

Lezione del 28/02/05

Due parametri fondamentali

La media è data dalla somma dei diversi valori diviso il numero dei casi che consideriamo. Ad esempio nelle età è la somma delle diverse età degli appartenenti al gruppo diviso il numero degli individui che compongono il gruppo. Questa è la definizione classica di media.

La media serve per dare una misura globale di un fenomeno fatto di tanti valori. Se do l'età media del mio campione, con 1 numero solo do una informazione (non tutte) che riguarda l'età dei diversi componenti.

È un parametro descrittivo che sintetizza in un solo valore avente certe caratteristiche i molti valori del mio campione. È un parametro di sintesi. (Media = \bar{x} soprassegnato)

$$\bar{x} = \sum_{1}^{n} \frac{x_i}{n}$$

La varianza esprime la variabilità attorno a questa media. Se dico che i miei studenti hanno in media 20 anni do una informazione sintetica ma non sufficiente ad avere un quadro reale del mio campione perché 20 anni potrebbe essere la media di persone che hanno età molto diverse tra loro. Se invece il mio gruppo è formato da persone che hanno una età compresa fra i 19 e i 21, hanno sempre 20 anni di media ma con una variabilità attorno a questi valori molto ridotta. Questo è il ruolo della varianza, ci dice quanto si discostano i valori in media dalla media. Questa è la definizione discorsiva della varianza e ne è anche la definizione più pregnante. Una formula che permette di calcolare la varianza dall'insieme dei valori è data dalla somma dei quadrati degli scarti divisi (n-1). Questa è la definizione di varianza di un campione sperimentale.

$$s^2 = \sum_{1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

Perché al quadrato? Perché quando calcolo la somma degli scarti dalla media questa è sempre identicamente uguale a zero

Se sommo solo gli scarti ottengo sempre zero, perché i valori positivi si compensano con i negativi. Al quadrato riesco a calcolare il modulo, cioè lo scarto con segno sempre positivo. Qui la dimensione è al quadrato. Se calcolo la media il risultato sarà espresso in anni, la varianza è in anni al quadrato, non è una dimensione semplice.

Perché non calcoliamo il modulo? In questo modo avremmo una dimensione alla prima potenza e non al quadrato. Si calcola il quadrato perché è facile da gestire matematicamente mentre il modulo non è una funzione lineare, non segue molte delle regole matematiche, otterrei sempre un numero positivo ma meno gestibile.

Perché n-1? Siccome la somma degli scarti è = 0 ciò significa che su n scarti uno di questi non ho bisogno di conoscerlo, l'ennesimo numero - non importa che sia il primo o l'ultimo- lo riesco a trovare anche senza che mi venga detto. Sarà quel valore che farà tornare a zero la cosa, è già determinato dal fatto che deve dare somma = 0. Questa numerosità che rappresenta il numero dei valori che possono essere stabiliti arbitrariamente, (l'ultimo non può essere stabilito arbitrariamente), cioè il numero dei valori che non ha vincoli, viene chiamato gradi di libertà. Molti manuali riportano la varianza divisa per n e non per (n-1). C'è grande confusione, non conosce bene la statistica chi divide per n.

19	19-20= -1	1
20	20-20= 0	0
20	20-20= 0	0
22	22-20= 2	4
19	19-20= -1	1
Media delle età = 20	Scarti= 0	Scarti al quadr.=6

Dividendo per (n-1) - i valori erano 5- allora divido per (5-1), quindi ho 6:4, la varianza quindi è = 1,5.

La media dà una informazione sintetica dell'età, da sola però non mi dice abbastanza, non so se sono tutti più o meno lì con l'età, se calcolo gli scarti mi dicono quanto oscillano i valori rispetto alla media. La varianza calcola per ogni scarto il quadrato, somma tutti gli scarti (sigma maiuscolo rappresenta la somma di tutti i valori compresi nel mio campione) e li divide per i valori meno 1.

La deviazione standard è uguale alla radice quadrata della varianza ed ha una dimensione alla prima potenza, quindi può essere confrontata con la media.

Per evitare confusione metodologica è stata proposta un' altra definizione sia per la media sia per la varianza. Nel caso discreto la media è definito valore atteso, ed è data dalla somma del prodotto di ciascun valore per la sua probabilità.

$$E (x) = \sum_{1}^{n} x_i \cdot p (x_i)$$

Se ho 5 casi, ogni valore ha 1/5 di probabilità quindi la somma dei valori per la loro probabilità corrisponde a moltiplicare ciascun valore per 1/5 o a sommare tutti i valori divisi per 5. Se utilizzo la prima definizione di media o questa, ottengo sempre lo stesso risultato. Se n è la dimensione del campione la probabilità di ogni valore è 1 ennesimo. Se in un campione tutti i casi sono equiprobabili. Sommare tutti i valori e dividerli per 5 o sommare ogni valore diviso per 1/5 dà lo stesso valore, Nel nostro caso, riferendoci alle età precedenti, il risultato è 20. In molti manuali viene chiamato valore atteso invece di media.

19*1/5	3.8
20*1/5	4
20*1/5	4
22*1/5	4.4
19*1/5	3.8
	20

Sono il prodotto della singola età x la sua probabilità.

La varianza è definita come il valore atteso dei quadrati degli scarti fra i valori e il loro valore atteso, il valore atteso dei valori è la media, lo scarto fra il valore e il suo valore atteso è lo scarto fra il valore e la media, elevato al quadrato, mi dà scarto al quadrato, il valore atteso dello scarto al quadrato mi dà la varianza.

Qui non c'è (n-1). Gli scarti hanno (n-1) gradi di libertà, cioè su 5 scarti 4 sono casuali ed equiprobabili, il 5° non è equiprobabile ed il suo valore è fisso. La probabilità dei 5 scarti non è 1/5 x ciascuno. Nel calcolo del valore atteso prendo in considerazione la probabilità non come 1/5 ma come ¼, ho 4 valori casuali.

Il valore atteso degli scarti è ottenuto moltiplicando ciascun scarto per ¼.

Usare la formula precedente come somma dei valori diviso n , o usare quella di valore atteso = alla somma dei singoli valori moltiplicati per loro probabilità, mi dà lo stesso risultato, per la media. Lo stesso ottengo se moltiplico la somma degli scarti al quadrato per la loro probabilità, per trovare la varianza. (in excel ci sono queste formule)

Le formule viste funzionano a calcolare la media e varianza di 1 campione sperimentale, queste formule non vanno bene per calcolare medie e varianze di valori teorici. Se voglio calcolare qual è il punteggio medio che ottengo nel lancio di un dado, posso ottenere 1 2 3 4 5 6. Sommo quindi $1+2+3+4+5+6 = 21$ che fa 3,5 se divido per 6.

Il problema nasce con la varianza, la somma degli scarti qui va divisa per 6. Se i valori sono teorici la varianza va divisa per n e non per $(n-1)$. La definizione di valore atteso toglie questo problema perché in una distribuzione teorica le probabilità dei diversi valori sono già definite teoricamente, non devo preoccuparmi se i valori sono sperimentali o teorici, nel caso sperimentale la probabilità è già $n-1$, nel caso teorico è già n .

Il calcolo della media e della varianza viene effettuato di solito su dei valori sperimentali, in genere si utilizzano le prime formule viste, il caso teorico non è di solito oggetto del nostro studio, c'è 1 solo caso –che vedremo– in cui studiamo la varianza calcolata su valori teorici, sennò noi parliamo sempre di varianza calcolata su casi sperimentali.

Esiste una legge importantissima per la probabilità, la **legge dei grandi numeri**, che dice: al crescere del numero delle prove, la probabilità che lo scarto in valore assoluto (in modulo) fra la probabilità dell'evento e la sua frequenza è piccolo a piacere; questa probabilità tende a 1 cioè non è vero che al crescere delle prove la probabilità e la frequenza tendono a coincidere. La probabilità che lo scarto tra probabilità dell'evento e sua frequenza diventi piccolissimo al crescere del numero delle prove tende a 1, ciò non vuol dire che diventa certezza, vuol dire che si avvicina a 1 la probabilità che lo scarto sia piccolissimo.

Gli errori dei giocatori sono 2: dicono che lo scarto è 0 e che la probabilità e la frequenza coincidono, cioè che la probabilità diventa 1, l'evento invece è molto probabile ma non è certo. La legge dice solo che al crescere delle prove la probabilità e la frequenza in qualche modo si avvicinano, possono anche coincidere, ma non è certo, è però alta la probabilità che ad es. nel lancio della moneta, il numero delle teste e il numero delle croci siano al 50%. È sbagliato quindi giocare sui numeri in ritardo da molto (es. 53)

Secondo la definizione frequentista la probabilità in un grande numero di eventi coincide con la frequenza. È in contrasto con la legge dei grandi numeri.

Tutte le formule che vedremo valgono se vale questa legge, sennò la maggior parte della statistica salta. Nessuno vieta di costruire una statistica che si basa su un'altra legge.

In realtà questa legge è non solo un fondamento della statistica, è anche abbastanza verificata in pratica, il 53 recente lo prova. La coincidenza fra probabilità e frequenza non è un legame stretto ma elastico, può funzionare in un arco di tempo anche molto lungo, non è un legame rigido, c'è un legame tra le due che più o meno va in quella direzione,.

(differenza tra metodo clinico e metodo sperimentale, nel clinico lo scopo non è quello di dimostrare una legge generale, ma di risolvere il problema di un singolo, nello sperimentale si cercano leggi generali che poi servono al metodo clinico. Per il prof. sono tutti e 2 sperimentali nel clinico lo psicologo ha suggerimenti ed indirizzi dal manuale, poi deve vedere se sono adatti, anche il clinico è sperimentale, sono lì a risolvere il problema del soggetto, sono 2 approcci diversi)

Mentre i parametri visti finora sono alla base dei metodi statistici, questi invece sono parametri descrittivi non alla base di nessun test statistica in particolare

Il percentile è il valore al disotto del quale ci sta un $n\%$ della popolazione. Se ci dicono che nostro figlio come altezza è al 90° percentile, significa che più alti di lui ce n'è solo un 10%. Il 90° percentile è il limite al di sotto del quale sta il 90% della popolazione.

La mediana è il 50° percentile, è il punto che ha esattamente il 50% della popolazione con valori al disotto di questo limite e il 50% al disopra, non è detto che sia la media. La mediana dice che sotto quel valore c'è il 50% popolazione. Nella media ad es. 20 non è il 50° percentile, sotto il 20 non c'è il 50% popolazione. La media non coincide necessariamente con la mediana in alcuni casi sì, però la media non ha necessariamente il 50% dei soggetti sopra e il 50% sotto. La media mi dice qual è il valore che ottengo sommandoli tutti e dividendo per il numero dei valori.

La mediana mi dice qual è il valore che divide in 2 la popolazione

La moda è il valore più probabile, prima la moda era 20, la moda dei voti universitari potrebbe essere 27, cioè quel valore che ha la frequenza più elevata, non è detto sia la media o la mediana.

Media moda e mediana misurano 3 indici di centralità diversi, interpretano diversamente la centralità. La media la interpreta in termini di probabilità, la mediana in termini di equipartizione della popolazione, la moda in termini di frequenza, hanno 3 meccanismi diversi.

01.03.05

Estrazioni del lotto dal 07.01.1871 fino a febbraio 2005. Sono 7248 estrazioni dove sono riportate anche le frequenze. Un numero dovrebbe uscire ogni 18 settimane, cioè 402. Sul totale delle estrazioni prese in esame il 48 è uscito 458 volte, il 38 è uscito meno di tutti 359 volte. Il campione è abbastanza esteso e se la legge grandi numeri fosse vera come un vincolo fisso che fa sì che le frequenze si eguagliano dovremmo avere che ogni numero è stato estratto 402, 401 o 403 volte, invece c'è una differenza di una cinquantina di volte in meno e una cinquantina di volte in più. La legge dei grandi numeri dice quindi che le frequenze non si eguagliano, tendono ad essere abbastanza vicine alla probabilità, ma non uguali.

La **distribuzione di probabilità** è la descrizione dell'insieme delle probabilità che competono a ciascun valore di una variabile. Esempio: nel lancio di un dado ad 1 compete $1/6$ a 2 $1/6$ etc...Le distribuzioni di probabilità sono l'insieme di tutte le probabilità che competono alle diverse uscite di un numero casuale. Se penso al lotto i numeri hanno tutti la stessa probabilità che è $1/18$. A cosa servono le distribuzioni? Sono il fondamento del test statistico. Nella prova sperimentale abbiamo un problema di fondo che è l'oscillazione casuale, scopo del test statistico è di tenere conto dell'oscillazione casuale, cioè di valutare se una differenza è importante oppure no, è significativa oppure no. Senza valutare l'importanza di una differenza misurata non possiamo dimostrare sperimentalmente nulla. Devo avere uno strumento che mi dice se queste differenze sono rilevanti o meno. Nella prima parte del corso cerchiamo di capire il meccanismo per stabilire se una differenza è significativa oppure no. Noi intuitivamente abbiamo questo meccanismo, lo applichiamo tutti i giorni senza studiare statistica, ad es. giudichiamo il ritardo di un treno di 5 minuti trascurabile, di 5 ore meno trascurabile. Noi abbiamo la capacità di capire quando una oscillazione è dentro i limiti che ci aspettiamo per puro caso e quando invece questa oscillazione è al di fuori di questi limiti e quindi è necessario fare qualcosa. La statistica stabilisce solo le regole per valutare numericamente queste differenze, crea una procedura per valutare un qualcosa che facciamo sempre. La nostra esperienza della realtà include il caso, l'oscillazione casuale, quindi queste valutazioni noi le facciamo tutti i giorni. Lo strumento di base per capire quali sono le differenze che stanno dentro l'oscillazione casuale e quelle che stanno fuori, che sono quindi ritenute dovute a uno specifico fattore, questo strumento è la distribuzione di probabilità.

