

Formulario e tavole



Ne è consentito l'uso all'esame scritto, ma ogni Studente deve consultare solo il proprio formulario, e nessun altro materiale!

Statistica Descrittiva

$$\text{densità} = \frac{\text{frequenza}}{\text{ampiezza}}$$

$$\Leftrightarrow \text{frequenza} = \text{ampiezza} \times \text{densità}$$

Media aritmetica – per tabelle di frequenze - pesata

$$\bar{x} = \frac{\sum x_i}{n} \quad \bar{x} = \frac{\sum_{i=1}^k x_i n_i}{n} \quad \bar{x}_p = \frac{\sum_{j=1}^K x_j p_j}{\sum_{j=1}^k p_j}$$

$$n \text{ dispari} \rightarrow \text{rango}(\text{Mediana}) = \frac{n+1}{2}$$

$$n \text{ pari} \rightarrow \text{rango}(\text{Mediana}) = \frac{n}{2}, \frac{n}{2} + 1$$

$$\text{Varianza} \quad \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \left(\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 \right) \cdot \frac{n}{n-1}$$

$$\text{dev.st.} = \sqrt{\text{varianza}}$$

$$\text{coeff. di variazione} = \frac{\text{dev.st.}}{\bar{x}} \cdot 100$$

Proprietà della Normale $N(\mu, \sigma)$

$$\Pr(\mu - \sigma < X < \mu + \sigma) = 68\%$$

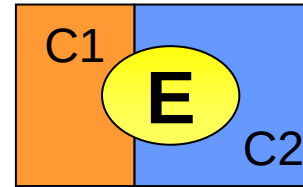
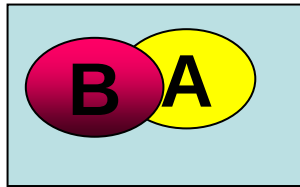
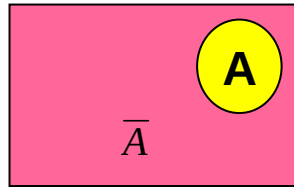
$$\Pr(\mu - 2\sigma < X < \mu + 2\sigma) = 95\%$$

$$\Pr(\mu - 3\sigma < X < \mu + 3\sigma) = 99.7\%$$

$$Q_1 = \mu - 0.67 \cdot \sigma$$

$$Q_3 = \mu + 0.67 \cdot \sigma$$

Regole elementari di calcolo delle probabilità



$$p(\bar{A}) = 1 - p(A)$$

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

$$\begin{aligned} p(A \cap B) &= p(A | B) \cdot p(B) \\ &= p(B | A) \cdot p(A) \end{aligned}$$

$$p(A | B) = \frac{p(A \cap B)}{p(B)}$$

$$p(A \cap B) \stackrel{\text{indipendenti}}{=} p(B) \cdot p(A)$$

Formula di Bayes

$$p(C1 | E) = \frac{p(E | C1) \cdot p(C1)}{p(E | C1) \cdot p(C1) + p(E | C2) \cdot p(C2)}$$

Calcolo con le distribuzioni Binomiale e di Poisson

prob(successo)= π , X = numero di successi in N prove

$$p(X = x) = \binom{N}{x} \pi^x (1 - \pi)^{N-x}$$

Il numero medio “atteso” di successi è $N \cdot \pi$

prob(successo) ≈ 0 ovvero tasso di successo = λ , X = numero di successi in $N \gg 0$ prove ovvero lungo un intervallo continuo di osservazione

$$p(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Il numero medio “atteso” di successi è $\lambda = N \cdot \pi$

$$\binom{N}{k} = \frac{N!}{(N-k)!k!}$$
$$k! = k \cdot (k-1) \cdot (k-2) \cdot \dots \cdot 2 \cdot 1$$
$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$$
$$0! = 1$$
$$e^{-\lambda} = \frac{1}{e^{\lambda}}$$
$$k^0 = 1$$

Calcolo con la distribuzione Normale

$$Area_{N(\mu, \sigma^2)}(X \leq x) = Area_{N(0,1)}\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

$$X_i \xrightarrow{\text{standardizzazione}} Z_i$$

Valori sulla scala
originaria, con
parametri μ e σ^2

Valori sulla
 $N(0,1)$

Standardizzazione:

$$Z = \frac{x - \mu}{\sigma}$$

Operazione inversa:

$$x = \mu + \sigma \cdot z$$

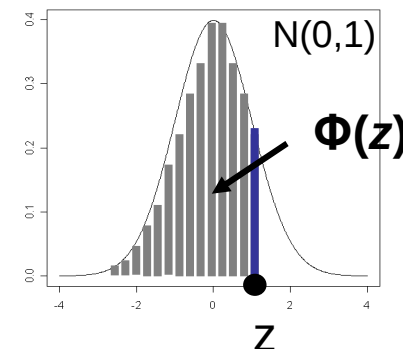
Per $z > 0$ $\Phi(z)$ si legge sulla **tabella fornita qui di seguito**
(attenzione: altri testi possono riportare altre aree)

Aree di forma diversa si ottengono
utilizzando la simmetria della curva.

