Statistica I, Laurea triennale in Ing. Gestionale, a.a. 2011/12
Registro delle lezioni

Lezione 1 (28/9, ore 11:30). Vedere la registrazione di Barsanti, disponibile alla pagina

http://users.dma.unipi.it/barsanti/statistica_2011/index.html.

Introduzione al corso. Introduzione alla statistica, statistica descrittiva. Frequenza assoluta e relativa. Modi di rappresentare graficamente i dati. Aggregazione di dati in classi. Campione, media campionaria e sue proprietà. Media pesata. Mediana e moda.

Lezione 2 (28/9, ore 16:30). Materiale e comunicazioni alla pag. di F. Flandoli:

http://www2.ing.unipi.it/~a008484/dispStatisticaGestionaleTriennale.html.

Inizio dello studio del capitolo 3, sui fondamenti del calcolo delle probabilità. Vengono date le definizioni di universo, eventi, eventi elementari o esiti; essi vengono esemplificati mediante affermazioni riguardanti il possibile risultato del lancio di un dato. Vengono discusse le operazioni (unione, intersezione e complementare) su eventi, collegandole alle operazioni logiche sulle relative affermazioni. Viene poi definita la probabilità (interpretata intuitivamente col concetto di massa). Vengono illustrazione alcune regole che discendono dagli assiomi, su $P(A^c)$ e $P(A \cup B)$ anche nel caso non disgiunto (interpretate tramite l'idea di massa). Viene spiegato che in uno spazio Ω finito è sufficiente avere le probabilità degli esiti, dalle quali si calcolano le probabilità di tutti gli eventi.

Lezione 3 (1/10, ore 10:30). Densità di probabilità ed esempio di probabilità associata, densità esponenziali, densità discrete ed esempio di probabilità associata, densità geometriche.

Probabilità condizionale, interpretazione grafica, osservazione sul fatto che è una probabilità (mentre $P(A|B)$ non ha particolari proprietà rispetto a B). Formula di fattorizzazione, dimostrazione, rappresentazione tramite albero degli eventi. Esercizio sulle vendite di vino, che esemplifica il concetto di probabilità condizionale e la formula di fattorizzazione.

Definizione di indipendenza tra eventi, a partire dalla condizione $P(A|B) = P(A)$.

Lezione 4 (5/10, ore 11:30). Vedere la registrazione di Barsanti. Varianza e deviazione standard campionaria con esempi e tecniche per abbreviare il calcolo. Percentili campionari e quartili. Box plot. Formula di Chebyshev con dimostrazione ed esempi di applicazioni. Campioni normali, asimmetrici, bimodali. Campioni bivariati e coefficiente di correlazione campionario.

Lezione 5 (5/10, ore 16:30). Dimostrazione del fatto che l'indipendenza tra due eventi A e B equivale alla proprietà $P(A) = P(A|B)$ ("A non dipende da B"), ed equivale a $P(B) = P(B|A)$ ("B non dipende da A"); queste ultime quindi equivalgono, pur non essendo simmetriche.

Esercizio. Un sistema S è composto da tre sottosistemi S_1, S_2, S_3. La probabilità che S_i si rompa è p_i. Supponendo che il funzionare o meno dei sottosistemi sia indipendente, calcolare la probabilità che si rompa S.

Osservazione: la proprietà che A e B siano indipendenti equivale a: A^c e B^c sono indipendenti; ed anche a: A e B^c sono indipendenti; ed infine a: A^c e B sono indipendenti.

Lezione 6 (8/10, ore 10:30). Formula di Bayes, un esempio, linguaggio cause-effetti, probabilità a priori ed a posteriori delle cause; problema della causa più probabile, anche nel caso di cause a priori equiprobabili, interpretazione del calcolo sull'albero degli eventi.

Funzione di distribuzione cumulativa nel caso continuo e discreto, sue proprietà, grafici, esempio dell'esponenziale. Concetto di variabile aleatoria, connessione con le densità e la cumulativa.

