

Metodi Matematici e Informatici per la Biologia

a.a. 2023-2024

(Agliari - Panati - Presilla - Simonella)

27 gennaio 2025

Prova di Verifica IV

Nome:

Cognome:

Matricola:

Esercizio 1 (2+2+2+2+2=10 punti)

Le misurazioni del consumo giornaliero di litri d'acqua di una famiglia eseguite in 40 giorni successivi forniscono i seguenti valori (misurati in litri/giorno):

13, 29, 85, 35, 0, 8, 2, 19, 45, 3, 37, 73, 46, 17, 29, 14, 0, 2, 31, 2, 108, 148, 59, 27, 32, 5, 31, 42, 31, 17, 205, 11, 97, 37, 68, 50, 21, 89, 64, 33.

1.1 Si calcolino media campionaria $\bar{X}$, mediana e deviazione standard campionaria S , scrivendo esplicitamente le relative definizioni.

Per il campione fornito avente $n = 40$ dati si ha:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = 41.6250 \text{ litri/giorno,}$$

$$\text{med}(X) = \frac{X_{20} + X_{21}}{2} = 31 \text{ litri/giorno, avendo ordinato i dati in maniera crescente } (X_1 \leq \dots \leq X_{40}),$$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} = 42.4032 \text{ litri/giorno}$$

1.2 Si dica se si tratta di un campione approssimativamente normale, fornendo almeno un argomento per giustificare la risposta.

Il campione non è approssimativamente normale per i seguenti possibili argomenti:

- a) l'istogramma dei dati presenta una evidente asimmetria;
- b) il normal-plot presenta una concavità positiva;
- c) la regola 68-95-99 è palesemente violata.

(Il campione è stato estratto da una distribuzione esponenziale)

1.3 Si rappresenti il boxplot del campione di dati, specificando tutti i relativi valori di riferimento.

minimo: 0 litri/giorno, primo quartile: 13.5 litri/giorno, mediana: 31 litri/giorno, terzo quartile: 54.5 litri/giorno, massimo: 205 litri/giorno

1.4 Si calcolino, illustrando il procedimento, il decimo ed il novantesimo percentile del campione di dati.

Dopo aver ordinato i dati in maniera crescente ($X_1 \leq \dots \leq X_{40}$), per il calcolo del decimo percentile si consideri $k = n \times \frac{10}{100} = 4$, essendo tale valore intero il percentile è

$$\frac{X_k + X_{k+1}}{2} = \frac{2 + 2}{2} = 2 \text{ litri/giorno.}$$

Analogamente, per il novantesimo percentile $k = n \times \frac{90}{100} = 36$, essendo tale valore intero il percentile è

$$\frac{X_k + X_{k+1}}{2} = \frac{89 + 97}{2} = 93 \text{ litri/giorno.}$$

1.5 Si determini un intervallo di valori che contenga il 68% dei dati del campione.

Per definizione il 68-esimo percentile corrisponde a circa il 68% dei dati, ordinati in modo crescente a partire dal loro valore minimo. Un intervallo come quello richiesto è quindi [minimo, 68-esimo percentile] che corrisponde approssimativamente a $[0, 44.1]$ litri/giorno.

Si noti che l'intervallo $[\bar{X} - S, \bar{X} + S] = [-0.7782, 84.0282]$ litri/giorno contiene invece circa l'86% dei dati a riprova che la distribuzione non è normale.

Esercizio 2 (2+2+2=6 punti)

Negli ultimi anni in Italia solo il 4% degli studenti che affrontano l'esame di maturità ha riportato una votazione di 100/100.

2.1 Si indichi una ragionevole distribuzione di probabilità per calcolare il numero di studenti, in una classe di 33, che si diplomano con il massimo punteggio.

Il numero di studenti che si diplomano con il voto massimo può essere considerato una variabile aleatoria binomiale $S_n = \sum_{k=1}^n X_k$, dove $X_k = 1$ se il k -esimo degli $n = 33$ studenti prende 100 e $X_k = 0$ altrimenti. Si ha $P(X_k = 1) = p$ con $p = 4/100$ e quindi

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

In alternativa, poiché n è abbastanza grande, possiamo approssimare i risultati della binomiale con la distribuzione di Poisson di una variabile aleatoria X , di media $\lambda = 33 \times 4/100 = 1.32$ e funzione di massa di probabilità

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

2.2 Utilizzando tale distribuzione, si calcoli la probabilità che nessuno studente della classe ottenga il massimo dei voti,

Usando la distribuzione binomiale

$$P(S_{33} = 0) = \binom{33}{0} p^0 (1-p)^{33} = (1-p)^{33} \simeq 0.260$$

Usando la distribuzione di Poisson

$$P(X = 0) = \frac{\lambda^0}{0!} e^{-\lambda} = e^{-1.32} \simeq 0.267$$

2.3 e la probabilità che almeno 3 studenti della classe ottengano il massimo dei voti.

