Statistica Matematica A

Docente: Dott. F. Zucca

III appello

Lecco 14/09/2005

<u>**Candidato**</u>

Cognome: .... **Nome:** ....

Corso di Laurea: ....

Matricola: .... **Anno di corso:** ....

tempo a disposizione **2h30m**.

1 Esercizi

Esercizio 1 Lanciamo varie volte una moneta non truccata.

1. Quante volte occorre tirare la moneta affinché la probabilità di non vedere mai uscire una testa sia compresa tra il 2% e il 10%?
2. Supponiamo di partire con un capitale uguale a 0 e di vincere 1 centesimo per ogni testa e perderne 2 per ogni croce. Quali sono la media e la varianza del capitale dopo il primo lancio? E dopo 10 lanci?
3. Considerando lo stesso gioco, qual è il minimo numero di lanci da fare perché la probabilità di essere almeno in pari sia inferiore al 3%?

Soluzione.

1. $0.02 < 2^{-n} < 0.1$, ovvero $10 < 2^n < 50$, ovvero $n = 4$ e $n = 5$.

2. La variazione di capitale nel lancio i è una variabile X_i con $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -2) = 1/2$, quindi:

$$E(X_i) = -\frac{1}{2}, \qquad V(X) = \frac{1^2 + (-2)^2}{2} - \left(-\frac{1}{2}\right)^2 = \frac{9}{4}.$$

   Dopo 10 lanci (indipendenti), il risultato è: media $10/2 = 5$, varianza $10 \cdot 9/4 = 22.5$.

3. Introduco $S_n = X_1 + \cdots X_n$, somma di variabili i.i.d. Per il TLC, se n è abbastanza grande, S_n è approssimativamente $N(-n/2, n \cdot 9/4)$. Essere almeno in pari significa $S_n \geq 0$. Quindi voglio porre $\mathbb{P}(S_n \geq 0) < 0.03$:

$$\mathbb{P}(S_n \geq 0) = \mathbb{P}\left(\frac{S_n + n/2}{\sqrt{n} \cdot 3/2} \geq \frac{n/2}{\sqrt{n} \cdot 3/2}\right) \approx \mathbb{P}\left(Z \geq \frac{\sqrt{n}}{3}\right) < 0.03$$

   Poiché $\Phi(1.88) \approx 0.97$, possiamo porre $\sqrt{n}/3 > 1.88$, ovvero $n > 31.8$, quindi occorre tirare almeno 32 volte.

Esercizio 2 Ad un corso di Laurea in Chimica si iscrivono 60 matricole. Poche di loro però frequentano le aule: la media campionaria dei presenti in 20 lezioni è solo 35. Se μ è il vero valore medio della variabile aleatoria "numero di studenti presenti alla lezione", si studi il test

$$H_0 : \mu \geq 40, \qquad H_1 : \mu < 40$$

nelle due seguenti situazioni (stimando anche il P-value):

1. deviazione standard nota $\sigma = 15$, $\alpha = 1\%$, 5%, 10%;

2. deviazione standard non nota, $s = 17.3$, $\alpha = 1\%$, 5%, 10%.

Soluzione.

1. Introduco la statistica test:

$$U := \frac{\overline{X} - 40}{15/\sqrt{20}}$$

   dove $\overline{X}$ è la media campionaria. Il valore numerico è

$$u = \frac{35 - 40}{15/\sqrt{20}} = -1.49.$$

Il criterio di rifiuto di H_0 è $u < q_\alpha$, quindi il P-value è $P(Z < u) = P(Z > |u|) = P(Z > 1.49) = 1 - \Phi(1.49) = 1 - 0.93 = 0.07 = 7\%$. Ne segue che al 10% rifiuto H_0 mentre al 5% e 1% accetto H_0.

2. Introduco la statistica test:

$$U := \frac{\overline{X} - 40}{S/\sqrt{20}}$$

Il valore numerico è

$$u = \frac{35 - 40}{17.3/\sqrt{20}} = -1.29.$$

Il criterio di rifiuto di H_0 è $u < -t_{1-\alpha,19}$, ovvero $|u| > t_{1-\alpha,19}$. Dalle tavole $t_{0.9,19} = 1.33 > |u|$, quindi al 10% accetto H_0 e perciò il P-value è almeno del 10%. Ne segue che accetto H_0 anche al 5% e 1%. Con un calcolatore si trova che il P-value esatto è $P(t > 1.29) = 11\%$, mentre con le tavole si può semplicemente osservare che $0.1 < \overline{\alpha} < 0.15$.

Esercizio 3 Due cinema registrano l'affluenza per 10 fine settimana successivi, ottenendo rispettivamente i seguenti dati:

650	620	580	640	650	590	570	570	640	580
660	630	610	640	620	670	630	600	620	650

1. Supponendo per le affluenze nei due cinema distribuzioni gaussiane *indipendenti* con medie μ_i e varianze $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ignote, si imposti e studi un test per verificare quale dei due cinema abbia l'affluenza maggiore, cercando di ottenere un'affermazione *forte* e stimando il P-value dei dati (si ricordi che si considera una conclusione *forte* consiste nel rifiutare H_0 a livelli di significatività bassi).

2. Stessa richiesta, supponendo però che le affluenze siano *non indipendenti*, ovvero *accoppiate*.

Soluzione.

