Esercizio 1
La tabella seguente mostra la distribuzione doppia di un collettivo rispetto ai caratteri sesso e condizione professionale.

	Disoccupato	Lavoratore dipendente	Libero professionista	Totale
Femmina	5	190	65	260
Maschio	7	143	108	258
Totale	12	333	173	518

a) Si calcolino le distribuzioni di frequenza relative del carattere condizione professionale, condizionatamente al carattere sesso.

	Disoccupato	Lavoratore dipendente	Libero professionista	Totale
Femmina	0,019	0,731	0,250	1,000
Maschio	0,027	0,554	0,419	1,000
Totale	0,023	0,643	0,334	1,000

b) Sulla base del punto a) si determini se c'è indipendenza tra i due caratteri motivando la risposta

Le frequenze condizionate sono diverse tra loro e diverse dalle frequenze marginali, quindi non c'è indipendenza.

c) Si costruisca la tabella di perfetta indipendenza

	Disoccupato	Lavoratore dipendente	Libero professionista	Totale
Femmina	$12 \cdot 260/518 = 6{,}023$	$333 \cdot 260/518 = 167{,}143$	$173 \cdot 260/518 = 86{,}834$	260
Maschio	$12 \cdot 258/518 = 5{,}977$	$333 \cdot 258/518 = 165{,}857$	$12 \cdot 173/518 = 86{,}166$	258
Totale	12	333	173	518

d) Si costruisca un indice opportuno per misurare il grado di dipendenza

Tabella osservata

	Disoccupato	Lavoratore dipendente	Libero professionista	Totale
Femmina	5	190	65	260
Maschio	7	143	108	258
Totale	12	333	173	518

Tabella di perfetta indipendenza

	Disoccupato	Lavoratore dipendente	Libero professionista	Totale
Femmina	$12 \cdot 260/518 = 6{,}023$	$333 \cdot 260/518 = 167{,}143$	$173 \cdot 260/518 = 86{,}834$	260
Maschio	$12 \cdot 258/518 = 5{,}977$	$333 \cdot 258/518 = 165{,}857$	$12 \cdot 173/518 = 86{,}166$	258
Totale	12	333	173	518

$$\chi^2 = \sum_{i=1}^{H}\sum_{j=1}^{K}\frac{\left(n_{ij}-n'_{ij}\right)^2}{n'_{ij}}$$
$$= \frac{(5-6{,}023)^2}{6{,}023} + \frac{(190-167{,}143)^2}{167{,}143} + \frac{(65-86{,}834)^2}{86{,}834} + \frac{(7-5{,}977)^2}{5{,}977} + \frac{(143-165{,}857)^2}{165{,}857} + \frac{(108-86{,}166)^2}{86{,}166} = 17{,}647$$

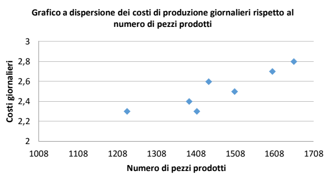
$$\text{Indice di Cramer} = \sqrt{\frac{17{,}647}{518 \cdot \min[(2-1),(3-1)]}} = \sqrt{\frac{17{,}647}{518 \cdot 1}} = 0{,}185$$

Esiste un grado di associazione piuttosto basso tra i due caratteri.

Esercizio 2
I seguenti dati riportano i costi giornalieri di produzione di un certo bene, in un determinato impianto, e il numero di pezzi prodotti giornalmente, in 7 giorni successivi

a) Si rappresenti graficamente la distribuzione doppia, considerando il numero di pezzi come variabile indipendente.

Costi giornalieri (in migliaia di euro)	Numero di pezzi
2,5	1506
2,3	1410
2,3	1232
2,6	1440
2,4	1390
2,8	1657
2,7	1602



b) Si determino i coefficienti della retta di regressione che esprime il costo in funzione del numero di pezzi e se ne interpreti il significato.

