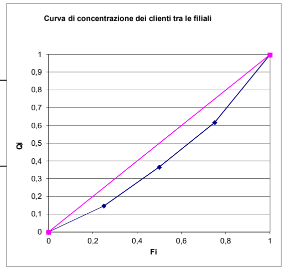
Esercizio 1

Il numero di clienti di 4 diverse filiali di una certa banca, situate in quartieri simili di una stessa città, è riportato nella tabella che segue:

Filiali	Numero di clienti
A	1.780
B	2.040
C	3.110
D	1.200

a) Si calcoli l'indice di concentrazione e si commenti il risultato.

Numero di clienti	A_i	Q_i	F_i	$F_i - Q_i$
1.200	1.200	0,148	0,250	0,102
1.780	2.980	0,367	0,500	0,133
2.040	5.020	0,617	0,750	0,133
3.110	8.130	1,000	~~1,000~~	~~0,000~~
8.130			1,500	0,368



$$G = \frac{\sum_{i=1}^{n-1}(F_i - Q_i)}{\sum_{i=1}^{n-1} F_i} = \frac{0,368}{1,5} = 0,246$$

Il grado di concentrazione è piuttosto basso.

b) Si rappresenti la curva di Lorenz.

c) Quale dovrebbe essere il numero di clienti in ciascuna filiale in caso di equidistribuzione?

$$\bar{x} = \frac{1}{n}\sum_{i}^{n} x_i = \frac{1200 + 1780 + 2040 + 3110}{4} = \frac{8130}{4} = 2032,5$$

Esercizio 2

In un certo anno accademico si è osservata la seguente distribuzione di laureati rispetto al voto di laurea e all'ateneo frequentato. Con riferimento agli atenei di una certa regione si sono rilevati i seguenti dati:

	Ateneo			
Voto di laurea	A	B	C	Totale
Sotto il 100	940	821	1100	2861
Da 100 a 105	1121	903	950	2974
Sopra il 105	376	428	306	1110
Totale	2437	2152	2356	6945

a) Si calcoli il voto medio alla laurea, condizionatamente all'ateneo.

$$\bar{y}_{x=A} = \frac{82,5 \cdot 940 + 102,5 \cdot 1121 + 108 \cdot 376}{2437} = 95,63$$

$$\bar{y}_{x=B} = \frac{82,5 \cdot 821 + 102,5 \cdot 903 + 108 \cdot 428}{2152} = 95,96$$

$$\bar{y}_{x=C} = \frac{82,5 \cdot 1100 + 102,5 \cdot 950 + 108 \cdot 306}{2356} = 93,88$$

	Ateneo			
Voto di laurea	A	B	C	Totale
Sotto il 100	940	821	1100	2861
Da 100 a 105	1121	903	950	2974
Sopra il 105	376	428	306	1110
Totale	2437	2152	2356	6945

$\bar{y}_{x=A} = 95,63$; $\bar{y}_{x=B} = 95,96$; $\bar{y}_{x=C} = 93,88$

b) Sulla base dei risultati al punto a) si dica se c'è indipendenza in media del voto medio dall'ateneo frequentato, motivando la risposta.

Essendo le medie condizionate diverse tra loro, dobbiamo concludere che l'ateneo frequentato influisce sul voto medio di laurea.

c) Si calcoli un opportuno indice per misurare il grado di dipendenza in media del voto di laurea dall'ateneo frequentato e se ne interpreti il risultato.

$$\bar{y} = \frac{95,63 \cdot 2437 + 95,96 \cdot 2152 + 93,88 \cdot 2356}{6945} = 95,14 \quad \text{oppure} \quad \bar{y} = \frac{82,5 \cdot 2861 + 102,5 \cdot 2974 + 108 \cdot 1110}{6945} = 95,14$$

$$\sigma^2_{Medie(Y|X)} = \frac{(95,63 - 95,14)^2 \cdot 2437 + (95,96 - 95,14)^2 \cdot 2152 + (93,88 - 95,14)^2 \cdot 2356}{6945} = 0,84$$

$$\sigma^2_Y = \frac{(82,5 - 95,14)^2 \cdot 2861 + (102,5 - 95,14)^2 \cdot 2974 + (108 - 95,14)^2 \cdot 1110}{6945} = 115,45$$

$\eta^2 = \frac{\sigma^2_{Medie(Y|X)}}{\sigma^2_Y} = \frac{0,84}{115,45} = 0,007$ Il grado di dipendenza in media del voto dall'ateneo è bassissimo.