La **funzione di distribuzione** è la funzione che rappresenta per ogni x la probabilità di ottenere un valore uguale o minore di X . Quindi invece di stabilire la probabilità dei diversi valori io vado a stabilire qual è la probabilità che ho di ottenere un valore minore o uguale a un certo valore X . Somiglia a ciò detto sui percentili. Questa rappresentazione si chiama funzione di distribuzione. Se se la **variabile è discreta** allora posso calcolare per ogni valore di x la sua probabilità, e la funzione di distribuzione è la somma di tutti i valori di x inferiore a quel valore, compresa la probabilità di quel valore. La funzione di distribuzione la ottengo sommando per ogni valore la probabilità di quel valore e di quelli precedenti. Ad es: se voglio calcolare la funzione di distribuzione del risultato del

lancio di un dado, so che la funzione di distribuzione relativa al valore 1, cioè la probabilità di ottenere 1 o meno di 1 è uguale alla probabilità di ottenere 1 quindi è 1/6, la funzione di distribuzione relativa al risultato/valore 2, cioè la probabilità di ottenere 2 o meno di 2 è la somma della probabilità di ottenere 1 + la probabilità di ottenere 2, quindi è 2/6. La funzione di distribuzione del risultato 3 è $1/6 + 1/6 + 1/6$. Nelle variabili discrete la funzione di distribuzione si calcola sommando le probabilità di tutti i valori della variabile uguali o inferiori a quello di riferimento.

La distribuzione uniforme è quando tutti i valori hanno la stessa probabilità, è la distribuzione delle probabilità relative al lancio di un dado. Il problema nasce quando la **variabile è continua**, perché **non è possibile** calcolare la probabilità di ogni singolo valore. Se parlo di altezza non sono in grado di calcolare la probabilità che un soggetto sia alto esattamente 170cm, perché se la variabile è continua esclude 170,1, 170,001 170,00001, cioè nelle variabili continue ho sempre un valore che considero rispetto a una infinità di valori possibili, quindi la sua probabilità è sempre zero, perché al suo denominatore ho sempre un numero infinito di eventi. L'altezza non ha estensione da $-\infty$ a $+\infty$, ha un limite, nonostante questo i possibili valori in quel range proprio perché la variabile è continua sono infiniti. La funzione di distribuzione invece è calcolabile facilmente anche con variabili continue. Quando parlo di funzione di distribuzione relativa al valore 170cm, io prendo in considerazione tutte le altezze inferiori a 170cm, e sono infiniti valori, diviso i valori possibili che sono sempre infiniti, infinito diviso infinito è possibile, (non sempre), ma è possibile che dia un valore finito come effettivamente dà. (qualsiasi numero finito diviso infinito = 0) La funzione di distribuzione permette di calcolare un valore finito in una variabile continua. In una variabile continua la probabilità è sempre uguale a zero, nella funzione di distribuzione invece no, si hanno infiniti casi al numeratore e infiniti al denominatore e questo dà un numero finito. La definizione di funzione di distribuzione vale comunque, sia che la variabile sia discreta che continua. La definizione di distribuzione di probabilità ha senso solo per le variabili discrete perché la probabilità non è calcolabile nelle variabili continue. Se la variabile è continua la funzione di distribuzione non è più definita dalla somma delle probabilità, ma come l'integrale di una certa funzione. L'integrale è una somma di infiniti termini, ognuno di valore infinitesimo, è la conversione nel caso continuo della sommatoria del caso discreto (sommatoria è somma di elementi definiti). Cos'è quella "f(x)" quando calcolo l'integrale? È la densità di probabilità, che rappresenta la probabilità che il valore x cada all'interno di un certo intervallo diviso la larghezza di questo intervallo. Se vado a valutare la probabilità di una certa altezza specifica io so già che questa è uguale a zero, perché nell'intervallo ho un numero infinito di valori e i casi possibili sono quindi infiniti, quindi la probabilità di un risultato è 1 diviso infinito che dà zero. Se invece di calcolare la probabilità che uno sia alto esattamente 170cm, calcolo la probabilità che uno sia alto tra 169,5cm e 170,5cm allora questa probabilità è calcolabile, perché in questo intervallo ci stanno infiniti valori favorevoli, diviso infiniti valori possibili, questo rapporto diventa finito e calcolabile. Nelle variabili continue quindi non posso calcolare la probabilità ma calcolo la densità di probabilità che è il rapporto tra la probabilità che il valore cada dentro un certo intervallo diviso la larghezza dell'intervallo. (Un numero infinito diviso infinito non dà zero ma un numero finito, un numero finito diviso infinito dà zero). Calcolo la densità che è la probabilità che il risultato stia all'interno dell'intervallo, diviso la larghezza di questo intervallo. La densità di probabilità calcola la probabilità di un intervallo, dentro l'intervallo ho infiniti valori che diviso infiniti valori possibili danno un numero finito, la differenza fra probabilità e densità di probabilità è che la probabilità è la probabilità di un valore, la densità è la probabilità di un intervallo divisa per la larghezza dell'intervallo.

La **distribuzione binomiale** è la distribuzione che compete a una variabile che può essere o SI o NO o successo o insuccesso o testa o croce, compete a una variabile dicotomica (cioè che ha 2 valori). Quando noi eseguiamo tante prove con il lancio di una moneta, se lanciamo 10 volte il numero di teste o croci che abbiamo come risultato segue una distribuzione binomiale. Suppongo di

lanciare una moneta 2 volte, i risultati che posso ottenere sono SS SI IS II cioè TT TC CT CC. Ciascuno di questi risultati ha la stessa probabilità degli altri cioè $\frac{1}{4}$, cioè $\frac{1}{2} * \frac{1}{2}$. Se mi interessa solo vedere quante teste sono risultate, ottengo che per 2 teste la probabilità è $\frac{1}{4}$, per 1 testa invece è $\frac{1}{4} + \frac{1}{4}$, va bene sia il 2° che il 3° risultato, quindi la probabilità di ottenere 1 testa è $\frac{1}{2}$. mentre quella di avere 2 croci è $\frac{1}{4}$. La mia distribuzione di probabilità è così: 2 teste probabilità $\frac{1}{4}$, 1 testa probabilità $\frac{1}{2}$ zero teste probabilità $\frac{1}{4}$. Se invece di avere 2 lanci io ne ho 10 calcolare tutte le combinazioni diventa lunga, per calcolare il numero dei successi in un esperimento in cui in un certo numero di prove ho solo la possibilità di successo o insuccesso, c'è una formula

$$p = \frac{1}{2^n} \binom{n}{s}$$

Dove

$$\binom{n}{s} = \frac{n!}{s!(n-s)!}$$

La probabilità di s successi in n prove è uguale a $\frac{1}{2^n}$ alla n che moltiplica n su s. E' la formula binomiale. Il fattoriale è la moltiplicazione del numero e di tutti gli interi minori di quel numero fino a uno. N su s è n fattoriale (che si indica con !) fratto s fattoriale che moltiplica il fattoriale di (n-s). N! è uguale a n*(n-1)*(n-2)*(n-3) fino a 1. Es. 5!=5*4*3*2*1=120

Vediamo con n=2, quindi in 2 lanci, che probabilità mi dà per 1 sola testa.

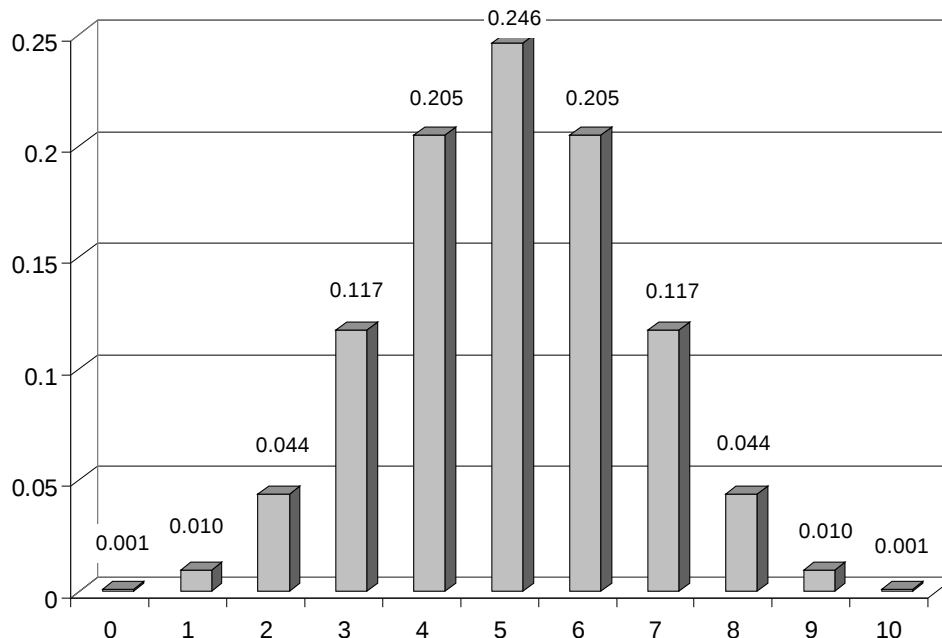
Ho $p = \frac{1}{4} * n$ su s, $p = \frac{1}{4} * 2 / (1 * 1) = 1/2$. Ottengo lo stesso risultato di prima, va bene sia il 2° che il 3° risultato.

Domande d'esame possibili: cos'è una distribuzione? Cos'è la densità di probabilità? Perché uso la densità e non la probabilità? Devo rispondere che la probabilità non è definita per variabili continue e usando la probabilità potrei calcolare la distribuzione solo per le variabili discrete. (Non ci chiederà nessuna formula.) Perché la varianza è il parametro usato per misurare l'oscillazione casuale? La varianza è tanto più grande quanto più i valori si allontanano dalla media, quindi, quanto più ho oscillazione attorno alla media tanto più alta sarà la varianza.

In una distribuzione binomiale posso elencare tutti i possibili risultati in n prove poi contare quanti di questi risultati sono caratterizzati dal numero di successi che voglio vedere, oppure se la probabilità di successo e di insuccesso sono uguali uso la formula binomiale.

Questa formula vale solo se la probabilità di successo p e di insuccesso q sono uguali ed ognuna è uguale ad $\frac{1}{2}$. Con questa formula se io lancio una moneta 10 volte l'insieme delle probabilità ha la seguente distribuzione.

Distribuzione binomiale
10 prove

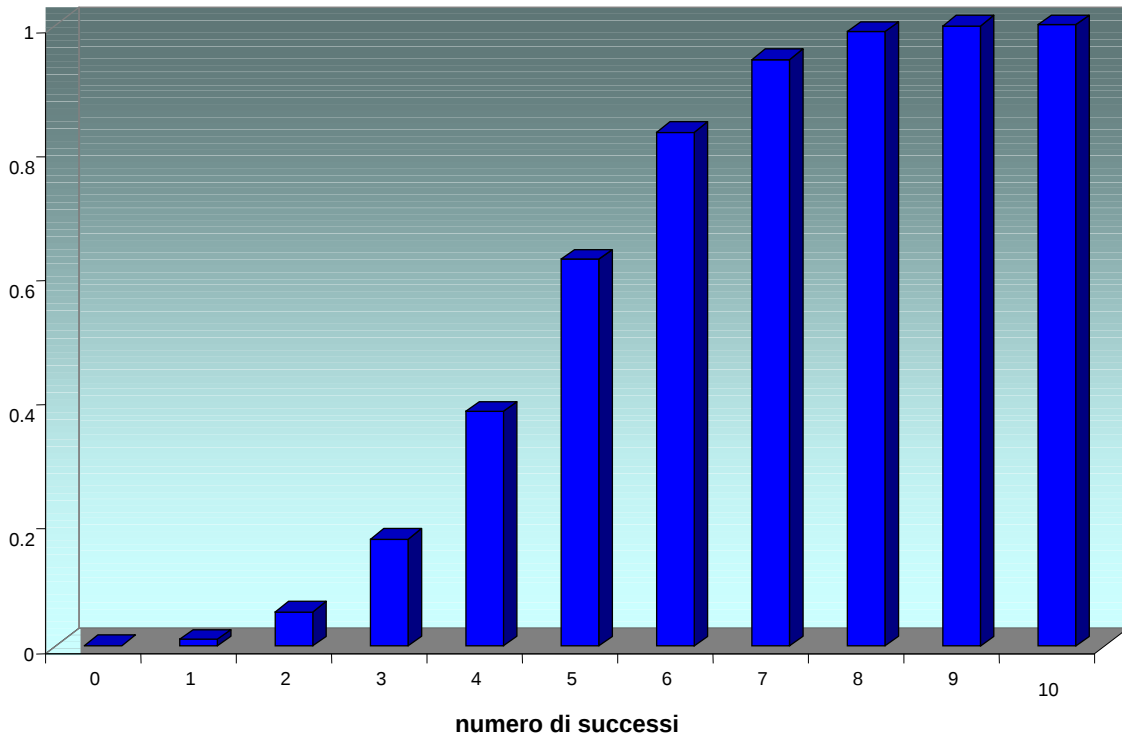


Questo è il grafico relativo alla distribuzione di probabilità di 10 lanci di una moneta. Rappresenta l'insieme delle probabilità che competono al diverso numero di successi, quindi la probabilità su 10 lanci di avere 0 teste è un numero prossimo allo zero, un po' più alta è la probabilità di ottenere 1 testa. Ottengo una distribuzione di valori simmetrica, attorno al massimo che è 5 abbiamo 2 code che cadono simmetricamente. Per distribuzione non intendo solo il grafico ma quasi sempre viene utilizzata solo questa rappresentazione perché si ha una informazione immediata di come si distribuiscono le diverse probabilità dei valori della variabile che sto considerando. Potrei fare undici caselle in cui scrivere le 11 diverse probabilità. La distribuzione binomiale ha 1 picco centrale se il numero delle prove è pari, ha 2 picchi se il numero prove è dispari.