Usando la distribuzione binomiale

$$\begin{aligned} P(S_{33} \geq 3) &= 1 - (P(S_{33} = 0) + P(S_{33} = 1) + P(S_{33} = 2)) \\ &= 1 - \left(\binom{33}{0} p^0 (1-p)^{33} + \binom{33}{1} p^1 (1-p)^{32} + \binom{33}{2} p^2 (1-p)^{31} \right) \\ &\simeq 0.144 \end{aligned}$$

Usando la distribuzione di Poisson

$$\begin{aligned}P(X \geq 3) &= 1 - (P(X = 0) + P(X = 1) + P(X = 2)) \\&= 1 - \left(1 + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!}\right) e^{-\lambda} \\&= 1 - \left(1 + 1.32 + \frac{1.32^2}{2}\right) e^{-1.32} \\&\simeq 0.147\end{aligned}$$

Esercizio 3 (2+2+2=6 punti)

Una macchina refrigerante in azione per X ore abbassa la temperatura di una cella frigorifera a Y °C come monitorato dalle seguenti 7 misurazioni:

$X(\text{ore})$	3.30	3.80	4.99	5.74	7.03	8.16	9.34
$Y(^{\circ}\text{C})$	-15.10	-17.67	-22.85	-26.05	-28.47	-34.77	-48.48

3.1 Determinare medie e deviazioni standard delle variabili X e Y del campione.

Il campione ha $n = 7$ dati e abbiamo i seguenti valori medi e deviazioni standard

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i = 6.05 \text{ ore}, & S_X &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} = 2.24 \text{ ore}, \\ \bar{Y} &= \frac{1}{n} \sum_{i=1}^n Y_i = -27.63 \text{ }^{\circ}\text{C}, & S_Y &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2} = 11.31 \text{ }^{\circ}\text{C}.\end{aligned}$$

3.2 Determinare l'equazione della corrispondente retta di regressione lineare e dell'associato coefficiente di correlazione r .

La retta di regressione lineare ha equazione $Y = aX + b$ con

$$a = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = -4.88 \text{ }^{\circ}\text{C/ore},$$

$$b = \bar{Y} - a\bar{X} = 1.93 \text{ }^{\circ}\text{C}.$$

Il coefficiente di correlazione è il numero

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{S_X} \right) \left(\frac{Y_i - \bar{Y}}{S_Y} \right) = -0.967.$$

3.3 In quanto tempo la cella dovrebbe raggiungere la temperatura $Y = -70$ °C?

$$X = \frac{Y - b}{a} = \frac{(-70 - 1.93) \text{ }^{\circ}\text{C}}{-4.88 \text{ }^{\circ}\text{C/ore}} = 14.74 \text{ ore}.$$

Esercizio 4 (2+2+2=6 punti)

I seguenti $n = 64$ valori costituiscono un campione di una variabile aleatoria

22.4068 21.4640 21.6448 20.8153 22.3582 21.8522 19.8134 20.3051
17.5286 21.5596 20.8131 24.1789 20.8267 17.8681 21.9928 16.5818
20.2513 20.6278 17.4116 20.5234 21.9711 17.2719 20.9209 22.7997
18.3895 22.8883 21.9153 20.4422 21.8812 19.3265 18.9496 21.8754
19.1839 21.6838 20.4167 19.9607 20.9080 21.7677 19.1629 20.4361
17.8960 19.2969 22.8496 21.7890 14.7062 16.5618 21.6812 19.9248
18.2798 17.4298 19.3612 21.3175 20.3548 18.0676 21.5738 18.2420
16.7587 20.5813 22.3904 21.1385 20.0346 20.0549 19.5419 19.8395

4.1 Determinare valore minimo, valore massimo, valore medio e deviazione standard del campione.

valore minimo = 14.7062, valore massimo = 24.1789,
valore medio = 20.1976, deviazione standard = 1.8814.

4.2 Calcolare le frequenze dei dati del campione usando classi di ampiezza 1 e quindi produrre il relativo istogramma.

classe	frequenza
(13,14]	0
(14,15]	1
(15,16]	0
(16,17]	3
(17,18]	6
(18,19]	5
(19,20]	10
(20,21]	16
(21,22]	16
(22,23]	6
(23,24]	0
(24,25]	1
(25,26]	0

4.3 Giudicare il carattere normale dei dati del campione realizzando un normal plot e riportando il coefficiente di determinazione della corrispondente retta di regressione lineare.

Il normal plot mostra che i dati hanno una distribuzione approssimativamente normale; il coefficiente di determinazione della retta di regressione lineare vale $r^2 = 0.964$.