Esercizio 3

Un certo processo produttivo produce 10 pezzi in un'ora. Il tasso di difettosità è dell'1%.

a) Quale è la probabilità che, osservando la produzione di un'ora, non si riscontri alcun pezzo difettoso?

$X = \{\text{Numero di pezzi difettosi su 10 prodotti}\}$ $\quad X \sim Bin(n\,;\,\pi)$

$n = 10$

$\pi = 0,01$

$$P(X=0) = \binom{n}{x}\pi^x(1-\pi)^{n-x} = \binom{10}{0}0,01^0(1-0,01)^{10-0} = \frac{10!}{0!\,(10-0)!} \cdot 1 \cdot 0,99^9 = 0,904$$

b) Quale è invece la probabilità di osservare al massimo 1 pezzo difettoso?

$$P(X \le 1) = P(X=0) + P(X=1) = \binom{10}{0}0,01^0(1-0,01)^{10-0} + \binom{10}{1}0,01^1(1-0,01)^{10-1} = 0,904 + 0,091 = 0,995$$

c) Considerato che il processo produttivo è in funzione 24 ore su 24, qual è il numero atteso di pezzi prodotti difettosi in una giornata?

$Y = X_1 + X_2 + \cdots + X_{24}$

$Y = \{\text{Numero di pezzi difettosi su 240 prodotti}\}$ $\quad Y \sim Bin(240\,;\,\pi)$ $\quad E(Y) = 240 \cdot 0,01 = 2,4$

d) Qual è la probabilità che in un giorno vengano prodotti più di 2 pezzi difettosi?

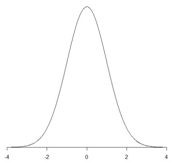
$$P(Y > 2) = 1 - [P(Y=0) + P(Y=1) + P(Y=2)]$$
$$= 1 - \left[\binom{240}{0}0,01^0(1-0,01)^{240-0} + \binom{240}{1}0,01^1(1-0,01)^{240-1} + \binom{240}{2}0,01^2(1-0,01)^{240-2}\right]$$
$$= 1 - [0,090 + 0,217 + 0,262] = 0,431$$

Esercizio 4

Il numero di nuovi contratti RC auto stipulati in un anno da una assicurazione si può considerare distribuito come una variabile aleatoria normale di media 450 e varianza pari a 300.
Assumendo che la distribuzione del numero di nuovi contratti rimanga approssimativamente invariata da un anno all'altro e che il numero di contratti stipulati in un anno sia indipendente da quello dei contratti stipulati in un altro anno:

a) qual è la probabilità che l'anno prossimo il numero di nuovi contratti sia inferiore a 400?

$X = \{\text{Numero di contratti stipulati in un anno}\}$



$X \sim N(450\,;300)$

$$P(X<400) = P\left(\frac{X-\mu}{\sigma} < \frac{400-\mu}{\sigma}\right) = P\left(\frac{X-450}{\sqrt{300}} < \frac{400-450}{\sqrt{300}}\right) =$$
$$P(Z<-2{,}89) = P(Z>2{,}89) = 1 - P(Z<2{,}89) = 1-\Phi(2{,}89) = 1-0{,}99087 = 0{,}00913$$

b) qual è la probabilità che l'anno prossimo il numero di nuovi contratti sia compresa tra 460 e 490?

$$P(460<X<490) = P\left(\frac{460-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{490-\mu}{\sigma}\right)$$
$$= P\left(\frac{460-450}{\sqrt{300}} < \frac{X-450}{\sqrt{300}} < \frac{490-450}{\sqrt{300}}\right) = P(0{,}58<Z<2{,}31)$$
$$= P(Z<2{,}31) - P(Z<0{,}58) = \Phi(2{,}31) - \Phi(0{,}58) = 0{,}98956 - 0{,}71904 = 0{,}27052$$

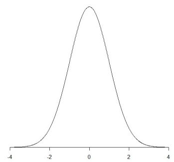
c) qual è la probabilità che in 3 anni il numero complessivo di contratti sia inferiore a 1300?