02.03.05

La distribuzione di probabilità di una variabile binomiale è simmetrica e lo è quando la probabilità di successo e di insuccesso sono uguali. Come si costruisce la funzione di distribuzione? (Che nel caso discreto si chiama anche cumulativa.) Si ottiene sommando per ogni valore tutte le probabilità dei valori precedenti quindi partiamo con un valore prossimo allo zero a cui è aggiunto un altro valore molto piccolo, poi con 2 successi il valore cresce, per il caso tre successi che è la somma della probabilità di 2 successi più la probabilità di 1 successo + la probabilità 0 successi. È la distribuzione a cui ad ogni valore associo la probabilità di quel valore + la probabilità di tutti i valori precedenti a quello considerato. È una funzione crescente, nel senso che ogni colonna è uguale o superiore alla colonna precedente, questa è la rappresentazione di una funzione di distribuzione per il caso di una distribuzione binomiale a 10 eventi. ($p=0,5$ significa che $p=1/2=q$)

Funzione di Distribuzione Binomiale con p=0.5



Abbiamo preso in considerazione solo il caso in cui la probabilità di successo e di insuccesso sono uguali, e ognuna è uguale a 0,5. Nel caso che ciò non sia vero la formula cambia leggermente.

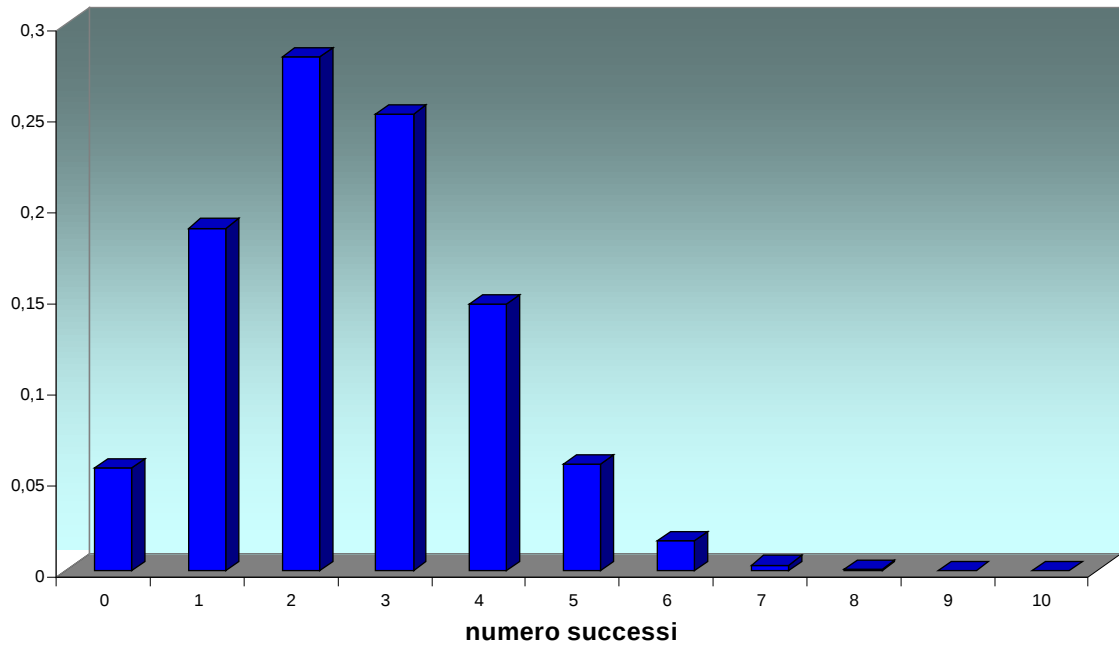
$$p \neq q$$

$$p = p^s q^{n-s}$$

La probabilità di s successi è data dal prodotto della probabilità dei successi elevata alla s (che qui è chiamata p) per la probabilità degli insuccessi elevata al numero degli insuccessi (che qui è q) per n su s, qui invece di avere 1/2 alla n, ho p cioè la probabilità dei successi elevata alla s, cioè al numero dei successi e q cioè la probabilità degli insuccessi elevata al numero degli insuccessi, dove il numero degli insuccessi è = n-numero successi. Questa è la formula + generale della distribuzione binomiale. È un'estensione della precedente, se p=q diventa quella di prima. Se sono = sono 0,5 cioè 1/2 cioè 1/2 alla s + 1/2 alla i cioè 1/2 alla s+i e s+i=n diventa quella di prima. La prima formula vista vale quindi con la restrizione che p+q deve essere =1, cioè la probabilità successi + la probabilità di insuccessi deve essere =1, cioè deve essere la probabilità totale perché siamo del caso di variabili dicotomiche, e vale anche con la restrizione che s+i = numero totale delle prove, perché un evento o è considerato successo o insuccesso.

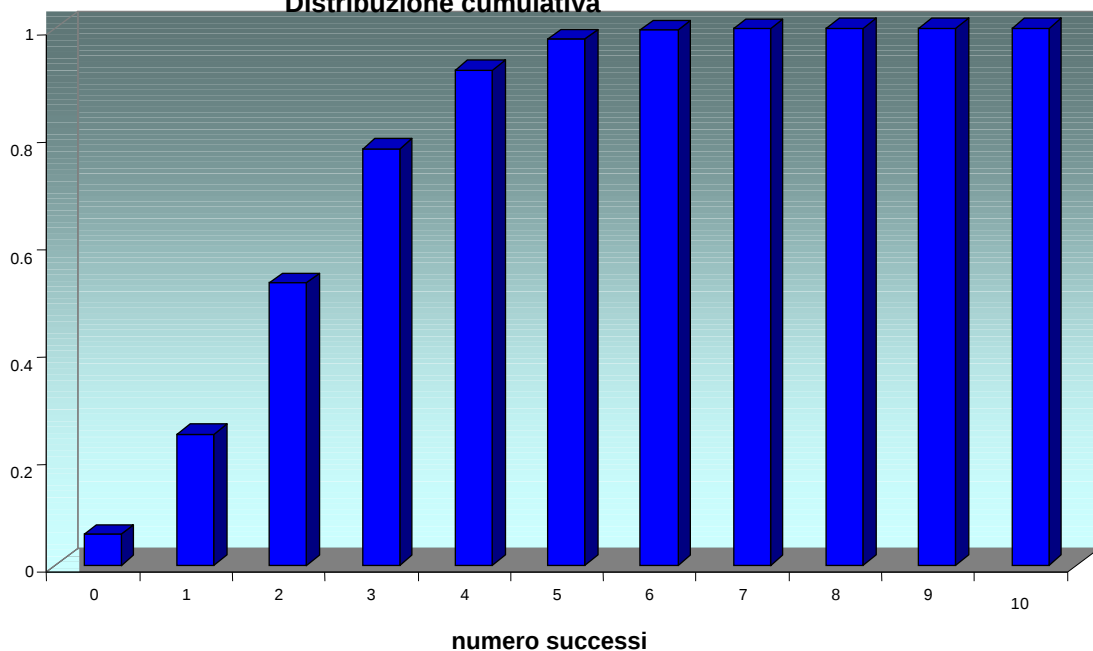
Nel caso che p sia diverso da q la distribuzione binomiale non è più simmetrica, si sbilancia dove la probabilità è più alta. Nel grafico la probabilità di successi è 0,25 quindi è più alta la probabilità di insuccessi, la barra più alta è verso gli insuccessi. La distribuzione binomiale è simmetrica solo se p=q. Sempre su 10 prove la prima colonna rappresenta la probabilità di zero successi quindi di 10 insuccessi.

Distribuzione Binomiale con $p=0.25$



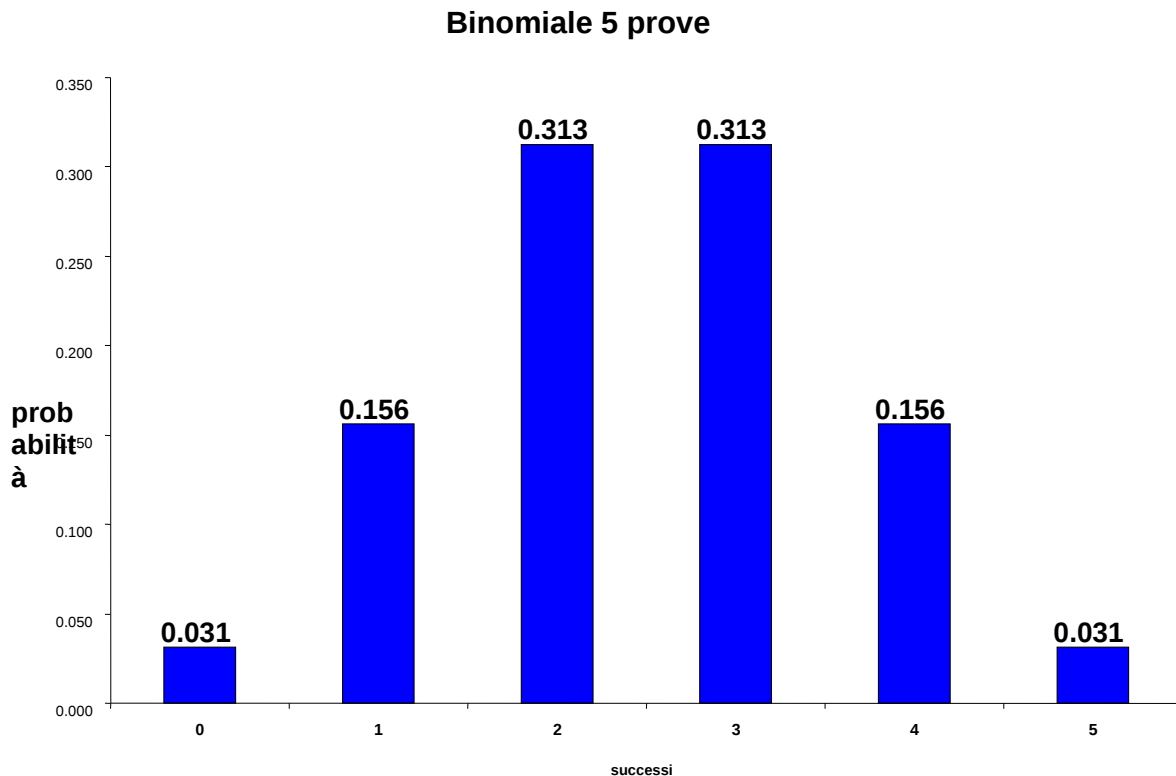
Nel caso di una distribuzione asimmetrica la funzione di distribuzione o distribuzione cumulativa sarà sempre una funzione crescente, o meglio, una funzione non decrescente, una funzione dove i valori successivi sono uguali o maggiori dei precedenti. Anche nella funzione asimmetrica la funzione cumulativa è non decrescente, ma il modo con cui cresce è diversa, se come in questo caso la probabilità degli insuccessi è maggiore di quella dei successi, cresce rapidamente, qui le colonne più alte sono inizio, nella precedente, la forma simmetrica cresce come un sigmoide, qui cresce rapida poi quasi stazionaria.

**Funzione di Distribuzione Binomiale con $p=0.25$,
Distribuzione cumulativa**



Un esempio pratico, semplice perché solo a cinque prove, se il numero prove è dispari e se $p=q=0,5$ la distribuzione è simmetrica ma le punte sono 2, avere 2 successi e 3 insuccessi o 3 successi e 2 insuccessi, nella formula vista deve dare lo stesso risultato. Se $p=q$ un numero dispari di prove dà sempre 2 picchi di uguale altezza. Vediamo come potremmo calcolare le diverse probabilità.

Abbiamo 5 prove, dove $p=q=0,5$ mi pongo il problema della probabilità di 0 successi. Sono dati da $\frac{1}{2}$ alla quinta cioè $\frac{1}{32}$
(Zero fattoriale è anomalo, fa 1)

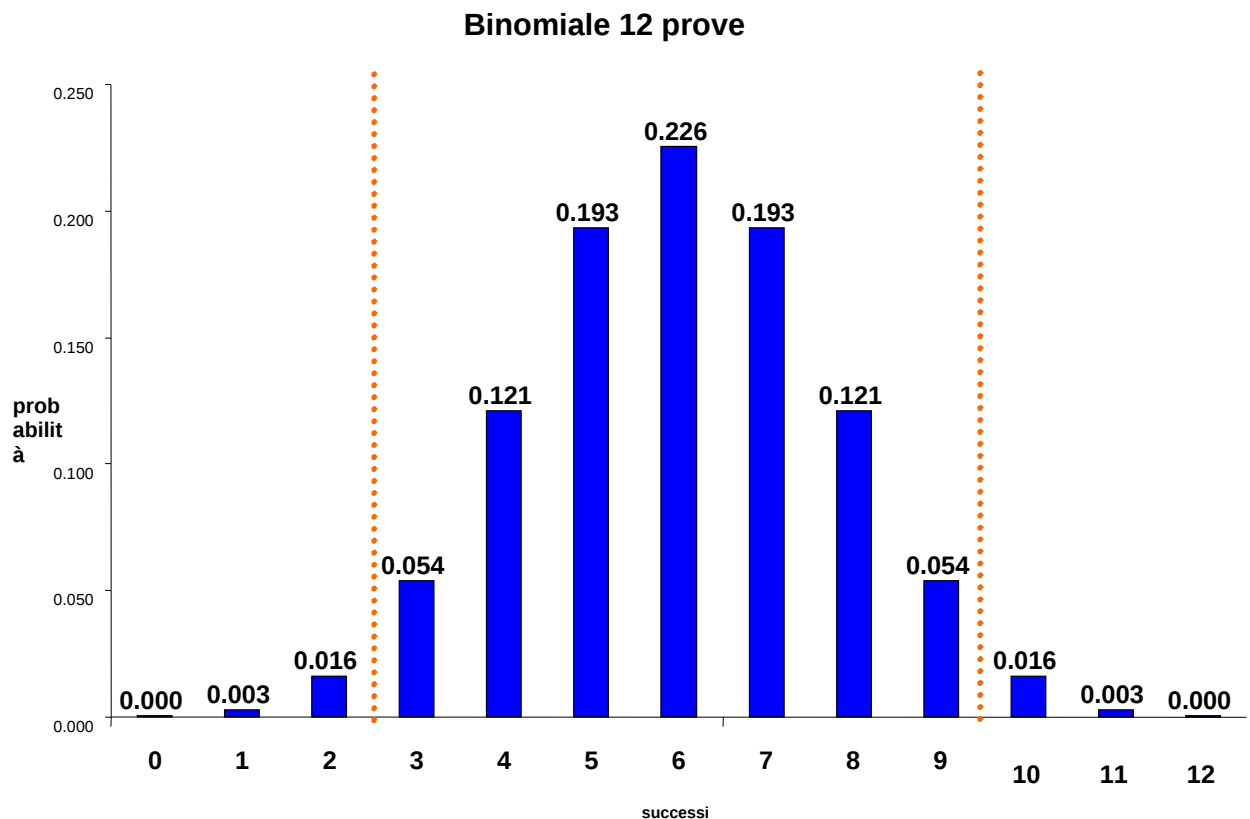


La probabilità di zero successi è 0,03125, questa è anche la probabilità di 5 successi, ho la stessa situazione, scambiamo i successi con gli insuccessi, ma abbiamo lo stesso risultato, poiché 3 teste e 2 croci hanno stessa probabilità di 2 teste e 3 croci.