1. Abbiamo: $\overline{x}_1 = 609$, $s_2^2 = 1165.56 = (34.14)^2$, $\overline{x}_2 = 633$, $s_2^2 = 490 = (22.13)^2$. Notiamo che $\overline{x}_2 > \overline{x}_1$. Quindi l'unica affermazione forte può venire dal rifiuto di H_0 nel seguente test:

$$H_0: \ \mu_1 \geq \mu_2, \qquad H_1: \ \mu_1 < \mu_2.$$

Infatti, nel test con $H_1: \ \mu_1 > \mu_2$ accettiamo per ogni grado di significatività inferiore a 0.5 l'ipotesi nulla $\mu_1 \leq \mu_2$, ma come affermazione debole, quindi il p-value di questo test sarà superiore a 0.5.

Introduciamo la statistica test:

$$U := \frac{\overline{X}_1 - \overline{X}_2}{S_p/\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

dove S_p^2 è lo stimatore pooled della varianza, con valori numerici:

$$S_p^2 = \frac{9 \cdot 1165.56 + 9 \cdot 490}{9+9} = 827.78 = (28.77)^2,$$

$$u = \frac{609 - 633}{28.77\sqrt{2/10}} = -1.86.$$

Il criterio di rifiuto di H_0 è $u < -t_{1-\alpha,18}$ e il P-value è $P(t < -1.86)$. Dalle tavole troviamo $t_{0.095,18} = 1.734 < 1.86$, quindi il P-value è inferiore al 5%. Abbiamo evidenza sperimentale di una maggiore affluenza nel secondo cinema.

2. Consideriamo un test t-accoppiato: le differenze tra i guadagni del primo studente e quelli del secondo nei 12 mesi del 2004 sono:

$$-10 \quad -10 \quad -30 \quad 0 \quad 30 \quad -80 \quad -60 \quad -30 \quad 20 \quad -70$$

Poniamo $Y = X_1 - X_2$, allora $Y \sim N(\mu_Y, \sigma_Y^2)$ con $\mu_Y = \mu_1 - \mu_2$. Studiamo il test:

$$H_0: \ \mu_Y \geq 0, \qquad H_1: \ \mu_Y < 0$$

Poiché σ_Y^2 è ignota, si tratta di un test t. La media e varianza campionarie sono $\overline{y} = -24$ e $s_Y^2 = 1382.22 = (37.17)^2$. Introduciamo la statistica test:

$$U := \frac{\overline{Y}}{S_Y/\sqrt{12}}$$

Il valore numerico di U è

$$u = \frac{-24}{37.17/\sqrt{10}} = -2.04.$$

Per questo test unilatero, rifiuto H_0 se $u < -t_{1-\alpha,n-1}$. Dalle tavole troviamo che $t_{0.95,9} = 1.833 < |u|$, quindi il P-value è inferiore al 5% e la conclusione è la stessa che per il caso precedente.

Esercizio 4 Si torni ai dati dell'esercizio precedente. Supponiamo che il primo cinema abbia un profitto se in un fine-settimana entrano almeno 600 persone, mentre per il secondo servono almeno 630 ingressi.

1. Se p_1 è la probabilità di ottenere un profitto in un fine-settimana per il primo cinema, si studi il test:

   $$H_0 : \; p_1 \geq 60\%, \qquad H_1 : \; p_1 < 60\%$$

   calcolandone il P-value.

2. Si imposti e studi un test per verificare se i due cinema hanno probabilità significativamente diverse di ottenere un profitto in un fine-settimana, calcolando il P-value dei dati.

Soluzione.

1. Introduciamo la statistica test:

   $$U := \frac{p_1 - p_0}{\sqrt{\frac{p_0(1-p_0)}{n_1}}} = \frac{p_1 - 0.6}{\sqrt{\frac{0.4 \cdot 0.6}{10}}}.$$

   Il valore numerico di U è

   $$\frac{0.5 - 0.6}{\sqrt{\frac{0.4 \cdot 0.6}{10}}} = -0.645.$$

   Il P-value di questo test è $P(Z < u) = P(Z < -0.64) = 1 - \Phi(0.64) \approx 26\%$. Il P-value alto non permette di affermare che la probabilità p_1 di avere un profitto sia inferiore al 60%.

2. Il test è:

$$H_0 : \ p_1 = p_2, \qquad H_1 : \ p_1 \neq p_2.$$

Con notazioni analoghe, $\overline{x}_2 = 6$ su $n_2 = 10$ volte, quindi $\overline{p}_2 = 6/10 = 0.6$. Introduciamo la statistica test:

$$U := \frac{p_1 - p_2}{\sqrt{\hat{p}(1-\hat{p})\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} = \frac{p_1 - p_2}{\sqrt{\frac{11}{20} \cdot \frac{9}{20} \cdot \frac{2}{10}}}, \qquad \hat{p} := \frac{\overline{x}_1 + \overline{x}_2}{n_1 + n_2} = \frac{11}{20}.$$

Il valore numerico di U è

$$\frac{0.5 - 0.6}{\frac{1}{20}\sqrt{99/5}} = -0.45.$$

Il P-value di questo test è $P(|Z| > |u|) = 2(1 - \Phi(0.45)) = 2(1 - 0.6736) \approx 66\%$. Il P-value alto non permette di valutare come significativamente differenti le due percentuali.