Uno studente di laurea triennale raccoglie misure ripetute di temperatura di una scheda elettronica di lettura di uno strumento, ottenendo una serie di misure. Lo studente le raggruppa in classi di frequenza come in tabella. T_i è il valore centrale della i-esima classe, n_i la frequenza misurata per gli eventi nell'i-esima classe.

T_i (° C)	n_i
10	3
11	8
12	12+A
13	40
14	15
15	6

Lo studente ha intenzione di valutare la natura centrale di questa serie di misure, attraverso un test del chi^2 secondo l'ipotesi gaussiana. Confrontatevi con lo studente svolgendo il test e motivando tutti i passaggi e indicando le formule utilizzate. In particolare:

- 1.1 Calcolando il **numero totale di eventi,** la media aritmetica la deviazione standard campionaria della distribuzione.
- 1.2 Calcolando le **frequenze attese** per ciascuna classe nell'ipotesi che i dati seguano la distribuzione gaussiana e la **varianza** associata ad ogni frequenza.
- 1.3 Calcolando il chi^2 misurato relativo all'ipotesi di cui sopra e si discuta il test con con un margine di confidenza del 99%. *Si usi la tabella sotto come riferimento riportando: numero di gradi di libertà, singoli termini del chi^2 e valore complessivo, valore di riferimento secondo il livello di confidenza e si discuta il confronto con le ipotesi*

N.B Alcune frequenze hanno un errore che non si può a rigore considerare distribuito in modo gaussiano. Quali ? Essendo occorrenze al di sotto del 10% degli eventi complessivi, si trascuri questo fatto nel computo del chi^2.

1.1 eventi / slot [A = 1]

Numero totale degli eventi è $\boxed{N = \sum_{i=1}^{6} n_i = 85}$

La media è $\boxed{\bar{T} = \sum_{i=1}^{6} \frac{T_i n_i}{N} = 12.8\,°C}$

La deviazione standard del campione [o della singola misura nel campione] è

$$\boxed{S_T = \sqrt{\frac{1}{N-1}\sum_{i=1}^{6}(T_i-\bar{T})^2 n_i} = 1.14\,°C}$$

e quella della media

$$\boxed{S_{\bar{T}} = \frac{S_T}{\sqrt{N}} = \frac{1.14}{\sqrt{6}} \simeq 0.465\,°C}$$

1.2 Frequenze attese

Per ogni valore, considerando ipotesi di andamento gaussiano, possiamo calcolare

$$p(T_i) = \frac{1}{\sqrt{2\pi}\,S_T}\, e^{-\frac{1}{2}\left(\frac{T_i-\bar{T}}{S_T}\right)^2}$$

e per un bin

$$p(bin_i) = \Delta T\, p(T_i) \quad \text{ma } \Delta T = 1\,°C$$

e considerando le FREQUENZE ATTESE

$$f(bin_i) = N\, p(bin_i)$$

quindi i valori attesi sono:

T_i	n_i	f_i
10	3	1.3
11	8	7.8
12	13	22.2
13	40	29.5
14	15	18.2
15	6	5.2
	85	84.2

→ contiene meno di 5 elementi ⇒ incertezza sarebbe non ben stimata ⇒ si potrebbe eliminare bin o raggruppare con il successivo, tuttavia, siccome contribuisce < 10% su N allora possiamo lasciare e usarlo così come è.

Le incertezze su f_i sono stimate con la statistica Binomiale. Usando Poisson vengo calcoli simili ma metodo è meno rigoroso.

$$\sigma_{f_i}^2 = N p_i (1 - p_i)$$

T_i	h_i	f_i	$\sigma_{f_i}^2$
10	3	1.3	1.24
11	8	7.8	7.06
12	13	22.2	16.40
13	40	29.5	19.26
14	15	18.2	14.31
15	6	5.2	4.90
	85	84.2	

Da questo calcoliamo $\chi_i^2 = \dfrac{(h_i - f_i)^2}{\sigma_{f_i}^2}$ ottenendo

T_i	h_i	f_i	$\sigma_{f_i}^2$	χ_i^2
10	3	1.3	1.24	2.43
11	8	7.8	7.06	0.01
12	13	22.2	16.40	5.17
13	40	29.5	19.26	5.72
14	15	18.2	14.31	0.72
15	6	5.2	4.90	0.12
	85	84.2		14.2

$\chi_m^2 = 14.2$

$NDOF = N - \nu = 6 - 3 = 3$

$\downarrow$

$\bar{x}, S_{\bar{x}}, N_{bin}$

$\chi_0^2 [NDof = 3, \alpha = 0.01] = 11.35$

Giudizio: $\chi_m^2 > \chi_0^2$ allora ipotesi di "gaussianità" dei dati è scartata con un margine di confidenza superiore al 99%.

NOTA: per A in $[0, 4]$ HP_0 SCARTATA

A in $[5, 9]$ HP_0 ACCETTATA

Ad un dottorando viene chiesto di valutare la compatibilità' tra la performance in termini di trasmittanza T (*) di fibre ottiche che andranno a sostituire 5 fibre usurate in un esperimento. L'esperimento ne contiene in totale 900 e, quando furono installate la prima volta, la trasmittanza era stata misurata per tutte le fibre ottenendo un valore medio $\langle T_1 \rangle = 0.9881 \pm 0.0004$.