Questi valori costituiscono la distribuzione di probabilità di una variabile binomiale con 5 prove e $p=q=0,5$, questa tabella è 1 rappresentazione della distribuzione.

A cosa servono le distribuzioni.

Come posso fare a dimostrare che una moneta è truccata? Lancio 12 volte, se è truccata verrà tante volte testa, se invece di venirmi 6 teste e 6 croci vuol dire che è truccata? Come faccio a decidere che è truccata? Stabilisco dei limiti, questo è ciò che fa la statistica in tutti i test, questo piccolo esempio rappresenta il cuore della statistica. So che c'è una oscillazione casuale da cui nessun esperimento è esente. Non esiste nessun metodo di misura sperimentale ad errore zero, esistono ad errore basso. Non esistono misure sperimentali ad errore zero neanche in fisica. So che c'è questa oscillazione casuale e che la considero influente fino a un certo punto: se faccio 12 lanci, l'oscillazione casuale è rappresentata dalla distribuzione di probabilità, mi dice in 12 lanci che probabilità ho di ottenere 6 teste 5 teste 4 teste etc.. fino a zero. Se c'è solo oscillazione casuale la distribuzione di probabilità mi dà la probabilità di ottenere diversi risultati.



Io prendo ad es. il 95% dei risultati casuali che è l'area che sta dentro sbarre rosse. Se ottengo un risultato lì dentro non so nulla, se invece di venirmi 6 teste me ne sono venute 5 questo fa parte dell'oscillazione casuale, se ottengo un risultato al di fuori di questi 2 limiti allora non è più oscillazione casuale, la probabilità di avere quel risultato per oscillazione è troppo bassa, è inferiore al 5%, e se il caso ha solo il 5% di probabilità di essere responsabile di quel risultato allora concludo che il risultato non è casuale, è dovuto ad altro, ad esempio al fatto che la moneta è truccata. Avere un errore di dimostrazione solo del 5% è accettabile per la scienza nella dimostrazione sperimentale delle idee scientifiche. Le distribuzioni rappresentano le oscillazioni casuali rispetto ad un valore di riferimento con i rispettivi gradi di probabilità (per ogni risultato). Attorno al valore teorico prendo il 95% e ritengo dimostrata la mia idea solo se il risultato è fuori da questa area, perché la probabilità che mi sbagli, dicendo che non è casuale, è solo del 5%.

Come calcolo la zona del 95%? Ho cominciato prendendo la probabilità del valore teorico, cioè del valore che otterrei se fosse solo caso, qui è di 6 teste, cioè 6 successi, quindi la mia probabilità è 0,226 ed a questa sommo le 2 colonne di destra e sinistra poi le altre più esterne fino a quando supero 0,95. Siccome le colonne sono discrete non è detto che ci sia un punto in cui la somma sia esattamente = 0.95, mi fermo quando ho superato quel limite, dentro ci deve essere almeno il 95%, quindi c'è un poco + del 95% dentro. In una distribuzione simmetrica ho lasciato fuori 2 aree uguali, ognuna che contiene il 2,5%. Posso anche fare l'inverso, parto dal fondo e dico, mi fermo prima di superare il 2,5% da ogni lato. Ci deve essere o il 2,5% o meno da ogni lato, 0003+0.016 e basta. Non ci deve essere più di 0,025.

Il 95% è una convenzione scientifica internazionale, quasi tutte le ricerche si considerano dimostranti l'idea se la probabilità di errore è inferiore al 5%. Solo in alcune ricerche, ad es. quelle sui tumori, il limite è ridotto all'1%.

Se faccio una prova che ha un margine d'errore del 5% posso considerarmi soddisfatto.

L'esempio a 5 prove serve per spiegare che non è vero che riesco a dimostrare un'idea con qualunque numerosità del mio campione. Cioè la numero del campione può essere insufficiente a dimostrare la mia idea, ad es. se io lancio la moneta 5 volte non ho nessun risultato che abbia meno del 2,5% di probabilità di succedere. Cioè se parto dall'esterno e vado a tagliare fino al 2,5% taglio subito perché la 1 colonna è già al 3,1% , se con 5 lanci ottengo 5 teste non sono ancora in grado di falsificare la correttezza della moneta (non posso dire che la moneta è truccata) con un margine d'errore inferiore al 5% perché nessuna colonna ha una probabilità inferiore a 0.025. Quindi 5 prove in una distribuzione binomiale non sono sufficienti a dimostrare nulla. Se faccio 6 lanci la probabilità di ottenere 6 teste è 0,016 che è inferiore a 0,025 se ottengo 6 teste o 0 teste posso dire che la moneta è truccata, perché questi 2 risultati insieme hanno la probabilità di accadere per caso inferiore al 5%. Ciò dice che 5 lanci sono insufficienti, non dice però che 6 sono buoni. Se faccio 12 lanci, se ottengo 10 teste o 11 o 12, oppure 0 o 1 o 2 posso dire che la moneta è truccata, la probabilità che non lo sia è inferiore al 5%.

CONCLUSIONI

1. le distribuzioni siccome sono in grado di rappresentare l'oscillazione casuale servono a distinguere quello che è un risultato casuale e quello che è un risultato dovuto al fenomeno che sto studiando. Servono per dirci di un certo risultato se lo devo considerare casuale o se invece è troppo strano per essere casuale e quindi è dovuto al meccanismo che sto studiando.
2. La procedura con cui questa dimostrazione viene effettuata è: io prendo il valore esatto del risultato casuale, in questo caso 6 teste su 12 lanci , da questo risultato mi allargo - in una distribuzione simmetrica - da una parte e dall'altra fino a quando inglobo il 95% o almeno il 95% delle possibilità. In questa zona la probabilità che il risultato sia dovuto al caso è del 95%, al di fuori di questi limiti questa probabilità è inferiore al 5%, allora scegliendo come soglia il 95% e il 5%, la dimostrazione sperimentale funziona così: se il mio risultato è dentro alla zona 95% non posso concludere nulla, se è fuori concludo che il fenomeno non è dovuto al caso.

Se ottengo 10 teste sono autorizzato a dire che questo risultato non è casuale, se ne ottengo 9? Non sono in grado di dire che è truccata ma la probabilità che non lo sia è bassa. Nella zona tra i 2 limiti c'è di tutto, ci sono anche 9 teste e 3 teste in cui non è sufficiente il valore per dire che è truccata, ma non mi rassicura sul fatto che non lo è. Se il mio risultato è dentro i limiti non posso concludere nulla, non ho dimostrato nulla, se è al di fuori ho dimostrato che invece è truccata. La statistica è asimmetrica, non biunivoca, sbilanciata, non è come la matematica, perché i fenomeni che stanno dentro non mi permettono di decidere nulla, quelli fuori invece mi permettono di trarre una conclusione. Ho fallito nel dimostrare che l'altro barava, ma ciò non vuol dire che ha giocato pulito.

07.03.05

La distribuzione binomiale è utilizzata in alcuni test statistici, anche se pochi, per fare dei confronti, ci sono dei test specifici, come quello del segno, che la utilizzano. Il problema è che la distribuzione binomiale funziona solo se io considero il verificarsi di 1 evento dicotomico i cui risultati possono essere solo il verificarsi o il non verificarsi di quell'evento. La realtà è che nella ricerca non è sempre così, ci sono volte in cui i valori possono essere tanti, ad esempio se prendo l'altezza. La distribuzione binomiale è la discretizzazione, cioè la conversione in valori discreti, di un'altra distribuzione relativa ad una variabile continua (che è fondamentale in statistica), ed è la distribuzione normale o distribuzione di Gauss. La distribuzione gaussiana è la distribuzione di una variabile continua, è l'estensione della binomiale al caso continuo. Ad es. se prendo la distribuzione a 6 prove dove ho 7 colonne (c'è anche quella a zero teste) aumentando il numero delle prove questa distribuzione, mantenendo la caratteristica di un picco centrale e di 2 parti simmetriche a

destra e sinistra del picco, diventa sempre più fitta e quando le prove diventano infinite si arriva alla distribuzione gaussiana.

Sulla definizione di infinito ci sono grandi discussioni in atto, è una cosa che noi non siamo in grado di sperimentare, richiederebbe un tempo infinito. Il fatto che diciamo che la distribuzione normale è l'estensione al numero infinito di prove della binomiale è un'affermazione teorica, non può avere una sperimentazione diretta, vuol dire che quando il numero delle prove cresce all'infinito, il numero delle barre che compongono la binomiale diventano infinite e la rappresentazione diventa una curva continua. La distribuzione normale è anche definita come curva degli errori. Ad es. se gioco alle freccette, lanciando posso prendere il centro o posso allontanarmi da questo, la probabilità che la distanza della freccetta rispetto al centro ha di assumere certi valori, ha un massimo ad una certa distanza che è diverso per ognuno di noi, poi cala molto andando sia verso il centro che verso la periferia. Questo è il comportamento di quasi tutti gli errori, se eseguo la misura su una bilancia ad altissima precisione, se mi peso un certo numero di volte non ho mai lo stesso valore, questi valori del peso si distribuiranno in un modo che ricorda la distribuzione binomiale, ed essendo valori continui si distribuiranno secondo una distribuzione gaussiana cioè si avrà la maggior parte dei valori vicino al nostro peso, poi un po' al di sotto e un po' al di sopra. La distribuzione normale è la distribuzione tipica degli errori casuali, cioè se non c'è nessun meccanismo specifico che produce l'oscillazione, ma è prodotta da un meccanismo casuale, si ottiene una distribuzione gaussiana. I matematici dicono: in un fenomeno in cui gli errori possono essere dati da piccoli incrementi o decrementi della stessa quantità e questi incrementi o decrementi sono fra loro indipendenti, allora la distribuzione degli errori è gaussiana, perché in sostanza ogni piccolo errore è come il successo o l'insuccesso in una distribuzione binomiale. Se ho tanti elementi tra loro indipendenti e ognuno mi fa diminuire o aumentare il valore di una certa entità, ottengo una distribuzione binomiale con infiniti elementi, e la curva degli errori non può che essere una curva normale. È esperienza di tutti che gli errori che noi vediamo, ad es. quelli di misura, hanno una distribuzione a campana, cioè con molti valori vicino a un picco centrale e allontanandoci da questo ho qualche valore ma con probabilità più bassa.

L' **entropia** è una misura, è nata in fisica, esportata in teoria delle comunicazioni e poi in statistica. È la misura del disordine, in fisica è nata perché per produrre energia ci vuole ordine. Quanto più un sistema è ordinato tanto più è a bassa entropia. Concetto esportato in teoria delle comunicazioni da Shannon: per rendere efficienti i sistemi di comunicazione devo trovare una misura per quantificare quante informazioni trasmettono, per farlo ha usato il concetto di entropia. Se trasmetto una sequenza di lettere "a" trasmetto 1 sola informazione, se trasmetto una sequenza di simboli, cioè cose diverse, trasmetto molte informazioni. L'informazione nasce dalla diversità dei simboli che vengono trasmessi, dipende dal disordine dei simboli. Ci sono lingue ridondanti, a parità di lettere le ridondanti trasmettono meno informazione, l'inglese è a bassissima ridondanza. Gli sms mostrano il fatto che la nostra lingua è ridondante, si abbrevia tutto, molte parole hanno una parte che non è indispensabile per capire il senso, e anche molte parole in una frase non servono per capire il senso. La ridondanza di una lingua non è un difetto ma una caratteristica, ci permette di coprire difetti di pronuncia, oppure evita che ad es. le vocali pronunciate aperte o chiuse pregiudichino la comprensione, ci permette di superare la non perfezione della comunicazione.

L'entropia in statistica, poiché misura il disordine, è una misura del caso. Ha cioè il senso opposto che nella teoria delle comunicazioni, perché lì è una cosa che vado a inseguire, più alta entropia ha il mio messaggio più informazioni trasmetto, in statistica invece più alta è l'entropia, più alto è il disordine, più alta è l'incidenza del caso. La distribuzione gaussiana è la distribuzione di una variabile caratterizzata dal fatto di avere la massima entropia possibile. La distribuzione normale è la distribuzione del caso per eccellenza, questo è il motivo per cui la gaussiana ha un ruolo fondamentale nella statistica, essendo la distribuzione di una variabile continua è molto più complessa della binomiale.

