Approfondimenti di Statistica e Probabilità

1. Richiami

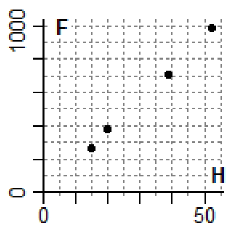
Nel punto **3b** del "**ripasso**" trovi come rivedere, sia sintetizzati che svolti per esteso, i principali argomenti di statistica e calcolo della probabilità finora affrontati. In questa scheda riprenderemo nel primo paragrafo alcuni concetti utili per studiare la relazione tra due variabili aleatorie e, nei successivi, svolgeremo alcuni ***approfondimenti*** per alcuni tipi di scuole.

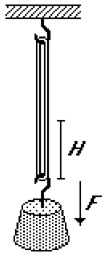
2. Relazioni tra variabili casuali

Rivediamo, rapidamente, su uno stesso esempio, alcuni concetti affrontati nella scheda Sistemi di variabili casuali.

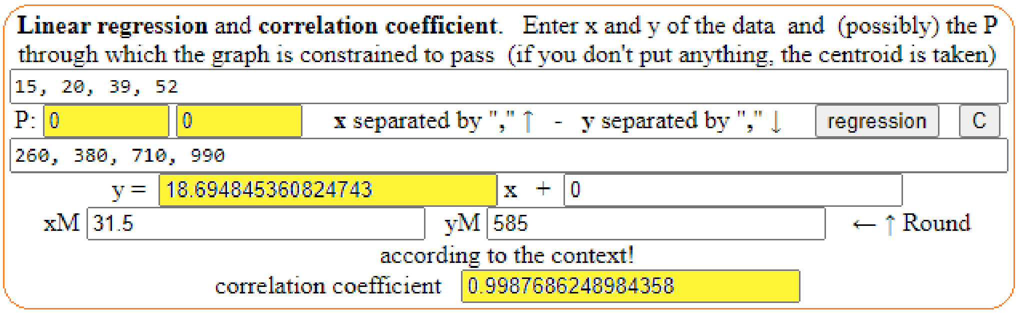
Per studiare le *caratteristiche di un elastico* lo tengo sospeso per un estremo e appendo all'altro diversi oggetti. Ogni volta misuro il peso dell'oggetto e il corrispondente allungamento dell'elastico. Il peso *F* degli oggetti lo misuro con una bilancia a molla con divisioni di 10 g (in modo che se l'ago si ferma vicino alla tacca 260 assumo che il peso sia 260±5 g). Le lunghezze dell'elastico sono misurate con la precisione di 1 mm, in modo che i valori dell'allungamento *H* (ottenuti come differenza di due lunghezze) hanno la precisione di 2 mm. Ottengo i valori riportati nella tabella a fianco, in cui il l'allungamento è espresso in millimetri e il peso in grammi.

H	*F*
15	260
20	380
39	710
52	990

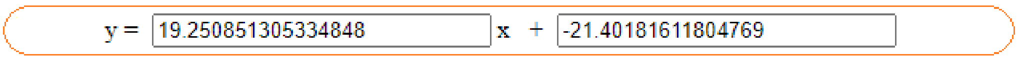




I valori sono stati rappresentati anche su un grafico, con dei pallini. Evidentemente H ed F sono *correlate,* ossia al variare dell'una anche l'altra tende a variare più o meno proporzionalmente. Per misurare quanto i punti che rappresentano le due variabili casuali tendono a disporsi lungo una retta obliqua (passante o no per l'origine) si usa il ***coefficiente di correlazione***, che è vicino a 0 se le variabili sono poco correlate, si avvicina ad 1 se i punti sono quasi allineati lungo una retta con pendenza positiva, si avvicina a −1 se sono quasi allineati lungo una retta con pendenza negativa. Nel nostro caso, facendo i calcoli con lo script **RegCorr**:



Lo script individua anche la retta che "meglio" approssima i dati, di cui richiameremo il signficato tra poco. Qui ho imposto che passi per (0,0). Se non imponevo che passasse per un particolare punto ottenevo un'altra retta, ma il coefficiente di correlazione rimaneva lo stesso:

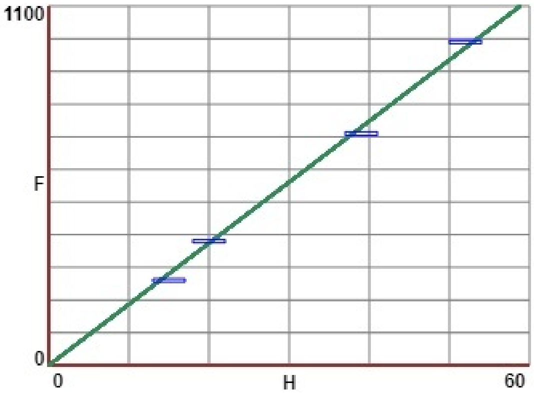


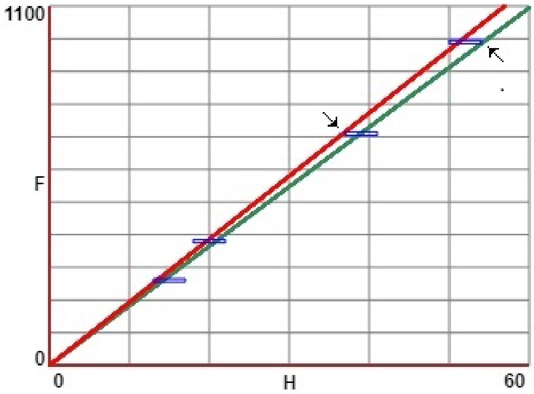
Ottengo, in questo caso particolare, un coefficente di correlazione (0.9987686) molto vicino ad 1. Non importa l'ordine con cui metto le due variabili: il valore è sempre lo stesso.

Ma studiare la relazione tra H ed F, in questa situazione, in cui evidentemente le due grandezze sono una dipendente dall'altra, non può limitarsi ad individuare il coefficiente di correlazione. Può essere utile cercare una funzione che esprima la relazione che c'è tra H ed F. Posso cercare di trovare la retta che approssima i punti rendendo ***minima*** *la somma dei* ***quadrati*** *degli scarti* tra i valori sperimentali di F e quelli che sarebbero stati associati ai corrispondenti valori di H dalla equazione della retta (questo metodo è chiamato dei *minimi quadrati*). Una retta ottenuta in questo modo, avendo o no fissato un punto attraverso cui passi, viene chiamata ***retta di regressione*** e il suo coefficiente direttivo è chiamato *coefficiente di regressione.* È quello che fa lo script appena visto.

Nel caso precedente, avendo imposto il passaggio per (0,0), il coefficiente di regressione è (arrotondando) 18.7.

Sotto a sinistra la rappresentazione della retta di regressione e dei punti sperimentali (H,F). In questo caso essi non li ho rappresentati con dei pallini in quanto di H e di F non conosco solo dei valori approssimativi ma conosco le misure approssimate e le relative precisioni, di 2 (mm) e di 5 (g): quindi ho rappresentato correttamente (sotto a sinistra) in punti sperimentali con dei rettangolini di base 4 e altezza 10.





Ma la retta è stata tracciata tenendo conto solo dei valori di H e di F, senza utilizzare le informazioni sulle loro precisioni! Devo cercare le rette passanti per l'origine e per i tutti i rettangolini di minima e di massima pendenza, come fatto nella figura sopra a destra. La prima ha pendenza (710+5)/(39-2) = 715/37 = 19.32432432..., la seconda (990-5)/(52+2) = 985/54 = 18.24074074... Quindi la relazione è $F = k \cdot H$ (F in g, H in mm) con k tra 18.24 e 19.33, ovvero tra 18.2 e 19.4, ovvero $k = 18.8 \pm 0.6$.

Qui e **qui** puoi vedere come sono stati realizzari i grafici precedenti.

Se non avessimo avuto informazioni sulle precisioni (ovvero avessimo rappresentato dei dati con dei pallini) alla retta di regressione avremmo dovuto comunque associare una precisione che tenesse conto sia dei valori dei dati che della loro quantità. Vi sono degli oggetti matematici (gli "intervalli di confidenza") che consentono di fare queste valutazioni, ma che (negli studi preuniversitari) non siamo in grado di affrontare.

