PRIMA PROVA INTERMEDIA DI STATISTICA
CLEA (COD. 5047/4038/371/377)
3 Novembre 2004

Cognome *Nome*

Numero di matricola

COMPITO A1

Ai fini della valutazione si terrà conto solo ed esclusivamente di quanto riportato negli appositi spazi. Al termine della prova, è OBBLIGATORIO consegnare il presente foglio ed il foglio di brutta (DI CUI NON SI TERRÀ CONTO AI FINI DELLA VALUTAZIONE).

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

I primi **8** gran premi della stagione di Formula1 del 2003 hanno dato i risultati che riportiamo in tabella (*notare che sono anche state riportate le somme dei valori e le somme dei quadrati dei valori*):

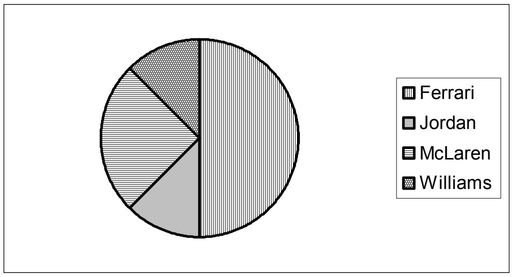
GP	Gran Premio disputato
SQUADRA	squadra di appartenenza del vincitore del GP
GOMME	gomme utilizzate dal vincitore del GP
TEMP	temperatura media dell'asfalto
PTI_M	punti totalizzati da squadre che utilizzano gomme Michelin
PTI_B	punti totalizzati da squadre che utilizzano gomme Bridgestone
PTI_F	punti totalizzati dalla squadra Ferrari
PTI_ML	punti totalizzati dalla squadra McLaren

GP	SQUADRA	GOMME	TEMP	PTI_M	PTI_B	PTI_F	PTI_ML
Australia	McLaren	Michelin	38	31	8	5	6
Malesia	McLaren	Michelin	40	25	14	11	10
Brasile	Jordan	Bridgestone	13	22	17	0	13
San Marino	Ferrari	Bridgestone	24	22	17	16	12
Spagna	Ferrari	Bridgestone	28	22	17	16	0
Austria	Ferrari	Bridgestone	25	18	21	16	12
Monaco	Williams	Michelin	27	32	7	7	10
Canada	Ferrari	Bridgestone	12	25	14	?	?
		Somme	207	197	115		
		Somme dei quadrati	6071	5011	1813		

1. (2 punti) Come si vede nella tabella, ci sono alcuni dati mancanti per il Gran Premio del Canada.
a. Sapendo che la media punti della McLaren (la cui variabile di riferimento è PTI_ML) è di 9,5 punti a Gran Premio, ricavare i punti conquistati dalla squadra nel GP del Canada.
b. Sapendo che la mediana dei punti Ferrari (PTI_F) registrati nei primi 8 Gran Premi è di 12,5 punti, ricavare quanti punti ha conquistato invece la Ferrari.

a. (1 punto)	**b. (1 punto)**
9,5=(6+10+13+12+0+12+10+x)/8, x=76-63. Quindi ***x=13***	Ordiniamo i dati. La Me sarà la media tra i valori corrispondenti alla 4° e alla 5° osservazione. (11+x)/2=12,5; quindi ***x=14***

2. (2 punti) Rappresentare con un grafico opportuno il carattere SQUADRA e indicare tutte le misure di sintesi calcolabili, il loro valore e significato.



Misure di sintesi (1 punto)
Moda= "Ferrari".
La moda è la modalità più frequente ed è l'unica misura di sintesi utilizzabile per caratteri qualitativi nominali.

3. (4 punti) Negli ambienti tecnici, si ha l'idea che le gomme Bridgestone abbiano migliori prestazioni a temperature più basse. Ipotizzando quindi una relazione lineare tra TEMP e PTI_B verificare se i punti totalizzati da Bridgestone dipendono dalla temperatura dell'asfalto e in che modo. In particolare:
a) Fornire l'equazione della retta di regressione, sapendo che la somma dei prodotti tra PTI_B e TEMP è pari a 2851.

Retta (2 punti)

PTI_B= 18,886 -0,1743 TEMP

Covarianza(TEMP; PTI_B)=(2851/8-25,875*14,375) = 356,375 - 371,9531 = -15,5781
Varianza(TEMP)= (6071/8-25,875*25,875) = 758,875-669,5156= 89,3594

b) Se si utilizzasse la retta di regressione appena calcolata, a quale temperatura le squadre che utilizzano le Bridgestone non conquisterebbero nemmeno un punto dei 39 in palio?