Cosa importante: una variabile a distribuzione binomiale ha una media e una varianza ben definite, la media è data da $n \cdot p$ dove n è il numero di prove e p è la probabilità ad es. di successo. La varianza invece è data da $n \cdot p \cdot q$, dove q è la probabilità dell'insuccesso. Questi due valori sono relativi a una distribuzione teorica, siamo in casi teorici. La media sia che riguardi valori teorici che sperimentali si calcola nello stesso modo, la varianza no, se la calcolo per valori teorici va divisa per n non per $(n-1)$. Se immaginiamo un esperimento teorico, avendo 100 prove di lancio di una moneta dove ogni prova è composta da 6 lanci, se calcolo la media come somma delle teste diviso 100 ottengo $n \cdot p$, perché $n=0,5$, mi aspetto di trovare testa 50 volte, perché la media è data da $n \cdot p$, la varianza è data da $100 \cdot 0,5 \cdot 0,5 = 100 \cdot 0,25 = 25$. Questo è un caso in cui noi calcoliamo una media e una varianza su valori teorici. (qui la varianza va divisa per 100 e non per 99)

LIMITE DELLA DISTRIBUZIONE BINOMIALE

Al crescere di n la distribuzione binomiale tende ad una distribuzione normale con media np e varianza npq

La distribuzione normale è la distribuzione alla quale tende la binomiale quando n cresce all'infinito, questa distribuzione ha media np e varianza npq . La curva degli errori è gaussiana cioè l'errore totale ha una distribuzione gaussiana, se gli errori hanno tutti la stessa ampiezza, se possono essere indifferentemente positivi o negativi e se l'errore totale è la somma di tutti gli errorini.

CURVA DEGLI ERRORI

Condizioni:

- un errore è la somma di molte componenti di uguale ampiezza
- le diverse componenti sono fra loro indipendenti
- ciascuna componente è positiva o negativa con uguale probabilità

allora l'ampiezza dell'errore ha una distribuzione normale

($p \cdot q$ è già una probabilità al quadrato, la varianza ha dimensione al quadrato)

Se la curva è bilanciata io ho 0,25, perché se è bilanciata devono essere 0,5 e 0,5, se è sbilanciata, ad es. 0,2 una e 0,8 l'altra la moltiplicazione diventa 0,16, è più bassa, quindi a parità di n la varianza è massima solo se $p=q$ perché se asimmetrica tende ad assottigliarsi se $p=1$ e $q=0$ diventa una linea.

Per la definizione di entropia, la distribuzione normale è la distribuzione a massima casualità. È alla base della statistica non solo perché è la distribuzione più diffusa, ma perché è quella che anche teoricamente rappresenta la massima casualità possibile. La distribuzione gaussiana è una distribuzione teorica, è un assioma, non è dimostrata né matematicamente né sperimentalmente. Non troveremo mai nessuna variabile distribuita in modo esattamente gaussiano. La distribuzione gaussiana è molto aderente alla nostra esperienza. Praticamente non è dimostrata ma è praticamente utilizzabile, ha un senso, è vicina a ciò che noi sperimentiamo.

FORMULE

Una generica variabile normale con media μ e varianza σ^2 è indicata con $N(\mu, \sigma^2)$ e la sua densità di probabilità è

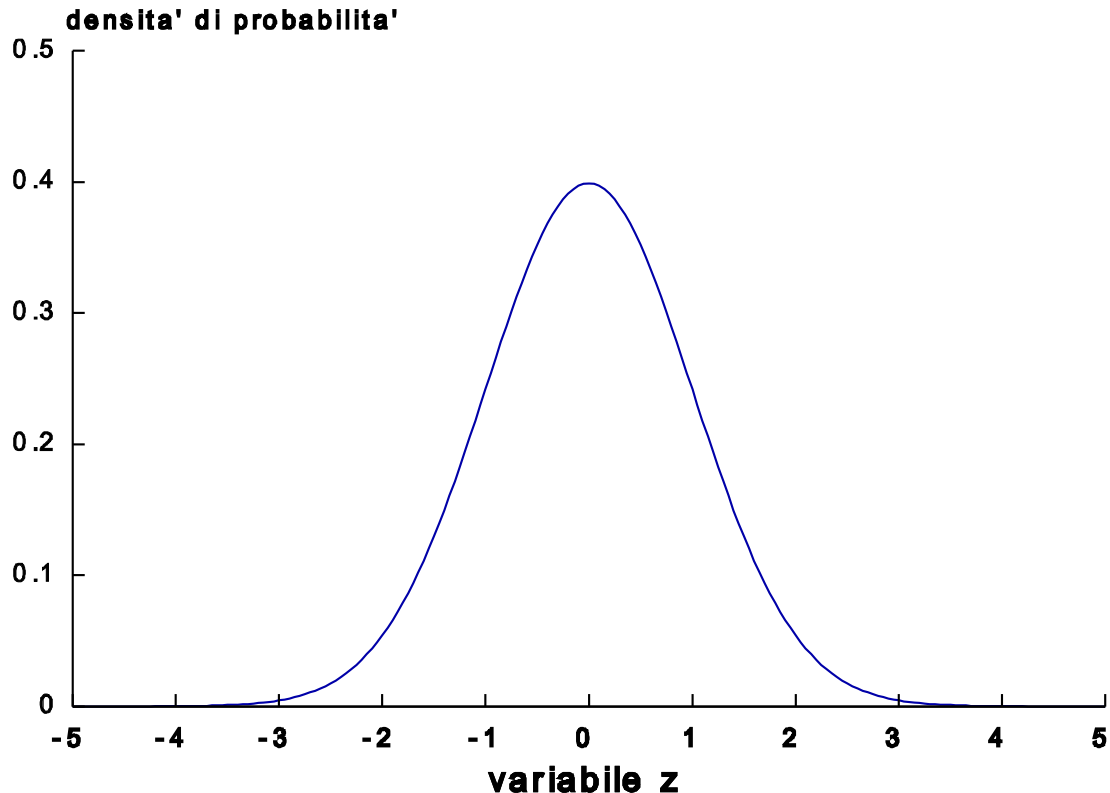
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Come già detto, per le variabili continue non posso rappresentare la distribuzione di probabilità (non c'è la probabilità che uno sia alto esattamente 170 cm, la probabilità è zero) quindi uso la densità di probabilità. L'equazione precedente mi dà la densità di probabilità della distribuzione gaussiana. È composta da numeri fissi (anche π lo è), poi ho sigma al quadrato $\sigma^2 = \text{varianza teorica}$ (per questo la esprimo con lettere greche) della mia variabile normale, il numero naturale e che è un numero fisso è 2,7..., poi ho il mio valore di x che è la variabile gaussiana, e la media teorica "mi" μ . Questa formula ci dice 2 cose importanti,

1. che la densità di probabilità e quindi la distribuzione di una variabile gaussiana è completamente determinata da 2 parametri: la sua media e la sua varianza. La distribuzione normale è quindi funzione di 2 soli parametri, media e varianza teorica.
2. siccome è fatta di una differenza tra la variabile e la media al quadrato, allora il quadrato toglie il segno. Quindi che la x sia sopra di 2 rispetto alla media o sotto di 2 rispetto alla media mi dà lo stesso valore perché quando faccio il valore al quadrato mi dà sempre 4. Quindi la mia distribuzione normale deve essere simmetrica attorno alla media perché andando sopra la media di un certo valore o sotto la media dello stesso valore il risultato non cambia. La densità di probabilità non cambia sia per una x che è di 2 sopra la media che per una x che è di 2 sotto la media. Ho un grafico simmetrico attorno alla media, somiglia moltissimo al grafico della distribuzione binomiale, solo che non ho più colonnine ma una linea perché le prove sono infinite.

Essendo la distribuzione di una variabile continua il suo valore per un dato x corrisponde alla densità di probabilità per quel valore.

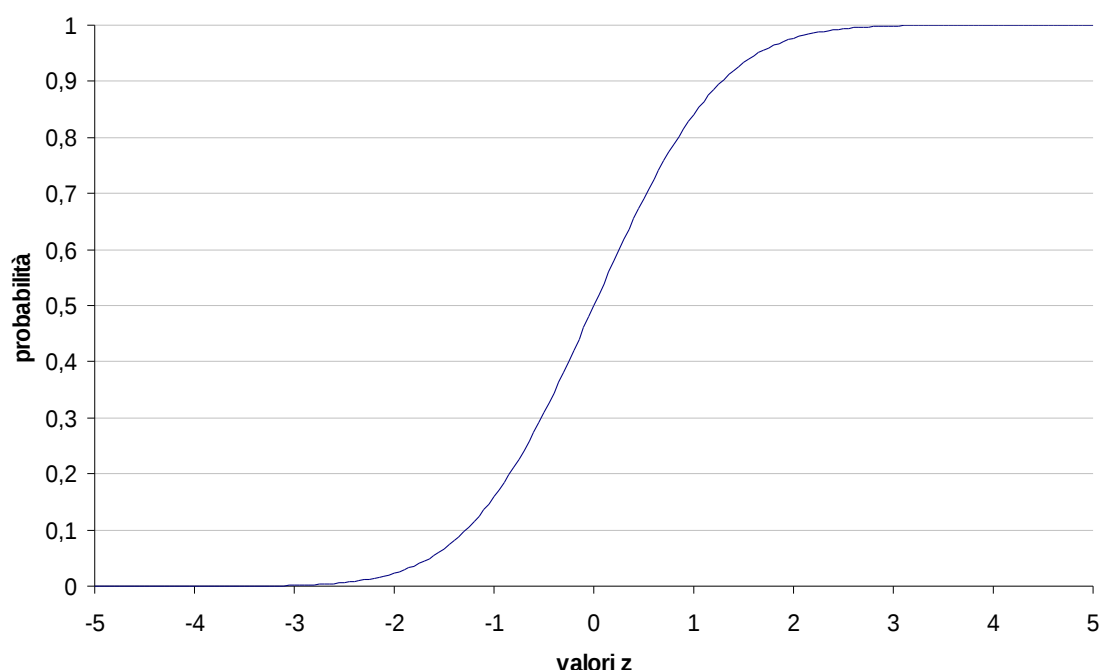
Distribuzione N(0,1)



La distribuzione normale si indica sinteticamente con N (media teorica, varianza teorica). Questa funzione di distribuzione è relativa a una distribuzione gaussiana che ha media zero e varianza unitaria. $N(0,1)$ è anche la gaussiana di riferimento.

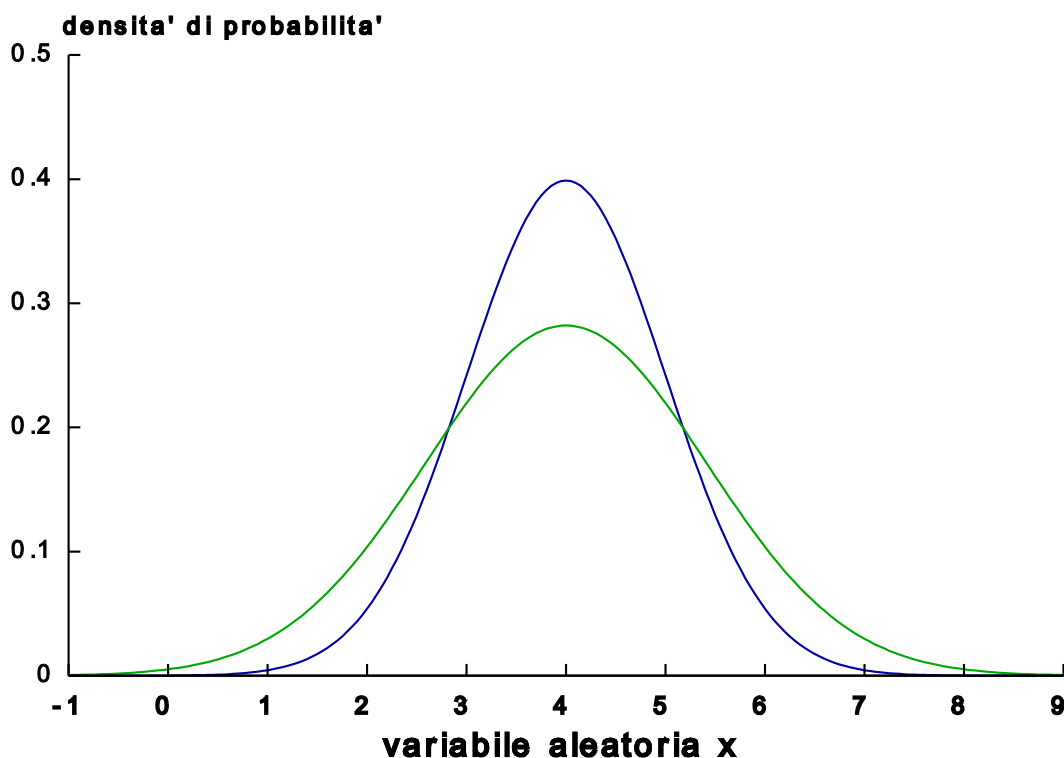
Ultimamente le distribuzioni vengono definite dalla funzione di distribuzione detta anche distribuzione cumulativa che non è altro che la somma di tutti i valori precedenti una certa x , quindi questa distribuzione può essere ottenuta da questa densità di probabilità calcolando l'area che sta a sinistra di una certa barra, quando sposto la barra, l'area che sta a sinistra non può che crescere poi però crescerà sempre meno, fino a ottenere il valore massimo che è uno, perché in una densità di probabilità l'area che sta sotto alla 1° metà della curva, cioè da $-\infty$ a zero rappresenta la possibilità di ottenere un valore tra meno infinito e zero, questa probabilità è uguale all'area che sta sotto la curva. Quando io mi sposto verso $+\infty$ la probabilità di avere un valore compreso fra $-\infty$ e $+\infty$ è 1, l'area totale deve essere = 1 perché quest'area rappresenta la probabilità di avere un valore qualsiasi, è la certezza. Se qualsiasi valore va bene abbiamo la certezza che il risultato sia in accordo. L'area sottesa alla densità di probabilità normalmente è =1, l'area sottesa alla densità di probabilità tra $-\infty$ e un certo valore è ciò che è rappresentato nella distribuzione cumulativa, che mi dice che la probabilità di avere un valore compreso tra $-\infty$ e zero è 0,5, è certo, perché è metà delle possibilità. La variabile gaussiana è una variabile continua compresa tra $-\infty$ e $+\infty$, la probabilità che il suo valore sia tra $-\infty$ e $+\infty$ è 1, la probabilità che il suo valore sia compreso tra $-\infty$ e zero (se zero è la sua media) è il 50% cioè 0,5 che è anche il punto dove la distribuzione cumulativa ha la massima pendenza, siamo al picco, quindi qualsiasi incremento fa aumentare l'area che sta sotto. Poi invece si adagia pian piano e cresce indefinitamente avendo come asintoto 1, cioè tende a 1 senza raggiungerlo mai.

