NUMERO: 2280A

ANNO: 2017

APPUNTI

STUDENTE: Locci Federico

MATERIA: Statistica + Teoria + Esercizi - Prof. Vicario

Statistica

Fenomeno casuale: fenomeno, o esperimento, i cui risultati non sono prevedibili a priori con certezza.

Definizione a priori di probabilità:

$$P[E] = \frac{s}{n}$$

La probabilità è il rapporto tra casi favorevoli s e casi possibili n.
Purché siano possibili egualmente e mutuamente escludenti.

ESEMPIO: lancio di un dado

- probabilità di avere un numero dispari?

$$P[E] = \frac{3}{6} = \frac{1}{2} = 50\ \%$$

- probabilità di avere la faccia con 4 pallini?

$$P[E] = \frac{1}{6} = 16,67\ \%$$

- probabilità di avere un risultato maggiore di 2?

$$P[E] = \frac{4}{6} = 66\ \%$$

ESEMPIO: estrazione carte da mazzo di 52 carte

- probabilità di estrarre carta di cuori?

$$P[E] = \frac{13}{52} = \frac{1}{4} = 25\ \%$$

- probabilità di estrarre una carta >5 e <10?

$$P[E] = \frac{16}{52} = \frac{4}{13} = 30.76\ \%$$

- probabilità di un asso o seme di fiori?

$$P[E] = \frac{16}{52} = \frac{4}{13} = 30.76\ \%$$

Nb: non si inserisce 2 volte asso di fiori

ESEMPIO: lancio 2 monete

- probabilità di due croci?

$$P[E] = \frac{1}{4} = 25\ \%$$

Qualcuno ebbe da dire su questa definizione di probabilità.

$$\boxed{0 \le P[E] \le 1}$$

LA PROBABILITÀ È SEMPRE COMPRESA TRA 0 e 1

$P[E] = 0$ non succede mai
$P[E] = 1$ succede sempre → CERTEZZA

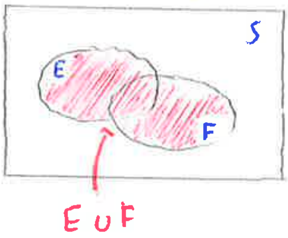
Algebra degli eventi:

Non tutti gli ~~eventi~~ insiemi sono eventi, lo sono quelli che soddisfano determinati criteri.

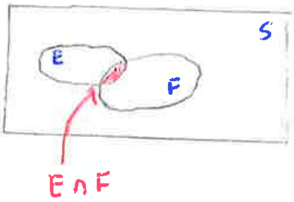
- Lo spazio dei campioni S appartiene allo spazio degli eventi e viene chiamato evento certo.

$$S \in A$$

- Per ogni evento $E \in A$, esiste $\bar{E}$ detto complementare, che si verifica se non si verifica E.
- Dati 2 eventi E ed F qualsiasi, con $E \in A$ ed $F \in A$, definiamo evento unione l'evento che si verifica quando si verificano E o F o entrambi.



Teorema: Definiamo l'evento intersezione $E \cap F$, EF, l'evento che si verifica quando si verifica sia E sia F.



ezione 4:

Esempio:

Lotto con M componenti, K difettosi.
Estraendo n componenti, qual'è la probabilità che ve ne siano K difettosi?

$$S = \{(z_1, z_2, \dots, z_n) : z_i = n° \text{ del componente estratto}\}$$

Quanto vale $\# S$?

- con reimmissione: M^n
- senza reimmissione: $M(M-1)(M-2)\dots(M-n+1)$

$A_K = \{$ estrazione di un campione di cui K difettosi $\} \subset S$

$\# A_K$?

- con reimmissione: $\binom{n}{K} K^K (M-K)^{n-K}$

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

$$P[A_K] = \frac{\# A_K}{\# S} = \frac{\binom{n}{k} K^k (M-K)^{n-K}}{M^n}$$

- senza reimmissione:

$$P[A_K] = \frac{\# A_K}{\# S} = \frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{M}{n}} \Bigg]$$

Vedi distribuzione ipergeometrica

Spazi campione finiti con punti non equiprobabili:

$P_j = P[\{s_j\}]$ $j = 1, \dots, N$ e $\sum_{j=1}^{N} P_j = 1$

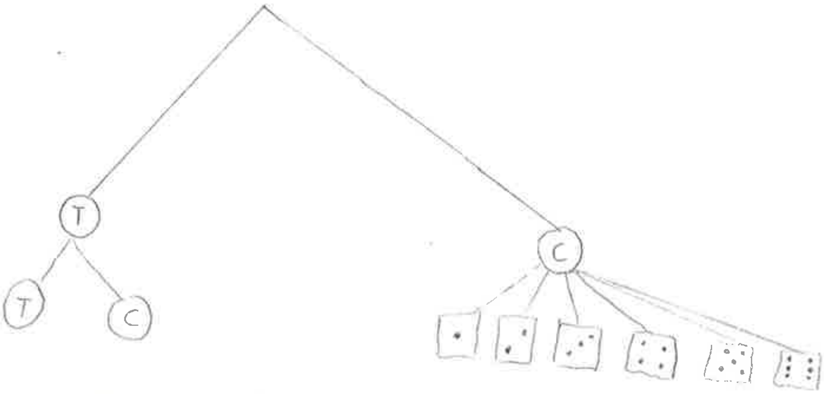
Se ho un evento $A = \{s_1, s_2, s_3\}$ $A = s_1 \cup s_2 \cup \dots \cup s_a$

$$P[A] = P\left[\bigcup_{j=1}^{a} s_j\right] = \sum_{j=1}^{a} P[s_j]$$

Diagramma ad albero:

1- lancio moneta

2- lancio moneta se esce testa, dado se esce croce.



Ovviamente i vantaggi sono palesi, in quanto permette un'elencazione degli elementi appartenenti allo spazio campione.

Nb: il risultato nodo finale di un ramo finale di un determinato evento è dato dalla moltiplicazione delle probabilità del singolo ramo.

Probabilità condizionata: probabilità che accada un evento A dato che è avvenuto un evento B

Esempio:

lotto di 300 pneumatici, 25 difetti battistrada, 12 nelle tracce, 7 entrambi i difetti.

$$\underbrace{\text{25 difetti battistrada}}_{E}, \underbrace{\text{12 nelle tracce}}_{F}$$

$$P[F|E] = \frac{7}{\underset{\uparrow \text{NO!}}{300}}$$

"estrazione pneumatico con difetti nelle tracce, dato che ha irregolarità nello spessore"

$$P[F|E] = \frac{7}{32}$$

perché restringo all'attenzione

In generale: $P[F|E] = \frac{P[FE]}{P[E]}$

Per cui, per definizione:

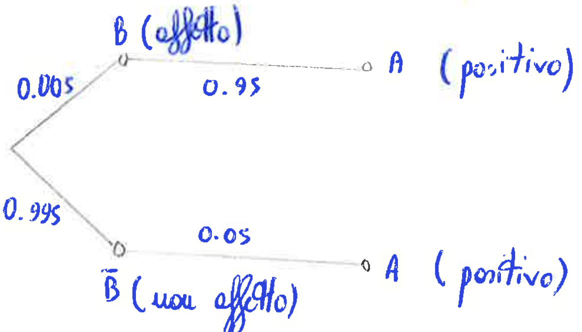
$$\boxed{P[A|B] = \frac{P[AB]}{P[B]}} \qquad P[B] \neq 0$$

Quella che abbiamo appena usato prende il nome di formula di Bayes:

$$P[B_k|A] = \frac{P[A|B_k]P[B_k]}{\sum_{i=1}^{n} P[A|B_i]P[B_i]}$$

Applicazioni:

- Abbiamo un test per la diagnosi di una malattia che fornisce una diagnosi corretta nel 95 % dei casi. Se l'incidenza di quella malattia è pari allo 0.5 % sulla popolazione esaminata, valutare la proporzione di diagnosi corrette di quella malattia.



$$P[B|A] = \frac{0.005 \cdot 0.95}{0.005 \cdot 0.95 + 0{,}995 \cdot 0.05} = 0.087 = 8.7\ \%$$

- Urna con due composizioni:
  1) Solo bianche $= B_1 \Rightarrow P[B_1] = 0.5$
  2) Metà e metà $= B_2 \Rightarrow P[B_2] = 0.5$

Pesco una palla ed è bianca, probabilità che vi siano solo bianche dato che l'ho pescata bianca?

A = "estraggo bianca"

Mi serve: $P[B_1|A] = \frac{\boxed{P[AB_1]}}{P[A]} = \frac{\overset{1}{P[A|B_1]}\,\overset{0.5}{P[B_1]}}{\underset{1}{P[A|B_1]}\underset{0.5}{P[B_1]} + \underset{0.5}{P[A|B_2]}\underset{0.5}{P[B_2]}} = 0.6667 = 66{,}67\ \%$

$P[A|B_1]P[B_1]$

Rimescolo ed estraggo bianca, "probabilità solo bianche dato che ho già estratto 2 volte bianca?"

$$P[B_1|A] = \frac{\overset{1}{P[A|B_1]}\,\overset{2/3}{P[B_1]}}{\underset{1}{P[A|B_1]}\underset{0.5}{P[B_1]} + \underset{0.5}{P[A|B_2]}\underset{1/3}{P[B_2]}}$$

C'è differenza!

Al 2° tentativo, aumenta la fiducia che l'urna abbia solo bianche.

Esempio: lancio 2 dadi

A_1 = "1° dado pari"

A_2 = "2° dado pari"

A_3 = "somma dispari"

A_1, A_2, A_3 sono indipendenti?

$P[A_1] = \frac{1}{2}$

$P[A_2] = \frac{1}{2}$

$P[A_3] = \frac{1}{2}$

a) $P[A_1 A_2] = \frac{9}{36} = \frac{1}{4}$ $\quad P[A_1] \times P[A_2] = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ Indipendenti

b) $P[A_1 A_3] = \frac{9}{36} = \frac{1}{4}$ $\quad P[A_1] \times P[A_3] = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ Indipendenti

c) $P[A_2 A_3] = \frac{1}{4}$ $\quad P[A_2] \times P[A_3] = \frac{1}{4}$ Indipendenti

Sono indipendenti a coppie, ma la terna?

$P[A_1 A_2 A_3] = 0$ $\quad P[A_1] \times P[A_2] \times P[A_3] = \frac{1}{8}$

Non sono indipendenti.

Regole di calcolo:

Moltiplicativa: A_1, A_2, A_n appartenenti allo stesso spazio degli eventi, si ha:

$$P[A_1 A_2 \dots A_n] = P[A_1] \times P[A_2 | A_1] \times P[A_3 | A_1 A_2] \times \dots \times P[A_n | A_1 A_2 \dots A_{n-1}]$$

Esempio: 10 palline, 5 bianche e 5 nere.

Se ne registro il colore di una estratta e la reinserisco con 2 dello stesso colore.

$$P[\text{"bianca nelle prime tre"}] = \frac{\cancel{5}^{1}}{\cancel{10}_{2}} \times \frac{\cancel{7}}{\cancel{12}_{4}} \times \frac{\cancel{9}^{3}}{\cancel{14}_{2}} = \frac{3}{16} = 0,1875 = 18,75\ \%$$

Funzione di densità discreta:

$$f_X(x) = \begin{cases} P[X = x_J] & se\ x = x_J \\ 0 & se\ x \neq x_J \end{cases}$$

Funzione che nei punti assunti è uguale alla probabilità che quella variabile casuale assuma quel valore; zero altrove.

Associa ad ogni risultato la probabilità con cui si verifica quell'evento.