Comunque, per dare un'idea di che cosa si tratta, vediamo (a "scatola nera") che cosa si potrebbe fare se (con la grande CT) calcolassimo i 4 rapporti F/H (260/15 = 17.333, 380/20 = 19, 710/39 = 18.205, 990/52 = 19.038), la loro media (18.394) e la loro standard deviation sperimentale (17.333, 19, 18.205, 19.038 → 0.804863) e in *WolframAlpha* cercassimo *confidence interval.* Mettendo *confidence level* = 0.9 (vedremo fra poco che cos'è) otteniamo:

input: confidence level 0.9
sample size 4
sample standard deviation 0.804863
sample mean 18.394
output: **17.73 to 19.06** (18.394 ± 0.661941)

L'intervallo ottenuto, [17.73, 19.06], è un ***intervallo di confidenza*** al 90%. In esso, con tale probabilità, cade il valore del rapporto F / H.
Sottolineamo che questo non è un intervallo certo, come invece lo è l'intervallo [18.2, 19.4] trovato disponendo dei 4 intervalli di indeterminazione.

Qui abbiamo visto solo come approssimare coppie di dati con funzioni lineari. Se i dati hanno andamento diverso occorre approssimarli con funzioni di tipo diverso, ma non disponiamo degli strumenti matematici per affrontare questo studio. Nel paragrafo successivo c'è qualche cenno al caso dei dati con andamento quadratico e cubico.

3. Relazioni non lineari

Nel paragrafo precedente abbiamo considerato solo il caso della approssimazione di dati mediante funzioni lineari. Nel caso di altri andamenti esistono varie tecniche, che dipendono dai modi in cui sono stati raccolti i dati di cui si dispone, dalle altre informazioni che abbiamo sulle funzioni che vogliamo li approssimino, …. Piuttosto che imparare alcune ricette è bene, quando si hanno una serie di dati da analizzare, rivolgersi a chi ha esperienze e conoscenze sull'argomento. L'obiettivo di questo paragrafo è solo quello di dare un'idea delle tecniche che si possono impiegare.

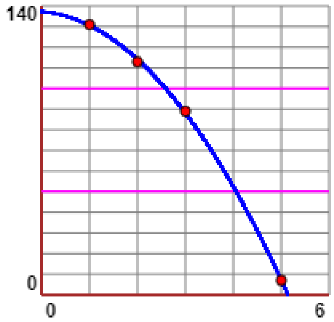
Consideriamo un esempio semplice, che, comunque, si riferisce a situazioni abbastanza diffuse.
Un oggetto, pesante e di forma compatta, viene lasciato cadere e ne viene misurata, mediante una successione di immagini fotografiche scattate ogni decimo di secondo, l'altezza in cm da terra. Supponiamo che in corrispondenza dei tempi di 1, 2, 3, 5 decimi di secondo (rilevati con errori trascurabili) si registrino, in ordine, le altezze da terra di 131, 113, 89 e 7 cm, arrotondate tutte con la stessa precisione, ad es. di 1 cm.

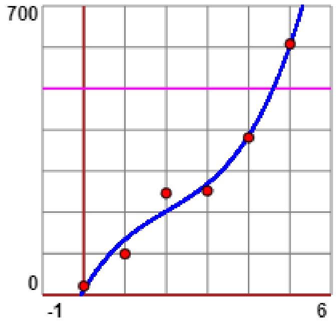
In situazioni di questo genere, in cui si conoscono le coordinate di N punti sperimentali con ascisse x_i note esattamente e con ordinate y_i della stessa indeterminazione, si può ricorrere ad un procedimento chiamato ***regressione polinomiale*** che generalizza le tecniche considerate per trovare le rette di regressione al caso della determinazione dei coefficienti di una funzione polinomiale che "con maggiore probabilità" approssima i punti noti. Noi consideremo solo i casi del grado 2 e del grado 3.

Nel caso del grado 2, senza entrare nei dettagli, si ottiene che si tratta della funzione $F: x \to A + B \cdot x + C \cdot x^2$ dove A, B e C sono le soluzioni del sistema seguente:

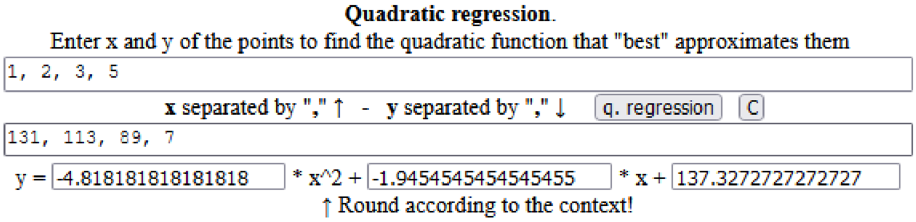
$$A \cdot N + B \cdot \Sigma x_i + C \cdot \Sigma x_i^2 = \Sigma y_i$$
$$A \cdot \Sigma x_i + B \cdot \Sigma x_i^2 + C \cdot \Sigma x_i^3 = \Sigma x_i y_i$$
$$A \cdot \Sigma x_i^2 + B \cdot \Sigma x_i^3 + C \cdot \Sigma x_i^4 = \Sigma x_i^2 y_i.$$