Previsione (1 punto)

108,3534

Infatti si risolve l'equazione: 0 = 18,886-0,1743*temp. Quindi, temp=(-18,886)/(-0,1743)

c) Valutare con un indice opportuno la bontà della retta di regressione stimata in a) e dire se si possa ritenere quindi affidabile la previsione fatta al punto b).

Bontà retta (1 punto)
$\rho = -0{,}3686$ **; R^2=0,1359**: l'R^2 è basso e quindi la previsione risulta poco affidabile.
[Varianza(PTI_B)=19,9844]

4. (4 punti) A partire dalla tabella contenente i dati relativi ai primi 8 Gran Premi della stagione 2003: ricavare una tabella per valutare se esiste indipendenza statistica tra SQUADRA vincitrice del Gran Premio e GOMME scelte da quella squadra, definire (scrivendone la formula) un indice di connessione e infine calcolarne il valore.

Tabella (1 punto)

	Bridgestone	Michelin
McLaren	0	2
Jordan	1	0
Ferrari	4	0
Williams	0	1

Indice (1 punto)

$$\varphi^2 = \sum_i \sum_j \frac{\left(p_{X,Y}(x_i, y_j) - p_X(x_i) \cdot p_Y(y_j)\right)^2}{p_X(x_i) \cdot p_Y(y_j)}$$

$\tilde{\varphi}^2 = \varphi^2 / \min(h-1, k-1)$ dove h, k sono le dimensioni della tabella.

Valore (2 punti)
In questo caso $\varphi^2 = \tilde{\varphi}^2 = 1$. Infatti, conoscendo la squadra, sappiamo con precisione che gomme vengono usate (perfetta connessione unilaterale).

5. (3 punti) Osservando i punti conquistati da 4 squadre minori in 4 Gran Premi la situazione è la seguente.

GP \ Punti per squadra	BAR	Sauber	Toyota	Jordan
Malesia	2	1	0	0
Brasile	3	4	0	10
San Marino	1	0	0	0
Austria	5	0	0	0

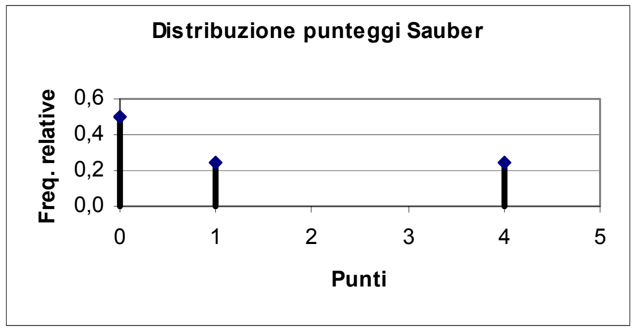
1. Quale squadra presenta il maggior scarto quadratico medio nella distribuzione dei punti?
2. Fornire la rappresentazione grafica dei punteggi della Sauber nei 4 GP considerati.

Maggior sqm (2 punti)

La distribuzione con maggior sqm è quella relativa alla ***Jordan***.

$$\begin{bmatrix} \sigma_{BAR} = 1{,}4790 \\ \sigma_{Sauber} = 1{,}6394 \\ \sigma_{Toyota} = 0 \\ \sigma_{Jordan} = 4{,}3301 \end{bmatrix}$$

Grafico(1 punto)



6. (3 punti) Valutare con un indice opportuno la concentrazione della distribuzione dei *39 punti in palio* conquistati dalle varie squadre nel Gran Premio d'Australia (vedi tabella) e scrivere la formula dell'indice spiegando cosa indicano nella formula le quantità coinvolte:

Punti	Numero squadre
0	4
3	1
5	1
6	1
9	1
16	1

Formula indice (1 punto)

$$R^* = \frac{\sum_{i=1}^{k-1}(F_i^* - Q_i^*)}{\sum_{i=1}^{k-1} F_i^*}$$

Dove F_i^* rappresenta la frazione di soggetti che detiene la quota relativa Q_i^*, e k è il numero di modalità osservate.

Calcolo indice (2 punti)

Punti	Numero squadre	F_i^*	Q_i^*
0	4	0,4444	0,0000
3	1	0,5556	0,0769
5	1	0,6667	0,2051
6	1	0,7778	0,3590
9	1	0,8889	0,5897
16	1	1,0000	1,0000
Somma (tranne ultima riga)		3,3334	1,2307

Quindi $R^* = \frac{3{,}3334 - 1{,}2307}{3{,}3334} = \boldsymbol{\mathit{0{,}6308}}$

7. (2 punti) La probabilità che la Ferrari riesca a salire sul podio è di 0,75. Supponendo che le gare dei Gran Premi siano tra loro indipendenti e che le chance della Ferrari siano sempre le stesse per ogni Gran Premio, *impostare i calcoli* per calcolare la probabilità che la Ferrari salga sul podio 10 volte su 16 (*senza calcolarne il valore finale*).

