

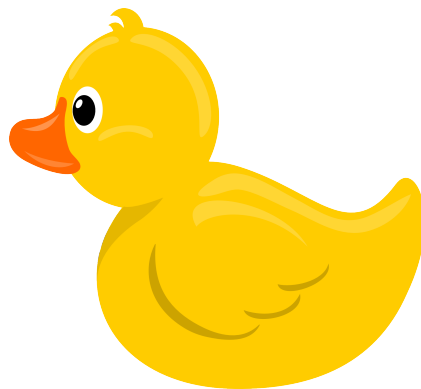
# PSI : Capitolo 6

Ver 1.0

Falbo Andrea

A.A 2022/2023

 LilQuacky



# Indice

<b>1</b>	<b>Verifica di Ipotesi</b>	<b>2</b>
1.1	Teoria . . . . .	2
1.1.1	Test di Ipotesi . . . . .	2
1.1.2	Considerazioni sugli errori . . . . .	4
1.1.3	Test sulla Media di una Popolazione . . . . .	4
1.1.4	Test sulla Differenza delle Medie di due Popolazioni . .	6
1.1.5	Test Non Parametrici . . . . .	6
1.1.6	Esercizi . . . . .	13

# Capitolo 1

## Verifica di Ipotesi

### 1.1 Teoria

#### 1.1.1 Test di Ipotesi

**Definizione:** Un'affermazione relativa ad una caratteristica di una popolazione è detta *ipotesi statistica* quando essa viene formulata sulla base dell'esperienza o sulla base di considerazioni teoriche.

**Osservazione:** Il problema di un ipotesi è la *verifica* della validità di un'ipotesi statistica. Per effettuare tali verifiche si utilizzano procedure statistiche dette *test di ipotesi* e si dividono in:

- *test parametrici*: si riferiscono ad ipotesi relative a parametri di distribuzione della popolazione (media e varianza)
- *test non parametrici*: si riferiscono al tipo di distribuzione ipotizzabile per la popolazione non esprimibili come parametri

**Definizioni:** Ogni test di ipotesi è caratterizzato da:

- una *popolazione statistica*  $X$  sulla quale viene effettuata il test
- un'*ipotesi nulla*  $H_0$  da convalidare o rifiutare sulla base dei valori assunti da un campione  $(X_i, \dots, X_n)$  estratto da  $X$
- un'*ipotesi alternativa*  $H_1$  da considerare valida quando si rifiuta  $H_0$
- una *statistica campionaria*  $T = T(X_i, \dots, X_n)$  di cui è nota la distribuzione quando  $H_0$  è vera

- una *regione di accettazione*  $\overline{C}$  che è l'insieme di valori assumibili dalla statistica  $T$  che portano ad un'accettazione dell'ipotesi  $H_0$
- una *regione critica*  $C$  che è l'insieme di valori assumibili dalla statistica  $T$  che portano ad un rifiuto dell'ipotesi  $H_0$
- un *livello di significatività*  $\alpha$  che permette di individuare la regione di accettazione, tale che se  $H_0$  vero allora  $T$  assume valori nella regione critica con probabilità  $\alpha$

**Definizione:** Se  $(X_i, \dots, X_n) \in C$ , ossia si rifiuta  $H_0$ , diremo che i dati sperimentali sono in **contraddizione significativa** con  $H_0$ . Altrimenti i dati non sono in contraddizione significativa con  $H_0$

**Osservazione:** La regione si chiama critica perché è improbabile che  $H_0$  sia vera quando la statistica campionaria assume valori appartenenti ad essa ma non possiamo comunque escluderlo.