Costi giornalieri (in migliaia di euro)	Numero di pezzi	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
2,5	1506	43,571	-0,014	-0,622	1898,469
2,3	1410	-52,429	-0,214	11,235	2748,755
2,3	1232	-230,429	-0,214	49,378	53097,327
2,6	1440	-22,429	0,086	-1,922	503,041
2,4	1390	-72,429	-0,114	8,278	5245,898
2,8	1657	194,571	0,286	55,592	37858,041
2,7	1602	139,571	0,186	25,920	19480,184
17,6	10237	0,000	0,000	147,857	120831,714

$$\bar{x} = \frac{10237}{7} = 1462{,}43 \qquad \bar{y} = \frac{17{,}6}{7} = 2{,}51$$

$$\beta_1 = \frac{147{,}857}{120831{,}714} = 0{,}0012 \qquad \beta_0 = 2{,}51 - 0{,}0012 \cdot 1462{,}43 = 0{,}725$$

Il coefficiente β_1 ci dice che per ogni pezzo prodotto in più, i costi giornalieri di produzione salgono di 1,2 euro. β_1 rappresenta dunque il costo marginale. β_0 rappresenterebbe il costo fisso che si dovrebbe sopportare anche nel caso in cui non si producesse nulla. Il valore 0 è tuttavia molto lontano dall'intervallo di definizione dei dati e quindi questo coefficiente non è interpretabile.

c) Si prevedano i costi giornalieri necessari per produrre 1700 pezzi al giorno.

Costi giornalieri (in migliaia di euro)	Numero di pezzi	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
2,5	1506	43,571	-0,014	-0,622	1898,469	0,000
2,3	1410	-52,429	-0,214	11,235	2748,755	0,046
2,3	1232	-230,429	-0,214	49,378	53097,327	0,046
2,6	1440	-22,429	0,086	-1,922	503,041	0,007
2,4	1390	-72,429	-0,114	8,278	5245,898	0,013
2,8	1657	194,571	0,286	55,592	37858,041	0,082
2,7	1602	139,571	0,186	25,920	19480,184	0,034
17,6	10237	0,000	0,000	147,857	120831,714	0,229

$\beta_1 = \dfrac{147{,}857}{120831{,}714} = 0{,}0012$ $\qquad \beta_0 = 2{,}51 - 0{,}0012 \cdot 1462{,}43 = 0{,}725$

$\hat{y} = 0{,}725 + 0{,}0012 \cdot x = 0{,}725 + 0{,}0012 \cdot 1700 = 2{,}805$

Il costo giornaliero previsto per produrre 1700 pezzi è di 2805 euro.

d) Si misuri la bontà dell'adattamento della retta ai dati e si commenti il risultato.

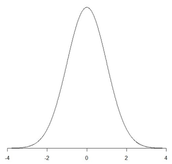
$R^2 = \dfrac{147{,}857^2}{120831{,}714 \cdot 0{,}229} = 0{,}79$

La retta si adatta bene ai dati osservati

Esercizio 3

L'entità giornaliera dei prelievi ad un certo sportello bancomat può considerarsi come una variabile aleatoria gaussiana di media pari a 1300 euro e deviazione standard pari a 150 euro. I prelievi risultano inoltre indipendenti da un giorno all'altro.

a) Qual è la probabilità che in un determinato giorno l'ammontare dei prelievi superi i 2000 euro?



$X = \{\text{Ammontare dei prelievi giornalieri}\}$

$X \sim N(1300\,;150^2)$

$P(X > 2000) = P\left(\dfrac{X-\mu}{\sigma} > \dfrac{2000-\mu}{\sigma}\right) = P\left(\dfrac{X-1300}{\sqrt{150^2}} > \dfrac{2000-1300}{\sqrt{150^2}}\right) =$

$P(Z > 4{,}67) = 1 - \Phi(4{,}67) \approx 1 - 1 = 0$

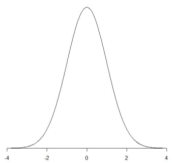
b) Qual è la probabilità che in un determinato giorno l'ammontare dei prelievi sia inferiore a 1000 euro?