Tornando all'esempio, risolvendo il sistema si ottengono le soluzioni seguenti, a cui corrisponde la rappresentazione grafica sottostante, a sinistra: $A = 137.327$, $B = -1.9455$, $C = -4.81818$





Lo script che calcola la regressione quadratica è **Regr2**. Ecco qui sotto le uscite:



Con un procedimento simile, ma più complesso, si può trovare la funzione polinomiale di grado 3 che approssima i punti di ascisse 0, 1, 2, 3, 4, 5 e ordinate 21, 99, 246, 252, 381, 608, ottenendo la funzione $x \rightarrow A + B \cdot x + C \cdot x^2 + D \cdot x^3$ con $A = 11.36508$, $B = 166.9709$, $C = -53.05159$, $D = 8.675926$, rappresentata sopra a destra.

Lo script analogo che calcola la regressione cubica è **Regr3**.

4. Il test χ^2

In questa voce ci occuperemo di una delle varie tecniche per prendere delle decisioni sulla base di dati statistici raccolti sperimentalmente.

Valutata la probabilità di un evento o individuata una legge di distribuzione o … solo sulla base di dati sperimentali, non abbiamo la certezza di questa conclusione. Possiamo, comunque, porci il problema di quanto sia ***verosimile*** l'ipotesi che il valore o la funzione o … individuata sia effettivamente una buona approssimazione, cioè valutare la probabilità che il suo scarto dall'oggetto (valore, funzione, …) "vero" rientri nei margini di aleatorietà dovuta alla limitatezza del materiale statistico a disposizione.

In base a questa valutazione, a seconda della situazione (con considerazioni pratiche, legate al contesto, ai rischi sociali, …), potremo stabilire se tale ipotesi è accettabile o è da rifiutare.

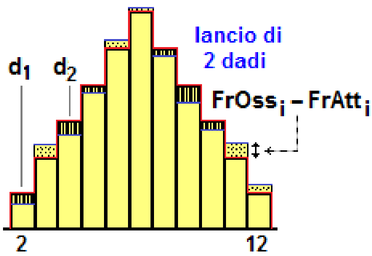
Abbiamo affrontato attività di questo genere quando abbiamo introdotto il teorema limite centrale: all'intervallo *frequenza* $\pm 3\sigma/\sqrt{n}$ associo la *probabilità di confidenza* del 99.7% che esso contenga la probabilità Pr(A) cercata; all'intervallo *frequenza* $\pm t\sigma/\sqrt{n}$ con $t < 3$ posso associare una probabilità di confidenza più piccola; in altre parole valuto la probabilità che tale l'intervallo sia veramente un intervallo di indeterminazione per Pr(A).

Se in base a qualche ragionamento ho *ipotizzato* che Pr(A) sia 3/7 e trovo che 3/7 sta nell'intervallo *frequenza* $\pm t\sigma/\sqrt{n}$, posso ritenere attendibile (con una opportuna confidenza) che lo scarto tra 3/7 e *frequenza* sia "normale", cioè ritenerlo non "significativo", non tale da giustificare il *rifiuto dell'ipotesi*. Questa terminologia (accettazione/rifiuto dell'ipotesi) in realtà non è fondata: la probabilità dell'ipotesi che Pr(A) sia "esattamente" 3/7 è nulla. Ma è un modo di esprimersi "convenzionale" ormai diffuso.

In questo paragrafo vedremo in particolare come valutare l'attendibilità di una legge di distribuzione o, meglio, di *valutare la conformità tra una distribuzione sperimentale e una teorica.*

Come valutare la discordanza tra gli esiti di ***n*** prove, classificati in ***nc*** classi, e una certa legge di distribuzione?

