


# PSI : Capitolo 5

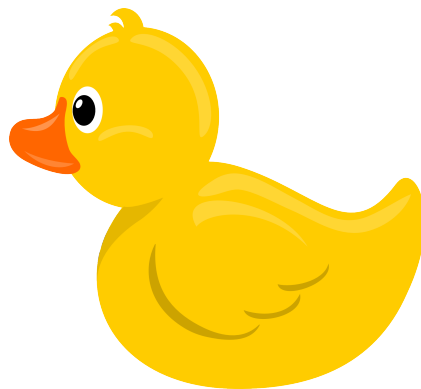
Ver 1.1

**Falbo Andrea**

A.A 2022/2023

Prof. Caravenna

 LilQuacky



# Indice

<b>1</b>	<b>Statistica Inferenziale</b>	<b>2</b>
1.1	Teoria . . . . .	2
1.1.1	Introduzione . . . . .	2
1.1.2	Stime Puntuali . . . . .	3
1.1.3	Distribuzione delle Statistiche Campionarie . . . . .	4
1.1.4	Stima per Intervalli . . . . .	5
1.2	Pratica . . . . .	13
1.2.1	Esercizi . . . . .	13

# Capitolo 1

## Statistica Inferenziale

### 1.1 Teoria

#### 1.1.1 Introduzione

**Definizione:** La *statistica inferenziale* consente di dedurre particolari caratteristiche di una popolazione limitandosi ad analizzare un numero finito e preferibilmente piccolo di suoi individui.

**Definizione:** Quando le caratteristiche che si vogliono individuare sono esprimibili numericamente allora esse sono dette *parametri*.

**Definizione:** Per *stima di parametri* si intende quindi il problema della deduzione di parametri di una popolazione facendo ricorso all'analisi di un suo sottoinsieme finito opportunamente scelto, detto *campione*.

**Osservazione:** Diverse tecniche possono essere utilizzate per effettuare delle stime di parametri. Noi ci limiteremo a considerare quelle classiche basate sulla conoscenza delle *distribuzioni campionarie*.

**Definizione:** Diverse ragioni possono portare a voler determinare le caratteristiche di una popolazione facendo ricorso esclusivamente ad un numero limitato di suoi individui: tempo, costo, disponibilità ecc. In questi casi occorre allora effettuare un *campionamento*, ovvero una scelta degli individui che verranno analizzati per effettuare le inferenze sull'intera popolazione.

**Osservazione:** Tutte le tecniche che verranno presentate in questo capitolo sono valide solo nel caso in cui il campione sia stato scelto secondo una pro-

cedura detta *campionamento casuale*.

**Definizione:** Denotiamo con  $X$  il *carattere* della popolazione su cui siamo interessati a fare dell'inferenza. Penseremo ad  $X$  come ad una *variabile aleatoria* la cui *distribuzione sconosciuta* corrisponde a quella che si otterrebbe facendo ricorso alle tecniche della statistica descrittiva sull'intera popolazione, e pensare invece ai valori assunti dai singoli individui come a delle *realizzazioni* di  $X$ . In forma matematica:

- Campione casuale di numerosità  $n$   $(X_1, X_2, \dots, X_n)$  : è una  $n$ -pla di v.a indipendenti aventi ognuna la stessa distribuzione del carattere  $X$  della popolazione.
- I valori  $(x_1, x_2, \dots, x_n)$  assunti dalla  $n$ -pla sono una realizzazione di  $(X_1, X_2, \dots, X_n)$ .

### 1.1.2 Stime Puntuali

**Definizione:** Possiamo pensare al carattere della popolazione su cui vogliamo fare delle inferenze come ad una variabile aleatoria  $X$ , avente una *funzione di ripartizione  $F$  sconosciuta*, ma corrispondente alla distribuzione di frequenza cumulata di tale carattere, che si potrebbe ottenere se fosse possibile analizzare per intero la popolazione.

**Definizione:** Una *stima* è una realizzazione di una statistica campionaria.

**Osservazione:** Per le prossime definizioni denoteremo con  $\mu$  il valore atteso e con  $\sigma^2$  la varianza della popolazione  $X$  con distribuzione  $F$  incognita.

**Definizione:** Uno stimatore si dice *non distorto* se il loro valore atteso è uguale al valore medio che vogliamo stimare:

$$E[T]_{\Theta} = E[g(X_1, \dots, X_n)]_{\Theta} = \Theta$$

Questa proprietà non è stabile a trasformazioni non lineari. Uno stimatore non distorto si dice *consistente* quando ha varianza che tende a 0 con  $N$  grande.