**Definizione:** Nei test di ipotesi si possono commettere due tipi di errori:

1. *Errore di prima specie:* si rifiuta  $H_0$  quando è vera. Coincide con il livello di significatività  $\alpha$
2. *Errore di seconda specie:* si accetta  $H_0$  quando è falsa. In genere non è nota la probabilità  $\beta$

**Definizione** La procedura per la formulazione di un test di ipotesi solitamente prevede nell'ordine:

1. Individuazione dell'ipotesi nulla  $H_0$  e dell'ipotesi alternativa  $H_1$
2. La scelta del livello di significatività  $\alpha$
3. La scelta della statistica campionaria  $T$
4. La determinazione delle regioni di accettazione  $\overline{C}$  e critica  $C$
5. L'accettazione o il rifiuto dell'ipotesi nulla  $H_0$

**Definizione:** Una statistica è detta *semplice* se il sottoinsieme di valori che essa assegna ad un parametro è costituito da un solo elemento, altrimenti è detta *composta*. Considerando  $\mu$  il parametro incognito, un esempio di statistica semplice è  $H_0 : \mu = x$  mentre le composte sono nella forma  $H_0 : \mu \in (x, y)$   $H_0 : \mu > x$

**Definizione:** L'appartenenza di  $X_i, \dots, X_n \in \overline{C}$  dipende dal livello di significatività  $\alpha$ . Esiste un valore  $\bar{\alpha}$  detto *p-value* del test t.c. :

- per  $\alpha > \bar{\alpha}$  si rifiuta  $H_0$
- per  $\alpha \leq \bar{\alpha}$  si accetta  $H_0$

**Osservazione:** Più piccolo è il p-value, più i dati sono in contraddizione con  $H_0$

**Definizione:** Un test statistico è detto *bidirezionale* se la regione critica è costituita dall'unione di due sottoinsieme disgiunti mentre diremo che è *unidirezionale* se è costituita da un solo sottoinsieme. Negli esempi del punto precedente, i primi due sono bidirezionali, mentre il terzo è unidirezionale.

### 1.1.2 Considerazioni sugli errori

Nel descrivere le caratteristiche di un test di ipotesi la probabilità di compiere errori di II specie va pensata come una funzione anziché come uno specifico valore numerico

**Definizione:** Sia  $\Theta$  il parametro a cui si riferisce il test e sia  $\Theta^*$  il valore specificato dall'ipotesi nulla. Denotiamo l'errore di II specie come

$$\beta(\hat{\Theta}) = \mathbb{P}(\text{accettare } H_0 | H_0 \text{ è falsa e } \Theta = \Theta^*)$$

allora viene detta *curva di potenza del test* la funzione

$$\pi(\hat{\Theta}) = 1 - \beta(\hat{\Theta})$$

Un test risulta tanto migliore quanto più la funzione  $\pi(\hat{\Theta})$  si avvicina ad 1 al variare di  $\hat{\Theta}$

**Osservazione:** Se nella costruzione di un test si desidera diminuire il livello di significatività  $\alpha$  (errori di I specie) si può ampliarne la regione di accettazione  $\bar{C}$ . In questo modo però è facile osservare che diminuisce anche la potenza del test  $\pi(\hat{\Theta})$ , ovvero aumenta la probabilità di compiere errori di II specie.

### 1.1.3 Test sulla Media di una Popolazione

Vediamo come si costruisce un test sulla media  $\mu$  di una popolazione  $X$  quando si formula l'ipotesi nulla che tale media sia un valore fissato:  $H_0 : \mu = \mu_0$

e quando si dispone di un campione casuale  $(X_i, \dots, X_n)$  estratto da  $X$ . Vediamo tre distribuzioni per la media, una normale con varianza nota, una normale con varianza incognita e una non normale.

**Popolazione normalmente distribuita e varianza  $\sigma^2$  nota:** Sia  $\mu$  il valore del parametro. Avremo che

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

$Z$  non è una statistica in quanto  $\mu$  non è noto. Allora

$$Z_n = \frac{\overline{X_n} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

è una statistica essendo  $\mu_0$  fissata. Il test rifiuta  $H_0$  se  $Z_n$  assume valori poco probabili per una  $N(0, 1)$ . Avremo 3 ipotesi alternative:

1.  $H_1' : \mu \neq \mu_0$
2.  $H_1'' : \mu < \mu_0$ : in tal caso  $\mu_0$  è una sovrastima della media per cui  $Z_n$  tende ad assumere valori negativi  $Z_n < Z \sim N(0, 1)$
3.  $H_1''' : \mu > \mu_0$ : in tal caso  $\mu_0$  è una sottostima della media per cui  $Z_n$  tende ad assumere valori positivi essendo  $Z_n > Z \sim N(0, 1)$