Distribuzione Cumulativa Normale



Ci servirebbe solo la distribuzione cumulativa, nei nostri test usiamo solo la cumulativa, la funzione di distribuzione però ci dà una immagine immediata del comportamento della variabile che dalla cumulativa non si evince subito.

Distribuzioni $N(4,1)$ e $N(4,2)$



- **Posizione massima** (giace sulla media)
- **Altezza del massimo** (tanto più è grande la varianza tanto più la curva è allargata)

Guardando queste due curve so che hanno tutte e due il massimo nello stesso punto, cioè hanno media nello stesso punto, 4, però so che quella verde ha una variabilità maggiore, si estende di più attorno alla media ha varianza maggiore, i punti attorno alla media sono molto più allargati. In quella blu ho un picco che va su abbastanza stretto, la probabilità di avere dei valori vicino alla media è ancora apprezzabile, se mi allontano però la probabilità si abbassa rapidamente, invece in quella verde succede l'opposto anche allontanandomi abbastanza dalla media ho ancora una probabilità alta di ottenere dei valori. In effetti la curva blu è relativa a una variabile $N(4,1)$ cioè una distribuzione normale che ha media 4 e varianza 1, la verde è relativa a una distribuzione normale $N(4,2)$ cioè a una distribuzione con media 4 e varianza 2. Che informazioni ci danno media e varianza in una distribuzione gaussiana? La media rappresenta l'orizzontale in cui io ho il massimo, la x , l'ascissa in cui io ho il picco di distribuzione, il punto in cui è più alta la probabilità di ottenere dei valori, o meglio il punto nell'intorno del quale è più alta la probabilità di ottenere dei valori. La media identifica la zona in cui è più alta la probabilità di ottenere dei valori, se dico che la media dell'altezza di un adulto maschio è 170,2 cm significa che attorno a quel valore ho la massima probabilità di trovare l'altezza di un individuo preso a caso. La varianza mi dice quanto attorno a questo valore centrale oscilla il mio fenomeno, se ho una varianza grande vuol dire che il mio fenomeno ha il massimo attorno a quel punto lì, ma anche 5 o 10 unità prima o dopo ho sempre una buona probabilità di trovare dei valori, se invece la mia varianza è bassa vuol dire che lì attorno ho una buona probabilità di trovare dei valori, ma se mi allontano anche di poco questa probabilità crolla quasi immediatamente. I 2 parametri della distribuzione normale mi dicono:

1. la media mi dice dove è centrata la mia distribuzione, dov'è il picco della curva
2. la varianza quanta oscillazione casuale c'è attorno a questa media, quanto attorno a questo picco la mia curva è allargata.

Questi 2 parametri sono fondamentali per confrontare qualsiasi cosa sperimentale, se voglio confrontare l'altezza degli uomini e delle donne non posso prendere un uomo e misurarlo e poi una donna e

misurarla, c'è oscillazione casuale so che posso aver trovato una differenza dovuta al caso che non rispecchia quello che è l'andamento generale. Se raccolgo 2 gruppi di 500 maschi e 500 femmine, misuro le altezze e ottengo 2 medie diverse, non mi basta questa informazione, devo sapere anche quanto i maschi e le femmine oscillano attorno alle loro altezze medie, se di molto o di poco. Questo permette di confrontare i 2 gruppi. Se ho 2 gruppi con 2 medie diverse e le oscillazioni attorno alla media sono modeste allora posso dire che i maschi hanno una altezza diversa dalle femmine, ma se le oscillazioni sono molto grandi non lo posso dire, perché non so se le medie sono dovute al caso oppure no. (La dimensione del campione è quella che mi permette o non mi permette di trovare delle differenze, l'abbiamo visto un po' con la distribuzione binomiale, la numerosità incide sulla capacità dimostrativa di un esperimento).

La distribuzione cumulativa in orizzontale ha i valori della x che in teoria dovrebbero andare da $+\infty$ a $-\infty$. Un esperimento che possiamo fare è questo: la capacità di intuire una distanza tra noi e un oggetto. Se non c'è 1 errore sistematico mi aspetto che un soggetto alcune volte dica una distanza + alta e alcune + bassa, se non ci sono errori sistematici queste previsioni hanno una distribuzione gaussiana, quindi chiedendo che distanza c'è dalla cattedra all'ultima fila, se sul grafico ho -5 invece di dire che ci sono 20 metri si è detto che ce ne sono 15, se ho +5 si è detto che ce ne sono 25. Questa è la variabile in orizzontale, in verticale ho la probabilità che il valore trovato sia inferiore o uguale a quella x , quindi se prendo -1 e vado su ho 0,18 che è la probabilità che il valore sia inferiore o uguale a -1, perché per ogni valore in orizzontale la verticale mi dice la probabilità di trovare un valore minore o uguale a quello, non mi dà la probabilità del valore perché non posso ottenerlo. Corrisponde a ciò che abbiamo visto per la distribuzione binomiale.

Se volessi descrivere le distribuzioni di probabilità di tutte le possibili variabili gaussiane dovrei costruire una curva per ogni coppia di media e varianza, e sarebbe una cosa molto briga. Si è pensato di fare così: si considera solo una coppia di distribuzione gaussiana, la $N(0,1)$ che si chiama anche **distribuzione z** o distribuzione normale standard, se ne fa il grafico, le tabelle, etc.... poi per gli altri casi con altre variabili trasformo la mia variabile in una variabile z , sottraendo la media teorica e dividendo per la deviazione standard teorica cioè per la radice quadrata della varianza teorica, σ^2 . Se sono alto 160cm che valore avrebbe la mia variabile z ? se la media teorica è 170 e la varianza teorica è 20 ho $160-170 = -10$, $-10/20 = -0,5$

Poi viene descritto anche il meccanismo inverso per passare da una variabile standard a una variabile grezza, soprattutto in psicologia l'uso dei punti z è diffusissimo, quando faccio misure psicometriche spesso queste hanno scale diverse tra di loro. Se ad es. misuro la depressione con test che hanno misure diverse, poi se li trasformo in punto z vengono ad avere tutte la stessa scala, sono centrate tutte sullo zero e tutte hanno la stessa variabilità casuale. Questo permette di standardizzare le variabili gaussiane che hanno medie e varianze diverse e quindi permette di confrontarle.

08.03.05

La distribuzione normale è una estensione al caso continuo della binomiale ed è completamente identificata da 2 parametri, media e varianza. Quanto più la varianza è bassa tanto più la curva è stretta, quindi necessariamente il picco è più elevato, perché l'area che sta sotto la curva rappresenta la probabilità di tutti gli eventi, quindi è sempre necessariamente =1, se la curva si allarga, affinché la sua area sia sempre =1 bisogna che il picco si abbassi. Lo vedo nel grafico sopra, la rappresentazione assomiglia alla distribuzione di probabilità della distribuzione normale, ma è la densità di probabilità perché con una variabile continua la distribuzione di probabilità non è calcolabile perché è sempre =0. Si è deciso di usare la densità di probabilità perché nel caso continuo è quello che assomiglia alla distribuzione di probabilità.

Il **punto z** è la trasformazione del punteggio grezzo x in un punteggio z che segue la distribuzione normale standard, il punto grezzo x deve avere una distribuzione normale, ma non a media 0 e varianza 1, con questa trasformazione portiamo una variabile gauss x con media e varianza qualsiasi a una distribuzione gaussiana con media 0 e varianza 1, ciò ci permette di utilizzare una sola distribuzione

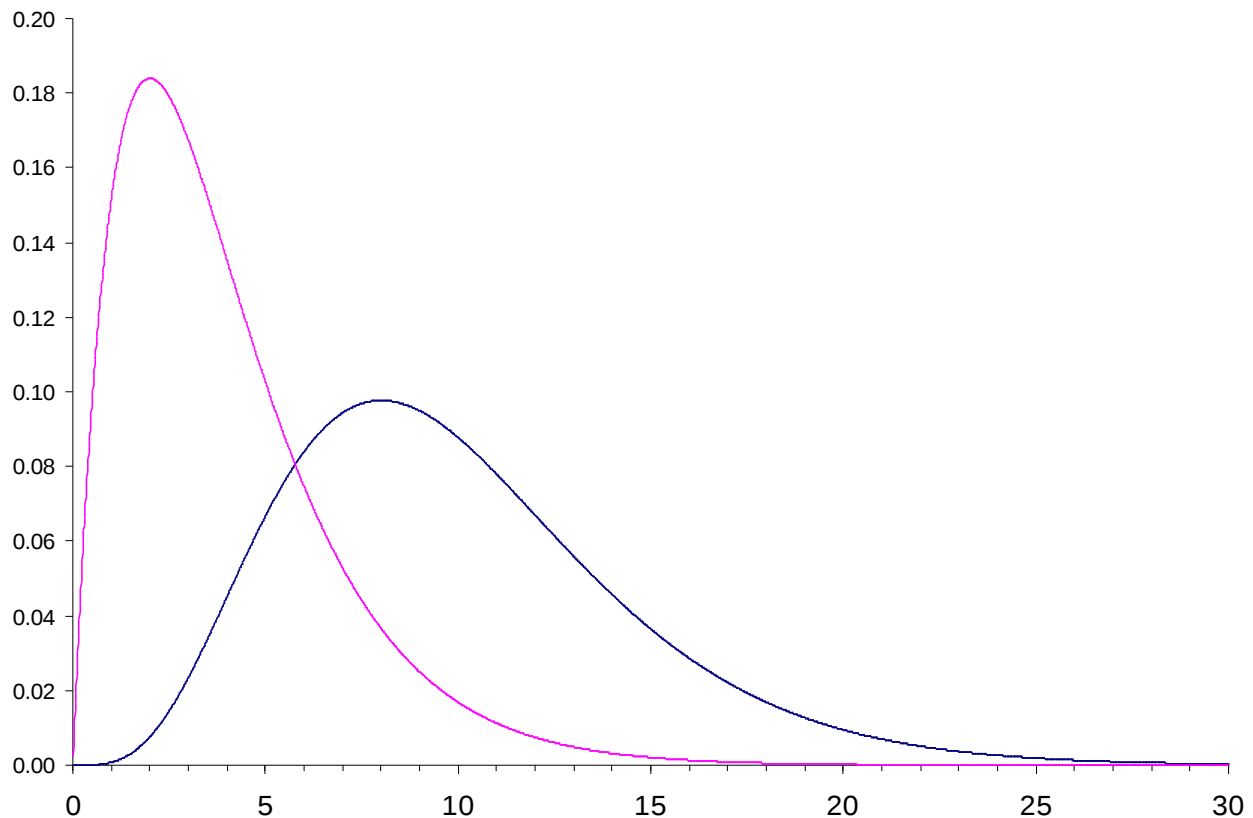
gaussiana che è la distribuzione gaussiana standard. Z è la distribuzione gauss $N(0,1)$, a noi conviene trasformare le variabili gaussiane, conoscendo μ e σ^2 riesco a trasformare in un punto z .

Per calcolare il punto z devo conoscere la media e la varianza teorica. Se ci viene sottoposto un test sul QI alla fine si ottengono per ognuno di noi 2 valori, posso fare la media, calcolare la varianza, non posso costruire il punto z ? No, la formula la posso applicare se conosco media e la varianza teorica. Nel caso del QI però conosco i valori teorici, questi vengono da esperimenti in cui il test sul QI è stato somministrato a popolazioni molto grandi, a campioni molto vasti di soggetti, per cui la media che ottengo è un'ottima stima della media teorica, che non conoscerò mai. Normalmente nel QI si usa la media 100 come se fosse una media teorica, non lo è ma è di sicuro molto vicina. In molti casi viene fornito il valore di media e deviazione standard teorici, perché il campione sul quale è stato fatto l'esperimento è vastissimo. Ad es. misuro il mio QI e ho 98, per trasformarlo in punto z so che la media teorica è 100 e la deviazione standard è 15, quindi faccio $98 - 100 = -2$, $-2/15 = -1.7$, questo è il valore z del mio QI.

Studiando test psicometrici in fondo al test spesso troveremo una formula per avere il punto z . Il valore z che ottengo alla fine di un test psicometrico è un valore che ha una distribuzione gaussiana con una media = 0 e una varianza = 1. Il vantaggio del punto z è che ha un riferimento fisso.

