# Politecnico di Milano
# STATISTICS (079086), 2009/2010, Prof. A.Barchielli
# Problem set n. 7

13th May 2010

**Exercise 1** A winemaker fears that the concentration of alcohol in the new vintage is decreased. He thus measures the quantity of alcohol in five $1l$ bottles finding respectively 10, 12, 9.5, 10.1 and 10.5 $cl$.

a) Suppose that the quantity $X$ of alcohol in a $1l$ bottle of wine is normally distributed. Verify the hypothesis that the mean quantity of alcohol per litre is less than $11cl$, at 5% significance level. *[We have to verify $H_0 : \mu \geq 11$ vs $H_1 : \mu < 11$. We do not know $\sigma$ and thus we estimate it by $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$. Under the null hypothesis that $\mu =$ is the true value of the mean, then $T_0 = \frac{\bar{X}-11}{\frac{S}{\sqrt{n}}}$ has a t-student distribution with $n-1$ degrees of freedom. We thus reject $H_0$ if $t_0 < t_{\alpha,n-1}$, where $t_{\alpha,n-1}$ is the appropriate quantile of t-student distribution. With the data above, $\bar{x} = 10.42$, $s = 0.952$, $t_0 = -1.36$, $t_{0.05,4} = -2.132$. We do not reject $H_0$ at level 5%.]*

b) Evaluate the power of the test considered at point (a) when $\mu = 10.5$. *[power=$1 - \beta \in [75\%, 90\%]$.]* Is the power increasing or decreasing for smaller values of $\mu$?

c) Now suppose that the standard deviation of $X$ is known and equal to 2. Repeat the test. *[In this case the test statistic has the the standard gaussian distribution. We do not reject $H_0$ at level 5%.]*

d) Compute the p-values of the tests performed at points (a) and (c). *[In case (a) we obtain p-value$=P(T_4 > 1.362) \in (10\%, 25\%)$, where $T_4$ is a t-student random variable with 4 degrees of freedom. In case (c) we obtain p-value$= 25.8\%$. The p-values of the two tests are both greater than 5% (in fact, in both cases we did not reject $H_0$ at level 5%).]*

**Exercise 2** A consumer association is carrying out a survey about the life span of two types of batteries: Duracell and Energizer. The standard deviations are assumed to be $\sigma_D = 1.8$ hours for Duracell batteries and $\sigma_E = 2$ hours for Energizer batteries. The sample mean of the life span of 100 Duracell batteries is 4.1 hours, and the sample mean of the life span of 120 Energizer batteries is 4.5 hours.

(a) Can the consumer association conclude, at significance level 5%, that Energizers batteries are better than Duracell batteries? *[$H_0 : \mu_E = \mu_D$ vs $H_1 : \mu_D < \mu_E$. Reject $H_0$ if $\bar{x}_D - \bar{x}_E < z_\alpha\sqrt{\sigma_D^2/n_D + \sigma_E^2/n_E} = -1.64\sqrt{2^2/100 + 1.8^2/120} = -0.42$. The null hypothesis is not rejected at significance level 5%.]*

(b) Compute the p-value. *[0.06]*

(c) Compute a confidence interval at level 99% for the difference between the mean life spans. *[$\bar{x}_D - \bar{x}_E \pm z_{1-\alpha/2}\sqrt{\sigma_D^2/n_D + \sigma_E^2/n_E} = (-1.06; 0.26)$.]*

**Exercise 3** We want to compare fuel consumptions of two car models, VW Polo 1.0 and Citroen Saxo 1.1 SPI, at constant speed of 120 km/h. We can assume that the fuel consumptions of the two car models are normally distributed, with equal variances. In 20 runs, VW Polo 1.0 has an average consumption of 6.5 l/100Km. In 22 runs, Citroen Saxo 1.1 SPI has an average consumption of 6.6 l/100Km. The sample variances are respectively 0.3 and 0.28.

(a) Can we conclude, at significance level 5%, that the two car models have the same mean consumption? *[$X_1$ = Polo's consumption, $X_1 \sim N(\mu_1, \sigma_1^2)$; $X_2$ = Saxo's consumption, $X_2 \sim N(\mu_2, \sigma_2^2)$. Assuming $\sigma^2 = \sigma_1^2 = \sigma_2^2$, we can estimate $\sigma^2$ by $S_P^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{(20-1)(0.3)+(22-1)(0.28)}{20+22-2} = 0.285$. Thanks* *to the fact that* $\frac{|\bar{X}_1 - \bar{X}_2|}{S_p\sqrt{1/n_1 + 1/n_2}} \sim T_{n_1+n_2-2}$ *we construct the following test for* $H_0 : \mu_1 = \mu_2$ *vs* $\mu_1 \neq \mu_2$*: reject* $H_0$ *if* $|\bar{x}_1 - \bar{x}_2| > t_{\alpha/2, n_1+n_2-2}\, S_p\sqrt{1/n_1 + 1/n_2}$. *With the above data (*$t_{0.025,40} = 2.02$ *e* $S_p\sqrt{1/n_1 + 1/n_2} = 0.167$*), we do not reject* $H_0$ *at level 5%.]*

(b) Compute a confidence interval of level 95% for the difference between the mean consumptions. *[*$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2, n_1+n_2-2} S_p\sqrt{1/n_1 + 1/n_2}$
$= (-0.44, 0.24)$*]*