La regione critica risulta essere

1.  $C' = (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, +\infty)$
2.  $C'' = (-\infty, -z_{1-\alpha})$
3.  $C''' = (z_{1-\alpha}, +\infty)$

**Osservazione:** Nel formulario, ci sono 3 ipotesi nulle

$$H_0 = \mu_0 \quad H_0 \geq \mu_0 \quad H_0 \leq \mu_0$$

invece che la singola  $H_0 = \mu_0$

**Popolazione normalmente distribuita e varianza  $\sigma^2$  incognita:** Si procede come nell'esempio precedente, ma essendo varianza incognita useremo la  $t$  di Student.

**Popolazione non normalmente distribuita:** Per poter definire un test occorre avere un campione di numerosità sufficientemente elevata ( $n \geq 30$ ) e deve valere  $np_0 \geq 5$  e  $n(1 - p_0) \geq 5$  Allora, possiamo ricondurci ad una normale con varianza nota.

### 1.1.4 Test sulla Differenza delle Medie di due Popolazioni

Molti problemi in statistica consistono nel confronto tra 2 variabili, come 2 popolazioni. Vediamo due test, uno sulla differenza di media di due campioni normali accoppiati, ed uno sulla differenza di media di due campioni normali indipendenti.

**Differenza Campioni Accoppiati:** Prendiamo due campioni della stessa numerosità  $(X_1, \dots, X_n)_x$   $(Y_1, \dots, Y_n)_y$ . Supponiamo essi siano normali, quindi con valore medio  $\mu_x$  e  $\mu_y$ . Definiamo una differenza  $D_i = X_i - Y_i$  e denotiamo con  $\overline{D_n}$  la media campionaria e con  $S_d^2$  la varianza campionaria del campione  $D_1, \dots, D_n$ . Allora troviamo la statistica come:

$$T = \frac{\overline{D_n} - \mu_0}{S_d} \sqrt{n}$$

Di conseguenza posso trovare le tre sezioni critiche. (non le riporto, sono presenti nella tabella)

**Differenza Campioni Indipendenti:** Prendiamo due campioni della stessa numerosità  $(X_1, \dots, X_n)_x$   $(Y_1, \dots, Y_n)_y$ . Supponiamo essi siano normali, quindi con valore medio  $\mu_x$  e  $\mu_y$ . Siano  $\overline{X}$  e  $S_x^2$  media e varianza campionaria di  $(X_1, \dots, X_n)_x$  e siano  $\overline{Y}$  e  $S_y^2$  media e varianza campionaria di  $(Y_1, \dots, Y_n)_y$ . Possiamo calcolare la varianza campionaria combinata come

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

Allora ottengo la mia statistica come:

$$T = \frac{\overline{XY}}{S_p \sqrt{\frac{1}{n_x} \frac{1}{n_y}}}$$

Di conseguenza posso trovare le tre sezioni critiche. (non le riporto, sono presenti nella tabella)

### 1.1.5 Test Non Parametrici

Nei test parametrici non si fanno inferenze sui parametri ma sulla distribuzione di una o più popolazioni. Un esempio è il test Chi-quadro di buon

adattamento, dove vogliamo sapere se la popolazione ha una certa distribuzione assegnata.

**Test  $\chi^2$  di buon adattamento:** prendiamo un insieme di dati che vogliamo analizzare e supponiamo una certa distribuzione che ci aspettata se i dati fossero distribuiti in modo casuale. Il test confronta la distribuzione effettiva dei dati con quella attesa e ci dice se ci sono differenze significative tra le due. In conclusione, se i dati che abbiamo raccolto sono simili a quelli che ci aspetteremmo in base alla nostra ipotesi, allora il test ci darà un valore basso, altrimenti ci darà un valore alto.

**Osservazione:** Si utilizzano due diverse formule, in base a se abbiamo la distribuzione della popolazione discreta o continua.