**Definizione:** Considerato un campione  $(X_1, X_2, \dots, X_n)$  estratto da una popolazione  $X$ , con distribuzione  $F$ , media  $\mu$  e varianza  $\sigma^2$  incognite. Definiamo *media campionaria* la variabile:

$$\overline{X_N} : \frac{X_1 + X_2 + \dots + X_n}{n}$$

Questo stimatore è non distorto in quanto  $E[\overline{X_N}] = \mu$ .

**Definizione:** Considerato un campione  $(X_1, X_2, \dots, X_n)$  estratto da una popolazione  $X$ , con distribuzione  $F$ , media  $\mu$  e varianza  $\sigma^2$  incognite. Definiamo *varianza campionaria* la variabile:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X_N})^2$$

Questo stimatore è non distorto in quanto  $E[S_n^2] = \sigma^2$ .

**Definizione:** Considerato un campione  $(X_1, X_2, \dots, X_n)$  estratto da una popolazione  $X$ , con distribuzione  $F$ , media  $\mu = E[X_i]$  nota e varianza  $\sigma^2$  incognita. Definiamo *varianza campionaria* la variabile:

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

Questo stimatore è non distorto in quanto  $E[S_n^2] = \sigma^2$ .

### 1.1.3 Distribuzione delle Statistiche Campionarie

**Definizione:** Prendiamo una v.a. i.i.d  $Z \sim N(\mu, \sigma^2)$  e  $\alpha \in (0, 1)$ . Si definisce  $z_\alpha$  quel valore tale che

$$\mathbb{P}(Z > z_\alpha) = \alpha$$

**Osservazione:** Vale anche  $z_\alpha = -z_{1-\alpha}$

**Definizione:** Siano  $Z_1, \dots, Z_n \sim N(0, 1)$ . Allora introduciamo  $Y$  come una distribuzione *chi quadrato con  $n$  gradi di libertà* tale che

$$Y = \sum_{i=1}^n Z_i^2 \sim \chi^2(n)$$

**Definizione:** Per  $\alpha$  si pone  $x_{n,\alpha}^2$  quel valore tale che:

$$\mathbb{P}(Y > x_{n,\alpha}^2) = \alpha$$

**Osservazioni:**

- si ha  $E[Y] = n, Var[Y] = 2n$

- per  $n = 2$  è la legge di  $\exp(1/2)$
- per  $n$  grande vale l'approssimazione della legge con una  $N(n, 2n)$

**Definizione:** Sia  $\overline{X}_N$  un campione casuale estratto da una popolazione  $N(\mu, \sigma^2)$ :

$$1. \sum_{i=1}^n \left( \frac{x_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

$$2. \sum_{i=1}^n \left( \frac{x_i - \overline{X}_N}{\sigma} \right)^2 \sim \chi^2(n-1)$$

$$3. \text{ se } S_n^2 = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \overline{X}_N}{\sigma} \right)^2 \text{ allora } (n-1) \frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

**Osservazione:** Osservando i punti 1) e 2), possiamo notare che ogni volta che stimiamo un parametro con chi quadrato, perdiamo un grado di libertà

**Definizione:** Siano  $Z \sim N(0, 1)$ ,  $Y \sim \chi^2(n)$  indipendenti, definiamo  $T$  come una distribuzione *t di Student con n gradi di libertà* come:

$$T = \frac{Z}{\sqrt{Y/n}} \quad T \sim t(n)$$

**Definizione:** Per  $\alpha$  si pone  $t_{n,\alpha}$  quel valore tale che:

$$\mathbb{P}(T > t_{n,\alpha}) = \alpha$$

**Osservazione:**  $T$  è simmetrica rispetto a 0. Quindi  $t_{\alpha,n} = t_{1-\alpha,n}$

### 1.1.4 Stima per Intervalli

Abbiamo visto come trovare un valore approssimato di un parametro incognito della popolazione per mezzo di una stima puntuale. Tali stime però non forniscono informazioni sul grado di approssimazione delle stesse. Per questo motivo alle stime puntuali vengono preferite quando possibile determinarle le *stime per intervalli* che sono stime espresse sotto forma di *intervalli fiduciari* all'interno dei quali con buona probabilità si trova il valore vero del parametro da stimare.

**Definizione:** Definiamo  $\alpha \in [0, 1]$  come *livello di confidenza* della stima ed il corrispondente intervallo è detto *intervallo di confidenza*. Spesso  $\alpha$  assume

come valori 0.1, 0.05 e 0.01 .

**Definizione:** La *stima intervallare della media* di un campione estratto da una popolazione normale con *media incognita e varianza nota pari a  $\sigma^2$*  ha come intervallo di confidenza:

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

**Osservazione:** L'ampiezza dell'intervallo è due volte l'errore, ovvero lo scarto dal valore centrato. Nell'esempio di stima della media con media incognita e varianza nota, l'ampiezza è:

$$2z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

**Osservazione:** La bontà della stima dipende dal livello di confidenza: maggiore è, più affidabile è la stima; ma all'aumentar di quest'ultimo, aumenta l'ampiezza dell'intervallo e quindi meno precisa sarà la stima.