Un'altra distribuzione molto importante è la **distribuzione χ^2** , ki quadro. È la distribuzione che compete a una variabile che è somma di variabili z elevate al quadrato. Se ho una variabile z e la sommo a un'altra variabile z ottengo di nuovo una variabile z , la somma di variabili normali produce una variabile che è anch'essa normale. Ma se io elevo una variabile normale al quadrato, questa variabile artificiale che ottengo non ha più una distribuzione normale, se prendo una z e la elevo al quadrato, la z è una variabile che va da $-\infty$ a $+\infty$, quando la elevo al quadrato i valori negativi scompaiono, quindi ottengo una curva che va da 0 a $+\infty$, quindi non è più simmetrica, avrà un valore medio che è vicino allo zero, otterrò una curva asimmetrica, che parte da zero, ha un picco, che poi decade e non ha niente nella parte negativa. Quindi: una z sola al quadrato non ha una distribuzione gaussiana ed ha una media che ottengo sommando tutti i valori di z elevati al quadrato, cioè la media di z^2 è = alla varianza di z . Perché la varianza di z è ogni singolo z meno zero (perché la media è zero) elevato al quadrato, sommato e diviso per il numero totale, quindi la varianza di z è la somma di tutti gli z elevati al quadrato, perché sottraendo 0 è come se non togliessi niente. Nella variabile z la media = 0 è la somma dei valori al quadrato diviso n , corrisponde con la media di z^2 al quadrato, quindi una variabile z al quadrato non ha distribuzione gaussiana, so che ha media = 1 cioè uguale alla variabile z che è 1. La distribuzione χ^2 quindi è quella che compete a una o più variabili z al quadrato sommate tra di loro. Anche una z solo elevata al quadrato ha una distribuzione χ^2 . Due variabili z elevate al quadrato e sommate tra di loro hanno una distribuzione χ^2 etc.. Le distribuzioni sono diverse nei gradi di libertà, nella χ^2 questi sono dati dal numero delle variabili z che compongono la mia χ^2 . Attualmente sono state messe a punto alcune distribuzioni la cui attuazione al momento non riusciamo a capire. Anche se partiamo da variabili gaussiane, purtroppo in molti test io non posso utilizzare la distribuzione gaussiana. Ad es. i quadrati della variabile gaussiana non sono più a distribuzione gaussiana, ma a distribuzione χ^2 . Se confronto la varianza di variabili gaussiane non hanno distribuzione gaussiana, il χ^2 serve per questo, è una distribuzione di una variabile elaborata da una variabile gaussiana. Se ho una variabile gaussiana questa ha una distribuzione normale, se faccio delle operazioni su questa variabile, la variabile che ottengo è ancora una funzione gaussiana? Se è una somma sì, se è una moltiplicazione o 1 quadrato no. La maggior parte delle situazioni nei test sono su variabili gaussiane elaborate.

La forma della distribuzione χ^2 è asimmetrica, esiste solo da zero in avanti, ha una media che è = 1 se ho una sola z elevata al quadrato, è = 2 se ho due z al quadrato e sommate, = 3 se ho 3 z etc., cioè ha la media che è = ai gradi di libertà. Qui ho 2 distribuzione χ^2 , una rosa con 4 gradi di libertà, l'altra blu con 10 gradi di libertà.



Cosa succede aumentando i gradi di libertà? Sappiamo che i gradi di libertà aumentano quando aumentano le z al quadrato e vengono sommate, e la media è = ai gradi di libertà, quindi si sposta verso i valori positivi. Qui il picco non è la media, perché media la ottengo sommando tutti i valori prima della media e tutti quelli dopo. Siccome qui la curva non è simmetrica il picco non è la media, però è vicino alla media, quindi crescendo la media cresce il picco, la curva tende a spostarsi verso i valori positivi.

Nella distribuzione gaussiana la media la moda e la mediana coincidono. Perché è una curva simmetrica, siccome la media sta sul picco perché alla destra e sinistra del picco io ho due andamenti esattamente simmetrici, allora quel picco lì è anche la mediana, ha alla sua sinistra esattamente il 50% dei casi, e se è il valore più probabile è anche la moda. Questo non succede sempre, generalmente non succede come ad es. nella distribuzione χ^2 , il picco non ha a sin e dex lo stesso numero di casi e non ha nemmeno la stessa probabilità perché l'area di sinistra è diversa da quella di destra. Il picco coincide con la moda perché la moda significa il picco di probabilità, ma non con la media e mediana, anche se è vicino a tutte e due. L'abbassamento significa che se sposti il picco più in avanti a sinistra hai intervallo maggiore, per fare in modo che l'area sottesa alla curva sia =1 la curva si deve necessariamente abbassare. La curva si allarga perché parte sempre da zero. Se i gradi di libertà sono = infinito, come sarà la curva? Si abbasserà molto ma non sarà una linea. Tutte le distribuzioni al crescere dei gradi di libertà all'infinito tendono alla distribuzione più casuale possibile, la gaussiana. La distribuzione χ^2 tende a una gaussiana al crescere dei gradi di libertà all'infinito. Necessariamente la distribuzione χ^2 va da 0 a + infinito, se a un certo punto diventa una gaussiana bisogna che il picco sia a metà, e come fa se va da 0 a + infinito? Al tendere infinito dei gradi di libertà il picco va a infinito mezzi, e infinito mezzi dista da 0 e da + infinito esattamente della stessa quantità (dal punto di vista matematico è così).

Anche la distribuzione χ^2 tende alla gaussiana con il crescere all'infinito dei gradi di libertà, nonostante però resti sempre definita solo da 0 a + infinito riesce a diventare simmetrica perché

possiamo immaginare il picco a infinito mezzi e ci sono da 0 a infinito mezzi gli stessi punti che ci sono da infinito mezzi a + infinito. Una curva si può schiacciare tanto mantenendo sempre il profilo, la dimensione diminuisce ma ciò non significa che deve appiattirsi e diventare una retta. Ciò che ci interessa è che quando crescono molto i gradi di libertà, siccome la distribuzione χ^2 tende ad essere come una gaussiana, potremmo usare anche una gaussiana. (oltre un certo numero di gradi di libertà, oltre i 120 gli uso sempre la gaussiana, anche se a 100 usiamo una gauss va bene comunque)

Un'altra distribuzione che si usa molto in statistica è la **distribuzione t**, che è quella che compete a un'altra variabile artificiale ottenuta dal rapporto tra una variabile normale e la radice quadrata di una variabile χ^2 divisa per i suoi gradi di libertà. Perché è stato calcolato questo rapporto? Per un motivo essenziale: se ho una variabile gauss di cui conosco media e varianza teoriche calcolo il punto z, quando lo conosco ho dei riferimenti precisi. Ad es. se io voglio sapere qual è l'intervallo in cui cadono il 95% dei casi per dire che al di fuori di lì le cose non tornano più, (come abbiamo fatto per la binomiale), se voglio calcolare questo intervallo è facilissimo, so già che la variabile z ha il limite del 95% a $+0 - 1,96$. Se ottengo un QI di un soggetto e lo trasformo in punto z e ho 2,4 so già che questo è un mostro d'intelligenza, non ho bisogno di vedere la media della popolazione perché una volta trasformato in punteggio z so già che il 95% della popolazione sta tra $-1,96$ e $+1,96$. Quindi se devo fare un test, come quello simulato per la moneta, con una variabile gaussiana, se posso trasformarlo in 1 punteggio z so subito se quel valore è dentro o fuori ai limiti del 95% perché conosco il limite fisso della variabile z che è $+0 - 1,96$, se i valori sono fuori da qs intervallo posso dire che il mio soggetto ha un QI che è al di fuori del 95% della popolazione. Il problema è che non sempre conosco media e varianza teorica. Se ho media e varianza teorica questi valori sono dei valori fissi. Se io invece misuro la media e la varianza sperimentalmente sul mio campione, questi non sono più dei valori fissi, sono a loro volta delle variabili casuali, cambiano da campione a campione.

Vediamo la formula: qui ho una variabile gaussiana che è la x meno una costante, diviso una costante che non mi cambiano la normalità della cosa, quindi il punto z nonostante parta da una gaussiana a cui sottrae un valore fisso e lo divide per un valore fisso rimane a distribuzione gaussiana, perché se sottraggo o moltiplico o divido una variabile gaussiana per un valore fisso il tipo di distribuzione resta sempre gaussiano, varia la larghezza, l'altezza etc.. ma il picco di distribuzione resta sempre gaussiano. Se invece uso la media e la varianza teoriche quando applico la formula per la variabile z non ho più due valori fissi di una variabile gaussiana ma ho 3 variabili casuali, x è gaussiana, la media è gaussiana, la deviazione standard non ha distribuzione gaussiana perché è la radice quadrata della varianza, (la varianza è somma di quadrati, è una variabile sempre positiva e non può avere una distribuzione gauss, anzi una variabile somma di quadrati di gaussiane ha una distribuzione χ^2). Questa formula che con una variabile z mi dava una distribuzione normale $N(0,1)$ se la applico con una media e una varianza misurata sul mio campione non ha più una distribuzione gaussiana ma ha una distribuzione t, perché è il rapporto tra una variabile gaussiana (perché la x è una gaussiana) la x medio (che è gaussiana) e la deviazione standard che ha dimensione di un χ^2 sotto segno di radice.

La distribuzione t è quella che useremo nei test statistici tutte le volte - quasi sempre- che non conosceremo media e varianza teorica. Il T- test o test di student è uguale a quello della distribuzione gaussiana con la differenza che uso la distribuzione t invece della gaussiana. (Questo quasi sicuramente c'è nell'esame). Anche la distribuzione t si calcola con certe funzioni matematiche, che a noi non interessano però.

La **distribuzione f**, di Fisher. È la distribuzione che compete al rapporto di 2 variabili χ^2 sia sopra che sotto divise per i rispettivi gradi di libertà. Se io voglio confrontare la variabilità di un punteggio in un corso con quello di un altro corso lo posso fare confrontando la varianza dei punteggi dei due corsi. Con che distribuzione le posso confrontare? Ne faccio il rapporto e le confronto con la distribuzione f perché le varianze hanno distribuzione χ^2 e il loro rapporto ha distribuzione f, in molti test la distribuzione f serve a confrontare il rapporto di 2 varianze, le variabili di partenza sono gaussiane le loro varianze no, hanno una distribuzione χ^2 , il rapporto di 2 varianze ha una distribuzione f. La distribuzione t ha una

distribuzione simmetrica come la gaussiana, la f ha una distribuzione asimmetrica come le variabili χ^2 e solo positiva. La variabile t al numeratore ha una gaussiana i cui valori possono essere sia negativi che positivi, la variabile f ha $2 \chi^2$ che possono essere solo positivi, quindi la f è una variabile che può andare da 0 a + infinito mentre la t da - infinito a + infinito.

Domande tipiche d'esame:

Perché in statistica vengono definite più distribuzioni di probabilità? Perché dipende da ciò che dobbiamo confrontare, se non ho dei valori di riferimento per la media e per la varianza, pur considerando la variabile misurata una variabile gaussiana, non possiamo utilizzare la distribuzione gaussiana. Se voglio confrontare varianze la distribuzione gaussiana non ci serve. Le diverse distribuzioni di probabilità servono ad affrontare o condizioni in cui noi non conosciamo i valori teorici oppure condizioni in cui andiamo a confrontare dei parametri che sono l'elaborazione delle variabili misurate gaussiane.

La distribuzione t è a campana come la gaussiana è solo più allargata, perché è più allargata della distribuzione normale? (NB. Sotto copre sempre un'area = 1) Perché nella distribuzione gaussiana ho una sola variabile casuale che è la x , nella variabile t ho 2 variabili casuali sopra perché anche la media varia casualmente, quindi ho 2 valori che cambiano (quello del soggetto e quello della media), ma soprattutto ho il denominatore -che ha un effetto amplificativo- che è casuale, quindi in una distribuzione t ho 3 sorgenti di variabilità, la x , la media e la varianza. Necessariamente la casualità, l'imprevedibilità del mio valore sarà più elevata, perché è basato non su un valore casuale ma su 3, che possono variare indipendentemente.

09.03.05

Le distribuzioni sono oggetti matematici nati per scopi pratici.

Quando si parla di statistica si intendono cose molto diverse.

La statistica descrittiva che vuole solo descrivere il campione che io ho analizzato. Ad es. la statistica descrittiva può essere la frequenza delle diverse altezze nell'aula, cioè dire quanti sono alti 160cm, quanti 170cm etc.. Questa statistica non vuole fare nessuna osservazione generale, descrive il campione, è relativa a questo anno di corso. È una descrizione contingente del campione che sto osservando, quindi è composta da un insieme di parametri che servono a descrivere questo campione, non la popolazione generale, non un principio, non una tendenza. Ad es. i risultati elettorali sono in genere dati con la statistica descrittiva. Nelle proiezioni c'è un broglio, non ci dicono che l'affidabilità ha la previsione che non è fatta su tutto il campione, qualche giornale comincia a dare la "forchetta" cioè dice: "sul piccolo campione esaminato ho ottenuto il 15% però potrebbe oscillare dal 9% al 20%". Questa informazione mi dice che l'affidabilità ha quella misura. Il 15% che oscilla tra il 14% e il 16% è una cosa, il 15% che oscilla tra il 2% e il 35% è un'altra cosa. L'affidabilità delle misure è scarsissima nelle proiezioni, il campione è troppo piccolo. Il valore è una previsione che ha un grado di affidabilità che bisogna esplicitare, non è un dato determinato.

La statistica descrittiva è una parte della statistica, è la meno informativa ed è la parte che non ci interessa, quando facciamo sperimentazione non intendiamo raccontare come si comportavano i 50 soggetti che abbiamo esaminato rispetto ad un indice di depressione, bensì vogliamo dire che quell'indice di depressione cambia a seconda di certi fattori. L'abbiamo stimato sul nostro campione ma intendiamo estenderlo a tutta la popolazione. Dal campione ci interessa poter estrapolare un meccanismo generale che riguardi tutti, ma a questo punto è un processo di statistica inferenziale, che è quella che interessa a noi. Quando facciamo ricerca si intende dire che si è preso un campione che ha dato dei risultati estensibili alla popolazione globale.

La differenza tra statistica descrittiva e inferenziale è questa. La descrittiva riguarda solamente e decisamente quel campione, ha un suo valore, ma descrive solo una situazione. L'inferenziale invece estende i risultati alla popolazione globale, prende un campione a caso e se in questo campione si verifica la mia legge allora è credibile che la mia legge funzioni, non che sia vero in assoluto.

