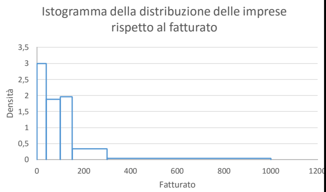
Esercizio 1
Un gruppo di imprese è stato classificato in base al fatturato (in migliaia di euro):

Classi di fatturato	0-40	40-100	100-150	150-300	300-1000
Frequenze	120	113	98	51	30

a) Si disegni l'istogramma per la distribuzione in classi

Classi di fatturato	0-40	40-100	100-150	150-300	300-1000	Totale
Frequenze	120	113	98	51	30	412
Ampiezza classi	40	60	50	150	700	
Densità di freq.	3	1,88	1,96	0,34	0,043	



b) Si calcoli il fatturato medio per impresa.

Valori centrali di classe	20	70	125	225	650	Totale
Frequenze	120	113	98	51	30	412
Freq·Val.centr.	2400	7910	12250	11475	19500	53535

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} \bar{x}_i \cdot n_i = \frac{53535}{412} = 129,9$$

c) Si calcoli il fatturato mediano.

Classi di fatturato	0-40	40-100	100-150	150-300	300-1000	Totale
Frequenze	120	113	98	51	30	412
Ampiezza classi	40	60	50	150	700	
Freq. relative	0,291	0,274	0,238	0,124	0,073	1
Freq.rel.cumulate	0,291	0,565	0,803	0,927	1	

Si individua la classe mediana $c_{i-1} - c_i$, tale che $F_{i-1} \leq 0,5 < F_i$ e si calcola la mediana come

$$Me = c_{i-1} + \frac{0,5 - F_{i-1}}{f_i} \cdot a_i = 40 + \frac{0,5 - 0,291}{0,274} \cdot 60 = 85,7$$

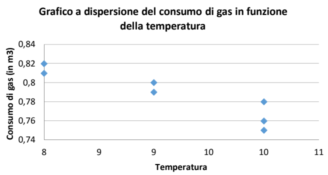
d) Si calcoli la percentuale di imprese con fatturato inferiore a 20 mila euro

Si può calcolare la funzione di ripartizione in $x = 20$ e moltiplicarla poi per 100.
$F(x) = F_{i-1} + \frac{(x-c_{i-1})}{a_i} f_i = 0 + \frac{(20-0)}{40}$0,291=0,146
La percentuale di imprese con fatturato inferiore a 20 mila euro è pari al 14,6%.

Esercizio 2
I seguenti dati riportano il consumo giornaliero di gas (in m3) in un certo comune, in funzione della temperatura (in gradi), in 7 giorni successivi.

Consumo giornaliero (in metri cubi)	Temperatura (in gradi)
0,8	9
0,78	10
0,75	10
0,81	8
0,79	9
0,76	10
0,82	8

a) Si rappresenti graficamente la distribuzione doppia.



b) Si misuri la correlazione tra i due caratteri e si commenti il risultato.

Consumo giornaliero (in metri cubi)	Temperatura (in gradi)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
0,8	9	-0,14286	0,01286	-0,00184	0,02041	0,00017
0,78	10	0,85714	-0,00714	-0,00612	0,73469	0,00005
0,75	10	0,85714	-0,03714	-0,03184	0,73469	0,00138
0,81	8	-1,14286	0,02286	-0,02612	1,30612	0,00052
0,79	9	-0,14286	0,00286	-0,00041	0,02041	0,00001
0,76	10	0,85714	-0,02714	-0,02327	0,73469	0,00074
0,82	8	-1,14286	0,03286	-0,03755	1,30612	0,00108
5,51	64	0,00000	0,00000	-0,12714	4,85714	0,00394

$\bar{x} = \frac{64}{7} = 9,14$ $\qquad \bar{y} = \frac{5,51}{7} = 0,787$

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\,/n}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/n \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2/n}} = \frac{-0,12714}{\sqrt{4,85714 \cdot 0,00394}} = -0,919$$

C'è una correlazione negativa molto elevata tra i due caratteri. All'aumentare della temperatura i consumi di gas decrescono.

c) Si determinino i coefficienti della retta di regressione e se ne interpreti il significato.

