Soluzioni del Tutorato di Statistica 1 del 27/11/2010
Docente: Prof.ssa Enza Orlandi
Tutore: Dott.ssa Barbara De Cicco

Esercizio 1.
I carichi di rottura espressi il libbre di 5 campioni di corda di manila con un diametro di 3/16 di pollici sono risultati essere $(6.60, 4.60, 5.40, 5.80, 5.50)$.

1. Stimate il carico di rottura medio con un intervallo di confidenza al 95%, assumendo la normalità.
   Sia X la variabile aleatoria che individua i carichi di rottura. $X = (6.60, 4.60, 5.40, 5.80, 5.50)$.
   Dobbiamo stimare μ al 95% assumendo che $X \sim N(\mu, \sigma^2)$
   La quantità pivotale è $Q = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
   dove $\bar{X}$ è la media campionaria e S^2 la varianza campionaria.
   $\bar{X} = \frac{1}{5}(6.60 + 4.60 + 5.40 + 5.80 + 5.50) = 5.58$
   $S^2 = 1/4 \sum_1^5 (X_i - \bar{X})^2 = 0.522$
   $P(q_1 < Q < q_2) = P(q_1 < \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} < q_2) = 0.95$
   q_1 e q_2 sono i quantili della t di Student con 4 gradi di libertà, e dalle tavole si ricava che $q_2 = 2.776$ e $q_1 = -q_2$. L'intervallo per μ è: $P(\bar{X} - q_2 \frac{S}{\sqrt{n}} < \mu < \bar{X} + q_2 \frac{S}{\sqrt{n}}) = P(4.69 < \mu < 6.47) = 0.95$

2. Stimate σ^2 con un intervallo di confidenza al 90%; stimate anche σ. L'intervallo di confidenza per σ^2 si trova considerando come quantità pivotale:
   $Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
   $P(q_1 < \frac{(n-1)S^2}{\sigma^2} < q_2) = 0.90$
   Dalle tavole si ricava che per una χ^2_4, $q_2 = 9.49$ e $q_1 = 0.711$
   $P(\frac{4S^2}{q_2} < \sigma^2 < \frac{4S^2}{q_1}) = P(1.24 < \sigma^2 < 16.61) = 0.90$

3. Tracciate una regione di confidenza all'81% per la stima congiunta di μ e σ^2; e tracciatene una anche per μ e σ. Per trovare l'intervallo di confidenza all'81% per μ e σ^2 si procede nel seguente modo:
   $P(-q_2 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < q_2, q^{'} < \frac{(n-1)S^2}{\sigma^2} < q^{''}) = \gamma_1\gamma_2 = 81\%$
   Dal punto precedente $\gamma_2 = 0.9$ quindi $\gamma_1 = 0.9$
   Per l'indipendenza:
   $P(-q_2 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < q_2) P(q^{'} < \frac{(n-1)S^2}{\sigma^2} < q^{''}) = 81\%$
   quindi: $P(-q_2 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < q_2) = 2\Phi(q_2) - 1 = 0.9$ da cui $q_2 = 1.65$.
   Per l'altra espressione invece dalle tavole della χ^2_4 si ricava che $q^{'} = 0.711$ e $q^{''} = 9.49$.

Esercizio 2.
Supponiamo che la variabile casuale $Y \sim \Gamma(2, \beta)$. Mostrate con il metodo della funzione generatrice dei momenti, che $Z = 2\beta Y \sim \chi^2_4$. Usate Z come quantitaà pivotale per trovare un intervallo di confidenza di livello $0,90$ per β.
Poichè $Y \sim \Gamma(2, \beta)$ si ricava subito che $f_Y(y) = \frac{\beta^2 y e^{-\beta y}}{\Gamma(2)}$ con $y > 0$.
Calcoliamo la distribuzione di Z:

$F_Z(z) = P(Z \leq z) = P(Y \leq \frac{z}{2\beta}) = \int_0^{\frac{z}{2\beta}} \beta^2 y e^{-\beta y} dy = 1 - e^{-z/2}(1 + \frac{z}{2})$
Da cui ricaviamo che $f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{z}{4} e^{-z/2}$
La funzione generatrice dei momenti di Z è quindi:
$E[e^{tZ}] = E[e^{2\beta t Y}] = \int_0^{\infty} e^{2\beta t Y} \beta^2 y e^{-\beta y} dy = \frac{\beta^2}{(2\beta t - \beta)^2} = \frac{1}{(1-2t)^2}$, $t > 1/2$. Questa è la fgm di una variabile χ^2_4.