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

- Test  $t$  sulla differenza delle media di due campioni normali **accoppiati**  $X_1, \dots, X_n$  di media  $\mu_X$  e  $Y_1, \dots, Y_n$  di media  $\mu_Y$

Test  $t$  sul campione delle differenze  $D_1 = X_1 - Y_1, \dots, D_n = X_n - Y_n$ , denotiamo con  $\bar{D}_n = \frac{1}{n}(D_1 + \dots + D_n)$  la media campionaria e con  $S_d^2$  la varianza campionaria del campione  $D_1, \dots, D_n$ .

$H_0$	$H_1$	Statistica	Regione critica
$\mu_X = \mu_Y + \mu_0 \rightsquigarrow \mu_d = \mu_0$	$\mu_X \neq \mu_Y + \mu_0 \rightsquigarrow \mu_d \neq \mu_0$	$T = \frac{\bar{D}_n - \mu_0}{S_d} \sqrt{n}$	$\left  \frac{\bar{d}_n - \mu_0}{s_d} \sqrt{n} \right  > t_{n-1, \alpha/2}$
$\mu_X \leq \mu_Y + \mu_0 \rightsquigarrow \mu_d \leq \mu_0$	$\mu_X > \mu_Y + \mu_0 \rightsquigarrow \mu_d > \mu_0$	$T = \frac{\bar{D}_n - \mu_0}{S_d} \sqrt{n}$	$\frac{d_n - \mu_0}{s_d} \sqrt{n} > t_{n-1, \alpha}$
$\mu_X \geq \mu_Y + \mu_0 \rightsquigarrow \mu_d \geq \mu_0$	$\mu_X < \mu_Y + \mu_0 \rightsquigarrow \mu_d < \mu_0$	$T = \frac{\bar{D}_n - \mu_0}{S_d} \sqrt{n}$	$\frac{d_n - \mu_0}{s_d} \sqrt{n} < -t_{n-1, \alpha}$

- Test  $t$  sulla differenza delle media di due campioni normali indipendenti  $X_1, \dots, X_{n_x}$  di media  $\mu_x$  e  $Y_1, \dots, Y_{n_y}$  di media  $\mu_y$

$\bar{X}$  e  $S_x^2$  media e varianza campionarie di  $X_1, \dots, X_{n_x}$ .

$\bar{Y}$  e  $S_y^2$  media e varianza campionarie di  $Y_1, \dots, Y_{n_y}$ .

varianza campionaria combinata:  $S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$

$H_0$	$H_1$	Statistica	Regione critica
$\mu_x = \mu_y$	$\mu_x \neq \mu_y$	$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$	$\left  \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \right  > t_{n_x + n_y - 2, \alpha/2}$
$\mu_x \leq \mu_y$	$\mu_x > \mu_y$	$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$	$\frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} > t_{n_x + n_y - 2, \alpha}$
$\mu_x \geq \mu_y$	$\mu_x < \mu_y$	$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$	$\frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} < -t_{n_x + n_y - 2, \alpha}$

- Test  $\chi^2$  di buon adattamento

$H_0$ : la popolazione ha una certa distribuzione assegnata.

si vuole decidere se accettare o rifiutare  $H_0$

$C_1, \dots, C_k$  classi

$N_1, \dots, N_k$  frequenze assolute delle classi;  $n_1, \dots, n_k$  frequenze assolute **osservate**;

$f_1, \dots, f_k$  frequenze assolute **attese**. Come le calcolo?

- se si vuole testare il buon adattamento a una distribuzione discreta, le classi coincidono con uno o più valori assunti dalla distribuzione incognita; sia  $C_1 = \{1\}, \dots, C_k = \{k\}$ , si assegna  $\pi = (\pi(1), \dots, \pi(k))$ , densità discreta, e

$$f_1 = n\pi(1), \dots, f_k = n\pi(k)$$

- se si vuole testare il buon adattamento a una distribuzione continua, si avrà  $C_1 = [a_0, a_1), C_2 = [a_1, a_2), \dots, C_k = [a_{k-1}, a_k)$  si assegna  $F$ , funzione di ripartizione, e

$$f_1 = n(F(a_1) - F(a_0)), \dots, f_k = n(F(a_k) - F(a_{k-1}))$$

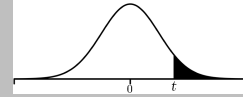
## TAVOLA DELLA DISTRIBUZIONE NORMALE

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

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

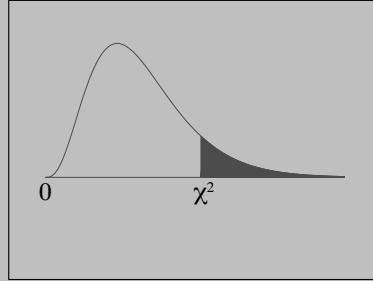
$z$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