$Y = \{\text{Numero di contratti stipulati in tre anni}\}$ $\qquad Y = X_1 + X_2 + X_3$

$Y \sim N(3\cdot 450\,;3\cdot 300) \sim N(1350\,;900)$

$$P(Y<1300) = P\left(\frac{Y-\mu}{\sigma} < \frac{1300-\mu}{\sigma}\right) = P\left(\frac{X-1350}{\sqrt{900}} < \frac{1300-1350}{\sqrt{900}}\right) =$$
$$P(Z<-1{,}67) = P(Z>1{,}67) = 1-P(Z<1{,}67) = 1-\Phi(1{,}67) = 1-0{,}95254 = 0{,}04746$$

d) qual è la probabilità che su 5 anni, il numero medio annuo di nuovi contratti sia inferiore a 440?



$Q = \{\text{Numero medio annuo di contratti stipulati su 5 anni}\}$ $\qquad Q = \dfrac{X_1+X_2+X_3+X_4+X_5}{5}$

$$Q \sim N\left(450\,;\frac{300}{5}\right) \sim N(450\,;60)$$

$$P(Q<440) = P\left(\frac{Q-\mu}{\sigma} < \frac{440-\mu}{\sigma}\right) = P\left(\frac{Q-450}{\sqrt{60}} < \frac{440-450}{\sqrt{60}}\right) =$$
$$P(Z<-1{,}29) = P(Z>1{,}29) = 1-P(Z<1{,}29) = 1-\Phi(1{,}29) = 1-0{,}90147 = 0{,}098$$

Esercizio 5

In un'indagine pre-elettorale, viene rilevata l'intenzione di voto in un campione di 1000 persone. Di queste, 520 dichiarano che voteranno per il partito A.

a) Sulla base del campione, si costruisca un intervallo di confidenza al livello del 99% per la proporzione di persone che voteranno per il partito A nell'intera popolazione.

$X \sim Ber(\pi)$ $\qquad \bar{X} \sim N\left(\pi;\frac{\pi(1-\pi)}{n}\right)$

$1-\alpha = 0{,}99 \quad \rightarrow \quad z_{\alpha/2} = 2{,}58$

$$P\left(-z_{\alpha/2} \le \frac{\bar{X}-\pi}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \le z_{\alpha/2}\right) = 1-\alpha$$

$$\bar{x} = \frac{520}{1000} = 0{,}52$$

$$\left[\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\,;\bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] = \left[0{,}52 - 2{,}58\sqrt{\frac{0{,}52\,(1-0{,}52)}{1000}}\,;0{,}52 - 2{,}58\sqrt{\frac{0{,}52\,(1-0{,}52\,)}{1000}}\right]$$
$$= [0{,}479\,;0{,}561]$$

Al livello di confidenza del 99%, la proporzione di elettori che voteranno per il partito A è compresa tra il 47,9% e il 56,1%.

b) A parità di altre condizioni, si determini la numerosità campionaria necessaria a ridurre di un quarto l'attuale ampiezza dell'intervallo.

$$A^* = \frac{3}{4}A \quad \rightarrow \quad 2\cdot z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n^*}} = \frac{3}{4}\cdot 2\cdot z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \quad \rightarrow \quad \sqrt{\frac{1}{n^*}} = \frac{3}{4}\sqrt{\frac{1}{n}} \quad \rightarrow \quad \frac{1}{n^*} = \frac{9}{16}\frac{1}{n} \rightarrow n^* = \frac{16}{9}n$$
$$= \frac{16}{9}1000 \approx 1778$$

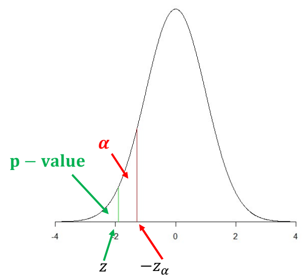
c) Si verifichi l'ipotesi che la percentuale di persone disposte a votare per il partito A sia pari al 55% nella popolazione, contro l'alternativa che sia minore, al livello di significatività del 10%.