$P(X < 1000) = P\left(\dfrac{X-\mu}{\sigma} < \dfrac{1000-\mu}{\sigma}\right) = P\left(\dfrac{X-1300}{\sqrt{150^2}} < \dfrac{1000-1300}{\sqrt{150^2}}\right) = P(Z < -2{,}00) = P(Z > 2{,}00)$

$= 1 - P(Z < 2{,}00) = 1 - \Phi(2{,}00) \approx 1 - 0{,}97725 = 0{,}02275$

c) Qual è la probabilità che su 7 giorni l'ammontare dei prelievi superi i 1500 euro almeno in 4 giorni?

$Y = \{\text{Numero di volte che l'ammontare dei prelievi supera i 1500 euro in 7 giorni}\}$ $\qquad Y \sim Bin(n\,;\pi)$



$n = 7$

$\pi = P(X > 1500) = P\left(\dfrac{X-\mu}{\sigma} > \dfrac{1500-\mu}{\sigma}\right) = P\left(\dfrac{X-1300}{\sqrt{150^2}} > \dfrac{1500-1300}{\sqrt{150^2}}\right) =$

$P(Z > 1{,}33) = 1 - P(Z < 1{,}33) = 1 - \Phi(1{,}33) = 1 - 0{,}90824 = 0{,}09176$

$P(Y \geq 4) = \sum_{y=4}^{7} \binom{n}{y} \pi^y (1-\pi)^{n-y}$

$= \binom{7}{4} 0{,}09176^4 (1-0{,}09176)^{7-4} + \binom{7}{5} 0{,}09176^5 (1-0{,}09176)^{7-5} + \binom{7}{6} 0{,}09176^6 (1-0{,}09176)^{7-6}$

$+ \binom{7}{7} 0{,}09176^7 (1-0{,}09176)^{7-7} = 0{,}00186 + 0{,}00011 + 0{,}000004 + 0 = 0{,}001974$

d) Qual è la probabilità che in 100 giorni l'ammontare totale dei prelievi superi i 135000 euro?

$Q = \{\text{Ammontare totale dei prelievi in 100 giorni}\}$ $\qquad Q = \sum_{i=1}^{100} X_i$

$X_i \sim N(1300\,;150^2)$ $\qquad Q \sim N(100 \cdot 1300\,; 100 \cdot 150^2) \sim N(130000\,;2250000)$

$P(Q > 135000) = P\left(\dfrac{Q-\mu}{\sigma} > \dfrac{135000-\mu}{\sigma}\right) = P\left(\dfrac{X-130000}{\sqrt{2250000}} > \dfrac{135000-130000}{\sqrt{2250000}}\right) = P(Z > 3{,}33)$

$= 1 - P(Z < 3{,}33) = 1 - \Phi(3{,}33) = 1 - 0{,}99957 = 0{,}00043$

Esercizio 4

Una banca, sulla base dei dati raccolti in passato, sa che tra i clienti che hanno rimborsato un prestito, il 60% ha un reddito alto. Tra coloro che non hanno rimborsato il prestito, invece, solo il 15% ha un reddito alto. Inoltre il 75% dei prestiti risulta rimborsato.

a) Qual è la probabilità che un potenziale cliente con un reddito alto rimborserà il prestito?

$A = \{\text{Il cliente ha un reddito alto}\}$

$R = \{\text{Il cliente ha rimborsato il prestito}\}$

$P(A|R) = 0{,}60$

$P(A|\bar{R}) = 0{,}15$

$P(R) = 0{,}75$

$P(R|A) = \dfrac{P(A|R)P(R)}{P(A)} = \dfrac{P(A|R)P(R)}{P(A|R)P(R) + P(A|\bar{R})P(\bar{R})} = \dfrac{0{,}60 \cdot 0{,}75}{0{,}60 \cdot 0{,}75 + 0{,}15 \cdot 0{,}25} = 0{,}923$

b) Qual è la percentuale di clienti della banca con reddito basso?