Divido lo spazio campione in 2 tipi:

Discreti

Continui

dove il x_J è il punto [illegible]

$f_X(x) : \mathbb{R} \to [0,1]$

I) $f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$

II $\Sigma_i f_X(x_J) = 1$

Esempio: lancio dado, X è la v.c. che indica il numero apparso

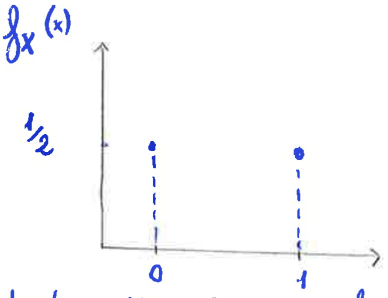
$S = \{1,2,3,4,5,6\}$

$$f_X(x) = \begin{cases} 1/6 & se\ x = i \\ 0 & se\ x \neq i \end{cases}$$

Esempio: testa o croce? X è la v.c. che conta il n° di teste

$S = \{T, C\}$

$$f_X(x) = \begin{cases} 1/2 & se\ X = 1 \\ 0 & altrimenti \end{cases}$$



Esempio: lancio di due dadi. X è la v.c. che indica la somma

$S = \{2,3,4,5,\dots,12\}$

$$f_X(x) = \begin{cases} 1/36 & se\ X = 2 \\ 2/36 & se\ X = 3 \\ 3/36 & se\ X = 4 \\ 4/36 & se\ X = 5 \\ 5/36 & se\ X = 6 \\ 6/36 & se\ X = 7 \\ 5/36 & se\ X = 8 \\ 4/36 & se\ X = 0 \\ 3/36 & se\ X = 10 \\ 2/36 & se\ X = 11 \\ 1/36 & se\ X = 12 \\ 0 & altrove \end{cases}$$

$$F_X(b) = P[X \le b] = \sum_{x_j : x \le b} f_X(x_j)$$
$$F_X(a) = P[X \le a] = \sum_{x_j : x \le a} f_X(x_j)$$

È chiaro che rimangono i valori tra a e b

$P[2 < X \le 6]$

Proprietà:

$F_X(x): R \to [0,1]$

I) $\lim_{x \to -\infty} F_X(x) = 0$ $\lim_{x \to +\infty} F_X(x) = 1$

II) $F_X(x)$ è una funzione monotona non decrescente.

$\forall x_1, x_2$ tali che x_1, x_2 : $F_X(x_1) \le F_X(x_2)$

III) $F_X(x)$ è continua da destra:

$$\lim_{h \to 0^+} F(x+h) = F(x)$$

Funzione di densità di probabilità:

X è una variabile casuale continua:

$$F_X(x) = \int_{-\infty}^{X} f_X(u)\,du$$

Proprietà:

$f_X(x) = R \to R^+$

I) $f_X(x) \ge 0$ $\forall x \in \mathbb{R}$

II) $\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$

III) $P[a < X \le b] = \int_a^b f_X(x)dx$

Lezione 10

Media (o valore atteso di X)

$E[X] = \sum_{x_j} x_j f_X(x_j)$ se X è discreta

$E[X] = \int_{-\infty}^{+\infty} x f_X(x_j) dx$ se X è continua

Per semplificare la notazione $E[X] = \mu$

Perché si chiama valore atteso? Perché il primo nome era speranza.

Esempio: lancio di un dado

$$E[X] = \sum_{i=1}^{6} i f_X(i) = \sum_{i=1}^{6} i P[X=x_i] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3,5$$

Nb: $\boxed{x_{min} \leq E[X] \leq x_{max}}$

Esempio:

X è una variabile casuale con: $f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{se } x \leq 0 \end{cases}$

↓
esponenziale

$E[X] \overset{def}{=} \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{0}^{+\infty} x \lambda e^{-\lambda x} dx$ integro per parti

↓

$\frac{1}{\lambda} \underbrace{\int_{0}^{+\infty} \lambda e^{-\lambda x} dx}_{1} = \frac{1}{\lambda}$

Nb: se X ha distribuzione esponenziale $E[X] = \frac{1}{\lambda}$

Esempio:

$f_X(x) = \begin{cases} \frac{1}{x^2} & \text{se } 1 < x < \infty \\ 0 & \text{altrove} \end{cases}$

$E[X] = \int_{1}^{+\infty} x \frac{1}{x^2} dx = \lim_{\xi \to \infty} \log \xi =$ non esiste

Varianza:

$V[X] = \sum_{x_j} (x_j - \mu_X)^2 f_X(x_j)$ se X è discreta

$V[X] = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx$ se X è continua

Per semplificare la notazione $V[X] = \sigma^2$

Proprietà dei valori attesi:

I) $E[c] = c$

II) $E[c\,g[X]] = c\,E[g(X)]$

III) $E[c_1 g_1(X) + \dots + c_n g_n(X)] = c_1 E[g_1(X)] + \dots + c_n E[g_n(X)]$

IV) $E[g_1(X)] \leq E[g_2(X)]$ se $g_1(X) \leq g_2(X)$

Esempio:

Se $y = a + bx$

$E[y] = E[a + bX] = a + bE[X]$

$V[y] = V[a + bX]$ come abbiamo detto $var[z] = E[(z - \mu_z)^2]$

$E[(a + bx - a - b\mu_x)^2] = E[b^2 (X - \mu_x)^2] = b^2 E[(X - \mu_x)^2] = b^2 V[X]$ è sparito a

- $y = a + 1X$ $b = 1$

  $var[y] = var[X]$

- $y = bX$ $a = 0$

  $var[y] = b^2 var[X]$

Le costanti escono dal segno di varianza elevandosi al quadrato

Teorema: formula per il calcolo

$V[X] = E[X^2] - (E[X])^2$ se $E[X^2]$ esiste

Esempio:

$$
\begin{aligned}
V[X] &= E[(X - E[X])^2] \\
&= E[X^2 - 2XE[X] + (E[X])^2] \\
&= E[X^2] - E[\boxed{2X}\boxed{E[X]}] + E[E[x]^2] \\
&= E[X^2] - 2\overset{=E[X]}{\mu_X} E[X] + \overset{\to E[X]^2}{\mu_X^2} \\
&= E[X^2] - (E[X])^2
\end{aligned}
$$

CVD

Nb: $E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx$ oppure $\sum_{x_j} x_j^2 f_X(x_j)$

Per il Teorema

$E[X^2] = V[X] + (E[X])^2$

Mediana: quantile per cui $q = 0,5$

$P[X \le med(X)] \ge \frac{1}{2}$ e $P[X \ge med(X)] \ge \frac{1}{2}$ se X è discreta

$$\int_{-\infty}^{med[X]} f_X(x)dx = \int_{med[X]}^{+\infty} f_X(x)dx = 0.5$$

Esempio: $med[X_{exp}]$?

$$\int_{-\infty}^{med[X]} f_X(x)dx = 0.5$$

↓

$$\int_{0}^{med[X]} \lambda e^{-\lambda x}dx = \frac{1}{2}$$

$$= -e^{-\lambda x}\Big|_{0}^{med(X)} = \frac{1}{2} \rightarrow -e^{-\lambda\, med[X]} + 1 = \frac{1}{2}$$

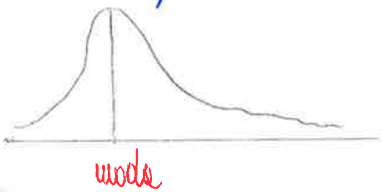
$$e^{-\lambda\, med[X]} = \frac{1}{2}$$

$$-\lambda\, med[X] = +\ln\frac{1}{2} = -\ln 2$$

$$med[X] = \frac{1}{\lambda}\ln 2 = E[X_{exp}]\ln 2$$

Quartili: dividono l'area sottesa in 4 parti uguali.

Moda: valore sull'asse x a cui compete l'ordinata massima.



Momenti di ordine n di X: valore atteso di X^n

$$\mu'_n = E[X^n]$$

$n=1 \rightarrow X^1 \rightarrow \mu'_1 = E[X] = \mu_X$ $\quad$ $var[X] = \mu'_2 - \mu'_1$

$n=2 \rightarrow X^2 \rightarrow \mu'_2 = E[X^2]$

Momento centrale di ordine n rispetto a μ_X

$$\mu_n = E[(X-\mu_X)^n]$$

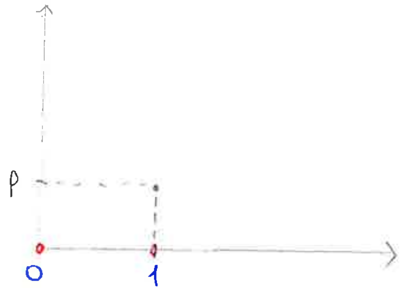
$n=1$ $\quad \mu_1 = E[(X-\mu_x)^1] = 0$

$n=2$ $\quad \mu_2 = E[(X-\mu_x)^2] = var[X]$

Distribuzione di Bernoulli:

$$f_X(x) = f_X(x;p) = \begin{cases} p^x(1-p)^{1-x} & \text{per } x=0,1 \\ 0 & \text{altrove} \end{cases}$$

solo 2 punti massa



$$f_X(0) = 1-p$$
$$f_X(1) = p$$

È funzione di densità?

$f_X(x) \geq 0$ OK

$\sum_{i=0}^{1} f_X(x_i) = f_X(0) + f_X(1) = (1-p) + p = 1$

$$\boxed{E[X] = p \qquad V[X] = p(1-p)}$$

Funzione di distribuzione cumulativa della distribuzione di Bernoulli:

$$F_X(x) = \begin{cases} 0 & \text{se } x<0 \\ 1-p & \text{se } 0 \leq x < 1 \\ 1 & \text{se } x \geq 1 \end{cases}$$

$S = \{s, f\}$

Modellizza gli spazi campione dicotomici (formati da 2 soli elementi).

Distribuzione binomiale:

$$f_X(x) = f_X(x;n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x=0,1,2 \dots n \\ 0 & \text{altrove} \end{cases}$$

$n \in \mathbb{N}, \; 0<p<1$

$\binom{n}{x} = \frac{n!}{(n-x)!\,x!}$

È funzione di densità?

$f_X(x) \geq 0$ OK

$\sum_{n=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + (1-p))^n = 1$ OK

$$\boxed{E[X] = np \qquad V[X] = np(1-p)}$$

Distribuzione ipergeometrica:

$$f_X(x) = \begin{cases} \dfrac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}} & \text{per } x = 0, 1, 2, \dots n, \dots \\ 0 & \text{altrove} \end{cases}$$

$$E[X] = n\frac{K}{M}$$

$$V[X] = n\frac{K}{M}\,\frac{M-K}{M}\,\frac{M-n}{M-1}$$

Esempio: estrazione senza reimmissione di un campione n palle da un'urna che ne contiene M di cui K difettose.

X = v.c. conta difettose

s = estrazione difettosa

$S = \{dif, non\ dif\}$

$$P[X = x] = \frac{\binom{K}{x}\binom{M-K}{n-x}}{\binom{M}{n}}$$

In questo tipo di esercizi (senza reimmissione) possiamo dire che $X \sim iper(M, K, n)$

Esempio: 800 pezzi, n = 150, a = 2 (max accettabile).

P[lotto con 5% difettose accettato] = ?

$M = 800$

$k = 40$

$n = 150$

$X \sim iper(800, 40, 150)$

$$P[accettato] = \sum_{x=0}^{2} f_X(x; 800, 40, 150) =$$

Importantissimo!

Con Reimmissione

↓

Binomiale

Senza Reimmissione

↓

Ipergeometrica.