Probabilità (2 punti)

P(Ferrari sul podio)=0,75 $\Leftarrow$ probabilità di successo costante in ogni prova
Prove indipendenti $\Bigg\} \Rightarrow$

La variabile aleatoria che rappresenta il numero di vittorie su 16 prove ha ***distribuzione binomiale***!

$$P(X=10)=\binom{16}{10}(0{,}75)^{10}(0{,}25)^{6}$$

8. (2 punti) La probabilità che Michael Schumacher vinca un Gran Premio quando il compagno di squadra Barrichello riesce a completare la gara è 0,4. Al contrario, se Barrichello non conclude la gara, la probabilità che Schumacher vinca scende a 0,25. Sapendo che Barrichello riesce a portare a termine la corsa l'80% delle volte, qual è la probabilità che Schumacher vinca un determinato Gran Premio?

Probabilità (2 punti)

Definiamo A = Schumacher vince GP; B = Barrichello finisce gara
P(A|B) = 0,4; P(A| non B) = 0,25; P(B) = 0,8 e quindi P(non B) = 0,2.
Quindi: P(A) = P(A|B)P(B) + P(A|non B)P(non B) = 0,32 + 0,05 = **0,37**.

9. (2 punti) Enunciate il teorema di Bayes, specificando le ipotesi:

Teorema (2 punti)

Si consideri uno spazio campionario Ω e un insieme di eventi $A_1, A_2, \ldots, A_n$, partizione di Ω (ovvero: $\bigcup_{j=1}^{n} A_j=\Omega$ e $A_i \cap A_j=\varnothing$ per ogni $i \neq j$, con $P(A_j)>0$ per ogni j). Quindi, preso un qualsiasi altro evento B con $P(B)>0$, si ha che

$$P(A_i|B)=\frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n}P(B|A_j)P(A_j)}$$

PRIMA PROVA INTERMEDIA DI STATISTICA
CLEA (COD. 5047/4038/371/377)
3 Novembre 2004

Cognome *Nome*

Numero di matricola

COMPITO A2

Ai fini della valutazione si terrà conto solo ed esclusivamente di quanto riportato negli appositi spazi. Al termine della prova, è OBBLIGATORIO consegnare il presente foglio ed il foglio di brutta (DI CUI NON SI TERRÀ CONTO AI FINI DELLA VALUTAZIONE).

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

Gli ultimi 8 gran premi della stagione di Formula1 del 2003 hanno dato i risultati che riportiamo in tabella (*notare che sono anche state riportate le somme dei valori e le somme dei quadrati dei valori*):

GP	Gran Premio disputato
SQUADRA	squadra di appartenenza del vincitore del GP
GOMME	gomme utilizzate dal vincitore del GP
TEMP	temperatura media dell'asfalto
PTI_M	punti totalizzati da squadre che utilizzano gomme Michelin
PTI_B	punti totalizzati da squadre che utilizzano gomme Bridgestone
PTI_F	punti totalizzati dalla squadra Ferrari
PTI_ML	punti totalizzati dalla squadra McLaren

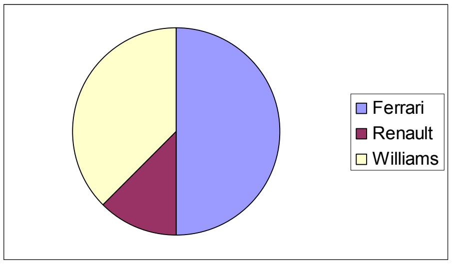
GP	SQUADRA	GOMME	TEMP	PTI_M	PTI_B	PTI_F	PTI_ML
Europa	Williams	Michelin	14	26	13	?	?
Francia	Williams	Michelin	24	31	8	8	9
Gran Bretagna	Ferrari	Bridgestone	23	23	16	15	10
Germania	Williams	Michelin	38	36	3	2	8
Ungheria	Renault	Michelin	36	38	1	1	12
Italia	Ferrari	Bridgestone	22	20	19	16	5
Usa	Ferrari	Bridgestone	9	17	22	10	8
Giappone	Ferrari	Bridgestone	24	19	20	11	14
		Somme	190	210	102		
		Somme dei quadrati	5182	5956	1744		