$P(\bar{A}) \cdot 100 = [1 - P(A)] \cdot 100 = [1 - (0{,}60 \cdot 0{,}75 + 0{,}15 \cdot 0{,}25)] \cdot 100 = [1 - 0{,}488] \cdot 100 = 51{,}2$

c) Gli eventi reddito alto e prestito rimborsato sono indipendenti?

Se fossero indipendenti si avrebbe, ad esempio, $P(A|R) = P(A)$. Poiché $P(A|R) = 0{,}60$ e $P(A) = 0{,}488$, i due eventi non sono indipendenti.

d) Qual è la probabilità che su 10 clienti scelti a caso almeno 1 non abbia rimborsato il prestito?

$Y = \{Numero\ di\ clienti\ che\ non\ hanno\ rimborsato\ il\ prestito\ sui\ 10\ estratti\}$ $\qquad Y \sim Bin(n\,;\,\pi)$

$n = 10$
$\pi = 1 - P(R) = 1 - 0{,}75 = 0{,}25$
$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{10}{0} 0{,}25^0 (1 - 0{,}25)^{10-0} = 1 - 0{,}0563 = 0{,}9437$

Esercizio 5
Per le vendite di un certo prodotto, la marca A possiede una quota di mercato che negli ultimi anni è stata pari al 10% delle vendite totali di quel prodotto. Al fine di aumentare tale quota di mercato, si investe in un'ampia campagna pubblicitaria. Per valutarne gli esiti, si estrae un campione di individui a cui viene chiesto se intendono acquistare quel tipo di prodotto ed eventualmente di che marca. Tra i 1130 individui che dichiarano di voler acquistare il prodotto, 130 scelgono la marca A.

a) Sulla base del campione, si costruisca un intervallo di confidenza al livello del 90% per la quota di mercato della marca A.

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

$1 - \alpha = 0{,}90 \quad \rightarrow \quad z_{\alpha/2} = 1{,}64$

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

$$\bar{x} = \frac{130}{1130} = 0{,}115$$

$$\left[\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\ ; \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] = \left[0{,}115\ - 1{,}64\sqrt{\frac{0{,}115\ (1 - 0{,}115)}{1130}}\ ; 0{,}115\ + 1{,}64\sqrt{\frac{0{,}115\ (1 - 0{,}115)}{1130}}\right]$$
$= [0{,}099\ ; 0{,}131]$

Al livello di confidenza del 90%, la quota di mercato della marca A è compresa tra il 9,9% e il 13,1%.

b) Sulla base del risultato ottenuto al punto a) e senza ulteriori calcoli, si può concludere che la quota di mercato sia rimasta invariata, al livello $\alpha = 0{,}01$?

I.C. al livello del 90% = [0,099 ; 0,131]

Il valore del 10% è incluso all'interno dell'intervallo al livello di fiducia del 90%. Sarà quindi incluso anche nell'intervallo di fiducia al 99%, essendo quest'ultimo più ampio. Non possiamo quindi escludere che la quota di mercato sia rimasta invariata.

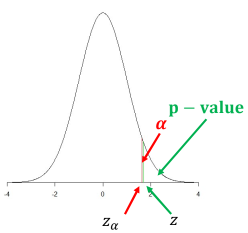
c) Si verifichi l'ipotesi che la campagna pubblicitaria abbia avuto l'effetto desiderato, al livello del 5%.

$\begin{cases} H_0: \pi = 0{,}10 \\ H_1: \pi > 0{,}10 \end{cases} \qquad Z = \frac{\bar{X} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} | H_0 \sim N(0\,;1)$

$\alpha = 0{,}05 \quad \rightarrow \quad z_\alpha = 1{,}64$

$\bar{x} = 0{,}115 \qquad z = \frac{0{,}115 - 0{,}10}{\sqrt{\frac{0{,}10(1 - 0{,}10)}{1130}}} = 1{,}69$

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



d) Si calcoli il p-value per la verifica di ipotesi al punto c)
$p - value = P(Z > 1{,}69 | H_0) = 1 - \Phi(1{,}69) = 1 - 0{,}95449 = 0{,}04331$
I dati contengono sufficiente evidenza per rifiutare l'ipotesi nulla al livello del 5%. I dati non permettono invece di rifiutare l'ipotesi nulla al livello dell'1%.