Consumo giornaliero (in metri cubi)	Temperatura (in gradi)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
0,8	9	-0,14286	0,01286	-0,00184	0,02041	0,00017
0,78	10	0,85714	-0,00714	-0,00612	0,73469	0,00005
0,75	10	0,85714	-0,03714	-0,03184	0,73469	0,00138
0,81	8	-1,14286	0,02286	-0,02612	1,30612	0,00052
0,79	9	-0,14286	0,00286	-0,00041	0,02041	0,00001
0,76	10	0,85714	-0,02714	-0,02327	0,73469	0,00074
0,82	8	-1,14286	0,03286	-0,03755	1,30612	0,00108
5,51	64	0,00000	0,00000	-0,12714	4,85714	0,00394

$\bar{x} = \frac{64}{7} = 9,14$ $\qquad$ $\bar{y} = \frac{5,51}{7} = 0,787$

$\beta_1 = \frac{-0,12714}{4,85714} = -0,026$ $\qquad$ $\beta_0 = 0,787 - (-0,026 \cdot 9,14) = 1,026$

Il coefficiente β_1 ci dice che se la temperatura aumenta di un grado, il consumo giornaliero di gas diminuisce in media di 0,026 metri cubi. β_0 rappresenterebbe il consumo medio nelle giornate in cui la temperatura è di 0 gradi. Il valore 0 è tuttavia lontano dall'intervallo di definizione dei dati e quindi questo coefficiente non è interpretabile.

d) Si preveda il consumo medio in una giornata con temperatura media di 10 gradi.

$\hat{y} = 1,026 - 0,026 \cdot x = 1,026 - 0,026 \cdot 10 = 0,765$

Esercizio 3

Il numero di persone che arrivano in fila ad un determinato sportello in un minuto nell'ora di punta può essere ben approssimato da una variabile di Poisson di parametro λ =2.

a) calcolare la probabilità che in un minuto si mettano in coda più di 3 persone.

$X = \{\text{Numero di persone che arrivano in fila in un minuto}\}$

$X \sim Poisson(2)$

$P(X > 3) = 1 - P(X \le 3) = 1 - \frac{e^{-2}2^0}{0!} - \frac{e^{-2}2^1}{1!} - \frac{e^{-2}2^2}{2!} - \frac{e^{-2}2^3}{3!} = 1 - 0,1353 - 0,2707 - 0,2707 - 0,1804 = 0,1429$

b) calcolare la probabilità che in un minuto le persone che si mettono in fila siano comprese tra 2 e 4.

$P(2 \le X \le 4) = P(X = 2) + P(X = 3) + P(X = 4) = \frac{e^{-2}2^2}{2!} + \frac{e^{-2}2^3}{3!} + \frac{e^{-2}2^4}{4!} = 0,2707 + 0,1804 + 0,0902 = 0,5413$

c) calcolare la probabilità che in un'ora si mettano in fila più di 100 persone.

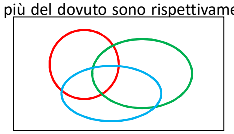
$Y = \{\text{Numero di persone che arrivano in fila in un'ora}\}$ $\qquad$ $Y = X_1 + X_2 + \cdots X_{60}$

$Y \sim Poisson(2 \cdot 60) \sim Poisson(120) \sim N(120\,;120)$

$P(Y > 100,5) = P\left(\frac{Y-\mu}{\sigma} > \frac{100,5-\mu}{\sigma}\right) = P\left(\frac{Y-120}{\sqrt{120}} > \frac{100,5-120}{\sqrt{120}}\right) = P(Z > -1,78) = P(Z < 1,78) = \Phi(1,78)$
$= 0,96246$

Esercizio 4

Un determinato oggetto viene prodotto attraverso un procedimento in tre fasi. Ciascuna delle fasi può essere conclusa entro il termine previsto oppure può protrarsi oltre il dovuto. Le probabilità che le 3 fasi durino più del dovuto sono rispettivamente 0,01, 0,03 e 0,02. La durata di una fase è indipendente dalla durata delle altre.



a) Qual è la probabilità che per produrre un oggetto serva più tempo del dovuto?