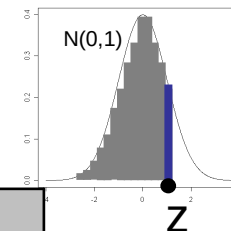
$$\Phi(0) = 0.5 \quad \Phi(+\infty) = 1$$

$$Area(Z < -z) = Area(Z > z)$$

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



Tavole della Normale (0,1), z da 0 a 1.99



z	0	1	2	3	4	5	6	7	8	9
0.0	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536
0.1	0.540	0.544	0.548	0.552	0.556	0.560	0.564	0.567	0.571	0.575
0.2	0.579	0.583	0.587	0.591	0.595	0.599	0.603	0.606	0.610	0.614
0.3	0.618	0.622	0.626	0.629	0.633	0.637	0.641	0.644	0.648	0.652
0.4	0.655	0.659	0.663	0.666	0.670	0.674	0.677	0.681	0.684	0.688
0.5	0.691	0.695	0.698	0.702	0.705	0.709	0.712	0.716	0.719	0.722
0.6	0.726	0.729	0.732	0.736	0.739	0.742	0.745	0.749	0.752	0.755
0.7	0.758	0.761	0.764	0.767	0.770	0.773	0.776	0.779	0.782	0.785
0.8	0.788	0.791	0.794	0.797	0.800	0.802	0.805	0.808	0.811	0.813
0.9	0.816	0.819	0.821	0.824	0.826	0.829	0.831	0.834	0.836	0.839
1.0	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.858	0.860	0.862
1.1	0.864	0.867	0.869	0.871	0.873	0.875	0.877	0.879	0.881	0.883
1.2	0.885	0.887	0.889	0.891	0.893	0.894	0.896	0.898	0.900	0.901
1.3	0.903	0.905	0.907	0.908	0.910	0.911	0.913	0.915	0.916	0.918
1.4	0.919	0.921	0.922	0.924	0.925	0.926	0.928	0.929	0.931	0.932
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944
1.6	0.945	0.946	0.947	0.948	0.949	0.951	0.952	0.953	0.954	0.954
1.7	0.955	0.956	0.957	0.958	0.959	0.960	0.961	0.962	0.962	0.963
1.8	0.964	0.965	0.966	0.966	0.967	0.968	0.969	0.969	0.970	0.971
1.9	0.971	0.972	0.973	0.973	0.974	0.974	0.975	0.976	0.976	0.977

[illegible]

IC al livello $(1-\alpha)$ % per la media μ

Ipotesi: nella popolazione il fenomeno si distribuisce secondo una Normale con media μ (incognita) e varianza σ^2 nota

Formula valida anche per il caso di una popolazione non Normale o se non si conosce* la varianza della popolazione σ^2 **SE** il campione è grande ($n \geq 30$)

*(al posto di σ^2 si usa la varianza calcolata nel campione)

Formula:

$$\left(\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Quantile: $N(0,1)$

$1-\alpha$	$z_{\frac{\alpha}{2}}$
90%	1.64
95%	1.96

Nota: quando non si conosce la varianza σ^2 e il campione è piccolo si usano i quantili della distribuzione T di Student con $(n-1)$ gradi di libertà

T-test su una media

$$H_0: \mu = \mu_0$$

Ipotesi: nella popolazione il fenomeno si distribuisce secondo una Normale con media μ (incognita) e varianza σ^2 nota

Formula valida anche per il caso di una popolazione non Normale o se non si conosce* la varianza della popolazione σ^2 **SE** il campione è grande ($n \geq 30$)

*(al posto di σ^2 si usa la varianza calcolata nel campione)

Statistica test: la media calcolata nel campione, standardizzata:

$$t = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

Calcolare il p-value sulla tavola della $N(0,1)$

oppure, per test al livello di significatività $\alpha=5\%$

Confrontare t con il limite della regione di rifiuto z:

H_1	z
$\mu > \mu_0$ ($\mu < \mu_0$)	+1.64 (-1.64)
$\mu \neq \mu_0$	± 1.96

Nota: quando non si conosce la varianza σ^2 e il campione è piccolo si usa la distribuzione T di Student con $(n-1)$ gradi di libertà

T-test per confrontare 2 medie da campioni indipendenti

$$H_0: \mu_1 = \mu_2 \text{ vs } H_1: \mu_1 \neq \mu_2$$

Ipotesi: i due gruppi provengono da due popolazioni rispettivamente con media μ_1 e μ_2 (incognite) e uguale varianza σ^2 incognita; entrambi i campioni sono grandi ($n_1, n_2 > 30$)

Calcolare:
$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Statistica test: la differenza delle medie calcolate nel campione, standardizzata:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Calcolare il p-value sulla tavola della $N(0,1)$

oppure, per test al livello di significatività $\alpha=5\%$

Confrontare t con il limite della regione di rifiuto $z = 1.96$

IC al livello $(1-\alpha)$ % e test per la prob. π

Ipotesi: il campione è grande ($n \geq 30$)