**Definizione:** La *stima intervallare della media* di un campione estratto da una popolazione normale con *media e varianza incognita* utilizza la t di Student con n-1 gradi di libertà. Ha come intervallo di confidenza:

$$IC = \left( \bar{x}_n - t_{n-1, \frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1, \frac{\alpha}{2}} \frac{s_n}{\sqrt{n}} \right)$$

**Definizione:** La stima proporzione-frequenza di una popolazione *Bernoulliana* con media e varianza incognite, valida se  $n\bar{x}_n > 5$  e  $n(1 - \bar{x}_n) > 5$  ha come intervallo di confidenza:

$$IC = \left( \bar{x}_n - z_{\alpha/2} \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}}, \bar{x}_n + z_{\alpha/2} \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}} \right) \quad \bar{x}_n(1 - \bar{x}_n) \leq \frac{1}{4}$$

**Definizione:** Stima intervallare della *varianza* su un campione estratto da una popolazione normale con media e varianza incognite è:

$$IC = \left( \frac{(n-1)s_n^2}{\chi_{n-1, \alpha/2}}, \frac{(n-1)s_n^2}{\chi_{n-1, 1-\alpha/2}} \right)$$

**Definizione:** Stima intervallare della *varianza* su un campione estratto da una popolazione normale con media nota e varianza incognita utilizza una  $\chi$  con n-1 gradi di libertà ed è:

$$IC = \left( \frac{n\bar{s}_n^2}{\chi_{n, \alpha/2}}, \frac{n\bar{s}_n^2}{\chi_{n, 1-\alpha/2}} \right)$$

## SECONDA PARTE

### TEOREMA DEL LIMITE CENTRALE

$X_1, \dots, X_n, \dots$  v.a. i.i.d. con media  $\mu$  e varianza  $\sigma^2$  e sia  $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$

$$P\left(\sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} \leq t\right) \rightarrow \Phi(t) \text{ se } n \rightarrow \infty$$

dove  $\Phi(t) = P(Z \leq t)$ ,  $Z \sim \mathcal{N}(0, 1)$

### STIMA PUNTUALE

- $X_1, \dots, X_n$  campione casuale estratto da una popolazione con media incognita.

Stimatore non distorto della media

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

- $X_1, \dots, X_n$  campione casuale estratto da una popolazione con media e varianza incognite .

Stimatore non distorto della varianza

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

- $X_1, \dots, X_n$  campione casuale estratto da una popolazione con media nota pari a  $\mu$  e varianza incognita .

Stimatore non distorto della varianza

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

### DISTRIBUZIONI UTILI PER LE STATISTICHE CAMPIONARIE

- $Z \sim \mathcal{N}(0, 1)$  e  $\alpha \in (0, 1)$ , si pone  $z_\alpha \in \mathbb{R}$  quel valore tale che  $\mathbb{P}(Z > z_\alpha) = \alpha$ . N.B:  $z_\alpha = -z_{1-\alpha}$ .

- $Z_1, \dots, Z_n$  i.i.d. normali standard

$$Y = Z_1^2 + \dots + Z_n^2, \quad Y \sim \chi^2(n)$$

$Y$  ha una distribuzione chi quadrato con  $n$  gradi di libertà:  $Y \geq 0$

Per  $\alpha \in (0, 1)$  si pone  $\chi_{n,\alpha}^2 \in \mathbb{R}$  quel valore tale che  $\mathbb{P}(Y > \chi_{n,\alpha}^2) = \alpha$ .

$$\mathbb{E}[Y] = n, \quad \text{var}[Y] = 2n$$

- Siano  $Z \sim \mathcal{N}(0, 1)$ ,  $Y \sim \chi^2(n)$  indipendenti

$$T = \frac{Z}{\sqrt{Y/n}}, \quad T \sim t(n)$$

$T$  ha una distribuzione  $t$  di Student con  $n$  gradi di libertà.  $T$  simmetrica rispetto a 0.

Per  $\alpha \in (0, 1)$  si pone  $t_{n,\alpha} \in \mathbb{R}$  quel valore tale che  $\mathbb{P}(T > t_{n,\alpha}) = \alpha$ . N.B:  $t_{n,\alpha} = -t_{n,1-\alpha}$



# STIMA PER INTERVALLI

Daremo formule per intervalli di confidenza, (estremi inferiori o superiori) al livello di  $100(1 - \alpha)\%$ , e daremo la realizzazione dell'intervallo sui dati campionari  $x_1, \dots, x_n$ .