$I = \{\text{La prima fase si protrae oltre il dovuto}\}$

$II = \{\text{La seconda fase si protrae oltre il dovuto}\}$

$III = \{\text{La terza fase si protrae oltre il dovuto}\}$

$P(I) = 0,01\ ; P(II) = 0,03\ ; P(III) = 0,02$

$P(I \cup II \cup III) = P(I) + P(II) + P(III) - P(I \cap II) - P(I \cap III) - P(II \cap III) + P(I \cap II \cap III) = 0,01 + 0,03 + 0,02 - 0,01 \cdot 0,03 - 0,01 \cdot 0,02 - 0,02 \cdot 0,03 + 0,01 \cdot 0,03 \cdot 0,02 = 0,06$

Oppure

$P(I \cup II \cup III) = 1 - P(\bar{I} \cap \overline{II} \cap \overline{III}) = 1 - (1 - 0,01)(1 - 0,03)(1 - 0,02) = 0,06$

b) Qual è la probabilità che nella produzione di un oggetto tutte e tre le fasi durino più del dovuto?

$P(I \cap II \cap III) = 0,01 \cdot 0,03 \cdot 0,02 = 0,000006$

c) Qual è la probabilità che la produzione dell'oggetto si concluda nei tempi previsti?

$P(\bar{I} \cap \overline{II} \cap \overline{III}) = (1 - 0,01)(1 - 0,03)(1 - 0,02) = 0,94$

d) Qual è la probabilità che nella produzione di un oggetto almeno due delle tre fasi abbiano una durata più lunga del previsto?

$P\left((I \cap II \cap \overline{III}) \cup (I \cap \overline{II} \cap III) \cup (\bar{I} \cap II \cap III) \cup (I \cap II \cap III)\right) = P(I \cap II \cap \overline{III}) + P(I \cap \overline{II} \cap III) + P(\bar{I} \cap II \cap III)$
$= 0,01 \cdot 0,03 \cdot 0,98 + 0,01 \cdot 0,93 \cdot 0,02 + 0,99 \cdot 0,03 \cdot 0,02 + 0,01 \cdot 0,03 \cdot 0,02 = 0,0011$

Esercizio 5

Una ditta molto grande vuole stimare la percentuale dei suoi dipendenti che si ritiene soddisfatta dell'ambiente lavorativo
Nel caso la percentuale fosse inferiore al 35%, un piano di miglioramento verrebbe messo in atto.
Su un campione di 100 dipendenti, solo 30 si dichiarano soddisfatti.

a) Sulla base del campione, si costruisca un intervallo di confidenza al livello del 95% per la percentuale di soddisfatti

$X = \{\text{Soddisfazione del dipendente}\}$ $\qquad$ $X \sim Ber(\pi)$

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

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

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

$\bar{x} = \frac{30}{100} = 0,3$

$\left[\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\ ; \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] = \left[0,3 - 1,96\sqrt{\frac{0,3\ (1-0,3)}{100}}\ ; 0,3 + 1,96\sqrt{\frac{0,3\ (1-0,3)}{100}}\right] = [0,21\ ; 0,39]$

Al livello di confidenza del 95%, la quota di lavoratori soddisfatti è compresa tra il 21% e il 39%.

b) Si calcoli la numerosità campionaria necessaria per ottenere un'ampiezza dell'intervallo pari a 0,1.

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

$$A = 2z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \quad \rightarrow \quad 0{,}1 = 2\cdot 1{,}96\sqrt{\frac{0{,}3(1-0{,}3)}{n}} \quad \rightarrow \quad n = \frac{(2\cdot 1{,}96)^2\cdot 0{,}3(1-0{,}3)}{0{,}1^2} \approx 323$$

Oppure, se non si avesse una stima preliminare di π:

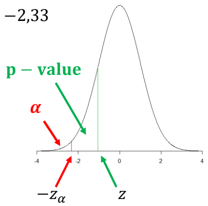
$$A = 2z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \quad \rightarrow \quad 0{,}1 = 2\cdot 1{,}96\sqrt{\frac{0{,}5(1-0{,}5)}{n}} \quad \rightarrow \quad n = \frac{(2\cdot 1{,}96)^2\cdot 0{,}5(1-0{,}5)}{0{,}1^2} \approx 385$$

c) Si verifichi la necessità di un piano di miglioramento, al livello dell'1%.

$$\begin{cases} H_0: \pi = 0{,}35 \\ H_1: \pi < 0{,}35 \end{cases} \qquad Z = \frac{\bar{X}-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} | H_0 \sim N(0\,;1) \qquad \alpha = 0{,}01 \quad \rightarrow \quad -z_\alpha = -2{,}33$$

$$\bar{x} = 0{,}3 \qquad z = \frac{0{,}3-0{,}35}{\sqrt{\frac{0{,}35(1-0{,}35)}{100}}} = -1{,}05$$

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



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

$p-value = P(Z < -1{,}05 | H_0) = 1 - \Phi(1{,}05) = 1 - 0{,}85314 = 0{,}14686$

I dati non contengono sufficiente evidenza per rifiutare l'ipotesi nulla che la percentuale di lavoratori soddisfatti sia del 35%.