Lezione 16

Esercizio: ditta produce componenti e ha potuto accertare che il 5% di un componente è difettoso.
Lotti da 100 pezzi, vengono accettati se non più di 2 pezzi sono difettosi.
P["accettato"] ?

s = trovo difetto $S = \{$ difettoso, non difettoso $\}$

$P[s] = 0,05$

$n = 100$

$P[\text{accettato}] = P[X \leq 2] = 0,1183$

X = v.c conta difettosi

$= \binom{n}{x} p^x (1-p)^{n-x}$

$= \binom{100}{0} 0.05^0 \, 0.95^{100} = 0,0059205$ $X = 0$

$= \binom{100}{1} 0,05^1 \, 0.95^{99} = 0,0311607$ $X = 1$

$= \binom{100}{2} 0.05^2 \, 0,95^{98} = 0,0811818$ $X = 2$

Nb: tutti questi esercizi si svolgono con le tavole della Distribuzione cumulativa binomiale $F_X(x; n, p) = \sum_{k=0}^{x} f_X(k; n, p)$

Esercizio: 10 % dei ~~lotti~~ pezzi prodotti sono difettosi. X indica i pezzi difettosi.
Scelti 12 pezzi, calcola la probabilità che:

a) 0,1,2 difettosi. $P[S] = 0.1$ $n = 12$

$P[X=0, X=1, X=2] = P[X \leq 2] = 0.889$

b) almeno 2 difettosi

$P[X \geq 2] = 1 - P[X \leq 1] = 1 - 0,659 = 0,341$

c) al più 3 difettosi

$P[X \leq 3] = 0,974$

d) Esattamente 4 difettosi

$P[X=4] = P[X \leq 4] - P[X \leq 3] = 0,022$

la cumulata somma 0,1,2,3,4   0,1,2,3

Lezione 17

Approssimazione ipergeometrica con binomiale

Teorema: se M e K sono sufficientemente grandi e se il rapporto $\frac{K}{M}$ si mantiene costante, posto $\frac{K}{M} = p$, la distribuzione ipergeometrica è approssimata abbastanza bene dalla binomiale di parametri n e p.

$$E[X]_{bin} \cong n\frac{K}{M} \equiv E[X]_{ipergeom}$$

$$\frac{M-n}{M-1} \; var[X]_{bin} \equiv npq\frac{M-n}{M-1} = n\frac{K}{M}\,\frac{M-K}{M}\,\frac{M-n}{M-1} = var[X]_{ipergeom}$$

Esempio: 30 libri, ce ne sono 6 con difetto nella rilegatura. 10 libri controllati.

$P[X=3]$?

X è la v.c che conta i successi

$$P[X=3] = \frac{\binom{6}{3}\binom{24}{7}}{\binom{30}{10}} = 0,23$$

Se facciamo l'esercizio ingrandendo K e M ma tenendo costante $\frac{K}{M}$ avremo:

$M = 150$ $\quad K=6$ $\quad X \sim$ ipergeom$(150, 30, 6)$

$K = 30$

$$P[X=3] = \frac{\binom{30}{3}\binom{120}{7}}{\binom{150}{10}} = 0,2065$$

Siccome M abbastanza grande $\rightarrow$ $p = \frac{K}{M} = 0,2$

$X \sim$ bin$(10, 0,2)$

$$P[X=3] = \binom{10}{3}(0,2)^3(0,8)^7 = 0,2013$$

Notiamo che i risultati hanno uno scarto dello 0,1% →

$$\boxed{n \leq M/10}$$

Esempio: 4 su 2500 sono allergiche.
Qual'è la probabilità che su 1500 vi siano:

a) 3 allergiche

b) meno di due allergiche

$p = 0{,}016$

$P[X=3] = \binom{1500}{3}(0{,}016)^3 (0{,}984)^{1497}$

$P[X<2] = \sum_{x=0}^{1} P[X=x] = 0{,}984^{1500}$

Approssimando la binomiale con $\lambda = np = 2{,}4$

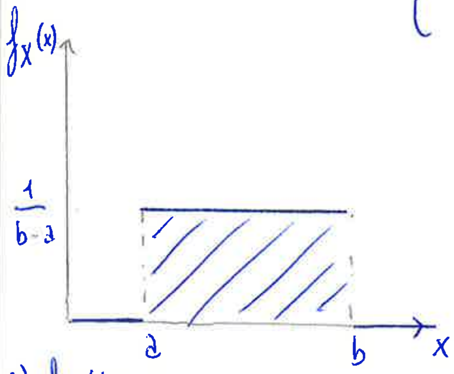
a) $P[X=3] = P[X \leq 3] - P[X \leq 2] = 0{,}779 - 0{,}570 = 0{,}209$

b) $P[X<2] = P[X \leq 1] = 0{,}308$

Distribuzioni continue:

Uniforme (o rettangolare)

$$f_X(x) = f_X(x; a, b) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{altrove} \end{cases}$$



$$\boxed{E[X] = \frac{(a+b)}{2} \qquad var[X] = \frac{(b-a)^2}{12}}$$

a) $f_X(x) \geq 0$ OK

b) $\sum f_X(x) = 1$ OK

Funzione di distribuzione cumulativa:

$$F_X(x) = \begin{cases} 0 & \text{se} \quad x < a \\ \frac{x-a}{b-a} & \text{se} \quad a < x < b \\ 1 & \text{per} \quad x \geq b \end{cases}$$

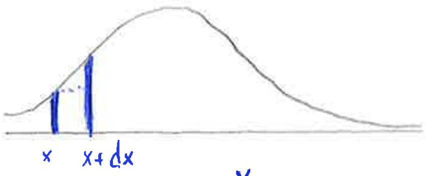
Lezione 20:

Standardizzazione:

Se $\mu = 0$ e $\sigma = 1$, la distribuzione assume il nome di distribuzione standardizzata

Con $Z = \frac{X-\mu}{\sigma}$ $Z \sim N(0,1)$

che è una v.c. tale da trasformare qualsiasi X in Z standard.



$P[x < X \leq x + dx]$

$= f_X(x)\,dx$

Se $X \rightarrow Z = \frac{X-\mu}{\sigma}$

x va a finire in z

dx in $z + dx$ → io voglio $z+dz$, ma se $dz << z$

$P[z < Z \leq z+dz] \doteq f_Z(z)dz$

$f_Z(z)\,dz = f_X(x)\,dx \rightarrow f_Z(z) = f_X(x)\frac{dx}{dz}$

Ma $z = \frac{X-\mu}{\sigma} \rightarrow X = \mu + \sigma X$ ↳ $f_Z(z) = f_X(x)\,\sigma$

$dx = \sigma dz$

$\frac{dx}{dz} = \sigma$

Teorema: $Z = \frac{X-\mu}{\sigma} \rightarrow X = \mu + \sigma Z$

$E[X] = E[\mu + \sigma z] = \mu + \sigma E[\underbrace{\cancel{Z}}_{0}] = \mu$

$V[X] = V[\mu + \sigma z] = \sigma^2 V[\underbrace{\cancel{Z}}_{1}] = \sigma^2$

$P[0 < Z \leq 0{,}09] = P[Z \leq 0{,}09] - P[Z \leq 0] = 0{,}53586 - 0{,}5 = 0{,}03586$

$P[-0{,}27 < Z \leq 0] = P[Z \leq 0] - P[Z \leq 0{,}27] = P[Z \leq 0] - (1 - P[Z \leq 0{,}27]) = 0{,}5 - (1 - 0{,}6064) = 0{,}10642$

Esempio: $P[-0{,}2 < X \leq 0.1]$ $X \sim N(-0{,}1,\ 0.004)$

$$P[-0{,}2 < X \leq 0{,}1] = P\left[\frac{-0{,}2 + 0{,}1}{0{,}02} < \frac{X - (-0{,}1)}{0{,}02} < \frac{0{,}1 - (-0{,}1)}{0.02}\right]$$

$$= P[-1{,}58 < Z \leq 3.16] = P[Z < 3.16] - (1 - P[Z \leq 1{,}58]) = 0{,}943$$

Distribuzione esponenziale:

$$f_X(x) \begin{cases} \lambda e^{-\lambda x} & \text{se } x>0 \\ 0 & \text{se } x \leq 0 \end{cases} \qquad \lambda \in \mathbb{R}^+$$

$$E[X] = \frac{1}{\lambda} \qquad V[X] = \frac{1}{\lambda^2}$$

È una variabile senza memoria:

$$P[X > s+t \mid X > s] = P[X > t]$$

La distribuzione cumulata dell'esponenziale vale:

$$F_X(x) = \begin{cases} 1-\lambda e^{-\lambda x} & \text{se } x>0 \\ 0 & \text{se } x \leq 0 \end{cases}$$

Esempio: X v.c durata pneumatico

$$P[X>25 \mid X>10] = P[X>15] \ ?$$

Non può essere applicata a componenti meccanici, perché hanno usura, in questo caso si usa la Weibull, se è elettronico va bene l'esponenziale.

Trasformazione di variabile casuale:

Dati $x, f_X(x)$, $Y = g(X)$, $f_Y(y)$?

Se X è una v.c discreta

$X, f_X(x)$ $y = g(X)$?

$$f_Y(y) = \begin{cases} \sum\limits_{J:g(x_J)=y_J} f_X(x_J) & y = y_J \\ 0 & \text{altrove} \end{cases}$$

x	-3	-2	-1	0	1	2	3	4
$f_X(x)$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$

Punti massa equiprobabili

$Y = X^2$

I punti massa sono $\{-3,-2,-1,0,1,2,3,4\}$

$\downarrow Y = X^2$

$\{9, \cancel{4}, \cancel{1}, 0, \cancel{1}, \cancel{4}, \cancel{9}, 16\}$

$\downarrow$

$S: \{0, 1, 4, 9, 16\}$

y	0	1	4	9	$\bigcirc\!\!\!\!16$
$f_Y(y)$	$1/8$	$1/4$	$1/4$	$1/4$	$1/8$

$\rightarrow$ È vero che si può ottenere da $+4$ e $\bigcirc\!\!\!\!-4$ $\rightarrow$ NON È UN PUNTO MASSA!

$f_Y(y)=0$ $P[Y=0] = P[X^2=0] = P[X=0] = 1/8$

$f_Y(y)=1$ $P[Y=1] = P[X^2=1] = P[X=\pm] = \frac{1}{4}$

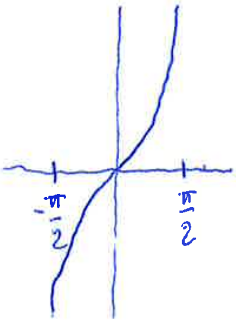
Esempio:

X v.c. con distribuzione uniforme su $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Qual'è la distribuzione di $Y = \tan(X)$

$$f_X(x) = \begin{cases} \frac{1}{\pi} & \text{se } -\frac{\pi}{2} < x \le \frac{\pi}{2} \\ 0 & \text{altrove} \end{cases}$$

$Y = \tan(X)$

$D_X = \left(-\frac{\pi}{2}, +\frac{\pi}{2}\right) \xrightarrow{y=\tan X} D_Y = (-\infty, +\infty)$



È univoca perché per ogni x c'è una sola y.

$$f_Y(y) = \frac{1}{1+y^2} \underbrace{\frac{1}{\pi}}_{f_X(g^{-1}(y))}$$

$y = \tan X \to X = \underbrace{\operatorname{arctg} Y}_{g^{-1}(y)}$

$$\frac{dg^{-1}(y)}{dy} = \frac{1}{1+y^2}$$

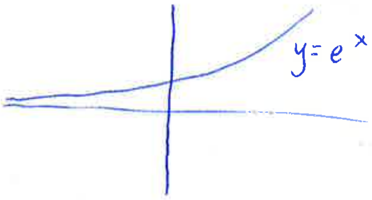
$E[Y] = ? \to E[\operatorname{tg}(X)] = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \operatorname{tg} x \cdot \frac{1}{\pi}\, dx$

$\to \int_{-\infty}^{+\infty} y \frac{1}{1+y^2} \frac{1}{\pi}\, dy =$

Esempio:

$X \sim N(\mu, \sigma^2)$. Distribuzione di $Y = e^X$

$D_X = (-\infty, +\infty) \xrightarrow{Y=e^X} D_Y = (0, +\infty)$



È biunivoca

$y = e^x \to x = \ln y$ $\quad g^{-1}(y)$

$$\frac{dx}{dy} = \frac{dg^{-1}(y)}{dy} = \frac{1}{y}$$

$$f_Y(y) = \begin{cases} \left|\frac{1}{y}\right| \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2} & y > 0 \\ 0 & y \le 0 \end{cases}$$

Distribuzione log-normale

Esempio:

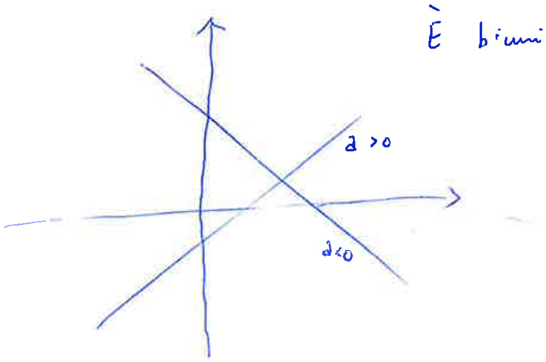
$X \sim N(\mu, \sigma^2)$

$Y = aX + b$

$E[Y] = a\mu_x + b$

$V[Y] = a^2 \sigma_X^2$

$f_Y(y) = ?$



È biunivoca

$D_X = (-\infty, +\infty) \xrightarrow{Y = aX+b} D_Y = (-\infty, +\infty)$

$y = aX + b \rightarrow X = \frac{y-b}{a} = g^{-1}(y) = \frac{y}{a} - \frac{b}{a}$

$\frac{dx}{dy} = \frac{1}{a}$

$$f_Y(y) = \left|\frac{1}{a}\right| \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\frac{y}{a} - \frac{b}{a} - \mu_x}{\sigma}\right)^2}$$

$\forall y \in \mathbb{R}$

$$= \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{1}{2}\left(\frac{y - b - a\mu_x}{a\sigma}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{1}{2}\left(\frac{y - \overbrace{(a\mu_x + b)}^{\mu_y}}{\underbrace{a\sigma}_{\sigma_Y}}\right)} = \frac{1}{\sqrt{2\pi}|a|\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)}$$

Se $y = aX + b$ e $X \sim N(\mu_x, \sigma_X^2) \rightarrow y \sim N(a\mu_x + b, a^2\sigma_X^2)$, cioè le trasformazioni lineari non variano la distribuzione.