Esercizio 5 (3+3=6 punti)

È stato selezionato un campione aleatorio di persone, alcuni fumatori e altri non fumatori. Queste persone sono state seguite per 10 anni, tenendo traccia di quanti tra loro hanno sviluppato un cancro ai polmoni. I risultati sono riportati nella seguente tabella:

	fumatori	non fumatori
cancro ai polmoni	6	2
niente cancro ai polmoni	994	1998

5.1 Si verifichi, a un livello di significatività dell'1%, l'ipotesi nulla H_0 secondo cui cancro ai polmoni e fumo sono indipendenti. Si riportino in dettaglio tutti i passaggi.

Si utilizza il test del χ^2 . Il campione è costituito da 3000 soggetti di cui 1000 fumatori e 2000 non fumatori, 8 ammalati e 2992 non ammalati, con frequenze relative:

$$\begin{aligned}p_f &= 1000/3000 \quad \text{frequenza relativa di fumatori,} \\p_{nf} &= 2000/3000 \quad \text{frequenza relativa di non fumatori,} \\p_a &= 8/3000 \quad \text{frequenza relativa di ammalati,} \\p_{na} &= 2992/3000 \quad \text{frequenza relativa di non ammalati.}\end{aligned}$$

In base all'ipotesi H_0 di indipendenza abbiamo delle frequenze attese:

$$\begin{aligned}E_1 &= p_f p_a 3000 \simeq 2.67 \quad \text{numero atteso di fumatori ammalati,} \\E_2 &= p_{nf} p_a 3000 \simeq 5.33 \quad \text{numero atteso di non fumatori ammalati,} \\E_3 &= p_f p_{na} 3000 \simeq 997.33 \quad \text{numero atteso di fumatori non ammalati,} \\E_4 &= p_{nf} p_{na} 3000 \simeq 1994.67 \quad \text{numero atteso di non fumatori non ammalati.}\end{aligned}$$

Le corrispondenti frequenze effettivamente registrate sono invece quelle indicate nella tabella sopra riportata: $F_1 = 6$, $F_2 = 2$, $F_3 = 994$, $F_4 = 1998$. Il valore di χ^2 corrispondente al test è quindi

$$\chi_{\text{test}}^2 = \sum_{i=1}^4 \frac{(E_i - F_i)^2}{E_i} = \frac{(6 - 2.67)^2}{2.67} + \frac{(2 - 5.33)^2}{5.33} + \frac{(994 - 997.33)^2}{997.33} + \frac{(1998 - 1994.67)^2}{1994.67} \simeq 6.26.$$

Poiché dalla tabella della distribuzione di χ^2 a 1 grado di libertà risulta $\chi_{0.01,1}^2 \simeq 6.63 > 6.26$, non possiamo rifiutare, a questo livello di significatività, l'ipotesi di indipendenza fumo-cancro ai polmoni

5.2 Si dica se il risultato cambia fissando il livello di significatività al 5%.

Dobbiamo ora confrontare χ_{test}^2 con $\chi_{0.05,1}^2 \simeq 3.841 < 6.26$. Pertanto, a questo livello di significatività, possiamo rifiutare l'ipotesi che se una persona a caso contrae un tumore ai polmoni questo sia indipendente dal fatto che fumi o meno.

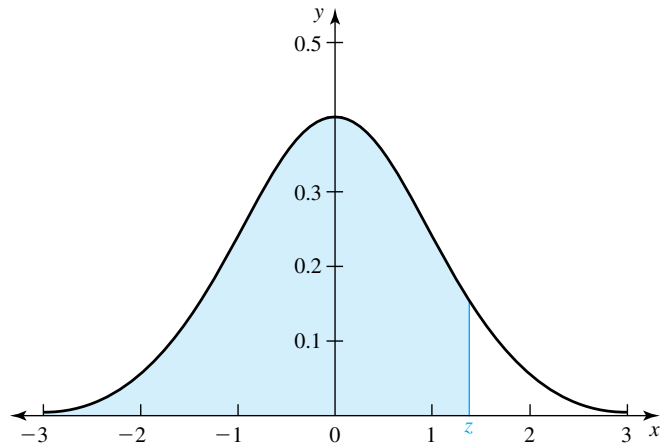
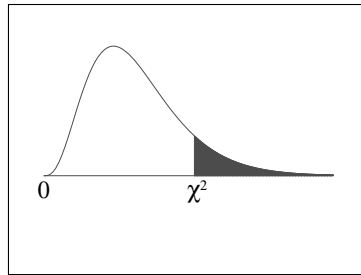


Figure B.1 Areas under the standard normal curve from $-\infty$ to z .

z	0	1	2	3	4	5	6	7	8	9
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5754
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7258	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7518	.7549
0.7	.7580	.7612	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7996	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986

Chi-Square Distribution Table



The shaded area is equal to α 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