Esercizio 3.
Siano $Y_1, ..., Y_n$ v.a. distribuite come una $Unif(0, \theta)$. Sia $Y_{(n)} = max\{Y_1, ..., Y_n\}$ e sia $U = \frac{1}{\theta} Y_{(n)}$.

1. Mostrate che U ha la seguente funzione di distribuzione:

$$F_X(x) = \begin{cases} 0 & u < 0 \\ u^n & 0 \leq u \leq 1 \\ 1 & u > 1 \end{cases}$$

   Calcoliamo la distribuzione di $Y_{(n)}$:
   $P(Y_{(n)} \leq y) = P(max\{Y_1, ..., Y_n\} \leq y) = P(X_1 \leq y, ..., X_n \leq y) = \prod_{i=1}^{n} P(X \leq y) = \prod_{i=1}^{n} \int_0^y \frac{1}{\theta} dx = \frac{y^n}{\theta^n}$
   $f_{Y_{(n)}}(y) = \frac{n}{\theta^n} y^{n-1}$.
   Calcoliamo ora le densità di U:
   $P(U \leq u) = P(\frac{1}{\theta} Y_{(n)} \leq u) = P(Y_{(n)} \leq u\theta) = \int_0^{u\theta} \frac{n}{\theta^n} y^{n-1} dy = u^n$ con $0 \leq u \leq 1$.

2. Poichè la distribuzione di U non dipende dal parametro θ, U è una quantità pivotale, usarla per trovare un intervallo di confidenza inferiore di livello $0,95$.
   $P(U \leq q) = q^n = 0,95$ da cui ricaviamo che $q = (0,95)^{1/n}$, quindi
   $0,95 = P(U \leq (0,95)^{1/n}) = P(\frac{Y_{(n)}}{\theta} \leq (0,95)^{1/n}) = P(\frac{Y_{(n)}}{(0,95)^{1/n}} \leq \theta)$

Esercizio 4.
É data una distribuzione di Bernoulli X, dove $P(X = 1) = \theta = 1 - P(X = 0)$.

1. Per un campione casuale di ampiezza $n = 10$, verificate $H_0 : \theta \leq 1/2$ in alternativa a $H_1 : \theta > 1/2$. Usate la regione critica $\{\sum_i^{10} x_i \geq 6\}$.
   Trovate la funzione di potenza e rappresentatela. Qual è l'ampiezza di questo test?

2. Per un campione di ampiezza $n = 10$ trovate il test più potente di ampiezza α ($\alpha = 0.0547$) per $H_0 : \theta = 1/2$ in alternativa a $H_1 : \theta = 1/4$.

   La funzione di potenza è data da:
   $\Pi_Y(\theta) = P_\theta(H_0 \ sia \ rifiutata | H_0 vera) = P_\theta(\sum_{i=1}^{6} \geq 6) = \sum_{j=6}^{10} \binom{10}{j} \theta^j (1 - \theta)^{10-j}$ con $\theta \in (0, 1)$
   L'ampiezza del test è data da: $sup_{\theta \leq 1/2}(\Pi_Y(\theta)) = sup_{\theta \leq 1/2} \sum_{j=6}^{10} \binom{10}{j} \theta^j (1 - \theta)^{10-j} =$
   Dallo studio della derivata prima di $\Pi_Y(\theta)$ si vede che la funzione è crescente sotto l'ipotesi H_0 per $\theta \leq \frac{1}{2}$ nel caso in cui $\frac{6}{10} > \theta$ infatti:

   $\sum_{j=6}^{10} \binom{10}{j} \theta^j (1 - \theta)^{10-j} =: g(\theta)$

$g'(\theta) = \sum_{j=6}^{10} \binom{10}{j} [\theta^{j-1}(1-\theta)^{9-j}(j-10\theta)]$ allora

