2_{.95}(5)}} = 0.24981\times 10^{-4}$. But this is false and we cannot reject H_0 at significance level 5%. The same conclusion could be obtained by using the duality CI/hypotheses tests.

3. Let us set $S_0^2 = \frac{1}{5}\sum_{i=1}^5 (X_i-\mu_0)^2$. Then, the power function is given by

$$\pi(\sigma) = P_\sigma\left(0.5\times 10^{-2}\le\sqrt{5S_0^2/11.07}\right) = P_{\sigma^2}\left(\frac{5S_0^2}{\sigma^2}\ge\frac{0.25\times 10^{-4}\times 11.07}{\sigma^2}\right) =$$
$$= 1 - F_{\chi^2_5}\left(\frac{2.7675\times 10^{-4}}{\sigma^2}\right),\quad \sigma > 0.5\times 10^{-2},$$

which is an increasing function of σ and $\lim_{\sigma\to+\infty}\pi(\sigma) = 1$.

4. $\beta(0.6\times 10^{-2}) = 1-\pi(0.6\times 10^{-2}) = F_{\chi^2_5}\left(\frac{2.7675\times 10^{-4}}{0.36\times 10^{-4}}\right) = F_{\chi^2_5}(7.6875) \gtrsim F_{\chi^2_5}(7.289) = 80\%$

Exercise 3 Let $X_1,\ldots,X_n$ be a random sample from a normal distribution with unknown mean μ and known variance equal to 900.

1. Construct a test for the mean μ such that the probability of committing type I error, concluding that $\mu > 5500$ when instead it is true that $\mu\le 5500$, is at most 6%.
2. If we have 100 observations and the sample mean $\overline{X}_{100}$ is equal to 5506.0, what is the p-value of the test above? What decision should we take?
3. Compute the probability of taking the right decision on the base of the test constructed at point 1., if the true value of μ is 5505 and $n = 100$.
4. We can now collect further observations. Compute the minimum number of observations to be collected such that the power of the test when $\mu = 5505$ increases by at least 50%.

Solution. We shall construct a unilateral z-test for the mean μ of a population with distribution $\mathcal{N}(\mu, 900)$, based on the random sample $X_1,\ldots,X_n$ extracted from this population.

1. We shall construct a test with significance level $\alpha = 6\%$ for the null hypothesis $H_0:\mu\le 5500$ vs the alternative $H_1:\mu > 5500$. We reject H_0 in favor of H_1 at level 6% if $\sqrt{n}(\overline{X}_n - 5500)/30 \ge z_{94\%}$, where $z_{94\%}\simeq 1.555$, or equivalently, if $\overline{X}_n \ge 5500 + 1.555\times 30/\sqrt{n}$.

2. If $n = 100$ and $\overline{X}_{100} = 5506.0$ then the p-value of the test constructed at point 1. is given by

$$P_{5500}\left(\overline{X}_{100}\ge 5506.0\right) = 1-\Phi(2)\simeq 1-0.9778 = 2.28\%$$

Since $\alpha = 6\% > 2.28\%$ we reject $H_0:\mu\le 5500$ in favor of $H_1:\mu > 5500$.

3. For $H_0:\mu\le 5500$ *versus* $H_1:\mu > 5500$, the probability of taking the right decision when $\mu = 5505$ is given by the power of the test in $\mu = 5505$. If $n = 100$ then we reject $H_0:\mu\le 5500$ in favor of $H_1:\mu > 5500$ if $\overline{X}_n \ge 5504.665$ and the power of the test in $\mu = 5505$ is given by

$$P_{5505}\left(\overline{X}_{100}\ge 5504.665\right) = 1-\Phi\left(\frac{5504.665-5505}{3}\right) = \Phi(0.111667)\simeq\Phi(0.11) = 0.5438$$

4. Increasing by 50% the power of the test we go from 0.5438 to $0.5438 \times 1.5 = 0.8197$. We are thus looking for the smallest n such that

$$P_{5505}\left(\overline{X}_{100} \geq 5500 + 1.555\frac{30}{\sqrt{n}}\right) \geq 0.8157$$

i.e., such that

$$1 - \Phi\left(\frac{5500 - 5505}{30/\sqrt{n}} + 1.555\right) \geq 0.8157$$

that has solution $n \geq (6(1.555 + 0.90))^2 = 14.73^2$ (using the approximation $z_{0.8157} \simeq z_{0.8159} = 0.90$). We thus obtain $n \geq 217$, i.e., we have to collect further 117 observations.

Exercise 4 Let X be a population with normal distribution with mean μ and knwon variance $\sigma^2 = 4$. To verify $H_0 : \mu = 8$ vs $H_1 : \mu > 8$, on the base of a random sample with size $n = 16$, it is proposed to use a test with critical region

$$R = \{(x_1, \ldots, x_n) : \overline{X} > k\}$$

with $k = 8.82$.

a) Compute the probability of type I error; *[The type I error is made when we reject H_0 but H_0 is true: $P_{H_0}(\overline{X} > k) = P_{H_0}(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} > \frac{k-\mu}{\frac{\sigma}{\sqrt{n}}}) = P(Z > \frac{k-\mu}{\frac{\sigma}{\sqrt{n}}})$. With the data above we obtain $P(\text{type I error}) = 0.05$]*

b) Compute the power function of the test and the probability of type II error when the true value of μ is 10. *[The power function of the test is the probability of rejection of H_0 as a function of the true mean μ: $\pi(\mu) = P_\mu(\overline{X} > k)$. With the data above we obtain $\pi(\mu) = 1 - \Phi(2(8.82 - \mu)) = 0.99$. The probability of type II error, when the true value of μ is 10, is given by $P_{\mu=10}(\overline{X} < k) = 1 - \pi(10) = 0.01$]*