Lezione 7 (12/10, ore 11:30). Cardinalità di un insieme. Spazi di esiti equiprobabili; $P(A) = \frac{|A|}{|\Omega|}$ (rapporto tra il numero dei casi favorevoli e quello dei casi possibili).

Elementi di calcolo combinatorico. Principio di enumerazione. Fattoriale e coefficienti binomiali. $|\{\text{permutazioni}\}| = n!$, $|\{\text{disposizioni}\}| = n(n-1)\cdots(n-k+1)$ (verificati col principio di enumerazione). Le funzioni biunivoche $f : \{1, ..., n\} \to \{1, ..., n\}$ sono le permutazioni; le funzioni iniettive $f : \{1, ..., k\} \to \{1, ..., n\}$ sono le disposizioni. Teorema: $|\{\text{combinazioni}\}| = \binom{n}{k}$. Dimostrazione usando classi di equivalenza di disposizioni. Esercizi sulle combinazioni (incluso esempio 3.5.4 del libro, detta distribuzione *ipergeometrica*).

Lezione 8 (12/10, ore 16:30). Variabili aleatorie discrete, loro valori, loro densità discreta e suo grafico. Esempi: Bernoulli e binomiale (verifica della somma 1 col binomio di Newton). Grafici tipici della binomiale. Enunciato del teorema che lega Bernoulli e binomiali. Esempio della banca e dei correntisti, in cui si vuole calcolare la probabilità che il numero di correntisti che si presenta superi una certa soglia.

Esercizi per casa su fattorizzazione e Bayes: es. 10 (H2) del 28/05/2010;

es. 1 del 20/07/2010; es. 1.i del 29/06/2010. Si consiglia anche la lettura degli esempi 3.5.3, 3.7.4, 3.7.7.

Lezione 9 (15/10, ore 10:30). Descrizione del foglio Excel da preparare a casa, contenente i valori della binomiale, densità e cumulata, che serve a calcolare probabilità e soglie. Ogni studente deve predisporre un tale foglio personalizzandolo con la risoluzione di un problema.

Dimostrazione rigorosa del teorema che lega Bernoulli e binomiali.

Valor medio aritmetico (empirico) di un campione e valor medio (valore atteso) teorico di una v.a. Legame tra i due: detti a_k i valori della v.a. X e $p(a_k)$ le loro probabilità, la def. di media è $E[X] = \sum a_k p(a_k)$, mentre vale

$$\overline{x} = \frac{x_1 + ... + x_n}{n} = \sum a_k \widehat{p}_k$$

dove $\widehat{p}_k$ è la frequenza relativa empirica con cui si osserva a_k nel campione.

Esempi elementari di calcolo del valor medio. Interpretazione grafica, casi simmetrici.

Esercizio per casa (probabilità che al massimo uno su sette abbia successo).

Lezione 10 (18/10, ore 11:30). Vedere la registrazione di Barsanti. Esercizi su probabilità discreta.

Lezione 11 (19/10, ore 11:30). Proprietà del valor medio: positività, monotonia, linearità (senza dimostrazione). Valor medio di una Bernoulli e di una binomiale.

V.a. di Poisson, verifica della somma 1, teorema degli eventi rari. Valor medio di una Poisson.

Schema di Bernoulli (schema successo-insuccesso). Numero di successi su n prove: v.a. $B(n,p)$; oppure approssimativamente $\mathcal{P}(\lambda)$ se n è grande e p è piccolo (successi rari); $\lambda = np =$ numero medio di successi. Primo istante di successo: geometrica (con la convenzione che si parte dal tempo 0).

Trasformazioni di v.a. discrete, valor medio di una trasformazione, esempi ed esercizi.

Varianza e deviazione standard. $Var[X] = E[X^2] - \mu_X^2$. Esempio della Bernoulli.

Lezione 12 (19/10, ore 16:30). Elenco provvisorio di formule generali (es. linearità del valor medio e $E[XY] = E[X]E[Y]$ nel caso di v.a. indipendenti che vedremo in seguito) e di valori medi particolari (media e varianza di Bernoulli, binomiali, Poisson, dando per buone le formule delle varianze che vedremo). Esercizi vari simili a quelli dei compiti.