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

### 1.1.6 Esercizi

#### Esercizio 1: Test z per popolazione gaussiana, varianza nota

**Traccia:** Azienda produce anelli. Il diametro di questi anelli è normalmente distribuito ed ha una deviazione standard pari a  $\sigma = 0.001mm$ . Campione di  $n=15$  anelli, si ricava  $\bar{x} = 74.036mm$

1. Si testi l'ipotesi che la media del diametro sia uguale a  $\mu = 74.035mm$  ad un livello di significatività pari a  $\alpha = 5\%$  e si calcoli il p-value del test.
2. Probabilità che a livello  $\alpha = 5\%$ , l'ipotesi che la media sia  $\mu = 74.035mm$  non venga rifiutata quando il valore della media è  $74.034$

**Soluzione punto 1:** Essendo il campione normalmente distribuito ed avendo varianza nota troviamo dal formulario il Test z sulla media di una popolazione normale con varianza nota pari a  $\sigma^2$

1. *Scrivo l'ipotesi nulla e l'ipotesi alternativa:*  $H_0 : \mu = 74.035$ ,  $H_1 : \mu \neq 74.036$
2. *Ricavo dalla tabella la regione critica e la calcolo:* Regione critica  $\bar{C}$  del test a livello  $\alpha$ :

$$\left| \frac{\bar{x}_n - \mu}{\sigma} \sqrt{n} \right| > z_{\alpha/2} \rightarrow \left| \frac{74.036 - 74.035}{0.001} \sqrt{15} \right| \simeq 3.873 > z_{\alpha/2}$$

3. *Confronto con il percentile della gaussiana:*  $\alpha = 0.05 \rightarrow z_{0.025}$  quindi cerco sulle tavole della normale  $0.975$  e trovo  $z_{0.025} = 1.96$
4. *Confronto:* Dopo aver calcolato regione critica e percentile della gaussiana ottengo  $3.873 > 1.96$ . Essendo questo vero, rifiuto  $H_0$  a livello  $5\%$ . I dati mi permettono di dimostrare statisticamente  $H_1$
5. *Formula p-value:* Calcolo  $\bar{\alpha}$ , ovvero la probabilità che la regione critica sia esattamente il percentile della gaussiana, ma  $\bar{\alpha}$  è incognita:  $\bar{C}$  del test a livello  $\alpha$ :

$$\left| \frac{\bar{x}_n - \mu}{\sigma} \sqrt{n} \right| = z_{\bar{\alpha}/2} \rightarrow 3.873 = z_{\bar{\alpha}/2}$$

6. *Calcolo p-value:* Per trovare il valore di  $\bar{\alpha}$  uso la fdr della gaussiana standard:

$$\Phi(3.873) = \Phi(z_{\bar{\alpha}/2}) \rightarrow \Phi(3.873) = 1 - z_{\bar{\alpha}/2} \rightarrow \bar{\alpha} = 2(1 - \Phi(3.873)) \rightarrow \bar{\alpha} \simeq 0$$

7. *Conclusion:* Più il p-value è piccolo, più i dati sono in forte contraddizione con  $H_0$ .