Esercizio 6
Il rendimento mensile percentuale di un certo titolo può considerarsi distribuito come una normale con media 2,3 e deviazione standard 0,6. I rendimenti mensili percentuali del titolo negli ultimi 6 mesi sono risultati i seguenti:

1,8	2,2	2,3	2,1	2,1	1,9

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

$X = \{\text{Rendimento mensile percentuale del titolo}\} \qquad X \sim N(2{,}3\,; 0{,}6^2)$

$n = 6 \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1{,}8 + 2{,}2 + 2{,}3 + 2{,}1 + 2{,}1 + 1{,}9}{6} = 2{,}05$

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

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

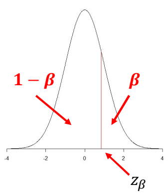
$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

$$\left[\bar{x} - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\ ; \bar{x} + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right] = \left[2{,}05 - 1{,}96\sqrt{\frac{0{,}6^2}{6}}\ ; 2{,}05 + 1{,}96\sqrt{\frac{0{,}6^2}{6}}\right] = [1{,}57\ ; 2{,}53]$$

Al livello di fiducia del 95% possiamo ritenere che il rendimento medio mensile del titolo sia compreso tra l'1,57% e il 2,53%

b) Si verifichi l'ipotesi che il rendimento medio sia rimasto invariato, contro l'alternativa che sia diminuito, al livello dell'1%.

$$\begin{cases} H_0: \mu = 2{,}3 \\ H_1: \mu < 2{,}3 \end{cases} \qquad Z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} | H_0 \sim N(0\,;1) \qquad \bar{x} = 2{,}05$$



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

$z = \frac{2{,}05 - 2{,}3}{\sqrt{\frac{0{,}6^2}{6}}} = -1{,}02$ 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 all'80%, nel caso in cui il rendimento medio sia effettivamente sceso a 2,1.

$$0{,}80 = P\left(\frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} < -z_\alpha | H_1\right) = P\left(\frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} - \frac{\mu_1 - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} < -z_\alpha - \frac{\mu_1 - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} | H_1\right) = P\left(Z < -z_\alpha - \frac{\mu_1 - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}\right) = P(Z < z_\beta)$$

$$z_\beta = -z_\alpha - \frac{\mu_1 - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \quad \rightarrow \quad 0{,}84 = -2{,}33 - \frac{2{,}1 - 2{,}3}{\sqrt{\frac{0{,}6^2}{n}}} \quad \rightarrow \quad n = \frac{(0{,}84 + 2{,}33)^2 \cdot 0{,}6^2}{(2{,}1 - 2{,}3)^2} \approx 91$$

d) Sulla base del campione, si può concludere che la volatilità del titolo si sia modificata, al livello di significatività del 10%?

$$\begin{cases} H_0: \sigma^2 = 0{,}6^2 \\ H_1: \sigma^2 \neq 0{,}6^2 \end{cases} \qquad C = \frac{(n-1) \cdot S^2}{\sigma_0^2} | H_0 \sim \chi^2_{n-1} \qquad \bar{x} = 2{,}05$$

$\alpha = 0{,}1 \quad \rightarrow \quad \chi^2_{1-\alpha/2} = 1{,}1455$ e $\chi^2_{\alpha/2} = 11{,}0705$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{(1{,}8 - 2{,}05)^2 + (2{,}2 - 2{,}05)^2 + \cdots + (1{,}9 - 2{,}05)^2}{5} = 0{,}035$$

$$c = \frac{(n-1) \cdot s^2}{\sigma_0^2} = \frac{5 \cdot 0{,}035}{0{,}6^2} = 0{,}486$$

Poiché $c < \chi^2_{1-\alpha/2}$, i dati contengono sufficiente evidenza per rifiutare l'ipotesi nulla che la volatilità sia rimasta invariata al livello del 10%.