$y = aX + b \qquad x = \frac{y - b \longrightarrow b}{a \rightarrow \sigma_X} \rightarrow z = \frac{X - \mu_x}{\sigma_X} = \frac{1}{\sigma_X} X - \frac{\mu_x}{\sigma_X}$

$X \sim N(\mu, \sigma^2)$

$E[Z] = E\left[\frac{1}{\sigma_x} X\right] + E\left[-\frac{\mu_x}{\sigma_X}\right] = \frac{1}{\sigma_x} E[X] - \frac{\mu_x}{\sigma_X} = 0$

Osservazione: il fatto che debba essere una trasformazione biunivoca può essere rimosso:

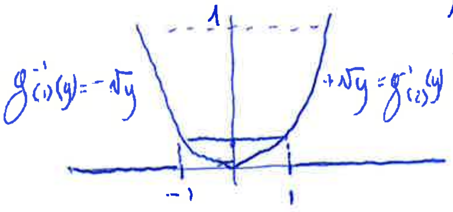
$$f_Y(y) = \sum_i \left|\frac{d}{dy} g^{-1}_{(i)}(y)\right| f_X(g^{-1}(y))$$

Se non ho una trasformazione biunivoca, divido il dominio in tanti sottodomini disgiunti in modo che sia biunivoca su essi e li sommo.

Esempio: $X \sim U(-1, 1)$, $Y = X^2$

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{se } -1 < x < 1 \\ 0 & \text{altrove} \end{cases}$$

$g^{-1}(y) = \pm\sqrt{xy}$



Non è biunivoca

$D_X = (-1, 0) \cup (0, 1]$

$\downarrow y = x^2$

$D_y = (0, 1)$

Lezione 25:

Processi stocastici:

Sono processi che si sviluppano secondo leggi probabilistiche (esempio di coda alle poste).

Noi studiamo i processi di Poisson:

- n° di difetti per unità di superficie
- n° di batteri presenti per unità di volume.
- n° di incidenti lungo un prefissato tratto stradale.

N(t) = n° di eventi poissoniani in un intervallo di lunghezza t.

Rappresentazione grafica del verificarsi di un evento in un intervallo di lunghezza t.

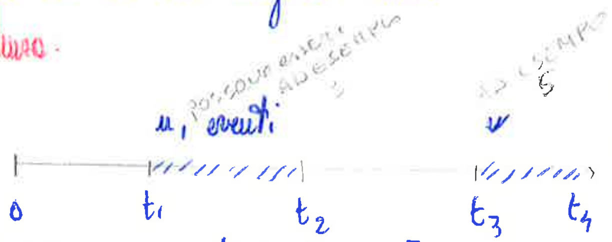
Tutti gli eventi possono essere definiti manifestazioni che soddisfano il processo di Poisson? → SE SODDISFANO LE 3 IPOTESI SI!

Ipotesi:

- esiste $\alpha > 0$ tale che la probabilità che si sviluppi un ~~pre~~ evento in un piccolo Δt è all'incirca uguale ad $\alpha \Delta t$.

$$P[\text{un evento in } \Delta t] = \alpha \Delta t + o(\Delta t)$$

- la probabilità che nello stesso Δt si verifichi più di un evento è trascurabile rispetto a quella che se ne verifichi uno.
- ipotesi di indipendenza.



$E_1 = n_1$ eventi in (t_1, t_2)
$E_2 = n_2$ eventi in (t_3, t_4)

$$P[n_1 \text{ eventi in } (t_1, t_2) \text{ e } n_2 \text{ eventi in } (t_3, t_4)] = P[E_1]P[E_2]$$

Teorema: se le ipotesi precedenti sono soddisfatte, $N(t) \sim \text{Poisson}(\alpha t)$

$\alpha t = \lambda$ dove t è la lunghezza dell'intervallo!

$$f_X(x) = P[N(t) = x] = \frac{e^{-\alpha t} \alpha^x}{x!}$$

Osservazioni:

$P[\text{in } \Delta t \text{ 1 evento}] = \alpha \Delta t$

$P[\text{in } \Delta t \text{ no eventi}] = 1 - \alpha t$

1ª IPOTESI ⇒ SE CONSIDERO UN INTERVALLO PICCOLO (1 min, ... eventi) la prob. che si verifichi una telefonata è = a αt

Da qualsiasi intervallo ~~è~~ uguale

Esempi:

$P[T>t] = P[N(t)=0] = e^{-\lambda}$

$F_T(t) = 1 - P[T>t] = 1 - e^{-\lambda}$

Nb: $P[T>t]$ = affidabilità, ossia la probabilità che un componente ha di superare t senza guasti.

Lezione 27:

Il tipo di campionamento che rispetta di più il concetto di efficienza è quello stratificato, che consiste nel considerare la popolazione e dividerla in n sottogruppi da cui vengono estratti dei minicampioni.

Gr 1	Gr 2	Gr 3	. . .	Gr K
x_{11}	x_{12}	⋮		
x_{21}	x_{22}	⋮		
⋮	⋮	⋮	⋱	
$x_{n_1 1}$	$x_{n_2 2}$			$x_{n_K K}$
$n_1, \bar{x}_1, s_1^2$	$n_2, \bar{x}_2, s_2^2$	. . .		$n_K, \bar{x}_K, s_K^2$

$$n = \sum_{j=1}^{K} n_j$$

Media totale
$$\bar{\bar{x}} = \bar{x}_{tot} = \frac{1}{n}\sum_{j=1}^{K} \bar{x}_j n_j = = \frac{1}{n}\sum_{j=1}^{K}\sum_{i=1}^{n_j} x_{ij}$$

Media parziale
$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{i1}$$
= la media del gruppo 1!

Variabilità interna agli elementi del campione.
$$s_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2$$
→ media della colonna 1

→ Misura qual è la variabilità interna agli elementi del campione!

Variabilità totale
$$s_{tot}^2 = \frac{1}{n}\sum_{j=1}^{K}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{tot})^2 = \frac{1}{n}\sum_{j=1}^{K} n_j s_j^2$$
→ ? = Non è uguale!

$$s_{tot}^2 = \underbrace{\frac{1}{n}\sum_{j=1}^{K} n_j s_j^2}_{\substack{s^2_W \text{ (within)} \\ \text{variabilità interna del} \\ \text{sottogruppo (colonne)}}} + \underbrace{\frac{1}{n}\sum_{j=1}^{K} n_j (\bar{x}_j - \bar{x}_{tot})^2}_{\substack{s^2_B \text{ (between)} \\ \text{variabilità tra sottogruppi} \\ \text{(tra colonne)}}}$$

Un'altra definizione di campione può essere che il campione è un insieme di n variabili casuali indipendenti $\{X_1, X_2, \dots, X_n\}$ e ciascuna X_i avente la stesso distribuzione.

Cosa significa indipendenza di n variabili casuali?

Se abbiamo A e B, sono indipendenti se $P[AB] = P[A]P[B]$.

Consideriamo X, Y:

$$F_X(x) = P[\underbrace{\{s: X(s) \le x\}}_{A}]$$

$$F_Y(y) = P[\underbrace{\{s: Y(s) \le y\}}_{B}]$$

X ind da Y se e solo se

$\Updownarrow$

A e B sono indipendenti $\forall x, \forall y$

Dunque:

$$F_X(x) F_Y(y) = P[\{X \le x \text{ et } Y \le y\}]$$

Cumulata della distribuzione congiunta

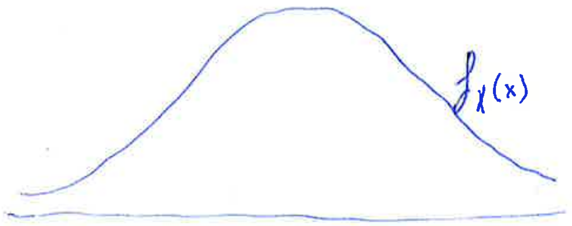
Se sono indipendenti $f_{XY}(x,y) = f_X(x) f_Y(y)$

Ma per n v.c?

$$F_{X_1, \dots X_n}(x_1, \dots x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

Cosa è la distribuzione campionaria?

È la distribuzione congiunta di n variabili casuali costituenti un campione.



$E[X_i] = \mu$

$V[X_i] = \sigma^2$

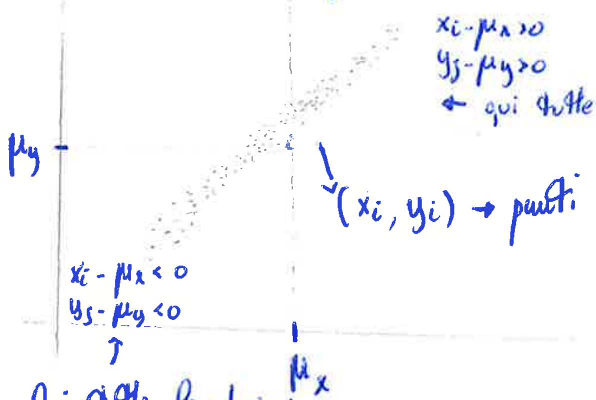
Lezione 29:

$$var[a_1X_1+a_2X_2+a_nX_n] = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \dots + a_n^2\sigma_n^2 + \sum_{i\neq j} a_ia_j \underbrace{cov[X_i,X_j]}_{?}$$

Che cosa è $cov[X_i, X_j]$?

È la covarianza.

$$cov[X,Y] \overset{definizione}{=} E[(X-\mu_X)(Y-\mu_Y)] \overset{calcolo}{=} E[XY]-\mu_X\mu_Y$$



$cov[X,Y]\geq 0$

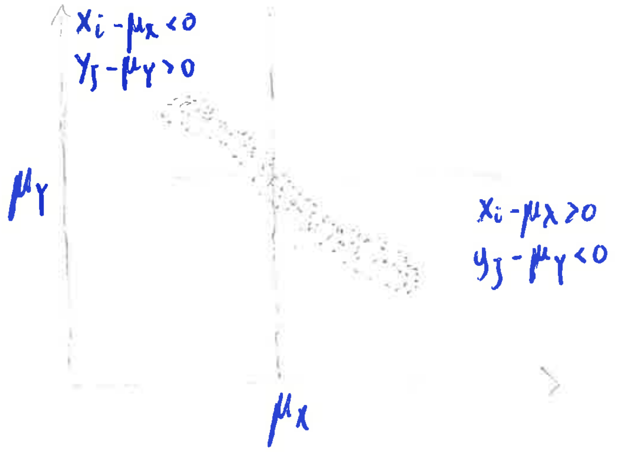
(x_i, y_i) → punti messi con disposizione sottile e allungata (sigaro Havana)

Qui tutte le deviazioni di x e y dalle medie sono negative

$$E[(X-\mu_X)(Y-\mu_Y)] = \sum_{(x_i,y_j)} \underset{>0}{(x_i-\mu_X)}\,\underset{>0}{(y_j-\mu_j)}\,\underline{P[X=x_i, Y=y_j]}$$

$f_{XY}(x_i,y_j)$ deve essere >0 perché è una probabilità.
I fattori del III quadrante contribuiscono alla covarianza con valori positivi, idem nel I
Per cui mi aspetto $cov[X,Y]\geq 0$.