Supponiamo di voler confrontare con la distribuzione $Pr(U=2) = 1/36$, $Pr(U=3) = 2/36$, … gli esiti del lancio di una coppia di dadi ripetuto n volte. Indichiamo con $FrOs_1, \ldots, FrOs_{11}$ le *frequenze assolute* ***osservate*** ($FrOs_1 + FrOs_{11} = n$) e con $FrAt_1, \ldots, FrAt_{11}$ le *frequenze assolute* ***attese***, cioè i valori $FrAt_1 = n \cdot 1/36$, $FrAt_2 = n \cdot 2/36$, …



Una idea è prendere $\Sigma_i (FrOs_i - FrAt_i)^2 / FrAt_i$:

- l'elevamento al quadrato fa sì che differenze positive (zone punteggiate nella figura a lato) e quelle negative (zone tratteggiate) non si compensino;
- la divisione per le frequenze attese fa sì che una differenza come $\mathbf{d_1}$ (nella figura a lato) pesi di più di una come d_2, che sarebbe leggermente maggiore ma relativa a dati molto più piccoli.

Tale valore viene indicato con χ^2 ("chi quadro"):

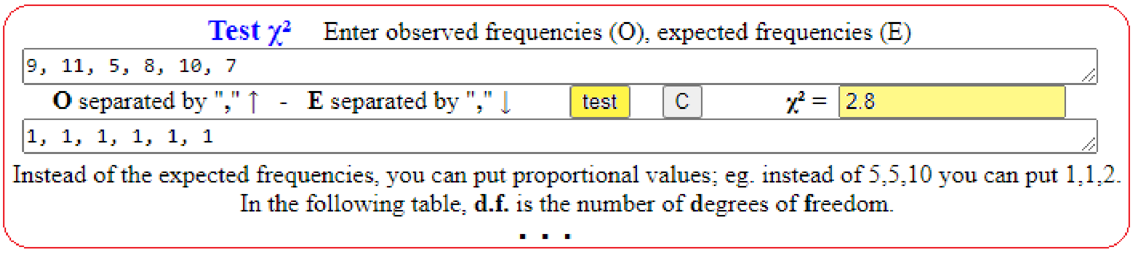
$$\chi^2 = \sum_{i=1}^{nc} \frac{(FrOs_i - FrAt_i)^2}{FrAt_i}$$

Lo stesso valore viene preso nel caso di un'altra variabile casuale. Cambia il modo in cui sono calcolate le frequenze attese nelle ***nc*** classi (intervallini o altri insiemi) I_i in cui ho classificato gli ***n*** dati; se la variabile casuale è continua sono calcolate integrando su I_i la funzione densità.

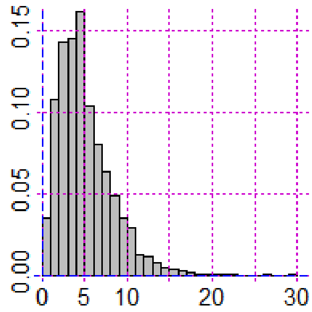
χ^2 è una variabile aleatoria (assume un valore diverso ogni volta che effettuo n prove) che, oltre che dal fenomeno studiato, dipende dal numero delle prove ***n***, dalle classi $I_1, \ldots, I_{nc}$ scelte e dalla legge di distribuzione teorica ***L*** di cui si vuole valutare la verosimiglianza, da cui dipendono le frequenze attese.

Consideriamo ad esempio un dado di cui si sono effettuati 50 lanci ottenendo 9 uno, 11 due, 5 tre, 8 quattro, 10 cinque e 7 sei. Per valutare la discordanza dalla distribuzione uniforme (distribuzione corrispondente ai dadi equi), calcoliamo il relativo χ^2. Poniamoci, poi, il problema di *individuare la distribuzione χ^2 teorica* nel caso in cui il *dado* sia realmente *equo*, in modo da poter confrontare con essa il χ^2 trovato e valutare la verosimiglianza dell'equità del nostro dado.

Potremmo svolgere il calcolo "a mano". Effettuiamolo mediante lo script Chi2. Otteniamo 2.8.