Lezione 13 (25/10, ore 10:30). Proprietà della varianza.

V.a. indipendenti: proprietà del valor medio del prodotto (senza dimostrazione), conseguenze sulla varianza della somma. Definizione di covarianza e di v.a. scorrelate, legame con l'indipendenza, teorema generale sulla varianza della somma.

Media e varianza di Bernoulli (dalla definizione) e binomiale (usando le proprietà di valor medio e varianza di una somma). Concentrazione delle binomiali, problema dei 10000 utenti.

Funzione generatrice dei momenti, teorema di legame coi momenti, teorema sulla generatrice della somma di v.a. indipendenti. Generatrice di Bernoulli, binomiale, Poisson e calcolo della varianza di una v.a. di Poisson. Idee sul fatto che varianza e generatrice di una Poisson si possono dedurre intuitivamente dal teorema degli eventi rari.

Media e varianza di una v.a. esponenziale, dalla definizione. Generatrice di v.a. con densità, esempio delle esponenziali, per le quali non è finita per tutti i valori del parametro.

Elenco di densità, media, varianza e generatrice di una gaussiana $N(\mu, \sigma^2)$, per ora senza tutti i calcoli.

Nota: se si escludono alcune domande basate su quantili e cumulativa gaussiana e relative tavole, e sul teorema limite centrale (TLC), si possono svolgere praticamente tutti gli esercizi teorici delle prove scritte (tali prove sono divise in una serie di domande teoriche ed una di domande di statistica, di solito raggruppate in due esercizi separati, quindi facilmente identificabili).

Lezione 14 (26/10, ore 11:30). Vedere la registrazione di Barsanti. Esercizi su valore atteso, probabilità di eventi e funzioni caratteristiche per v.a. bernoulliane, binomiali, poissoniane e gaussiane.

Lezione 15 (2/11, ore 11:30). Vedere la registrazione di Barsanti. Disuguaglianze di Markov e di Chebishev, convergenza in probabilità, legge debole dei grandi numeri. Esercizi su probabilità discreta, variabili Bernoulliane e binomiali.

Lezione 16 (8/11, ore 10:30-12:30). Vedere la registrazione di Barsanti. Standardizzazione e riproducibilità della variabile gaussiana. Uso delle tabelle della funzione di ripartizione della gaussiana standard. Quantili gaussiani (standard e non).

Somma di n gaussiane indipendenti e identicamente distribuite.

Teorema limite centrale. Enunciato ed esempio di utilizzo.

Lezione 16 (8/11, ore 12:30-13:30). Dimostrazione del teorema limite centrale. Esercizi dai compiti.

Lezione 17 (9/11, ore 11:30). Inizio della Statistica Matematica. Definizione di campione $X_1, ..., X_n$ e di $\overline{X}$, legame e differenze rispetto a $x_1, ..., x_n$ e di $\overline{x}$, anche esemplificato in un problema di vendite e dati giornalieri, passati e futuri. Teorema su media e varianza di $\overline{X}$, e sulla sua gaussianità in ipotesi di gaussianità delle X_i.

Legami tra $\overline{X}$ (o $\overline{x}$) e μ. Oltre ad uno già visto alla lezione 9, il teorema precedente indica il legame $E\left[\overline{X}\right] = \mu$ (stimatore corretto) e $E\left[\left|\overline{X} - \mu\right|^2\right] = \frac{\sigma^2}{n}$ che quantifica la vicinanza tra $\overline{X}$ e μ (interpretazione grafica). Si pone un problema più stringente di quantificazione: discutere la validità di una disuguaglianza del tipo

$$\left|\overline{X} - \mu\right| \leq \delta.$$

Viene discussa intuitivamente, poi viene calcolata $P\left(\left|\overline{X} - \mu\right| \leq \delta\right)$ trovando

$$P\left(\left|\overline{X} - \mu\right| \leq \delta\right) = \Phi\left(\frac{\delta}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{\delta}{\sigma}\sqrt{n}\right).$$