II caso:



$E[(X-\mu_X)(y_j-\mu_y)]$: apporto di termini negativi

$cov[X,Y]\leq 0$

Estraggo un campione che dipende da θ, l'obiettivo è stimare θ o una funzione campionaria.

$f(\cdot;\theta)$ ↳ è un vettore nella normale $\theta = (\mu, \sigma)$

Passiamo all'inferenza vera e propria.

Definizioni:

- statistiche: $g(X_1, X_2 \dots X_n)$ sono funzioni di variabili casuali campionarie che non contengono parametri INCOGNITI

Esempio:

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ è statistica

$\bar{X}_n - \mu$ non lo è. → perché contiene un parametro incognito

Momenti campionari: Def di ordine R - la media aritmetica comune e campionaria

$M'_r = \frac{1}{n} \sum_{i=1}^{n} X_i^r$ → Media aritmetica delle variabili elevate r

$M_r = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^r$

Utile: $M'_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2$

Esempio:

$\{X_1, X_2, \dots X_n\}$ da unif (a, b)

$$\begin{cases} \mu_1' = M_1' \\ \mu_2' = M_2' \end{cases}$$

$$\begin{cases} \mu_1' = \frac{(a+b)}{2} \xrightarrow{} M_1' \\ \mu_2' = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12} + (M_1')^2 \end{cases} \quad M_2'$$

$$\begin{cases} a = 2M_1' - b \\ \frac{(b - 2M_1' + b)^2}{12} = M_2' - (M_1')^2 \end{cases}$$

$$\begin{cases} \frac{4(b - M_1')^2}{12} = M_2' - (M_1')^2 \end{cases} \qquad \begin{cases} a = \cancel{2}M_1' - \cancel{M_1'} \mp \sqrt{3}\sqrt{M_2' - (M_1')^2} \\ b = M_1' \pm \sqrt{3}\sqrt{M_2' - (M_1')^2} \end{cases} \qquad a < b$$

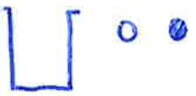
$$\begin{cases} A = \bar{X}_n - \sqrt{3}\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_n)^2} \\ B = \bar{X}_n + \sqrt{3} \end{cases}$$

$\{x_1, x_2, x_n\} \nearrow \hat{a}, \hat{b}$

METODO DELLA MASSIMA VEROSIMIGLIANZA:

$f(x_1; \theta) f(x_2; \theta) \dots f(x_n, \theta) = L(\theta, x_1, x_2; x_n)$ è detta di massima verosimiglianza

Esempio:



4/1 $\quad p = P[\text{"nera"}] = \begin{cases} \frac{1}{5} & \text{se 1 nera} \\ \frac{4}{5} & \text{se 1 bianca} \end{cases}$

Estraggo con reimmissione $\quad X \sim bin(4, p)$

Esito x	0	1	2	3	4
$p = \frac{1}{5}$	$\left(\frac{4}{5}\right)^4$	$4\frac{1}{5}\left(\frac{4}{5}\right)^3$	$6\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^2$	$4\left(\frac{1}{5}\right)^3\frac{4}{5}$	$\left(\frac{1}{5}\right)^4$
$p = \frac{4}{5}$	$\left(\frac{1}{5}\right)^4$	$4\frac{4}{5}\left(\frac{1}{5}\right)^3$	$6\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^2$	$4\left(\frac{4}{5}\right)^3\left(\frac{1}{5}\right)$	$\left(\frac{4}{5}\right)^4$

La funzione di massima verosimiglianza consiste nella probabilità che si realizzi il campione estratto.

$L(\theta; x_1, x_2, x_n) = f(x_1, \theta) f(x_2, \theta) \dots f(x_n, \theta)$ se il campione si è realizzato, esso era il più probabile; se la situazione era 1 nera "avere 0 nere era il più probabile", quindi se ottengo 0 zero o 1 pallina nera, sono più propenso a pensare che ci fosse 1 nera, se ne ottengo 0 3 o 4 devo pensare il contrario.

Lezione 31

$$\{x_1, x_2, \dots, x_n\} \to N(\mu, \sigma^2)$$

$$f_X(x;\sigma,\mu) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$L(\mu,\sigma; x_1,\dots x_n) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2} \times \dots \times \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$\ln L(\mu,\sigma) = \ln \frac{1}{(2\pi)^{n/2}\sigma^n} - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$\frac{\partial \ln L(\mu,\sigma)}{\partial \mu} = -\frac{1}{\cancel{2}}\sum_{i=1}^{n}\cancel{2}\left(\frac{x_i-\mu}{\sigma}\right)\cancel{\frac{1}{\sigma}} = 0$$

$$\cancel{\frac{1}{\sigma}}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\sum_{i=1}^{n} x_i - n\mu = 0 \to \boxed{\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}_n}$$ Stimatore di μ

PER σ?

$$\ln L(\mu,\sigma^2) = \ln \frac{1}{(2\pi^{n/2}\sigma^n)} - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$= -\ln (2\pi)^{n/2}\sigma^n - \frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$= -\ln(2\pi)^{n/2} - \ln(\sigma^2)^{n/2} - \frac{1}{2}\sum(\ \ )^2$$

$$\frac{\partial L(\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{\cancel{2}}\frac{1}{\cancel{\sigma^2}} + \frac{1}{\cancel{2}}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^{\cancel{4}2}} = 0$$

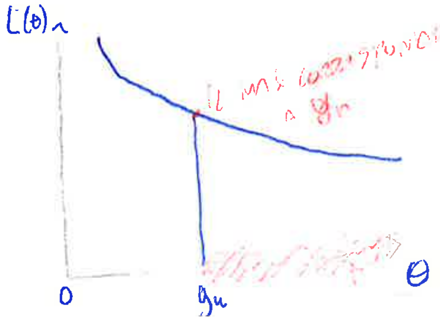
$$\boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2}$$ Stimatore per σ^2

Esempio: $\{x_1, x_2 \dots x_n\}$ da unif$(0,\theta)$

$$f_X(x) = \begin{cases} \frac{1}{\theta} & \text{se } 0 < x \leq \theta \\ 0 & \text{altrove} \end{cases}$$

$$L(\theta; x_1,\dots x_n) = \underbrace{\frac{1}{\theta}}_{0<x_1<\theta} \times \underbrace{\frac{1}{\theta}}_{0\leq x_2<\theta} \times \dots \times \underbrace{\frac{1}{\theta}}_{0\leq x_n<\theta} = \frac{1}{\theta^n}$$

$\to \theta \geq \max(x_1, x_2 \dots x_n) = y_n$



il max corrisponde a y_n

$$\theta_{ML} = \max(X_1,\dots,X_n) = y_n$$

→ posso prendere solo i valori > y_n

Teorema limite centrale

Se $f_X(x) \sim (\mu, \sigma^2)$

$f_{\bar{X}_n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, al tendere di n ad ∞.

con $n \to \infty$

Allora sono in grado di fare molti calcoli:

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

E se la distribuzione è normale?

↓

Proprietà riproduttive delle normale:

Date $X_1, X_2 \dots X_n$ indipendenti

$X_i \sim N(\mu_i, \sigma_i^2)$

$\sum_{i=1}^{n} a_i X_i = W \sim N(\mu_W, \sigma^2_W)$

$\mu_W = E[\sum a_i X_i] = \sum_{i=1}^{n} a_i \mu_i$

$\sigma^2_W = V[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i^2 \sigma_i^2$

È la distribuzione χ^2. $\to$ Distribuzione della varianza campionaria

z_i: n variabili casuali normali standard indipendenti.

$Z_i \sim N(0,1)$ $\qquad E[\chi^2] = n$

$$\chi^2 = \sum_{i=1}^{n} z_i^2$$

χ^2 ha una distribuzione chi-quadro con n gradi di libertà.

Se le v.c. non fossero standard ma qualsiasi? Standardizzo la variabile

$$U = \sum_{i=1}^{n} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2$$ che segue una χ^2 con n gradi di libertà.

ESERCIZIO:

$\{x_1, x_2, \dots, x_n\}$ estratto da $N(\mu, \sigma^2)$

Se X_i sono indipendenti e $E[X_i] = \mu$, $V[X_i] = \sigma^2$

$$z_i = \frac{X_i - \mu}{\sigma} \to z_i^2 \sim \sum_{i=1}^{n} z_i^2 = \chi^2_{(n-1)}$$

Se non conosco μ? Possiamo usare la media campionaria $\bar{X}_n$, dove sta la differenza? la nostra χ^2 ha n-1 gradi di libertà.

Perché?

Se ho un campione di dimensione n, esso ha n g.d.l.

Se calcolo la media $\bar{X}_n = \frac{1}{n} \sum x_i$, perdo un grado di libertà perché sto imponendo un vincolo.

Vediamolo chiaramente: voti studente

$\{24, 22, 27, 28\} \to 4$ g.d.l

$\{24, 22, 27, V\}$ $\quad \bar{X}_4 = 25.6 \to 3$ g.d.l perché i primi sono valori liberi e il 4° dipende dagli altri.

Consideriamo:

$\nearrow S_m^2 \sim \chi^2_{m-1}$

$$\frac{\frac{1}{m-1}\sum_{i=1}^{m}(X_i - X_m)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - Y_n)^2} \sim F_{m-1,\,n-1} = \frac{\frac{\chi^2_{m-1}}{m-1}}{\frac{\chi^2_{n-1}}{n-1}}$$

$\hookrightarrow S_n^2 \sim \chi^2_{n-1}$

Distribuzione t di Student

Z v.c standardizzata $\sim N(0,1)$

U v.c $\sim \chi^2_n$

Z e U indipendenti

$$T = \frac{Z}{\sqrt{\frac{U}{n}}} \sim T \text{ di Student con } n \text{ g.d.l}$$

Se $n > 30 \rightarrow t_{n,\alpha} \cong z_\alpha$

$\{X_1, X_2 \dots X_n\}$ campione dalla normale

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$U = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim X^2_{n-1}$$

$$T = \frac{Z}{\sqrt{\frac{U}{\boxed{n-1}}}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\frac{1}{\sqrt{n-1}}\sqrt{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2}} = \frac{\bar{X}_n - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$$

$\downarrow$ gdl della U

Consistenza

Sia $T_1, T_2, \ldots T_n$ una successione di stimatori di θ. La successione è consistente se:

$$\lim_{n\to\infty} MSE_{T_n} = 0$$

Mentre se $T_1, T_2, \ldots, T_n$ è una successione di stimatori di θ, si dice consistente debolmente se:

$$\lim_{n\to\infty} P[\,|T_n-\theta| < \varepsilon\,] = 1$$

Se lo stimatore è corretto, allora è anche consistente se:

$$\lim_{n\to\infty} V[T_n] = 0$$

Osservazione

Se ho uno stimatore non distorto per θ, allora $\varphi(T_n)$ è uno stimatore corretto per $\varphi(\theta)$ solo se φ è una funzione lineare.

Osservazione:

$E[S_n^2] = \sigma^2$

$\sqrt{S_n^2}$ non è stimatore corretto per σ perché la radice non è lineare.