2. (2 punti) Come si vede nella tabella, ci sono alcuni dati mancanti per il Gran Premio d'Europa.
a. Sapendo che la media punti della McLaren (la cui variabile di riferimento è PTI_ML) è di 8,75 punti a Gran Premio, ricavare i punti conquistati dalla squadra nel GP d'Europa.
b. Sapendo che la mediana dei punti Ferrari (PTI_F) registrati negli ultimi 8 Gran Premi è di 10 punti, ricavare quanti punti ha conquistato invece la Ferrari.

a. (1 punto)	**b. (1 punto)**
8,75=(x+9+10+8+12+5+8+14)/8; x=70-66. Quindi ***x=4***	Ordiniamo i dati. La Me sarà la media tra i valori corrispondenti alla 4° e alla 5° osservazione. (10+x)/2=10, quindi ***x=10***

2. (2 punti) Rappresentare con un grafico opportuno il carattere SQUADRA e indicare tutte le misure di sintesi calcolabili, il loro valore e significato.

Grafico (1 punto)



Misure di sintesi (1 punto)

Moda= "Ferrari".

La moda è la modalità più frequente ed è l'unica misura di sintesi utilizzabile per caratteri qualitativi nominali.

3. (4 punti) Negli ambienti tecnici, si ha l'idea che le gomme Bridgestone abbiano migliori prestazioni a temperature più basse. Ipotizzando quindi una relazione lineare tra TEMP e PTI_B verificare se i punti totalizzati da Bridgestone dipendono dalla temperatura dell'asfalto e in che modo. In particolare:

a) Fornire l'equazione della retta di regressione, sapendo che la somma dei prodotti tra PTI_B e TEMP è pari a 1988.

Retta di regressione (2 punti)

PTI_B= 28,1638 – 0,649 TEMP

Covarianza(TEMP; PTI_B)=(1988/8-23, 75*12, 75) = 248,5 – 302,8125 = -54,3125

Varianza(TEMP)= (5182/8-23, 75*23, 75) = 647,75 – 564,0625 = 83,6875

b) Se si utilizzasse la retta di regressione appena calcolata, a quale temperatura le squadre che utilizzano le Bridgestone non conquisterebbero nemmeno un punto dei 39 in palio?

Previsione (1 punto)

43,3957

Infatti si risolve l'equazione: 0 = 28,1638 - 0,649*temp. Quindi, temp=(28,1638)/(0,649)

c) Valutare con un indice opportuno la bontà della retta di regressione stimata in a) e dire se si possa ritenere quindi affidabile la previsione fatta al punto b).

Bontà retta (1 punto)

$\rho = -0{,}7974$ **; R^2=0,6358**: l'R^2 è relativamente alto e quindi la previsione risulta abbastanza affidabile.

[Varianza(PTI_B)=55,4375]

4. (4 punti) A partire dalla tabella contenente i dati relativi agli ultimi 8 Gran Premi della stagione 2003: ricavare una tabella per valutare se esiste indipendenza statistica tra SQUADRA vincitrice del Gran Premio e GOMME scelte da quella squadra, definire (scrivendone la formula) un indice di connessione e infine calcolarne il valore.

Tabella (1 punto)

	Bridgestone	Michelin
Renault	0	1
Ferrari	4	0
Williams	0	3

Indice (1 punto)

$$\varphi^2 = \sum_i \sum_j \frac{\left(p_{X,Y}(x_i, y_j) - p_X(x_i)\cdot p_Y(y_j)\right)^2}{p_X(x_i)\cdot p_Y(y_j)}$$

$\tilde{\varphi}^2 = \varphi^2 / \min(h-1, k-1)$ dove h, k sono le dimensioni della tabella.

Valore (2 punti)

In questo caso $\varphi^2 = \tilde{\varphi}^2 = 1$. Infatti, conoscendo la squadra, sappiamo con precisione che gomme vengono usate (perfetta connessione unilaterale).

5. (3 punti) Osservando i punti conquistati da 4 squadre minori in 4 Gran Premi la situazione è la seguente.