campione numeroso  $\rightsquigarrow n \geq 30$

- campione estratto da una popolazione normale con media incognita e varianza nota pari a  $\sigma^2$  (vale anche per campioni numerosi non necessariamente normali): stima intervallare della media

$$\text{Intervallo di confidenza } \left( \bar{x}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

$$\text{Estremo inferiore } \bar{x}_n - z_{\alpha} \frac{\sigma}{\sqrt{n}}, \quad \text{intervallo destro } \left( \bar{x}_n - z_{\alpha} \frac{\sigma}{\sqrt{n}}, +\infty \right)$$

$$\text{Estremo superiore } \bar{x}_n + z_{\alpha} \frac{\sigma}{\sqrt{n}}, \quad \text{intervallo sinistro } \left( -\infty, \bar{x}_n + z_{\alpha} \frac{\sigma}{\sqrt{n}} \right)$$

- campione estratto da una popolazione normale con media e varianza incognite (vale anche per campioni numerosi non necessariamente normali): stima intervallare della media

$$\text{Intervallo di confidenza } \left( \bar{x}_n - t_{n-1, \alpha/2} \frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1, \alpha/2} \frac{s_n}{\sqrt{n}} \right)$$

$$\text{Estremo inferiore } \bar{x}_n - t_{n-1, \alpha} \frac{s_n}{\sqrt{n}}, \quad \text{intervallo destro } \left( \bar{x}_n - t_{n-1, \alpha} \frac{s_n}{\sqrt{n}}, +\infty \right)$$

$$\text{Estremo superiore } \bar{x}_n + t_{n-1, \alpha} \frac{s_n}{\sqrt{n}}, \quad \text{intervallo sinistro } \left( -\infty, \bar{x}_n + t_{n-1, \alpha} \frac{s_n}{\sqrt{n}} \right)$$

- campione numeroso estratto da una popolazione Bernoulliana con media e varianza incognite (vale anche per campioni numerosi non necessariamente normali): stima intervallare della proporzione-frequenza: ok se  $n\bar{x}_n > 5$ ,  $n(1 - \bar{x}_n) > 5$ .

$$\text{Intervallo di confidenza } \left( \bar{x}_n - z_{\alpha/2} \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}}, \bar{x}_n + z_{\alpha/2} \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}} \right) \text{ N.B.: } \bar{x}_n(1 - \bar{x}_n) \leq \frac{1}{4}.$$

$$\text{Estremo inferiore } \bar{x}_n - z_{\alpha} \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}}, \quad \text{intervallo destro } \left( \bar{x}_n - z_{\alpha} \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}}, +\infty \right)$$

$$\text{Estremo superiore } \bar{x}_n + z_{\alpha} \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}}, \quad \text{intervallo sinistro } \left( -\infty, \bar{x}_n + z_{\alpha} \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}} \right)$$

- campione estratto da una popolazione normale con media e varianza incognite: stima intervallare della varianza

$$\text{Intervallo di confidenza } \left( \frac{(n-1)s_n^2}{\chi_{n-1, \alpha/2}^2}, \frac{(n-1)s_n^2}{\chi_{n-1, 1-\alpha/2}^2} \right)$$

$$\text{Estremo inferiore } \frac{(n-1)s_n^2}{\chi_{n-1, \alpha}^2}, \quad \text{intervallo destro } \left( \frac{(n-1)s_n^2}{\chi_{n-1, \alpha}^2}, +\infty \right)$$

$$\text{Estremo superiore } \frac{(n-1)s_n^2}{\chi_{n-1, 1-\alpha}^2}, \quad \text{intervallo sinistro } \left[ 0, \frac{(n-1)s_n^2}{\chi_{n-1, 1-\alpha}^2} \right)$$

- campione estratto da una popolazione normale con media nota e varianza incognita: stima intervallare della varianza

$$\text{Intervallo di confidenza} \left( \frac{n\bar{s}_n^2}{\chi_{n,\alpha/2}^2}, \frac{n\bar{s}_n^2}{\chi_{n,1-\alpha/2}^2} \right)$$

$$\text{Estremo inferiore } \frac{n\bar{s}_n^2}{\chi_{n,\alpha}^2}, \quad \text{intervallo destro } \left( \frac{n\bar{s}_n^2}{\chi_{n,\alpha}^2}, +\infty \right)$$

$$\text{Estremo superiore } \frac{n\bar{s}_n^2}{\chi_{n,1-\alpha}^2}, \quad \text{intervallo sinistro } \left[ 0, \frac{n\bar{s}_n^2}{\chi_{n,1-\alpha}^2} \right)$$

### TEST DI IPOTESI

$\alpha$  = livello di significatività

- Test  $z$  sulla media di una popolazione normale con varianza nota pari a  $\sigma^2$  ( vale anche per campioni numerosi estratti da popolazioni non necessariamente normali )