Efficienza:

CRITERIO 1

Se per θ esistono più stimatori corretti, quale scegliere?

$$P[\,|T_1 - \tau(\theta)| < \varepsilon\,] \geq P[\,|T_2 - \tau(\theta)| < \varepsilon\,]$$

T_1 è più concentrato di T_2

CRITERIO 2

$$P[\,|T_1 - \theta| < |T_2 - \theta|\,] \geq 0.5$$

T_1 è più efficiente

CRITERIO 3

Scelgo T_1 (corretto) invece di T_2 (corretto) se:

$$V[T_1] = E[(T_1-\theta)^2] \leq E[(T_2-\theta)^2] = V[T_2]$$

Si confrontano gli MSE

T_1 è più efficiente di T_2

perché $MSE_{T_1} \leq MSE_{T_2}$

Lezione 36

Per una distribuzione normale è impossibile calcolare il valore di una stima puntuale, per cui passiamo ad una stima intervallare, in quanto quella puntuale è sempre pari a zero.

Stima per intervalli

Sia $\{X_1, X_2 \dots X_n\}$ un campione casuale da $f(\cdot, \theta)$ e $T_1 \le T_2$ statistiche e per le quali $P[T_1 \le \theta < T_2] = 1-\alpha$. Nb: T_1 e T_2 maiuscole perché è un intervallo casuale.

(T_1, T_2) si chiama intervallo di fiducia al livello $(1-\alpha)\%$

$1-\alpha \to$ livello di fiducia

Stima per intervalli della media μ

1- Varianza della popolazione nota

$\{X_1, X_2 \dots X_n\}$ da popolazione con distribuzione normale con μ incognita e σ^2 nota.

$$\hat{\mu} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$\downarrow$ TLC o proprietà riproduttiva

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$\downarrow$

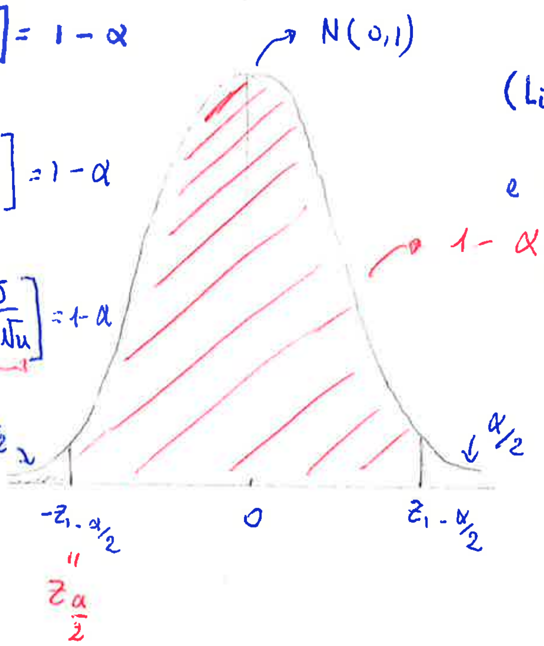
$$P\left[-z_{1-\alpha/2} < Z \le z_{1-\alpha/2}\right] = 1-\alpha$$

$\downarrow$

$$P\left[-z_{1-\alpha/2} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

$\downarrow$

$$P\left[\underbrace{\bar{X}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}_{L_i} < \mu < \underbrace{\bar{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{L_s}\right] = 1-\alpha$$



Quindi:

$$(L_i, L_s) = \left(\bar{X}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}}\right)$$

e l'intervallo di fiducia di μ è:

$$(l_i, l_s) = \left(\bar{x}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

Varianza della popolazione non nota:

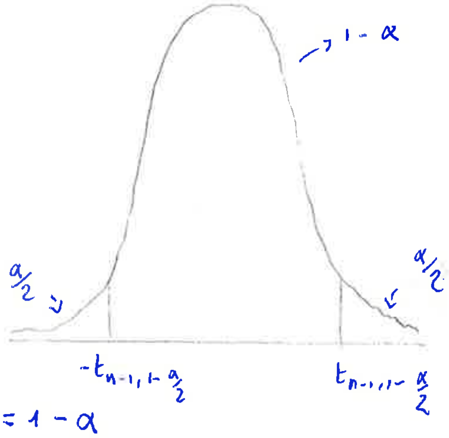
Se σ^2 non è nota possiamo usare la varianza campionaria

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$$

$$\downarrow$$

$$P\left[-t_{n-1,1-\alpha/2} < T \leq t_{n-1,1-\frac{\alpha}{2}}\right] = 1-\alpha$$



sostituisco la T e risolvo rispetto a μ

$$P\left[\bar{X}_n - t_{n-1,1-\alpha/2} \frac{\sqrt{S_n^2}}{\sqrt{n}} < \mu < \bar{X}_n + t_{n-1,1-\alpha/2} \frac{\sqrt{S_n^2}}{\sqrt{n}}\right] = 1-\alpha$$

$$(L_i, L_s) = \left(\bar{X}_n - t_{n-1,1-\alpha/2} \frac{\sqrt{S_n^2}}{\sqrt{n}}, \bar{X}_n + t_{n-1,1-\alpha/2} \frac{\sqrt{S_n^2}}{\sqrt{n}}\right)$$

Nb: se la varianza non è nota e $n-1 \geq 30$, la t è approssimabile con una standard.

Esempio:

8 prove di un motore di macchina che consuma $\{14, 12, 11, 13, 15, 12, 16, 13\}$ l per 100 km
Costruire un'intervallo di fiducia al 99 %.

$$\bar{X}_8 = \frac{1}{8}\sum_{i=1}^{8} X_i = 13.25$$

$$S_8^2 = \frac{1}{8-1} \sum_{i=1}^{8} (X_i - 13.25)^2 = 2.786 \ l^2 \Rightarrow \sqrt{S^2} = 1.67 \ l$$

$1 - \alpha = 0.99$
$\alpha = 0,01$
$\frac{\alpha}{2} = 0,005$

$$t_{7,0.995} = 3.499$$

Intervallo di fiducia per μ

$$(l_i, l_2) = (\bar{x}_8 - t_{7,0.995}, \bar{x}_8 + t_{7,0.995}) = (11,186, \ 15.314$$

Lezione 38

Stima per intervalli della differenza di due medie:

$\{x_{11}, x_{12}, \dots x_{1n_1}\}$ campione di dimensione con μ_1 e σ_1^2 distribuita normalmente

$\{x_{21}, x_{22} \dots x_{2n_2}\}$ campione di dimensione con μ_2 e σ_2^2 distribuita normalmente, indipendente dal precedente

$$\bar{X}_1 - \bar{X}_2 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i} - \frac{1}{n_2}\sum_{i=1}^{n_2} X_{2i}$$ miglior stimatore

E la varianza?

I) σ_1^2 e σ_2^2 note

$$\sigma^2_\Delta = \sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2}$$

$$P\left[-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}\right] = 1-\alpha \qquad Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\boxed{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}} \rightarrow \sigma_\Delta$$

$$P\left[(\bar{X}_1 - \bar{X}_2) - \sigma_\Delta z_{1-\alpha/2} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + \sigma_\Delta z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

Per cui $(l_i, l_s) = \left((\bar{X}_1 - \bar{X}_2) - z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{X}_1 - \bar{X}_2) + z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$

II) σ_1^2 e σ_2^2 non note, ma possono ritenersi uguali

Siccome non sono note, sostituiamo σ_1^2 e σ_2^2 con S_1^2 e S_2^2 con (n_1-1) e (n_2-1) g.d.l rispettivamente

Siccome σ_1^2 e σ_2 sono ritenibili uguali $\sigma_1^2 = \sigma_2^2 = \sigma^2$, quale prendo tra le 2?

Prendo la media pesata:

$$\hat{\sigma}^2 = S^2_{pool} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \sim \chi_{n_1+n_2-2}$$

$$P\left[-t_{n_1+n_2-2,1-\frac{\alpha}{2}} < T < t_{n_1+n_2-2,1-\frac{\alpha}{2}}\right] = 1-\alpha \qquad T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{pool}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

$$(l_i, l_s) = \left((\bar{X}_1 - \bar{X}_2) - t_{n_1+n_2-2,1-\frac{\alpha}{2}} S_{\bar{X}_1 - \bar{X}_2}, (\bar{X}_1 + \bar{X}_2) + t_{n_1+n_2-2,1-\frac{\alpha}{2}} S_{\bar{X}_1 - \bar{X}_2}\right)$$

Stima per intervalli di varianza

$\{X_1, X_2 \dots X_n\}$ da $N(\mu, \sigma^2)$

$$V = (n-1)\frac{S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2 \sim \chi^2_{n-1}$$

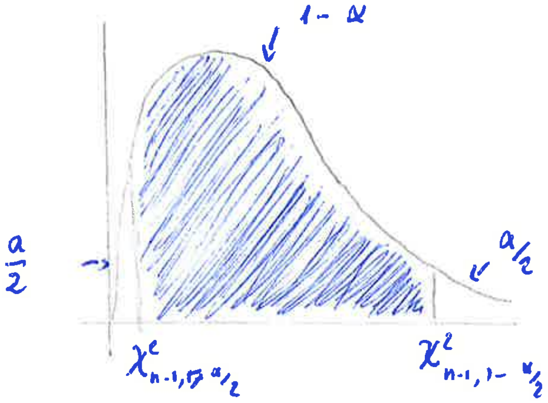
$$P\left[\chi^2_{n-1,\frac{\alpha}{2}} < \chi^2_{n-1} \leq \chi^2_{n-1,1-\frac{\alpha}{2}}\right] = 1-\alpha$$

$$\downarrow$$

$$P\left[\chi^2_{n-1,\frac{\alpha}{2}} < \frac{(n-1)S_n^2}{\sigma^2} < \chi^2_{n-1,1-\frac{\alpha}{2}}\right] = 1-\alpha$$

$$\downarrow$$

$$(L_i, L_s) = \left(\frac{(n-1)S_n^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right)$$



Osservazione

Se $n>30$, la distribuzione è approssimabile con una standard $z_{1-\frac{\alpha}{2}}$ con $\left(\sigma, \frac{\sigma}{\sqrt{2n}}\right)$

$$P\left[-z_{1-\alpha/2} < \frac{S-\sigma}{\sigma/\sqrt{2n}} \leq z_{1-\frac{\alpha}{2}}\right] = 1-\alpha$$

$$\downarrow$$

$$P\left[\frac{S_n}{1+\frac{z_{1-\alpha/2}}{\sqrt{2n}}} < \sigma \leq \frac{S_n}{1-\frac{z_{1-\alpha/2}}{\sqrt{2n}}}\right] = 1-\alpha$$

$$(L_i, L_s) = \left(\frac{S_n}{1+\frac{z_{1-\alpha/2}}{\sqrt{2n}}}, \frac{S_n}{1-\frac{z_{1-\alpha/2}}{\sqrt{2n}}}\right)$$

Intervallo di fiducia per le proporzioni

Vale se $n \to \infty$

$X \sim Bern(\hat{p})$

$\sum_{i=1}^{n} X_i \sim N$ se $n \to \infty$

$\sum_{i=1}^{n} X_i \sim bin(n, p)$

Non conoscendo p, mi serve uno stimatore $P = \bar{X}_n$

$P \to N(\mu_p, \sigma_p)$

$$\mu_p = E\left[\frac{\sum X_i}{n}\right] = \frac{1}{n}\overset{np}{\cancel{\sum E[X_i]}} = \frac{1}{\cancel{n}} \cdot \cancel{n}p = p$$

$$\sigma_p = \sqrt{V\left[\frac{\sum X_i}{n}\right]} = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2} = \sqrt{\frac{1}{n^{\cancel{2}}}\cancel{n}(p(1-p))} = \sqrt{\frac{p(1-p)}{n}}$$

$$P \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

$$P\left[-z_{1-\alpha/2} < Z < z_{1-\frac{\alpha}{2}}\right] = 1 - \alpha$$

Con $Z = \frac{P - p}{\sqrt{\frac{p(1-p)}{n}}}$ da cui $(L_i, L_s) = \left(\hat{P} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}, \hat{P} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

Test di ipotesi:

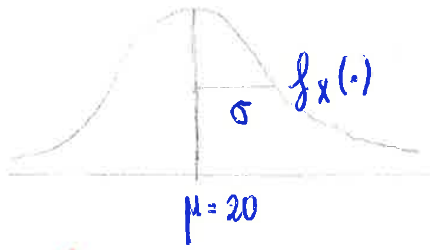
Introduciamo con esempio:

Ditta che produce cosmetici afferma che sostanza si è abbassata da 25 $\frac{mg}{dm^3}$ a 20 $\frac{mg}{dm^3}$

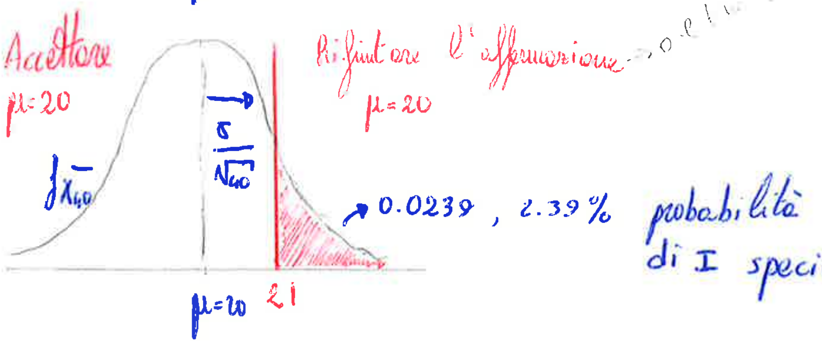
Come fare per accettare o confutare l'ipotesi?