GP \ Punti per squadra	BAR	Sauber	Jaguar	Toyota
Ungheria	0	0	3	0
Italia	3	0	2	0
Usa	0	10	0	0
Giappone	8	0	1	0

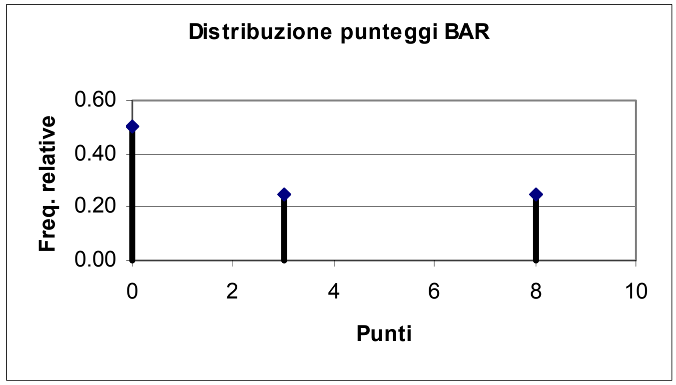
3. Quale squadra presenta il maggior scarto quadratico medio nella distribuzione dei punti?
4. Fornire la rappresentazione grafica dei punteggi della BAR nei 4 GP considerati.

Maggior sqm (2 punti)

La distribuzione con maggior scarto quadratico medio è quella relativa alla ***Sauber***.

$$\begin{bmatrix} \sigma_{BAR} = 3{,}2692 \\ \sigma_{Sauber} = 4{,}3301 \\ \sigma_{Jaguar} = 1{,}1180 \\ \sigma_{Toyota} = 0 \end{bmatrix}$$

Grafico(1 punto)



6. (3 punti) Valutare con un indice opportuno la concentrazione della distribuzione dei *39 punti in palio* conquistati dalle varie squadre nel Gran Premio d'Ungheria (vedi tabella qui sotto) e scrivere la formula dell'indice spiegando cosa indicano le quantità coinvolte (sia nella formula che nel calcolo).

Punti	Numero squadre
0	4
3	1
5	1
6	1
9	1
16	1

Formula indice (1 punto)

$$R^* = \frac{\sum_{i=1}^{k-1}(F_i^* - Q_i^*)}{\sum_{i=1}^{k-1} F_i^*}$$

Dove F_i^* rappresenta la frazione di soggetti che detiene la quota relativa Q_i^*, e k è il numero di modalità osservate.

Calcolo indice (2 punti)

Punti	Numero squadre	F_i^*	Q_i^*
0	4	0,4444	0,0000
3	1	0,5556	0,0769
5	1	0,6667	0,2051
6	1	0,7778	0,3590
9	1	0,8889	0,5897
16	1	1,0000	1,0000
Somma(tranne ultima riga)		3,3334	1,2307

Quindi $R^* = \frac{3{,}3334 - 1{,}2307}{3{,}3334} = \boldsymbol{0{,}6308}$

7. (2 punti) La probabilità che la Ferrari riesca a salire sul podio è di 0,6. Supponendo che le gare dei Gran Premi siano tra loro indipendenti e che le chance della Ferrari siano sempre le stesse per ogni Gran Premio, *impostare i calcoli* per calcolare la probabilità che la Ferrari salga sul podio 9 volte su 16 (*senza calcolarne il valore finale*).

Probabilità (2 punti)
P(Ferrari sul podio)=0,6 $\Leftarrow$ probabilità di successo costante in ogni prova
Prove indipendenti
La variabile aleatoria che rappresenta il numero di vittorie su 16 prove ha ***distribuzione binomiale***!

$$P(X=9)=\binom{16}{9}(0,6)^9(0,4)^7$$

8. (2 punti) La probabilità che Michael Schumacher vinca un Gran Premio quando il compagno di squadra Barrichello riesce a completare la gara è 0,4. Si sa inoltre che Barrichello riesce a portare a termine la corsa l'80% delle volte, e che Schumacher vince un Gran Premio con probabilità 0,6. Sapendo che Schumacher ha vinto il Gran Premio, qual è la probabilità che anche Barrichello abbia terminato la corsa?

Probabilità (2 punti)
Definiamo A = Schumacher vince GP; B = Barrichello finisce gara
P(A|B) = 0,4; P(B) = 0,8 e P(A) = 0,6.
Quindi: P(B|A) = P(A|B)P(B)/P(A) = $(0,4\cdot 0,8)/0,6$ = 0,32 / 0,6 = **0,5333**.

9. (2 punti) Enunciate il teorema delle probabilità totali, specificando le ipotesi:

Teorema (2 punti)
Si consideri uno spazio campionario Ω e un insieme di eventi A_1, A_2, …, A_n, partizione di Ω (ovvero: $\bigcup_{j=1}^{n} A_j=\Omega$ e $A_i\cap A_j=\varnothing$ per ogni $i\neq j$, con $P(A_j)>0$ per ogni j). Quindi, preso un qualsiasi altro evento B, si ha che

$$P(B)=\sum_{j=1}^{n}P(B|A_j)P(A_j)$$