$H_0$	$H_1$	Statistica	Regione critica
$\mu = \mu_0$	$\mu \neq \mu_0$	$Z = \frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n}$	$\left  \frac{\bar{x}_n - \mu_0}{\sigma} \sqrt{n} \right  > z_{\alpha/2}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$Z = \frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n}$	$\frac{\bar{x}_n - \mu_0}{\sigma} \sqrt{n} > z_{\alpha}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$Z = \frac{\bar{X}_n - \mu_0}{\sigma} \sqrt{n}$	$\frac{\bar{x}_n - \mu_0}{\sigma} \sqrt{n} < -z_{\alpha}$

- Test  $t$  sulla media di una popolazione normale con varianza incognita ( vale anche per campioni numerosi estratti da popolazioni non necessariamente normali )

$H_0$	$H_1$	Statistica	Regione critica
$\mu = \mu_0$	$\mu \neq \mu_0$	$T = \frac{\bar{X}_n - \mu_0}{S_n} \sqrt{n}$	$\left  \frac{\bar{x}_n - \mu_0}{s_n} \sqrt{n} \right  > t_{n-1, \alpha/2}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$T = \frac{\bar{X}_n - \mu_0}{S_n} \sqrt{n}$	$\frac{\bar{x}_n - \mu_0}{s_n} \sqrt{n} > t_{n-1, \alpha}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$T = \frac{\bar{X}_n - \mu_0}{S_n} \sqrt{n}$	$\frac{\bar{x}_n - \mu_0}{s_n} \sqrt{n} < -t_{n-1, \alpha}$

- Test  $z$  approssimato sulla proporzione con  $n \geq 30$ ,  $np_0 \geq 5$ ,  $n(1 - p_0) \geq 5$ .

$H_0$	$H_1$	Statistica	Regione critica
$p = p_0$	$p \neq p_0$	$Z = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}$	$\left  \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} \right  > z_{\alpha/2}$
$p \leq p_0$	$p > p_0$	$Z = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}$	$\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} > z_{\alpha}$
$p \geq p_0$	$p < p_0$	$Z = \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n}$	$\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} \sqrt{n} < -z_{\alpha}$

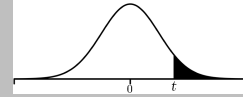
## TAVOLA DELLA DISTRIBUZIONE NORMALE

La tabella seguente riporta i valori di  $\Phi(z) := \int_{-\infty}^z \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx$ , la funzione di ripartizione della distribuzione normale standard  $N(0, 1)$ , per  $0 \leq z \leq 3.5$ .

I valori di  $\Phi(z)$  per  $z < 0$  possono essere ricavati grazie alla formula

$$\Phi(z) = 1 - \Phi(-z).$$

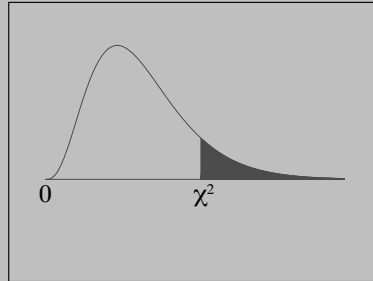
[illegible]

Critical Values for Student's  $t$ -Distribution.