$= \sum_{j=6}^{10} \binom{10}{j} \frac{1}{2^{10}}$

Il test piú potente dato $n = 10$ ed $\alpha = 0.0547$ si trova applicando il lemma di Neyman-Pearson (teorema 9.2):
$\lambda(x_1, .., x_{10}) = \frac{L(\theta_0)}{L(\theta_1)} \geq k$ allora

$\frac{\prod_{i=1}^{10} \theta_0^{x_i}(1-\theta_0)^{1-x_i} 1_{(0,1)}(x_i)}{\prod_{i=1}^{10} \theta_1^{x_i}(1-\theta_1)^{1-x_i} 1_{(0,1)}(x_i)} = \frac{[\frac{1}{2}]^{10}}{[\frac{1}{4}]^{\sum_i^{10} x_i} [\frac{3}{4}]^{10-\sum_i^{10} x_i}} \leq k$

quindi $(\frac{1}{2})^{10} \leq k(\frac{1}{4})^{\sum_i^{10} x_i} (\frac{3}{4})^{10-\sum_i^{10} x_i}$
passando al log:
$(log1/4 + log4/3) \sum_{i=1}^{10} \geq -logk - 10log3/4 + 10log1/2$ quindi
$\sum_{i=1}^{10} x_1 \leq k'$
allora l'ampiezza del test è:
$0.0547 = \alpha = P(rifiutare\ H_0|H_0) = P(\sum_{i=1}^{10} x_i \leq k')$
dove k' è il qualtile della $Bin(10, \frac{1}{2})$.

Esercizio 5.

Sia $X_1, ..., X_n$ un campione casuale da $N(\mu, 25)$. Si vuole testare in seguente test di ipotesi: $H_0 : \mu = 10$ contro $H_1 : \mu = 5$. Trovare l'ampiezza n per cui il test più potente ha $\alpha = \beta = 0,025$ dove α e β sono rispettivamente gli errori di I e II specie.
Il test più potente è dato dal lemma di Neyman Pearson:
$\frac{L(\mu_0)}{L(\mu_1)} = \frac{\prod_{i=1}^n \frac{e^{-(x-\mu_0)^2/50}}{\sqrt{50\pi}}}{\prod_{i=1}^n \frac{e^{-(x-\mu_1)^2/50}}{\sqrt{50\pi}}} =$
$e^{-1/50\{\sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\mu_1)^2\}}$
Passando al logaritmo si ottiene:
$-\{\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x - \mu_1)^2\} \leq 50k \Longleftrightarrow \frac{\sum_i x_1}{n} \leq k^*$
L'ampiezza del test è:
$\alpha = 0,025 = P(RifiutareH_0|\mu_0) = P(\bar{X} \leq k^*|\mu_0) = P(\frac{\bar{X}-\mu_0}{5/\sqrt{n}} \leq \frac{k^*-\mu_0}{5/\sqrt{n}}) = \Phi(\frac{k^*-\mu_0}{5/\sqrt{n}})$
Dalle tavole si ricava che $\frac{k^*-\mu_0}{5/\sqrt{n}} = -1,96$ da cui $k^* = 10 - \frac{9,8}{\sqrt{n}}$.
$\beta = P(AccettareH_0|\mu_1) = P(\bar{X} > k^*|\mu_1) = P(\frac{\bar{X}-\mu_1}{5/\sqrt{n}} > \frac{k^*-\mu_1}{5/\sqrt{n}}|\mu_1) = 1 - \Phi(\frac{k^*-\mu_1}{5/\sqrt{n}})$
Dalle tavole si ricava che $\frac{k^*-\mu_1}{5/\sqrt{n}} = 1,96$ quindi $k^* = 5 + \frac{9,8}{\sqrt{n}}$.
Ora $\alpha = \beta$ quindi risolvendo otteniamo:
$10 - \frac{9,8}{\sqrt{n}} = 5 + \frac{9,8}{\sqrt{n}}$
Da cui $n = 16$.