(Ricordiamoci che non siamo in grado di dimostrare la verità, non abbiamo il termine di confronto per

sapere se quello che abbiamo trovato corrisponde alla verità oppure no. Possiamo solo ipotizzare che una certa cosa funzioni.) La statistica inferenziale vuole solo dimostrare che è credibile, non certo, che ci sia una relazione tra un certo fenomeno e un altro fenomeno. In psicologia qualche volta si distinguono esperimenti correlazionali, cioè quelli in cui faccio vedere la correlazione tra una variabile ed un'altra, ed esperimenti veri dove modifico una variabile e vado a vederne gli effetti. Il prof (e molti altri) non è d'accordo con questa visione perché tutte le cose che noi vediamo sono cose di correlazione, noi sappiamo ad es. che correlato all'inverno c'è l'influenza ma al fatto che l'inverno la faccia venire, non ci crede nessuno. Una relazione di causa effetto non è dimostrabile sperimentalmente. Popper, lavori fondamentali 1° metà 900, qual è il compito della scienza? Non può essere la ricerca della verità perché non c'è nessun modo per sapere se una certa legge corrisponde alla verità, non abbiamo il meccanismo di controllo per farlo, per misurare la verità di qualcosa. Una legge scientifica per lui è una legge generale falsificabile. Compito della scienza è di formulare delle leggi falsificabili, che non vuole dire fare leggi false, ma che se per caso sono false ci sia il modo di vederlo. Popper fa un lavoro di umiltà, la scienza non può dire la sua parola su tutto. La scienza formula delle ipotesi che poi conferma con esperimenti anche se mai in modo definitivo, e formula leggi scientifiche che sono falsificabili. Ciò che non è falsificabile non fa parte della scienza, l'essere umano non fa parte della scienza perché il suo credo non è falsificabile.

Falsificabile vuol dire che esiste un esperimento che può rendere quella cosa falsa, se questo esperimento non esiste è un'altra cosa, non è scientifico ma non è da buttar via. Uno scienziato deve avere la posizione di umiltà di Popper deve sapere che ciò a cui arriva non è la verità assoluta ma può essere tenuto perché può servire in certe cose (i farmaci aiutano in alcuni casi). Ciò non vuol dire che se non possiamo conoscere la verità qualsiasi cosa diciamo va bene, noi dobbiamo seguire certi paradigmi che sono condivisi dal genere umano, e soprattutto dal mondo scientifico, attraverso questi paradigmi noi confermiamo in qualche modo una nostra ipotesi o non la confermiamo e i risultati che otteniamo sono disponibili a tutto il mondo scientifico. La scienza va avanti con dei processi inferenziali che servono a dimostrare delle ipotesi scientifiche (scientifiche nel senso che sono falsificabili.) Calcolare una media non vuol dire fare della statistica. La certezza, l'assoluto, la verità non è cosa che ci appartiene, soprattutto allo scienziato, gli appartiene aver dimostrato credibili certe ipotesi, poi se queste sono da considerarsi vere questo è da vedere. Nemmeno la fisica è una scienza esatta, le sue leggi generali sono andate in crisi, non erano vere in certe condizioni, si è cominciato a mettere in dubbio che la scienza fosse un modo per conoscere la verità. All'inizio del 900 molti filosofi ne hanno discusso, la scienza non è produttrice di verità. La statistica lavora per affermare con un certo grado di attendibilità una certa ipotesi. Il margine d'errore zero non esiste, ecco perché nella binomiale prendiamo il 95%. Il fatto che il margine d'errore 0 non esista non significa che qualsiasi dimostrazione è buona, c'è un paradigma di dimostrazione scientifica che deve essere dimostrato, si condivide ciò che è stato dimostrato con dei meccanismi che io trovo accettabili per la dimostrazione. Il ruolo delle distribuzioni è che le uso per rendere più o meno credibile un certo risultato sperimentale.

La statistica descrittiva dal punto vista della scienza ha questo ruolo: raccoglie i dati di un certo campione, li descrive e al massimo permette la formulazione di una ipotesi scientifica. La descrizione dei risultati in sé e per sé però non dimostra nulla, può essere la base per la formulazione di una ipotesi, che è ciò che succede normalmente nella ricerca, si raccolgono i dati, si pensava venisse una certa cosa che invece non viene e ciò dà lo stimolo per dimostrare un'altra ipotesi.

La statistica inferenziale sostanzialmente studia degli effetti generali, in campo medico si occupa ad es. dell'effetto di un certo farmaco su certe malattie, in campo psicologico si occupa dell'effetto di un certo ambiente socio economico su certi disturbi psicologici, dell'effetto dell'età su un certo stato depressivo, la statistica inferenziale si occupa di effetti. Ciò che dimostriamo sono degli effetti, non sono delle caratteristiche di un gruppo o di un altro, se dico che c'è una certa differenza tra maschi e femmine, in realtà dimostro che c'è un certo effetto su una variabile che misuro legata al sesso di appartenenza, dimostro che il sesso di appartenenza influisce, questo è un effetto

generale. Dico che c'è un effetto legato al sesso di appartenenza, se confronto le prestazioni nella velocità di risposta ad uno stimolo fra maschi e femmine e trovo una differenza, qs differenza la attribuisco al sesso di appartenenza, ed è un effetto generale, non dico che ho trovato dei maschi rimbambiti e delle femmine sveglie, se ho raccolto un campione casuale concludo che appartenere ad un sesso o ad un altro incide sulla velocità di risposta ad uno stimolo. La statistica inferenziale ha anche altri scopi marginali, meno forti di questo e meno scientifici, ad es. ha lo scopo creare un meccanismo di classificazione dei soggetti, crea meccanismi con cui posso classificare dei soggetti in un gruppo o in un altro in base a certe cose. È una caratteristica secondaria come quella della previsione, qs due caratteristiche non fanno parte del settore psicologico, la prima riguarda più la sociologia, la psicologia è più attenta agli effetti, la seconda riguarda più il mondo dell'economia. La psicologia si occupa sostanzialmente di studiare gli effetti produce delle terapie, dei sistemi di protezione dell'individuo, crea degli ambienti adatti a non sviluppare certe patologie etc... Gli effetti generali che si studiano hanno poi una ricaduta pratica, vengono studiati in pratica. Il seguente grafico è la struttura principale della dimostrazione sperimentale.



È un punto nodale, non averla presente significa non capire i test statistici che faremo. Come procede la dimostrazione sperimentale? Procede così, c'è una idea generale, cioè scientifica, che non si sa come nasce, Popper ha lavorato su questo e conclude che non si può sapere, è un meccanismo che sfugge alla nostra mente, non abbiamo gli elementi per capirlo. I passi successivi invece li analizzo, ho l'idea scientifica, generale e falsificabile (generale significa che deve riguardare un effetto e non un singolo caso), e da questa idea scientifica nasce l'ipotesi sperimentale. Se io sostengo che i tempi di reazione diminuiscono con l'età e prendo un campione a caso, un gruppo di giovani e un gruppo di anziani e vado a misurarne i tempi di reazione a uno stimolo di qualche tipo io posso dimostrare l'idea generale, formulo l'ipotesi sperimentale cioè che il gruppo dei giovani abbia dei tempi di reazione diversi, non più bassi ma diversi. L'ipotesi è l'ipotesi di un effetto non il senso dell'effetto. Poi arrivo a formulare il punto centrale della dimostrazione sperimentale cioè l'ipotesi nulla, cioè H_0 , (hypothesis) che significa ipotesi di assenza di effetti. Noi formuliamo prima l'ipotesi sperimentale che è l'esistenza di un effetto, poi l'ipotesi nulla, che è l'ipotesi che se fosse vera falsificherebbe quella sperimentale, quindi è la complementare dell'ipotesi sperimentale, siccome l'ipotesi sperimentale è l'ipotesi di esistenza di un effetto, quella nulla è necessariamente l'ipotesi dell'assenza di questo effetto. Questo perché non siamo in grado di dimostrare delle cose ma siamo in grado di falsificarle. Ad es. non sono in grado di dimostrare l'uguaglianza di 2 fogli bianchi perché anche se trovo dei valori uguali non è detto che i 2 fogli lo siano, perché l'uguaglianza che io trovo è relativa al sistema di misure che io ho

usato. Se prendo un sistema di misura più raffinato del doppio decimetro è possibile che i fogli siano diversi. Siccome non posso mai arrivare alla precisione assoluta, io non sono in grado di dimostrare l'uguaglianza di nulla, però se io prendo 2 fogli e li metto uno sopra l'altro e uno è 2 cm più corto, su questo non c'è ombra di discussione, sono diversi. In sostanza siamo in grado di dimostrare delle diversità. Così funziona la teoria statistica, non è in grado di dimostrare la presenza di un effetto ma è in grado di falsificarne l'assenza. Ad es. se io voglio dimostrare che la capacità di apprendere la matematica dipende dai sistemi di insegnamento, il modo più diretto per farlo è questo: insegnare a due gruppi diversi con due metodi diversi, se a fine corso ottengo dei risultati diversi non c'è dubbio che c'è un effetto della tecnica di insegnamento sulla capacità di apprendimento. Nel dimostrare il falso, cioè l'ipotesi nulla io vengo ad affermare la presenza dell'effetto. Siccome dimostro che l'assenza dell'effetto è insostenibile perché l'esperimento ha dimostrato che non è così, io ho in questo modo dimostrato la mia idea. La scienza sperimentale funziona con un meccanismo tipo la dimostrazione per assurdo, non è in grado di dimostrare direttamente l'idea ma è in grado di falsificare il fatto che non ci sia l'effetto. Se dimostro che ci sono differenze nella capacità di risposta ad uno stimolo tra un gruppo di anziani e uno di giovani, non è più sostenibile l'ipotesi nulla, cioè che l'età non ha effetto, un effetto ce l'ha perché ho ottenuto risultati diversi. (L'ipotesi nulla era che l'età non incideva sui tempi di risposta). Bisogna trovare il campione idoneo per fare l'esperimento. È importante che sia casuale perché devo dimostrare al mondo che l'età influisce sui tempi di risposta, e devo prendere un campione casuale per farlo. Questa tecnica sperimentale è quella nata attorno al 1500 introdotta sostanzialmente da Galileo, che si rifaceva a questo principio: un esperimento è dimostrativo di una idea se ripetuto in qualsiasi laboratorio con le stesse metodiche produce gli stessi risultati. Questo era il suo concetto di prova sperimentale che si è evoluto nel concetto che abbiamo adesso, noi diciamo: se prendo un campione casuale e trovo queste differenze non c'è nessun motivo per cui se prendo un altro campione casuale non abbia le stesse differenze. Il concetto di riproducibilità che è alla base della teoria sperimentale di Galileo è stato trasformato nel concetto di casualità, cioè se faccio una certa prova su un campione casuale non c'è nessun motivo per cui non debba ottenere le stesse cose con un altro campione, soprattutto per il modo in cui faccio l'esperimento. Prima formulo l'ipotesi e una volta fatto questo eseguo l'esperimento e lo eseguo per confermare la mia ipotesi. La credibilità di un risultato sperimentale segue dalla procedura, credo ad un esperimento se qualcuno mi dice cosa ci si aspetta di trovare e se questa cosa poi la trovo in un campione casuale. Crediamo a una verifica il cui risultato è stato ipotizzato prima di fare l'esperimento. Teoricamente un campione è casuale se tutti i soggetti che appartengono alla popolazione globale hanno la stessa probabilità di essere inclusi nel campione. Stiamo parlando di effetti, cioè di meccanismi che funzionano su tutta la popolazione globale (la mia popolazione globale può anche essere tutti gli italiani). Gli elementi che compongono il campione devono essere scelti a caso nel senso che qualsiasi italiano deve avere la stessa probabilità di essere scelto. Il campione casuale è teorico, mai si può realizzare un campione così fatto. Questo non vuol dire che siccome non posso avere la perfezione qualsiasi imperfezione va bene allo stesso modo. Dobbiamo cercare di limitare l'imperfezione, mi posso avvicinare il più possibile al campione casuale e soprattutto so quali sono i criteri con i quali posso giudicare l'imperfezione del mio campione. Se ad es. somministro un questionario agli studenti questi non sono un campione casuale, è poco rappresentativo delle età, dei diversi livelli scolastici, siamo in grado di vedere il difetto del campione. Individuato un campione corretto eseguo il mio esperimento, raccolgo i dati e comincio a descrivere i risultati che ho ottenuto, ancora non ho dimostrato nulla, comunque io abbia scelto i gruppi ottengo medie diverse, so che c'è l'oscillazione casuale che fa sì che tra i diversi individui ci siano differenze. A questo punto per poter dimostrare la mia idea applico un test statistico che tende a questo scopo: dimostrare che l'ipotesi nulla è falsa, sono fatti esclusivamente per dimostrare questo i test statistici, per falsificare l'ipotesi nulla. (È ciò che abbiamo fatto con l'esperimento del lancio della moneta a 12 prove nella binomiale.) Considero falsa l'ipotesi nulla, non la dimostro falsa, non esistono ipotesi nulle false, siamo noi che le consideriamo false, (se esistessero, allora esisterebbe la

verità.) le riteniamo talmente improbabili da considerarle false. Il risultato del test statistico può essere di 2 tipi, o falsifico l'ipotesi nulla, oppure non la falsifico e non concludo nulla. Il difetto di alcuni ricercatori è che non falsificando l'ipotesi nulla arrivano a certe conclusioni che il test statistico non includeva, perché se non falsifico l'ipotesi nulla non vuol dire che è vera, ma che io non sono riuscito a falsificarla, o perché era vera, o perché era falsa ma il mio strumento non era idoneo. Nella statistica sperimentiamo una cosa completamente nuova rispetto alla matematica, l'asimmetria, la non biunivocità delle conclusioni. L'esperimento divide l'ipotesi in 2 parti una dove è falsa l'ipotesi nulla e una dove non sono stato in grado di falsificarla. Se non falsifico l'ipotesi nulla può darsi che abbia scelto un campione troppo eterogeneo o piccolo, rifaccio l'esperimento, faccio misure più accurate, scelgo un campione più eterogeneo e più numeroso, in questo modo magari riesco a falsificare l'ipotesi nulla, modifico. Il test statistico ha due possibili uscite, o la dimostrazione dell'ipotesi generale o l'abbandono, oppure ci si rende conto che si è usato un sistema di misura o un campione inadeguato, quindi modifichiamo e ripetiamo. Se ripeto l'esperimento lo devo fare con un campione nuovo, non devo usare le stesse persone. Se il mio campione è di 50 persone, ma non è stato sufficiente, si moltiplica, si va a 100 o 150 quando si ripete l'esperimento.