**Soluzione punto 2:** Stiamo lavorando con un errore di II specie: "accetto  $H_0$  quando è falsa"

1. *Formula statistica:* Ricavo dalla tabella la formula

$$\mathbb{P}_{\mu=74.034} = \left( \left| \frac{\bar{x}_n - \mu}{\sigma} \sqrt{n} \right| > z_{\alpha/2} \right) = \left( \left| \frac{\bar{x}_{15} - 74.035}{0.001} \sqrt{15} \right| > 1.96 \right)$$

2. *Riscrivere come normale standard:* Essendo sotto ipotesi che  $\mu \neq \bar{x}_n$  il nostro corpo della probabilità non è approssimabile come una  $N(0, 1)$ . Aggiungo allora (+ 0.001 - 0.001):

$$\left( -1.96 \leq \frac{\bar{x}_{15} - 74.035 + 0.001 - 0.001}{0.001} \sqrt{15} \leq 1.96 \right)$$

3. *Faccio i calcoli:*

$$\begin{aligned} & \left( -1.96 \leq \frac{\bar{x}_{15} - 74.034}{0.001} \sqrt{15} - \frac{0.001}{0.001} \sqrt{15} \leq 1.96 \right) = \\ & = \left( -1.96 \sqrt{15} \leq \frac{\bar{x}_{15} - 74.034}{0.001} \sqrt{15} \leq 1.96 \sqrt{15} \right) \simeq \\ & \simeq \mathbb{P}(1.91 \leq z \leq 5.83) = \Phi(5.83) - \Phi(1.11) = 0.0281 \end{aligned}$$

**Traccia:** Un produttore di batterie ha messo sul mercato un nuovo modello sostenendo che la durata media è superiore a quella del vecchio modello che era pari a 14 ore. Su un campione di 10 batterie sono state osservate le seguenti durate: 18 15 14 16 15 12 13 15 17. Supponendo che il tempo di durata sia normale:

1. Sottoporre a verifica l'affermazione del produttore a livello  $\alpha = 5\%$
2. Calcolare il p-value del test

**Soluzione punto 1:** Essendo il campione normalmente distribuito ed avendo varianza incognita troviamo dal formulario il Test t sulla media di una popolazione normale con varianza incognita

1. *Ricavo l'ipotesi nulla e l'ipotesi alternativa:* In questo caso la traccia non specifica direttamente qual'è l'ipotesi nulla. Possiamo scegliere che l'affermazione del produttore sia l'ipotesi alternativa  $H_1 : \mu > 14$ , in quanto se rifiuto  $H_0 : \mu \leq 14$  posso dire con certezza che il produttore abbia torto. Se avessimo messo l'ipotesi del produttore come ipotesi nulla  $H_0$  e la rifiutassimo i dati non ci permetterebbero di escludere che il produttore abbia ragione (conclusione meno forte).

2. *Ricavo dalla tabella la regione critica e la calcolo:* Regione critica  $\bar{C}$  del test a livello  $\alpha$ :

$$\left| \frac{\bar{x}_n - \mu}{s_n} \sqrt{n} \right| > t_{n-1, \alpha}$$

3. *Trovo i dati e risolvo:*  $\alpha = 0.05$ ,  $n = 10$ ,  $\bar{x}_n = \frac{18+\dots+13}{10} = 15$ ,  $s_n = \frac{1}{9}((18-15)^2 + \dots + (13-15)^2) = 4$ ,  $s_n^2 = \sqrt{4} = 2$  dunque la mia regione critica sarà:

$$\frac{15 - 14}{2} \sqrt{10} \simeq 1.58 > t_{n-1, \alpha}$$

4. *Confronto con  $t$  di Student e conclusioni:* trovo che  $t_{9, 0.005} = 1.833$  quindi  $1.58 > 1.833$ . Essendo questo falso accetto  $H_0$  a livello 5%, ovvero i miei dati non mi permettono di rifiutare  $H_0$ , ovvero non mi permettono di dimostrare che il produttore ha ragione. I dati non sono in contraddizione significativa con  $H_0$

**Soluzione punto 2:** P-value, so già che  $\bar{\alpha} > 5\%$  dal punto prima, quindi p-value alto. So che  $1.58 = t_{9, \alpha}$  quindi cerco i numeri che includono 1.58 e li trovo ai valori  $0.05 < \bar{\alpha} < 0.1$  della tabella, ovvero  $5\% < \bar{\alpha} < 10\%$

### Esercizio 3: Test sulla proporzione

**Traccia:** Un'inserzione pubblicitaria per un prodotto contro il mal di testa dichiara che almeno un 90% delle persone che soffrono di questo disturbo otterrebbe beneficio se lo usasse. L'associazione dei consumatori, considerando tale pubblicità tendenziosa, ottiene un campione di 100 individui di cui 88 dichiarano che il prodotto è stato efficace. Siete d'accordo con l'associazione?