df	Upper Tail Probability: $\Pr(T > t)$									
	0.2	0.1	0.05	0.04	0.03	0.025	0.02	0.01	0.005	0.0005
1	1.376	3.078	6.314	7.916	10.579	12.706	15.895	31.821	63.657	636.619
2	1.061	1.886	2.920	3.320	3.896	4.303	4.849	6.965	9.925	31.599
3	0.978	1.638	2.353	2.605	2.951	3.182	3.482	4.541	5.841	12.924
4	0.941	1.533	2.132	2.333	2.601	2.776	2.999	3.747	4.604	8.610
5	0.920	1.476	2.015	2.191	2.422	2.571	2.757	3.365	4.032	6.869
6	0.906	1.440	1.943	2.104	2.313	2.447	2.612	3.143	3.707	5.959
7	0.896	1.415	1.895	2.046	2.241	2.365	2.517	2.998	3.499	5.408
8	0.889	1.397	1.860	2.004	2.189	2.306	2.449	2.896	3.355	5.041
9	0.883	1.383	1.833	1.973	2.150	2.262	2.398	2.821	3.250	4.781
10	0.879	1.372	1.812	1.948	2.120	2.228	2.359	2.764	3.169	4.587
11	0.876	1.363	1.796	1.928	2.096	2.201	2.328	2.718	3.106	4.437
12	0.873	1.356	1.782	1.912	2.076	2.179	2.303	2.681	3.055	4.318
13	0.870	1.350	1.771	1.899	2.060	2.160	2.282	2.650	3.012	4.221
14	0.868	1.345	1.761	1.887	2.046	2.145	2.264	2.624	2.977	4.140
15	0.866	1.341	1.753	1.878	2.034	2.131	2.249	2.602	2.947	4.073
16	0.865	1.337	1.746	1.869	2.024	2.120	2.235	2.583	2.921	4.015
17	0.863	1.333	1.740	1.862	2.015	2.110	2.224	2.567	2.898	3.965
18	0.862	1.330	1.734	1.855	2.007	2.101	2.214	2.552	2.878	3.922
19	0.861	1.328	1.729	1.850	2.000	2.093	2.205	2.539	2.861	3.883
20	0.860	1.325	1.725	1.844	1.994	2.086	2.197	2.528	2.845	3.850
21	0.859	1.323	1.721	1.840	1.988	2.080	2.189	2.518	2.831	3.819
22	0.858	1.321	1.717	1.835	1.983	2.074	2.183	2.508	2.819	3.792
23	0.858	1.319	1.714	1.832	1.978	2.069	2.177	2.500	2.807	3.768
24	0.857	1.318	1.711	1.828	1.974	2.064	2.172	2.492	2.797	3.745
25	0.856	1.316	1.708	1.825	1.970	2.060	2.167	2.485	2.787	3.725
26	0.856	1.315	1.706	1.822	1.967	2.056	2.162	2.479	2.779	3.707
27	0.855	1.314	1.703	1.819	1.963	2.052	2.158	2.473	2.771	3.690
28	0.855	1.313	1.701	1.817	1.960	2.048	2.154	2.467	2.763	3.674
29	0.854	1.311	1.699	1.814	1.957	2.045	2.150	2.462	2.756	3.659
30	0.854	1.310	1.697	1.812	1.955	2.042	2.147	2.457	2.750	3.646
31	0.853	1.309	1.696	1.810	1.952	2.040	2.144	2.453	2.744	3.633
32	0.853	1.309	1.694	1.808	1.950	2.037	2.141	2.449	2.738	3.622
33	0.853	1.308	1.692	1.806	1.948	2.035	2.138	2.445	2.733	3.611
34	0.852	1.307	1.691	1.805	1.946	2.032	2.136	2.441	2.728	3.601
35	0.852	1.306	1.690	1.803	1.944	2.030	2.133	2.438	2.724	3.591
36	0.852	1.306	1.688	1.802	1.942	2.028	2.131	2.434	2.719	3.582
37	0.851	1.305	1.687	1.800	1.940	2.026	2.129	2.431	2.715	3.574
38	0.851	1.304	1.686	1.799	1.939	2.024	2.127	2.429	2.712	3.566
39	0.851	1.304	1.685	1.798	1.937	2.023	2.125	2.426	2.708	3.558
40	0.851	1.303	1.684	1.796	1.936	2.021	2.123	2.423	2.704	3.551
41	0.850	1.303	1.683	1.795	1.934	2.020	2.121	2.421	2.701	3.544
42	0.850	1.302	1.682	1.794	1.933	2.018	2.120	2.418	2.698	3.538
43	0.850	1.302	1.681	1.793	1.932	2.017	2.118	2.416	2.695	3.532
44	0.850	1.301	1.680	1.792	1.931	2.015	2.116	2.414	2.692	3.526
45	0.850	1.301	1.679	1.791	1.929	2.014	2.115	2.412	2.690	3.520
46	0.850	1.300	1.679	1.790	1.928	2.013	2.114	2.410	2.687	3.515
47	0.849	1.300	1.678	1.789	1.927	2.012	2.112	2.408	2.685	3.510
48	0.849	1.299	1.677	1.789	1.926	2.011	2.111	2.407	2.682	3.505
49	0.849	1.299	1.677	1.788	1.925	2.010	2.110	2.405	2.680	3.500
50	0.849	1.299	1.676	1.787	1.924	2.009	2.109	2.403	2.678	3.496
60	0.848	1.296	1.671	1.781	1.917	2.000	2.099	2.390	2.660	3.460
70	0.847	1.294	1.667	1.776	1.912	1.994	2.093	2.381	2.648	3.435
80	0.846	1.292	1.664	1.773	1.908	1.990	2.088	2.374	2.639	3.416
90	0.846	1.291	1.662	1.771	1.905	1.987	2.084	2.368	2.632	3.402
100	0.845	1.290	1.660	1.769	1.902	1.984	2.081	2.364	2.626	3.390
120	0.845	1.289	1.658	1.766	1.899	1.980	2.076	2.358	2.617	3.373
140	0.844	1.288	1.656	1.763	1.896	1.977	2.073	2.353	2.611	3.361
180	0.844	1.286	1.653	1.761	1.893	1.973	2.069	2.347	2.603	3.345
200	0.843	1.286	1.653	1.760	1.892	1.972	2.067	2.345	2.601	3.340
500	0.842	1.283	1.648	1.754	1.885	1.965	2.059	2.334	2.586	3.310
1000	0.842	1.282	1.646	1.752	1.883	1.962	2.056	2.330	2.581	3.300
$\infty$	0.842	1.282	1.645	1.751	1.881	1.960	2.054	2.326	2.576	3.291
	60%	80%	90%	92%	94%	95%	96%	98%	99%	99.9%
	Confidence Level									

Note:  $t(\infty)_{\alpha/2} = Z_{\alpha/2}$  in our notation.