Si prende un campione di 40 barattoli e si misurano. Si accetta se il livello medio è inferiore a 21 $^{mg}/_{dm^3}$ e di rifiutarlo se è maggiore di $\boxed{21}$ $\frac{mg}{dm^3}$.
↳ si dà un po' di margine

X: misura livello



Analizzo



probabilità di fare un errore di I specie

Scenario: il livello si è ridotto davvero

1) Il campione ha fornito un livello medio < 21 $\frac{mg}{dm^3}$. → Accetto affermazione
2) Il campione ha fornito un livello medio > 21 $\frac{mg}{dm^3}$ → Rifiuto affermazione

Questo è un errore di I specie: rifiuto affermazione quando è vera.

Scenario: il livello non si è ridotto

1) Il campione ha fornito un livello medio < 21 $\frac{mg}{dm^3}$ → Accetto affermazione
2) Il campione ha fornito un livello medio > 21 $^{mg}/_{dm^3}$ → Rifiuto affermazione

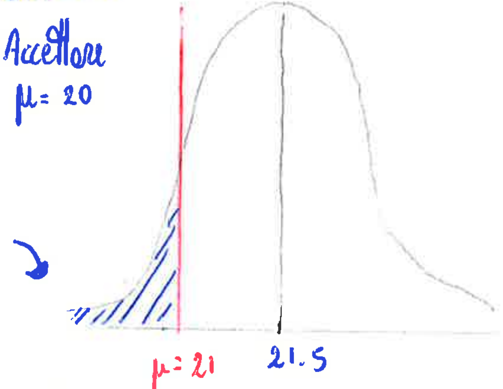
Questo è un errore di II specie: accetto affermazione quando è falsa.
↳ l'azienda non ha detto il vero

Lezione 41

È se sbagliassi al contrario?

Accettare
$\mu = 20$

Errore di II specie ↘

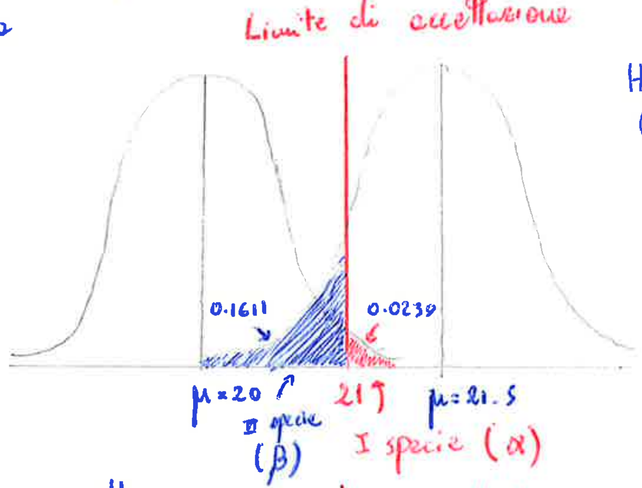


$$P[\text{"II specie"}] = P[X < 21 \mid \mu = 21.5]$$

$$= P\left[Z < \frac{21 - 21.5}{\frac{3.2}{\sqrt{40}}}\right] = P[Z < -0.988] = 0.1611$$

affermazione azienda
H₀ è vera
(H₁ falsa)

H₀ è falsa
(H₁ vera)



Accettazione	Rifiuto
Accetto H₀ Rifiuto H₁	Rifiuto H₀ Accetto H₁

Situazione incognita	Decisione presa in base risultati campione	
	Accetto H₀	Rifiuto H₀
H₀ vera	Decisione corretta $P = 1 - \alpha$	Errore di I specie $P = \alpha$
H₀ falsa	Errore di II specie $P = \beta$	Decisione corretta $P = 1 - \beta$

p- value

Probabilità che la statistica assuma valori maggiori o uguali del valore calcolato. Quindi per l'esercizio precedente è la probabilità di trovare un valore maggiore di 1.77 nel caso di test monolaterale.

$P[Z > 1.77] = 0.0384 = $ p-value

Nel caso di bilaterale:

$P[|Z| > 1.77]$

Se :

$0 <$ p- value < 0.01	altamente significativo, differisce da z_{calc} con alta probabilità e il test va rifiutato con alta probabilità.
$0.01 <$ p-value < 0.05	significativo
$0.05 <$ p- value < 0.1	debolmente significativo.
p-value > 0.1	non significativo

(Per affermare che un test di ipotesi nullo sia vero serve un campione con n = n° elementi popolazione

↳ Per rifiutarlo basta campione con n = 1 se la varianza nota)

Lezione 44:

Applicazioni sui test d'ipotesi:

$X_1 \sim (\mu_1, \sigma_1^2)$

$X_2 \sim (\mu_2, \sigma_2^2)$

Output minitab

Variable	Total Count	Mean	StDev	Variance
A	$n_1 = 23$	17.796	3.504	12.279
B	$n_2 = 18$	18.239	2.972	8.832

$\bar{X}$ S_n^2

$H_0: \mu_1 = \mu_2 \quad (\mu_1 - \mu_2 = 0)$

$H_A: \mu_1 \neq \mu_2$ (test bilaterale)

$\alpha = 0.05$

$\bar{X}_{1(23)} \sim N\left(\mu_1, \frac{\sigma_1^2}{23}\right)$

$\bar{X}_{2(18)} \sim N\left(\mu_2, \frac{\sigma_2^2}{18}\right)$

$$\bar{X}_{1(23)} - \bar{X}_{2(18)} \sim N(\mu_1 - \mu_2, \sigma_\Delta)$$

$$\sigma_\Delta = \sqrt{\frac{\sigma_1^2}{23} + \frac{\sigma_2^2}{18}}$$

PRIMA:

$H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 \neq \sigma_2^2$ (test bilaterale)

$$F = \frac{\dfrac{\sigma_1^2}{S_{1(n_1)}^2}}{\dfrac{\sigma_2^2}{S_{2(n_2)}^2}} \sim F_{n_1-1, n_2-1}$$

$$P\left[f_{n_1-1, n_2-1, \alpha/2} < F_{n_1-1, n_2-1} < f_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}\right] = 1-\alpha$$

sotto $H_0 = 1$

$$P\left[f_{n_1-1, n_2-1, \alpha/2}\, \boxed{\frac{\sigma_2^2}{\sigma_1^2}} < \frac{S_1^2}{S_2^2} < f_{n_1-1, n_2-1, 1-\frac{\alpha}{2}}\right] = 1-\alpha$$

Nelle stime intervallari al centro c'era $\frac{\sigma_1^2}{\sigma_2^2}$, ora S^2, perché ora calcoliamo la probabilità che le medie campionarie stiano nell'intervallo.

$$\frac{S_1^2}{S_2^2} = \frac{12.270}{8.832} = 1,39$$

$$f_{22,17,0.025} = \frac{1}{f_{17,22,0.975}} = 0.408$$

$$f_{22,17,0.975} = 2.58$$

$\rightarrow 0.408 < 1.39 < 2.58$

Non posso rifiutare $H_0: \sigma_1^2 = \sigma_2^2$

Calcolare i valori della curva caratteristica per $H_A: \mu_1-\mu_2 = 1.15$, $\mu_1-\mu_2=-1.15$, $H_A: \mu_1-\mu_2 = 0$] Coincide con $H_0: \mu_A-\mu_B = 0$, per cui $\beta_{H_A: \mu_1-\mu_2=0} = 1-\alpha = 0.95$ Significa calcolare β.

Per 1.15 e -1.15 vedo che sono simmetrici, rispetto ad H_0, quindi ne calcolo uno

$H_0 = \mu_1-\mu_2 = 0$

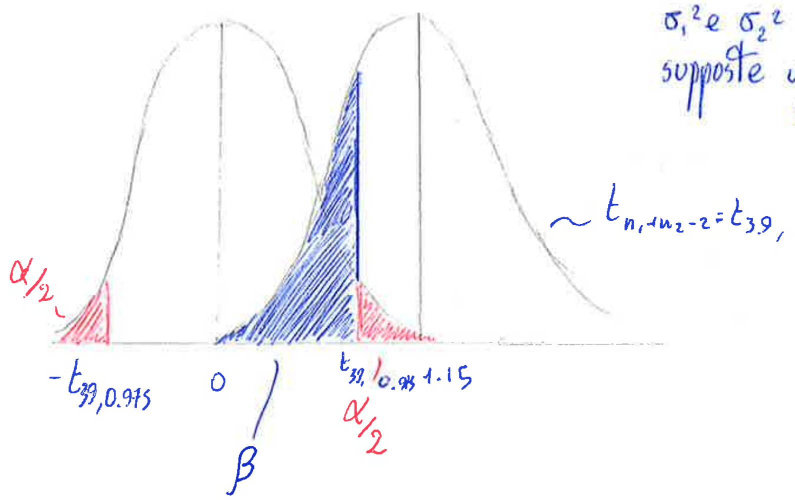
$H_A: \mu_1-\mu_2 = 1.15$

$\alpha = 5\%$

$n_1 = 23$

$n_2 = 18$

σ_1^2 e σ_2^2 non note ma supposte uguali



$\beta = H_A: \mu_1-\mu_2 = 1.15$

$P\left[(\bar{X}_1-\bar{X}_2)_{\Delta,min} < \bar{X}_{1(23)} - \bar{X}_{2(18)} \leq (\bar{X}_1-\hat{X}_2)_{\Delta,max}\right]$ supponendo $\mu_1-\mu_2 = 1.15]$

Accetto H_0 ogni qualvolta $\bar{X}_1-\bar{X}_2$ è in quell'intervallo

$\tilde{X}_{\Delta,min} : -t_{n_1+n_2-2,1-\frac{\alpha}{2}} = \dfrac{\bar{X}_{\Delta,min} - 0}{\sqrt{S^2_{pool}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \rightarrow \bar{X}_{\Delta,min} = -t_{n_1+n_2-2,1-\frac{\alpha}{2}} \sqrt{S^2_{pool}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

$= -2.087$

Quindi i limiti di accettazione di H_0 vengono dettati dai quantili della distribuzione sotto H_0, però possono essere anche dettati dalla caratteristica sotto ipotesi e quindi saranno quei valori che una volta standardizzati danno i quantili.