Esercizio per casa: risolvere un problema inverso, cioè di trovare δ tale che

$$P\left(\left|\overline{X} - \mu\right| \leq \delta\right) = 0.95.$$

Lezione 18 (15/11, ore 10:30). Intervalli di confidenza. Basandosi sul risultato della lezione precedente, si risolve il problema inverso, di trovare il numero δ tale che $P\left(\left|\overline{X} - \mu\right| \leq \delta\right) = 1-\alpha$. Si trova $\delta = \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}$. Di passaggio, viene sottolineato che l'equazione $\Phi(x) = \beta$ equivale a $x = q_\beta$.

Risultato espresso nella forma: $\mu = \overline{X} \pm \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}$ con probabilità $1-\alpha$. Quando dai dati sperimentali si trova $\overline{x}$, scriveremo: $\mu = \overline{x} \pm \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}$ a livello di confidenza $1-\alpha$.

Analisi della formula $\delta = \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}$ al variare di α ed n.

Esercizio dai compiti, in cui si applica la formula, incontrando il problema della scelta di α, che necessita di averne capito il significato, e del calcolo di σ.

Problema dell'ipotesi di gaussianità, aiuto dal TLC per quanto riguarda la formula dell'intervallo di confidenza.

Problema del calcolo della quantità da tenere in magazzino, differenze e legami relativamente al problema precedente (calcolo della soglia più cautelativo, usando come media il valore estremo dell'intervallo di confidenza).

Lezione 19 (16/11, ore 11:30). Soluzione molto dettagliata dell'esercizio 1 del 17/9/2011, domande i, ii, iii, iv.

Lezione 20 (22/11, ore 10:30). Teoria dei test statistici (verifica delle ipotesi). Problema, partendo dall'esercizio 1 del 17/9/2011, domanda v, della differenza tra un nuovo campione ed uno vecchio, o meglio (ipotizzando che il campione vecchio avesse determinato i parametri con esattezza), della differenza tra un nuovo campione ed una vecchia media μ_0. Elementi di teoria: ipotesi nulla $\mathcal{H}_0$, regione di rifiuto e di accettazione, probabilità α che la grandezza statistica su cui si basa il test cada nella regione di rifiuto quando $\mathcal{H}_0$ è vera, schema del test. Esemplificazione con caso del test sulla media. La base è il teorema sugli intervalli di confidenza. Riformulazione canonica del test (dati $\to z$, $\alpha \to q_{1-\frac{\alpha}{2}}$, confronto tra $|z|$ e $q_{1-\frac{\alpha}{2}}$).

Valore p, definito come il valore di demarcazione tra gli α per cui il test rifiuta $\mathcal{H}_0$ (si verifica che sono quelli che verificano $\alpha > p$) e quelli per cui il test non rifiuta $\mathcal{H}_0$ (quelli che verificano $\alpha < p$). Calcolo del valore p nel test per la media. Test in quest'ottica: dati $\to z \to p$, valutazione della sua piccolezza.

Lezione 21 (23/11, ore 10:30). Valore p definito come probabilità che i valori teorici siano più estremi di quelli sperimentali: $P(|Z| > |z|)$ nel caso del test per la media. Giustificazione intuitiva e significato negli esempi. Esercizio 1 di Luglio 2011, domande (i) e (iii).

Lezione 22 (29/11, ore 10:30). Errori di prima e seconda specie, loro probabilità, potenza di un test, esempio di controllo della qualità.

Lezione 23 (30/11, ore 11:30). Test unilaterali, modalità, convenienza, esercizi, p-value.

Lezione 24 (6/12, ore 10:30). Esercizi su stime della media e intervalli di confidenza, test per la media, soglie.

Lezione 25 (7/12, ore 11:30). Vedere la registrazione di Barsanti. Variabili aleatorie chi quadro, legame con la varianza campionaria, elementi di statistica legati alle chi quadro.

Lezione 25 (13/12, ore 10:30). Esercizi proposti di statistica.

Lezione 26 (14/12, ore 11:30). Esercizi proposti di statistica.