# Chi-Square Distribution Table



The shaded area is equal to  $\alpha$  for  $\chi^2 = \chi^2_{\alpha}$ .

$df$	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

## 1.2 Pratica

### 1.2.1 Esercizi

#### Esercizio 1: Intervalli di Confidenza, Stime Media

**Traccia:** La concentrazione di PCB nel latte materno ha approssimativamente una distribuzione Normale con media  $\mu$  e varianza  $\sigma^2$  entrambe incognite. Si misura un campione di 20 individui, ottenendo  $\bar{x}_n = 5.8$  e  $s_n = 5.085$

1. IC per  $\mu$  a livello 95%
2. IC per  $\mu$  a livello 99%

#### Soluzione punto 1:

1. *Trovo la formula da usare:* sono nel caso di media e varianza incognite e voglio trovare l'intervallo di  $\mu$ , controllo nel formulario delle stime per intervalli e trovo:

$$IC = \left( \bar{X}_N - t_{n-1, \frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}, \bar{X}_N + t_{n-1, \frac{\alpha}{2}} \frac{s_n}{\sqrt{n}} \right)$$

2. *Calcolo  $\alpha$ :* Abbiamo  $100(1 - \alpha)\% = 95\%$ . Trovo  $\alpha = 0.05$
3. *Riscrivo i miei dati:* Ho  $n = 20$ ,  $\bar{x}_n = 5.8$ ,  $s_n = 5.085$ ,  $\alpha = 0.05$ .  
Riguardando la formula mi manca conoscere  $\frac{\alpha}{2} = 0.025$
4. *Uso la tavola di  $t$  di Student:* Incrocio  $n - 1 = 19$  e  $\frac{\alpha}{2} = 0.025$  e trovo 2.093
5. *Riscrivo la formula:*

$$IC = \left( 5.8 - 2.093 \frac{5.085}{\sqrt{20}}, 5.8 + 2.093 \frac{5.085}{\sqrt{20}} \right) \simeq (3.12, 8.18)$$

6. *Conclusione:* L'intervallo di confidenza per  $\mu$  a livello 95% è (3.12, 8.18)

#### Soluzione punto 2:

1. *Trovo la formula da usare:* Sono nello stesso caso di prima in quanto cambia solo il livello di confidenza

$$IC = \left( \bar{X}_N - t_{n-1, \frac{\alpha}{2}} \frac{s_n}{\sqrt{n}}, \bar{X}_N + t_{n-1, \frac{\alpha}{2}} \frac{s_n}{\sqrt{n}} \right)$$

2. *Calcolo  $\alpha$* : A differenza del precedente punto, adesso ho un livello di 99% quindi  $100(1 - \alpha)\% = 99\%$ . Trovo  $\alpha = 0.01$
3. *Riscrivo i miei dati*: Ho  $n = 20$ ,  $\bar{x}_n = 5.8$ ,  $s_n = 5.085$ ,  $\alpha = 0.01$ . Riguardando la formula mi manca conoscere  $\frac{\alpha}{2} = 0.005$
4. *Uso la tavola di  $t$  di Student*: Incrocio  $n - 1 = 19$  e  $\frac{\alpha}{2} = 0.005$  e trovo 2.861
5. *Riscrivo la formula*:

$$IC = \left( 5.8 - 2.861 \frac{5.085}{\sqrt{20}}, 5.8 + 2.861 \frac{5.085}{\sqrt{20}} \right) \simeq (2.55, 9.05)$$

6. *Conclusione*: L'intervallo di confidenza per  $\mu$  a livello 95% è (2.55, 9.05)

## Esercizio 2: Intervalli di Confidenza, Stime Proporzioni

**Traccia:** Voglio stimare la proporzione di donne tra gli insegnanti della scuola secondaria. Su un campione di 1000 insegnanti ci sono 518 donne.

1. Stima puntuale della popolazione tramite uno stimatore non distorto
2. IC al 95% della proporzione
3. IC al 99% la cui ampiezza non sia maggiore di 0.03. Quanto dovrebbe essere numeroso il campione?