Effettuato questo calcolo, *che cosa possiamo concludere sulla equità del dado?*



Studiamo la distribuzione teorica χ^2, cioè come si distribuirebbe il valore di χ^2 se il dado fosse equo. Realizziamo questo studio *sperimentalmente,* con una simulazione dei 50 lanci. Ecco l'esito della generazione di 10000 valori, analizzato col software (a destra l'istogramma, sotto le uscite numeriche):

```
Min.   1st Qu.  Median   Mean   3rd Qu.  Max.
0.160   2.800    4.240   4.972   6.640  29.200
```

Posso osservare dall'istogramma che 2.8 è un valore abbastanza centrale. Esaminando le uscite numeriche osservo che 2.8 è il 25° percentile (o primo *quartile* o *quantile* di ordine 0.25), cioè il limite sinistro del 50% centrale dei dati.

Quindi posso ritenere *plausibile* (cioè *non rifiutare*) *l'* ***ipotesi*** che il dado sia equo. Se avessi ottenuto un valore verso la coda sinistra o quella destra avrei invece dovuto avere dubbi su tale ipotesi: se è verso la coda destra si tratta di un valore molto alto, che fa supporre che il dado non sia equo; se è verso la coda sinistra si tratta di un valore molto basso, ma un po' troppo "perfetto", che può far supporre che ci sia stato qualche errore (o qualche "imbroglio") nel riportare le frequenze. In tali casi, prima di scartare l'ipotesi, sarebbe stato opportuno, se possibile, ripetere i lanci, ricalcolare χ^2 e valutare la posizione del nuovo valore rispetto all'istogramma sopra riprodotto (o rispetto ai quantili).

Nel caso di un'altra legge di distribuzione L, altre classi o un altro numero n di prove , ad esempio nel caso dell'esempio iniziale, potrei procedere analogamente. In genere, tuttavia, si preferisce utilizzare un procedimento standard di tipo generale, che ha come retroterra il fatto che si può dimostrare che, se il numero n delle prove è sufficientemente grande, *la legge di distribuzione teorica di χ^2* è praticamente indipendente dalla legge di distribuzione L:

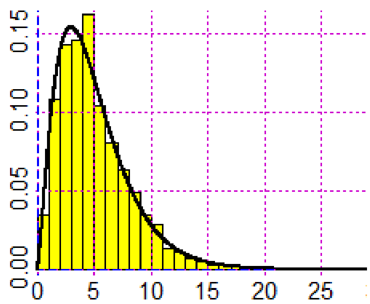
per n tendente all'infinito tende a una legge $\boldsymbol{\chi^2(r)}$ che dipende solo dal numero r dei ***gradi di libertà***, cioè dalla quantità delle frequenze sperimentali che devo conoscere direttamente.

Spieghiamo meglio il concetto di "grado di libertà" (nello script precedente compariva denominato in inglese, "**d**egrees of **f**reedom") con alcuni esempi.

Nel caso del **dado** sopra considerato, sappiamo che le frequenze $FrOs_1$, $FrOs_2$, …, $FrOs_6$ devono avere come somma n. Questa *connessione* fa sì che note **5** frequenze la rimanente sia determinata automaticamente. Questa connessione è presente in tutti i casi. Quindi i gradi di libertà sono in ogni caso al più $nc-1$ (nc = numero delle classi).

Nel caso di una variabile casuale a valori reali positivi, se $FrOs_1$,…, $FrOs_{16}$ sono le frequenze osservate nei 16 intervalli [0,5), [5,10), … e voglio operare il confronto con la densità esponenziale $x \to \mathbf{w}\cdot e^{-\mathbf{w}x}$ con **w** pari al reciproco della media sperimentale M, come connessione imposta, ulteriore rispetto a $\Sigma FrOs_i = n$, ho che $\Sigma x_i FrOs_i / n = M$ (x_i centro dell'intervallo i-esimo: x_1=2.5, x_2=7.5, …). I gradi di libertà sono 14. Note 14 tra le $FrOs_i$, ad es. $FrOs_1$, $FrOs_2$, …, $FrOs_{14}$, potrei determinare $FrOs_{15}$ e $FrOs_{16}$ risolvendo il sistema (avente $FrOs_{15}$ e $FrOs_{16}$ come incognite): $\Sigma FrOs_i = n$ AND $\Sigma x_i FrOs_i / n = M$.

Se, invece, come **w** avessi scelto un valore stabilito a priori, non dipendente dai dati sperimentali, non avrei avuto la connessione $\Sigma x_i FrOs_i / n = M$, e i gradi di libertà sarebbero stati 15.