$$\begin{cases} H_0: \pi = 0{,}55 \\ H_1: \pi < 0{,}55 \end{cases} \qquad Z = \frac{\bar{X}-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}\,|H_0 \sim N(0\,;1)$$

$\alpha = 0{,}10 \quad \rightarrow \quad -z_\alpha = -1{,}28$

$$\bar{x} = 0{,}52 \qquad z = \frac{0{,}52-0{,}55}{\sqrt{\frac{0{,}55(1-0{,}55)}{1000}}} = -1{,}91$$

Poiché $z < -z_\alpha$, possiamo rifiutare l'ipotesi nulla.



d) Si calcoli il p-value per la verifica di ipotesi al punto c)

$$p-value = P(Z<-1{,}91|H_0) = 1-\Phi(1{,}91) = 1-0{,}97193 = 0{,}02807$$

I dati contengono sufficiente evidenza per rifiutare l'ipotesi nulla.

Esercizio 6
Il fatturato annuo delle aziende di un certo settore e di una certa dimensione può essere considerato distribuito in maniera normale con media pari a 450 mila euro e deviazione standard pari a 80 mila euro. In una recente rilevazione su un campione di 100 aziende di questa tipologia, si è rilevato un fatturato medio annuo pari a 445 mila euro.

a) Si trovi l'intervallo di confidenza per il fatturato medio di tutte le aziende di questa tipologia a livello di confidenza del 95%.

$X = \{\text{Fatturato annuo delle aziende}\}$ $\qquad X \sim N(450; 80^2)$

$n = 100$ $\qquad \bar{x} = 445$

$\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$

$1 - \alpha = 0{,}95 \quad \rightarrow \quad z_{\alpha/2} = 1{,}96$

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

$$\left[\bar{x} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\ ; \bar{x} + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right] = \left[445 - 1{,}96\sqrt{\frac{80^2}{100}}\ ; 445 + 1{,}96\sqrt{\frac{80^2}{100}}\right] = [429{,}3\ ; 460{,}7]$$

b) Quale deve essere la numerosità campionaria n se si vuole che l'ampiezza dell'intervallo si dimezzi?

$$A^* = \frac{1}{2}A \quad \rightarrow \quad 2 \cdot z_{\alpha/2}\sqrt{\frac{\sigma^2}{n^*}} = \frac{1}{2} \cdot 2 \cdot z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \quad \rightarrow \quad \sqrt{\frac{1}{n^*}} = \frac{1}{2}\sqrt{\frac{1}{n}} \quad \rightarrow \quad \frac{1}{n^*} = \frac{1}{4}\frac{1}{n} \quad \rightarrow n^* = 4n = 4 \cdot 100 \approx 400$$

c) Senza effettuare un test (bilaterale) d'ipotesi, sulla base del punto a), si valuti se il fatturato medio di tutte le aziende può ritenersi uguale a 450 mila euro al livello di significatività dell'1%. Motivare accuratamente la risposta.

$I.C. = [429{,}3\ ; 460{,}7]$ al livello $1 - \alpha = 0{,}95$

Sulla base dell'intervallo al punto a) non possiamo rifiutare l'ipotesi nulla che il fatturato medio delle aziende sia pari a 450 mila euro, in favore dell'alternativa che sia diverso da 450, ad un livello dell'1%. Infatti, il valore 450 è contenuto nell'intervallo di confidenza al 95% e quindi sarà sicuramente contenuto nell'intervallo di confidenza al 99%.

d) Si verifichi l'ipotesi al punto c) rispetto all'alternativa che il fatturato medio sia diminuito al livello del 1% e si confrontino le conclusioni con quelle al punto c)

$$\begin{cases} H_0: \mu = 450 \\ H_1: \mu < 450 \end{cases} \qquad Z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} | H_0 \sim N(0\ ; 1)$$

$\alpha = 0{,}01 \quad \rightarrow \quad -z_\alpha = -2{,}33$

$z = \frac{445-450}{\sqrt{\frac{80^2}{100}}} = -0{,}625$ Poiché $z > -z_\alpha$, non possiamo rifiutare l'ipotesi nulla.

Anche sulla base di un test unilaterale (più potente di quello bilaterale) non possiamo rifiutare l'ipotesi nulla.