**Soluzione Punto 1:** Siamo nel caso campione numero estratto da una popolazione Bernoulliana  $Be(p)$  con  $p$  incognito e quindi media e varianza incognite. Uno stimatore non distorto di una  $Be(p)$  è la media  $p$ . Dunque la stima puntuale richiesta è  $\bar{x}_n = 518/1000 = 0.518$

### Soluzione Punto 2:

1. *Trovo la formula da usare*: Dal formulario,

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

2. *Trovo  $\alpha$* :  $z_{\alpha/2}$  è il  $100(1 - \alpha/2)$ esimo percentile quindi  $\alpha = 0.05$  e  $\alpha/2 = 0.025$

3. *Tavola Gaussiana*: A differenza della t Student, non troviamo la coda, ma la fdr. Quindi dobbiamo trovare  $P(Z \leq z_{0.025}) = 1 - 0.025 = 0.975$  e quindi trovo che  $z_{0.025} = 1.96$

4. *Riscrivo la formula*:

$$IC = \left( 0.518 - 1.96 \sqrt{\frac{0.518(1 - 0.518)}{1000}}, 0.518 + 1.96 \sqrt{\frac{0.518(1 - 0.518)}{1000}} \right) \\ \simeq (0.487, 0.549)$$

5. *Conclusione*: L'intervallo di confidenza per la proporzione a livello 95% è (0.487, 0.549)

### Soluzione Punto 3:

1. *Formula Ampiezza*: L'ampiezza di un IC per definizione è 2 volte lo scarto:

$$2z_{\alpha/2} \sqrt{\frac{\bar{x}_n(1 - \bar{x}_n)}{n}}$$

2. *Ricavare i dati*: Non conosciamo  $\bar{x}_n(1 - \bar{x}_n)$ , ma è sicuramente  $\leq 1/4$  (vedi da formulario). La formula diventa

$$2z_{\alpha/2} \sqrt{\frac{1/4}{n}} = z_{\alpha/2} \frac{1}{\sqrt{n}}$$

3. *Tavola Gaussiana*: Dato  $\alpha/2 = 0.005$  devo trovare nella tavola 0.995 e lo incastro tra i valori (2.57, 2.58) quindi  $z_{0.005} = 2.575$

4. *Riscrivo la formula*:

$$2.575 \frac{1}{\sqrt{n}} \leq 0.03 \rightarrow \sqrt{n} \geq \frac{2.575}{0.03} \rightarrow n \geq 7373.08$$

5. *Conclusione*: Per avere IC al 99% con ampiezza non maggiore di 0.03 ho bisogno di 7374 professori

### Esercizio 3: Intervalli di Confidenza, Stime Varianza

**Traccia**: Si considera il campione

1.752.251.92.32.11.7

proveniente da una legge normale con media e varianza incognite.



1. Stima puntuale della varianza usando stimatore non distorto
2. IC al 99% per  $\sigma^2$
3. Come cambiano le risposte se  $\mu$  è nota pari a 2

**Soluzione Punto 1:** Stimatore non distorto di  $\sigma^2$  con media e varianza incognite è

$$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

Conosco  $n=6$ ,  $\bar{x}_n = 2$  e quindi la stima puntuale di  $\sigma^2$  è:

$$s_n^2 = \frac{1}{5} [(1.75 - 2)^2 + \dots + (1.7 - 2)^2] = 0.065$$

**Soluzione Punto 2:**

1. *Trovo la formula da usare:* Stima intervallare della varianza con media e varianza incognita utilizza una  $\chi$  con  $n-1$  gradi di libertà ed è:

$$IC = \left( \frac{(n-1)s_n^2}{\chi_{n-1, \alpha/2}}, \frac{(n-1)s_n^2}{\chi_{n-1, 1-\alpha/2}} \right)$$

2. *Calcoliamo e Usiamo la Tavola del chi quadro:* Abbiamo  $\alpha = 0.01$  e quindi cerchiamo  $\chi_{5,0.005}^2$  e  $\chi_{5,0.995}^2$  e trovo rispettivamente 16.75 e 0.412
3. *Riscrivo la formula:*

$$IC = \left( \frac{5 \cdot 0.065}{16.75}, \frac{5 \cdot 0.065}{0.412} \right) \simeq (0.019, 0.79)$$

**Soluzione Punto 3:**

1. *Trovo lo stimatore non distorto:* Stimatore di  $\sigma^2$  incognito conoscendo  $\mu = 2$  è  $\bar{s}_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = 0.054$
2. *Trovo la formula da usare:* Stima intervallare varianza con media nota e varianza incognita è:

$$IC = \left( \frac{n\bar{s}_n^2}{\chi_{n, \alpha/2}}, \frac{n\bar{s}_n^2}{\chi_{n, 1-\alpha/2}} \right)$$

3. *Calcoliamo e Usiamo la Tavola del chi quadro:*  $\alpha = 0.01$  e cerchiamo  $\chi_{6,0.005}$  e  $\chi_{6,0.995}$  e trovo rispettivamente 18.548 e 0.676
4. *Riscrivo la formula:*

$$IC = \left( \frac{6 \cdot 0.054}{18.548}, \frac{6 \cdot 0.054}{0.676} \right) \simeq (0.017, 0.479)$$