L'espressione di $\chi^2(r)$ non è facile da descrivere né da comprendere (non è una delle cosiddette funzioni "elementari": vedi Complementi di analisi matematica). Comunque nel software matematico in genere è definita e richiamabile con degli opportuni comandi.

Ecco, a destra, la sovrapposizione all'istogramma ottenuto sopra per il dado equo del grafico della densità di $\chi^2(5)$.

Ecco, qua sotto, arrotondati, i valori del 5°, del 10°, …, del 95° percentile per diversi gradi di libertà, una parte di quelli che appaiono elencati azionando lo script precedente:

d.f.	5	10	25	50	75	90	95
1	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84
2	0.103	0.211	0.575	1.39	2.77	4.61	5.99
3	0.352	0.584	1.21	2.37	4.11	6.25	7.81
4	0.711	1.06	1.92	3.36	5.39	7.78	9.49
5	1.15	1.61	2.67	4.35	6.63	9.24	11.1
6	1.64	2.20	3.45	5.35	7.84	10.6	12.6

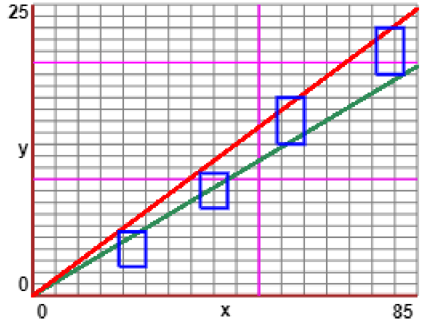
8	2.73	3.49	5.07	7.34	10.2	13.4	15.5
9	3.33	4.17	5.90	8.34	11.4	14.7	16.9
10	3.94	4.87	6.74	9.34	12.5	16.0	18.3
15	7.26	8.55	11.0	14.3	18.2	22.3	25.0
20	10.9	12.4	15.5	19.3	23.8	28.4	31.4
30	18.5	20.6	24.5	29.3	34.8	40.3	43.8
50	34.8	37.7	42.9	49.3	56.3	63.2	67.5
75	56.1	59.8	66.4	74.3	82.9	91.1	96.2
100	77.9	82.4	90.1	99.3	109	118	124

Tornando al nostro dado (5 gradi di libertà), se come χ^2 invece di 2.8 (che è intorno al 25° percentile e quindi è abbastanza "normale") avessimo ottenuto 13 avremmo dovuto manifestare qualche dubbio sul fatto che il dado sia equo: c'è una discordanza molto alta rispetto alla legge uniforme: il 95° percentile è 11.1.

Ma anche se avessimo ottenuto una discordanza molto bassa, ad esempio $\chi^2<1$, avremmo dovuto avere dei dubbi sulla equità del dado o sulla attendibilità dei dati fornitici: è improbabile che si ottenga un valore inferiore a 1 (la probabilità è inferiore al 4%, infatti `pchisq(1, 5)` = 0.037).

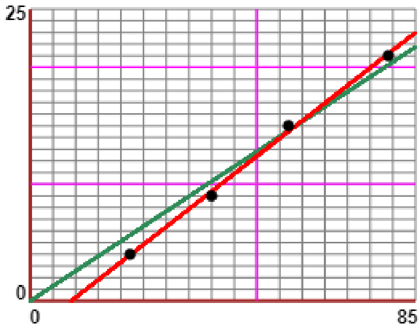
Nell'usare la distribuzione del χ^2 limite occorre prestare qualche attenzione: occorre che le prove siano numerose (diciamo, almeno un centinaio); occorre, inoltre, che in ogni classe cadano abbastanza valori (diciamo, almeno 5); se in qualche classe cadono poche osservazioni è opportuno unire questa classe ad un'altra.

Accanto a questi usi del test χ^2 esitono altri impieghi dei test statistici. **Qui** puoi vederne qualche esempio.



5. Esercizi

e1 A lato sono rappresentati i dati relativi allo studio sperimentale del legame tra due grandezze x e y che sappiamo essere direttamente proporzionali. Le misure sono effettuate con strumenti a bassa sensibilità. Le basi dei rettangolini rappresentano gli intervalli di indeterminazione per le misure di x, le altezze per quelle di y. Che cosa rappresentano le rette che vengono tracciate? Che cosa posso concludere (utilizzando solo il grafico) sul valore del coefficiente di proporzionalità? (motiva la risposta; tieni presente che gli intervalli delle x hanno estremi espressi in unità e che quelli delle y hanno estremi espressi in mezze unità)



e2 I pallini nel grafico qui a sinistra (x: 22, 40, 57, 79; y: 4, 9, 15, 21) rappresentano gli esiti dello studio sperimentale del legame tra due grandezze x e y che sappiamo essere legate da una relazione lineare. Trova le equazioni delle due rette di regressione tracciate, quella vincolata a passare per l'origine e quella libera.

e3 Considera i dati x e y dell'esercizio precedente, che sono anche i valori centrali degli intervalli che rappresentano le approssimazioni dell'esercizio 1. Determina il coefficiente di regressione tra x ed y.