Ora, sono state acquistate 7 fibre (2 in più' per sicurezza). Lo studente prepara il sistema di misura e ottiene come valori in questo ordine (sostituire la cifra A nel primo dato).

$$0.977+0.001\text{x}A, \quad 0.983, \quad 0.988, \quad 0.991, \quad 0.992, \quad 0.991, \quad 0.993$$

Lo studente vuole valutare con un margine di significanza del 99.5% la compatibilita delle nuove fibre con quelle precedenti per garantire che la performance dello strumento non cambi significativamente. Lo studente decide di applicare un test di Student.

Si proceda come segue.

- 2.1 Calcolare media $\langle T_2 \rangle$ e errore della media della serie di misura di trasmittanza effettuate sulle nuove fibre.

Nel seguito, per il confronto con la ipotesi, riteniamo trascurabile l'incertezza su $\langle T_1 \rangle$ rispetto a quella su $\langle T_2 \rangle$.

- 2.2 Applicare il testi di t-Student per valutare se le 7 fibre nuove hanno la stessa trasmittanza delle precedenti. In particolare si riporti se il test è a una o due code, quanti sono i gradi di libertà, come si calcola formalmente la variabile t e qual e' il suo valore, quale è il valore di riferimento con il livello di confidenza del 99.5%, Infine se il test è accettato o rifiutato. Si faccia uso della tabella riportata sotto.

- 2.3 [facoltativa] Era giustificabile trascurare l'incertezza su $\langle T_1 \rangle$ rispetto a quella su $\langle T_2 \rangle$.

- 2.4 [facoltativa] Osservando l'andamento temporale delle serie di misure, che sospetto può' avere lo studente? Discutere a parole un possibile test di controllo.

(*) rapporto tra l'intensità luminosa trasmessa rispetto a quella entrante nella fibra

Risolvo per $A=4$.

Anzitutto ricalcoliamo i T_i : [0.981, 0.983, 0.988, 0.991, 0.992, 0.991, 0.993]

- Da questo:

$$N = 7$$

$$\overline{T}_2 = \sum_{i=1}^{7} \frac{T_i}{7} = 0.9884$$

$$S_{T_2} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{7} (T_i - \overline{T})^2} = 0.0047$$

$$S_{\overline{T}_2} = S_{T_2}/\sqrt{7} = 0.0018 \Rightarrow T_2 = 0.9884 \pm 0.0018$$

- T_1 ha errore considerato trascurabile
- Imposto TEST DI COMPATIBILITA' ASSOLUTA, ⇒ A 2 CODE

$$t_m = \frac{\overline{T}_2 - T_1}{S_{\overline{T}_2}} = 0.19$$

- Per il test $NDOF = N - 1 = 6$, $t^\circ[2\text{ code}, NDOF = 6, \alpha = 0.01] = 4.032$

$\downarrow$ N misure

- giudizio $t_m < t_0$ allora accetto ipotesi con un livello di significanza superiore al 1%.

NOTA: se si fosse considerato che T_1 ottenuto con 900 misure allora

$$t = \frac{\overline{T}_1 - \overline{T}_2}{\sqrt{\frac{1}{N} + \frac{1}{M}} \sqrt{\frac{1}{N+M-2}\left[\sum (T_i - \overline{T}_1)^2 + \sum (T_{2i} - \overline{T}_2)^2\right]}}$$

e si sarebbe ottenuto giudizio simile.

Nel mese di aprile a Padova si verificano in media 2 eventi piovosi a settimana (per semplicità si considerino gli eventi scorrelati, casuali e molto brevi). Si risponda alle seguenti domande indicando oltre ai valori numerici, procedimenti, nome e formule delle statistiche utilizzate.

- 3.1 Che probabilità c'è' che non piova per 10 giorni consecutivi ?

- 3.2 Che probabilità c'è' di avere un giorno di pioggia in aprile ? (ovvero un giorno con almeno un evento piovoso)

- 3.3 Quanti giorni di pioggia (ovvero con almeno un evento piovoso) ci aspettiamo in media nel mese di aprile.

- 3.4 Che probabilità c'e' di avere esattamente A+3 giorni di pioggia in aprile.

3.1

Il numero medio di eventi piovosi al giorno è

$$\lambda = \frac{2 \text{ eventi/giorno}}{7 \text{ giorni}} = 0.286$$

mentre n° medio eventi in 10 giorni è

$$\lambda^{10} = 10\lambda = 2.86 \frac{\text{eventi piovosi}}{10 \text{ giorni}}$$

Evento è di tipo Poisson, quindi probabilità che NON PIOVA

$$P(0;\lambda^{10}) = \frac{(\lambda^{10})^0 e^{-\lambda^{10}}}{0!} = 0.0574 \simeq 5.74\%$$

3.2

$$\lambda = 0.286 \frac{\text{eventi}}{\text{giorno}}$$

$$P(1+;\lambda) = 1 - P(0;\lambda) = 1 - \frac{\lambda^0 e^{-\lambda}}{0!} = 0.2485 \simeq 24.85\%$$

3.3 Giorni con eventi piovosi in aprile

In un mese [APRILE] $p^* = 30 \cdot P(1+;\lambda) = 7.46$ giorni

3.4 prob di A+3 giorni di pioggia in Aprile

Bernoulli:

$$P(30; A+3; p^*) = \binom{30}{A+3} p^{*(A+3)} (1-p^*)^{(30-A-3)}$$

per $A = 4$ $\quad p = 16.68\%$

Attenzione: n° eventi piovosi $\neq$ n° giorni di pioggia