Formula IC95%:

$$\left(\hat{\pi} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}, \hat{\pi} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \right)$$

Quantile: N(0,1)	$1-\alpha$	$z_{\frac{\alpha}{2}}$
	90%	1.64
	95%	1.96

Statistica test per $H_0: \pi = \pi_0$: la proporzione calcolata nel campione, standardizzata:

$$t = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}}$$

Calcolare il p-value sulla tavola della N(0,1)

oppure, per test al livello di significatività $\alpha=5\%$

Confrontare t con il limite della regione di rifiuto z:

H_1	z
$\pi > \pi_0$ ($\pi < \pi_0$)	+1.64 (-1.64)
$\pi \neq \pi_0$	± 1.96

Nota: in sostanza, si tratta la probabilità come la media di un campione, SE il campione è grande. Si può usare anche il T-test per confrontare 2 probabilità da campioni indipendenti e GRANDI, l'alternativa non-parametrica è il test Chi-Quadrato (e, per campioni piccoli, il test F di Fisher)

Test Chi-Quadrato per l'associazione fra due caratteri qualitativi (tabella doppia rxc)

H_0 : X (r modalità) e Y (c modalità) indipendenti vs H_1 : X,Y associati

Ipotesi: nella tabella doppia, tutte le frequenze attese sono >5

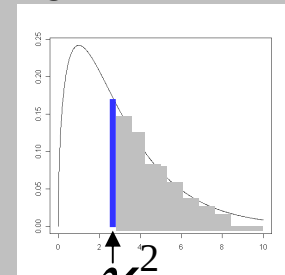
Calcolare le frequenze attese sotto l'ipotesi H_0 di indipendenza:

$$\tilde{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n_{..}} \quad \text{ovvero, per ogni cella: } \textit{Attesa} = \frac{\text{tot riga} \cdot \text{tot colonna}}{\text{tot generale}}$$

Statistica test:

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}} = \sum \frac{(\textit{Osservate} - \textit{Attese})^2}{\textit{Attese}}$$

Calcolare il p-value sulla tavola del Chi-quadrato con gradi di libertà $(r-1)(c-1)$

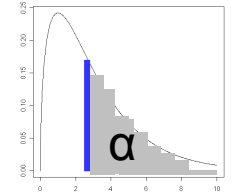


oppure

Confrontare χ^2 con il limite della regione di rifiuto z: solo per tabella 2x2:

α	z
0.05	3.841
0.01	6.635

Tavole del Chi-Quadrato



gdl	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	gdl	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	gdl	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$
1	2.706	3.841	6.635	21	29.615	32.671	38.932	41	52.949	56.942	64.950
2	4.605	5.991	9.210	22	30.813	33.924	40.289	42	54.090	58.124	66.206
3	6.251	7.815	11.345	23	32.007	35.172	41.638	43	55.230	59.304	67.459
4	7.779	9.488	13.277	24	33.196	36.415	42.980	44	56.369	60.481	68.710
5	9.236	11.070	15.086	25	34.382	37.652	44.314	45	57.505	61.656	69.957
6	10.645	12.592	16.812	26	35.563	38.885	45.642	46	58.641	62.830	71.201
7	12.017	14.067	18.475	27	36.741	40.113	46.963	47	59.774	64.001	72.443
8	13.362	15.507	20.090	28	37.916	41.337	48.278	48	60.907	65.171	73.683
9	14.684	16.919	21.666	29	39.087	42.557	49.588	49	62.038	66.339	74.919
10	15.987	18.307	23.209	30	40.256	43.773	50.892	50	63.167	67.505	76.154
11	17.275	19.675	24.725	31	41.422	44.985	52.191				
12	18.549	21.026	26.217	32	42.585	46.194	53.486				
13	19.812	22.362	27.688	33	43.745	47.400	54.776				
14	21.064	23.685	29.141	34	44.903	48.602	56.061				
15	22.307	24.996	30.578	35	46.059	49.802	57.342				
16	23.542	26.296	32.000	36	47.212	50.998	58.619				
17	24.769	27.587	33.409	37	48.363	52.192	59.893				
18	25.989	28.869	34.805	38	49.513	53.384	61.162				
19	27.204	30.144	36.191	39	50.660	54.572	62.428				
20	28.412	31.410	37.566	40	51.805	55.758	63.691				

Misure di associazione

Chi-Quadrato:

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - \tilde{n}_{ij})^2}{\tilde{n}_{ij}}$$

Frequenze attese:

$$\tilde{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n_{..}}$$
$$= \frac{\text{tot riga} \cdot \text{tot colonna}}{\text{tot generale}}$$

Covarianza:

$$\text{COV}_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum_{i=1}^n x_i y_i}{n} - \bar{xy}$$

Coefficiente di Correlazione Lineare:

$$r = \frac{\text{COV}_{xy}}{\text{std}_x \cdot \text{std}_y}$$

Retta di regressione:

$$b = \frac{\text{COV}_{xy}}{\text{var}_x} = r_{xy} \frac{\text{std}_y}{\text{std}_x}$$

$$a = \bar{y} - b \cdot \bar{x}$$