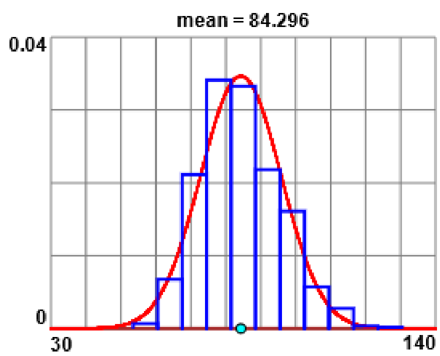
e4 In un esperimento sulla crescita del grano durante l'inverno si registrano la temperatura media (in °C) del suolo ad una profondità di 8 cm e i giorni necessari per la germogliazione. Qual è la correlazione tra temperatura del suolo e tempo necessario per la germogliazione?

`T: 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5`
`g: 40, 36, 32, 27, 23, 19, 19, 20`

e5 `x:  0.1,  0.9,  1.6,  2.0,  2.9` e
`y: -29.9,-11.4,-7.1, -3.8, 10.2`
sono i dati relativi ad una grandezza (y) in funzione di un'altra (x). Determina, mediante regressioni polinomiali, una funzione polinomiale di 2° grado ed una di 3° grado che approssimino i dati.

e6 Un amico mi dice: *Questa moneta è equa. Infatti su 1000 lanci ho ottenuto 499 "testa" e 501 "croce".* Che cosa posso concludere sulla verosimiglianza di quanto raccontato dall'amico?

e7 Un amico mi dice: *Questa moneta è equa. Infatti su 1000 lanci ho ottenuto 470 "testa" e 530 "croce".* Come valuti la conclusione di questo amico?



e8 Un'indagine sul peso dei maschi tra i 45 e i 55 anni di un certo stato dà gli esiti grafici rappresentati graficamente a lato e quelli numerici riportati qui sotto:
median=85 − 1º,3º quartile: 78 92 − mean=84.296 − experim. standard dev. = 11.512
All'istogramma (costituito da **11** colonne) è stato sovrapposto il grafico della gaussiana (di **media 84.296** e **sigma** 11.512). Utilizzando i dati sperimentali e quelli ottenuti integrando la gaussiana negli intervallini considerati nell'istogramma si è ottenuto 27.1 come **chi quadro**.

Perché i gradi di libertà sono 8?

È da accettare l'ipotesi che la distribuzione sia gaussiana?

[se interessato, vedi QUI elaborazioni grafiche e numeriche]

1) Segna con l'evidenziatore, nelle parti della scheda indicate, frasi e/o formule che descrivono il significato dei seguenti termini:

coeffciente di correlazione (§2) *minimi quadrati* (§2) *retta di regressione* (§2) *intervallo di confidenza* (§2) *regressione polinomiale* (§3) χ^2 (§4) *gradi di libertà* (§4)

2) Su un foglio da "quadernone", nella prima facciata, esemplifica l'uso di ciascuno dei concetti sopra elencati mediante una frase in cui esso venga impiegato.

3) Nella seconda facciata riassumi in modo discorsivo (senza formule, come in una descrizione "al telefono") il contenuto della scheda (non fare un elenco di argomenti, ma cerca di far capire il "filo del discorso").

script: piccola CT grande CT isto isto con % boxplot striscia 100 ordina Grafici GraficD divisori Indet distanza Triang eq.polinomiale eq.nonPolin sistemaLin moltPolin sempliciEq divisori fraz/mcd opFraz SumPro sin LenArc Poligono Circ3P Inscr3P Istogramma RandomNum IntegrPol IntGauss Integral AB3dim distOnSphere TabFun Deriv Det3 DaTabella RegCorr Regr2 Regr3 Chi2 ALTRO WolframAlpha