Esercizio 6

Un processo produttivo produce pezzi con peso che può assumersi distribuito normalmente con media 5 g e deviazione standard 0,1 g. In un controllo di qualità sono stati rilevati i seguenti pesi

5,1 5,2 5,0 5,1 4,9

a) Sulla base dei dati osservati, si costruisca un intervallo di confidenza al 95% per il peso medio

$X = \{\text{Peso dei pezzi prodotti}\} \qquad X \sim N(5\,;0{,}1^2)$

$$n = 5 \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{5{,}1+5{,}2+5{,}0+5{,}1+4{,}9}{5} = 5{,}06$$

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

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

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

$$\left[\bar{x} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\,;\bar{x} + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right] = \left[5{,}06 - 1{,}96\sqrt{\frac{0{,}1^2}{5}}\,;5{,}06 + 1{,}96\sqrt{\frac{0{,}1^2}{5}}\right] = [4{,}97\,;5{,}15]$$

Al livello di fiducia del 95% possiamo ritenere che il peso medio dei pezzi sia compreso tra 4,97 e 5,15 g.

b) Si verifichi l'ipotesi che il peso medio sia rimasto invariato, contro l'alternativa che sia aumentato, al livello di significatività dell'1%.

$$\begin{cases} H_0: \mu = 5 \\ H_1: \mu > 5 \end{cases} \qquad Z = \frac{\bar{X}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} | H_0 \sim N(0\,;1) \qquad \bar{x} = 5{,}06$$

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

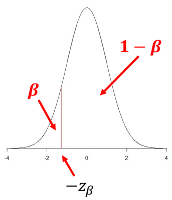
$z = \frac{5{,}06-5}{\sqrt{\frac{0{,}1^2}{5}}} = 1{,}34$ Poiché $z < z_\alpha$, non possiamo rifiutare l'ipotesi nulla.

c) Si calcoli la numerosità campionaria necessaria per avere una potenza del test al punto b) pari al 90%, nel caso in cui il peso medio sia effettivamente passato a 5,01 grammi.

$$0{,}90 = P\left(\frac{\bar{X}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} > z_\alpha | H_1\right) = P\left(\frac{\bar{X}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} - \frac{\mu_1-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} > z_\alpha - \frac{\mu_1-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} | H_1\right) =$$

$$P\left(Z > z_\alpha - \frac{\mu_1-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}\right) = P(Z > -z_\beta)$$

$$-z_\beta = z_\alpha - \frac{\mu_1-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} \quad \rightarrow \quad -1{,}28 = 2{,}33 - \frac{5{,}01-5}{\sqrt{\frac{0{,}1^2}{n}}} \quad \rightarrow \quad n = \frac{(1{,}28+2{,}33)^2\cdot 0{,}1^2}{(5{,}01-5)^2} \approx 1302$$



d) Sulla base del campione, si può concludere che la precisione del processo si sia modificata, al livello di significatività del 5%?

5,1 5,2 5,0 5,1 4,9

$$\begin{cases} H_0: \sigma^2 = 0{,}1^2 \\ H_1: \sigma^2 \neq 0{,}1^2 \end{cases} \qquad C = \frac{(n-1)\cdot S^2}{\sigma_0^2} | H_0 \sim \chi^2_{n-1} \qquad \bar{x} = 5{,}06$$

$\alpha = 0{,}5 \quad \rightarrow \quad \chi^2_{1-\alpha/2} = 0{,}4844 \quad \text{e} \quad \chi^2_{\alpha/2} = 11{,}1433$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{(5{,}1-5{,}06)^2 + (5{,}2-5{,}06)^2 + (5{,}0-5{,}06)^2 + (5{,}1-5{,}06)^2 + (4{,}9-5{,}06)^2}{4} = 0{,}013$$

$$c = \frac{(n-1)\cdot s^2}{\sigma_0^2} = \frac{4\cdot 0{,}013}{0{,}1^2} = 5{,}2$$

Poiché $\chi^2_{1-\alpha/2} < c < \chi^2_{\alpha/2}$, i dati non contengono sufficiente evidenza per rifiutare l'ipotesi nulla che la precisione sia rimasta invariata al livello del 5%.