$\bar{X}_{\Delta,max} : t_{n_1+n_2-2,0.975} = \dfrac{X_{\Delta max} - 0}{\sqrt{S^2_{pool}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \rightarrow X_{\Delta max} = 2.087$

simmetrici se $H_0 = 0$

$P[-2.087 < \bar{X}_1 - \bar{X}_2 < 2.087]$ $\mu_1-\mu_2 = 1.15]$

Lezione 45:

$\beta_{H_0:\mu=0.335} = P[\bar{X}_{min} < \bar{X}_{35} < \bar{X}_{max}]$

$\bar{X}_{min}$: $\dfrac{\bar{x}_{min} - 0.34}{\frac{0.01}{\sqrt{35}}} = -z_{0.975} \rightarrow \bar{X}_{min} = 0.337$

$\bar{X}_{max}$: $\dfrac{\bar{x}_{max} - 0.34}{\frac{0.01}{\sqrt{35}}} = z_{0.975} \rightarrow \bar{x}_{max} = 0.343$

$= P\left[\dfrac{0.337 - \overset{\rightarrow H_A}{0.335}}{\frac{0.01}{\sqrt{35}}} < z \leq \dfrac{0.343 - 0.335}{\frac{0.01}{\sqrt{35}}}\right] = P[1.18 < z < 4.73] \cong 12\%$

Esercizi di test d'ipotesi

$\beta = 0.15$

$H_0: \mu = 1$ (test bilaterale)

$H_A: \mu = 0.9$ $\alpha = 0.1$

n minimo affinché il campione da estrarre abbia $\sigma^2_x = 0.81$

SVOLGIMENTO:

$H_0: \mu = 1$

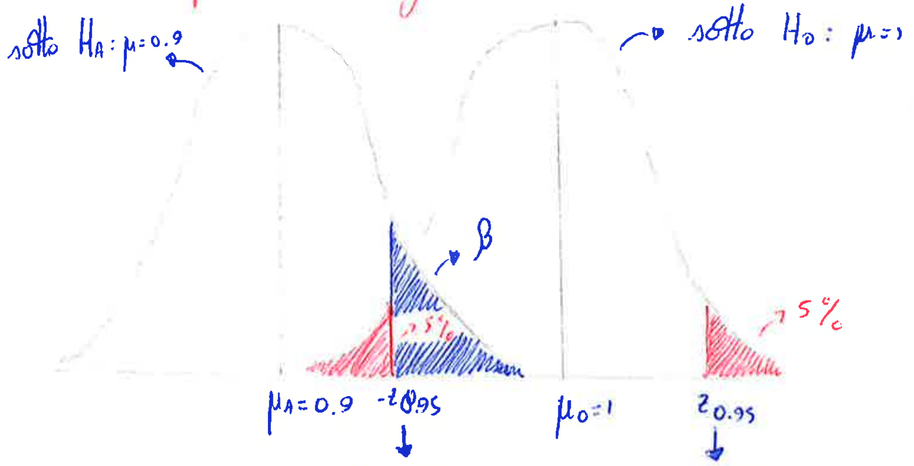
$H_A: \mu \neq 1$

$\sigma^2 = 0.81$

$\alpha = 10\%$

n ? $\beta_{H_A:\mu=0.9} = 0.15$

Normali perché n incognito



$0.15 = \beta_{H_A:\mu=0.9} = P[\bar{X}_{min} < \hat{\bar{X}}_n < \bar{X}_{max}]$ sotto $H_A: \mu = 0.9]$

$\bar{x}_{min}$: $-1.645 = \dfrac{\bar{x}_{min} - 1}{\frac{0,9}{\sqrt{n}}} \rightarrow \bar{X}_{min} = 1 - 1.645 \dfrac{0.9}{\sqrt{n}}$

$\bar{X}_{max}$: $1.645 = \dfrac{\bar{x}_{max} - 1}{\frac{0.9}{\sqrt{n}}} \rightarrow \bar{x}_{max} = 1 + 1.645 \dfrac{0.9}{\sqrt{n}}$

$P\left[1-1{,}645 \dfrac{0.9}{\sqrt{n}} < \bar{X}_n < 1+1{,}645 \dfrac{0.9}{\sqrt{n}}\right] = P\left[\dfrac{\left(1-1.645\frac{0.9}{\sqrt{n}}\right) - 0.9}{\frac{0.9}{\sqrt{n}}} < z < \dfrac{\left(1+1.645\frac{0.9}{\sqrt{n}}\right) - 0.9}{\frac{0.9}{\sqrt{n}}}\right]$

Lezione 47:

Modifica dell'esercizio dei test d'ipotesi:

$S^2_{25} = 0.81$

In questo caso cambia la distribuzione, avremo una $\pm t_{24, 0.95} = \pm 1.71$

Quindi:

$0.15 = \beta_{H_0: \mu = 0.9} = P[\text{Accetto } H_0: \mu = 1 \mid \mu = 0.9] = P[\bar{x}_{min} < \bar{X}_n < \bar{x}_{max} \mid H_A \text{ vera}]$

$\bar{X}_{min} = -1.71 = \frac{\bar{x}_{min} - 1}{\sqrt{0.81/n}} \rightarrow \bar{x}_{min} =$

$\bar{X}_{max} = +1.71 = \frac{x_{max} - 1}{\sqrt{0.81/n}} \rightarrow X_{max} =$

Per cui

$P\left[-1.71 + 0.1\frac{\sqrt{n}}{0.9} < T_{24} < 1.71 + 0.1\frac{\sqrt{n}}{0.9}\right] \cong P\left[T_{24} \geq -1.71 + 0.1\frac{\sqrt{n}}{0.9}\right]$

$-1.71 + 0.1 \cdot \frac{\sqrt{n}}{0.9} = 1.0593 \rightarrow \boxed{n \geq 622}$

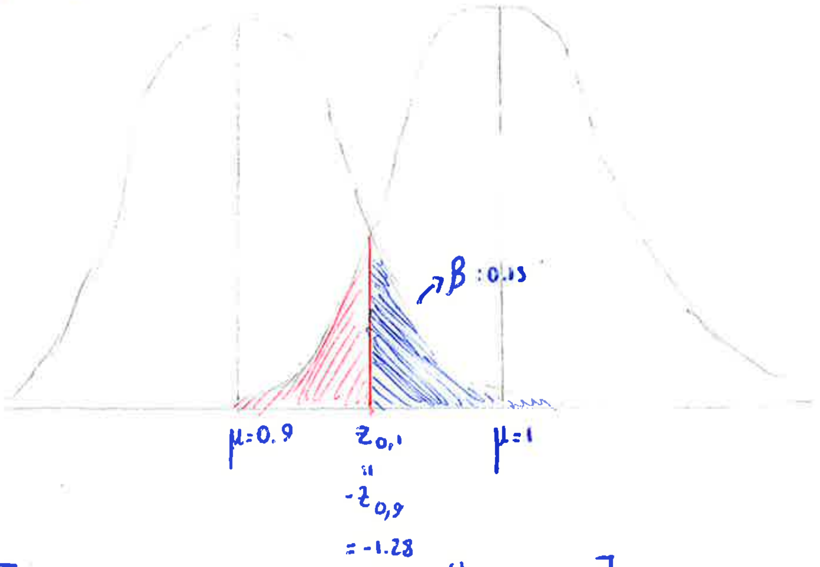
Test monolaterale (inferiore perchè $H_A < H_0$)

$H_0: \mu = 1$

$H_A: \mu = 0.9$

$\sigma^2 = 0.81$

$n? \beta_{H_A: \mu = 0.9} = 0.15$



$0.15 = \beta_{H_A: \mu = 0.9} = P[\bar{X}_n > \bar{x}_{min} \mid \text{dato che è vera } H_A: \mu = 0.9]$

$\bar{x}_{min} : \frac{\bar{x}_{min} - 1}{\sqrt{\frac{0.81}{n}}} = -z_{0.9} = -1.28 \rightarrow \bar{x}_{min} = 1 - 1.28\sqrt{\frac{0.81}{n}}$

$= P\left[Z > \frac{\left(1 - 1.28\sqrt{\frac{0.81}{n}}\right) - 0.9}{\sqrt{\frac{0.81}{n}}}\right] = P\left[Z > 0.1\sqrt{\frac{n}{0.81}} - 1.28\right] = 0.15$ per cui $z_{0,85}$

$\Rightarrow 1.036 = -1.28 + 0.1\sqrt{\frac{n}{0.81}}$

$\boxed{n \geq 435}$

Definizioni:

$$\frac{1}{K(n-1)}\boxed{\sum_{j=1}^{K}\sum_{i=1}^{n}(y_{ij}-\bar{y}_j)^2} = \frac{SS_W}{K(n-1)} = \boxed{MS_W} \rightarrow \text{quadrati medi}$$

SS_W: somma dei quadrati

Lezione 42:

$$\hat{\sigma}^2_{\bar{y}} = s^2_{\bar{y}} = \frac{\sum_{j=1}^{K}(\bar{y}_{\cdot j}-\bar{y}_{\cdot\cdot})^2}{k-1}$$ Somma di quadrati di colonna.

È una stima corretta solo sotto H_0 di $\sigma^2_{\bar{y}}$, altrimenti sarà corretta da un bias.

Ma $\sigma^2_{\bar{y}} = \frac{\sigma^2_y}{n}$

$$\hookrightarrow S^2_y = n S^2_{\bar{y}} = \frac{n\sum_{j=1}^{K}(\bar{y}_{\cdot j}-y_{\cdot\cdot})^2}{k-1} = \frac{SS_B}{k-1} = MS_B$$

Altra misura è la somma dei quadrati totali corretti

$$SS_{TC} = \sum_{i=1}^{n}\sum_{j=1}^{K}(y_{ij}-\bar{y}_{\cdot\cdot})^2 = \Sigma_i\Sigma_j\underset{\substack{\uparrow\\ \text{~~fattore~~ caso}}}{(y_{ij}-\bar{y}_{\cdot j})} + \Sigma_i\Sigma_j\underset{\substack{\uparrow\\ \text{fattore}}}{(\bar{y}_{\cdot j}-y_{\cdot\cdot})^2}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{K}(y_{ij}-\bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{n}\sum_{j=1}^{K}y_{ij}^2 + \sum_{i=1}^{n}\sum_{j=1}^{K}\bar{y}_{\cdot\cdot} - 2\sum_{i=1}^{n}\sum_{j=1}^{K}y_{ij}\bar{y}_{\cdot\cdot}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{K}y_{ij}^2 + nK\bar{y}^2_{\cdot\cdot} - 2\bar{y}_{\cdot\cdot}^2 nk\cancel{K}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{K}y_{ij}^2 - nk\bar{y}^2_{\cdot\cdot}$$

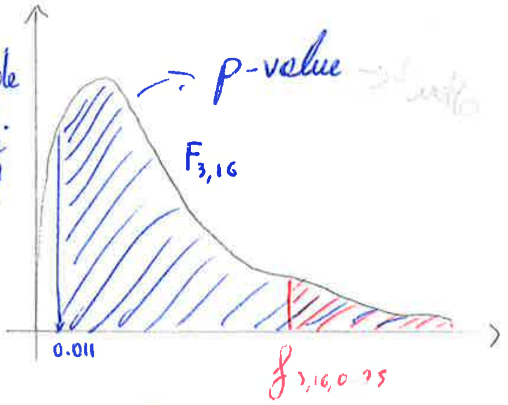
$$= SS_T - SS_M$$

Caso dei fertilizzanti:

	GdL	Somme quadrati	Quadrati medi	F_{calc}
Gruppi	3 (4 livelli)	$SS_B = 0.546$	$MS_B = 0.182$	$\frac{MS_B}{MS_W} = 0.115$
Errore	16	$SS_W = 253.698 - 0.546$ $= 253.152$	$MS_W = 15.822 = \hat{\sigma}^2$	
Totali	19	$SS_{TC} = 253.698$		

~~Siccome~~ p-value è molto grande

Siccome 0.011 è nella regione di accettazione, non posso rifiutare l'affermazione che le medie sono uguali.



Quindi, a rigor di logica, si prende quello che costa meno. Però guardando i dati ci accorgiamo che MS_B è grande, per cui potrebbe esserci anche un altro fattore, la varietà dei semi.

Il nostro obiettivo ora è capire quale parte è dovuta all'errore sperimentale.

Ovviamente c'è un modello statistico

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij}$$

Con $\sum \alpha_j = \sum \mu_j - \sum \mu$

oppure

$$Y_{ij} = \mu + \underbrace{(\mu_j - \mu)}_{\alpha_j \to \text{dipende dalla colonna}} + \varepsilon_{ij}$$

Quindi:

$H_0 : \mu_1 = \mu_2 = \dots = \mu_K$

Oppure

$H_0 : \alpha_j = 0$

$H_A : \alpha_j \neq 0$