**Soluzione:** Essendo il campione una proporzione usiamo il Test z sulla proporzione. Le ipotesi ( $n \geq 30$ ),  $np_0 \geq 5$  e  $n(1 - p_0) \geq 5$  sono verificate con i nostri dati  $p_0 = 0.9$  e  $n = 100$  quindi posso procedere.

1. *Trovo ipotesi nulla ed alternativa:* Metto come ipotesi alternativa  $H_1 : p < 0.9$  in modo tale che se rifiuto  $H_0 : p \geq 0.9$  posso dire con certezza che la pubblicità sta mentendo.



2. *Scelta di svolgimento:* Per i test sulla proporzione possiamo scegliere se prendere un livello di significatività a piacere oppure trovare il p-value. Facciamo il primo.
3. *Trovo Area Critica:* Scelgo  $\alpha = 0.05$  e la mia area critica usando il formulario è

$$\frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} < -z_\alpha \rightarrow \frac{0.88 - 0.9}{\sqrt{0.9(1-0.1)}}\sqrt{100} < -1.645$$

4. *Conclusioni:* Ottengo  $-0.667 < -1.645$  che è falso quindi accetto  $H_0$  a livello 5% e quindi i dati non sono in contraddizione significativa con  $H_0$  e quindi non mi permettono di dimostrare statisticamente che la pubblicità sia tendenziosa. Se calcolassi il p-value ottengo 0.2524 quindi p-value grande

#### Esercizio 4: Dati Accoppiati

**Traccia:** I dati seguenti mettono in relazione la frequenza cardiaca di 12 individui prima e dopo aver masticato tabacco. Verifica a livello 5% l'ipotesi che masticare tabacco non provochi un aumento della frequenza cardiaca e calcolare p-value

Masticare Tabacco		
Individuo	Frequenza prima	Frequenza Dopo
1	73	77
2	67	69
3	68	73
..	..	..
12	78	80

**Soluzione:** Usiamo il test t sulla differenza della media di 2 campioni normali accoppiati  $X_i, \dots, X_n$  di media  $\mu_x$  e  $Y_1, \dots, Y_n$  di media  $\mu_y$ .  $D_1 = X_1 - Y_1$ ,  $D_n = X_n - Y_n$ , media  $\bar{D}_n$ , varianza  $S_d^2$ .

1. *Scelgo ipotesi nulla ed ipotesi alternativa:*  $H_0 : \mu_d \leq 0$ ,  $H_1 : \mu_d > 0$  ovvero metto come ipotesi alternativa che ci sia differenza così se rifiuto l'ipotesi nulla posso dire con certezza che c'è differenza tra prima e dopo masticare tabacco.
2. *Regione Critica:* Dal formulario:

$$\frac{\bar{d}_n - \mu_0}{S_d}\sqrt{n} > t_{n-1, \alpha}$$

3. *Calcolo i valori:* Prendo i valori  $n = 12$ ,  $\mu_0 = 0$ ,  $\overline{d}_n = \frac{(77-73)+\dots+(80-78)}{12} = 3.75$ ,  $s_d^2 = \frac{1}{11}((4 - 3.75)^2 + \dots + (2 - 3.75)^2) = 9.477$ ,  $s_d = 3.078$ ,  $t_{11,0.05} = 1.796$

4. *Confrontiamo:*

$$\frac{3.75 - 0}{3.078} \sqrt{12} = 4.22 > 1.796$$

I dati sono in contraddizione significativa con  $H_0$  dimostro statisticamente che masticare tabacco provoca un aumento della frequenza cardiaca.

5. *P-value:*  $4.22 = t_{11,\bar{\alpha}}$  e trovo  $0.0005 < \bar{\alpha} < 0.001$  forte evidenza empirica contro  $H_0$  in quanto p-value molto piccolo