1. Una macchina confezionatrice di una certa bevanda è programmata in modo tale che il contenuto delle bottiglie abbia distribuzione normale con media μ = 50 cl e deviazione standard σ = 1 cl. Un campione di 100 bottiglie, estratto casualmente dalla linea di produzione, ha un contenuto medio pari a 49 cl. Sulla base del valore dell'intervallo di confidenza al 95% della media, ritenete che la macchina confezionatrice della bevanda debba essere riprogrammata?

   La deviazione standard è nota

   $\Rightarrow$ la distribuzione campionaria di $\overline{X}$ è normale

   $IC95\%(\mu) = \bar{x} \pm z_{0.025} * \sigma/\sqrt{n} = 49 \pm 1.96 * 1/\sqrt{100} = (48.8\ cl, 49.2\ cl)$

   La macchina deve essere riprogrammata in quanto viene introdotta in media una quantità di bevanda inferiore al valore dichiarato (50 cl).

2. Il costo medio annuale per paziente affetto da una particolare malattia ha distribuzione normale. In un campione di 9 soggetti, il costo medio è risultato essere pari a 1200 euro con deviazione standard pari a 250 euro. Calcolate l'intervallo di confidenza al 95%.

   La deviazione standard non è nota e la dimensione campionaria è limitata

   $\Rightarrow$ la distribuzione campionaria di $\overline{X}$ è t di Student con 8 gdl

   $IC95\%(\mu) = \bar{x} \pm t_{8,0.025} * s/\sqrt{n} = 1200 \pm 2.306 * 250/\sqrt{9} = (1007.8\ euro, 1392.2\ euro)$

3. Qual è l'intervallo di confidenza al 95% della probabilità di avere l'asma in una popolazione, se la frequenza relativa di asma in un campione formato da 225 soggetti è pari a 0.05?

   np = 225*0.05 = 11.25 > 5

   $\Rightarrow$ la distribuzione campionaria di P è approssimativamente normale

   $IC95\%(\pi) = p \pm z_{0.025} * \sqrt{p*(1-p)/n} = 0.05 \pm 1.96 * \sqrt{0.05*0.95/225} = (0.0215, 0.0785)$

4. Si vuole stimare la probabilità di avere l'asma in una popolazione. Dati preliminari provenienti dalla letteratura suggeriscono che la probabilità di avere l'asma sia pari a 0.05. Qual è la numerosità campionaria necessaria per ottenere un intervallo di confidenza al 95% di ampiezza uguale a 0.02?

$$\mathbf{L_{sup} - L_{inf} = (p + z_{0.025} * \sqrt{p*(1-p)/n}) - (p - z_{0.025} * \sqrt{p*(1-p)/n})}$$
$$\mathbf{= 2 * z_{0.025} * \sqrt{p*(1-p)/n} = 0.02}$$

$$\mathbf{2 * 1.96 * \sqrt{0.05*0.95/n} = 0.02 \quad \Rightarrow \quad n \cong 1825}$$

5. In uno studio sull'età al menarca condotto negli USA, si ottennero le seguenti informazioni per le donne di età 21-30 anni e 31-40 anni:

	donne di età 21-30 anni	donne di età 31-40 anni
	D_1	D_0
n	78	66
Età media al menarca	12.42 anni	13.88 anni

Si può affermare che l'età media al menarca sia inferiore nelle donne più giovani? Verificate l'ipotesi mediante un test d'ipotesi e mediante il p-value. Assumete che l'età al menarca abbia distribuzione normale e ripete l'esercizio assumendo che:

i) la deviazione standard dell'età al menarca sia nota e sia 1.075 anni nel gruppo di donne di età 21-30 e 1.387 anni nel gruppo di donne di età 31-40:

Definisco il sistema d'ipotesi:

$$\begin{cases} H_0: \mu_0 = \mu_1 \\ H_1: \mu_0 \neq \mu_1 \end{cases}$$

Calcolo la statistica test:

$$z_{oss} = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{\sigma^2_1/n_1 + \sigma^2_0/n_0}} = \frac{12.42 - 13.88}{\sqrt{1.075^2/78 + 1.387^2/66}} = -6.96$$

Definisco la regione di rifiuto e formulo la decisione (test d'ipotesi):

La distribuzione campionaria della statistica test è normale

Regione di rifiuto di H_0 ($\alpha = 0.05$): $\{Z < -z_{0.025} = -1.96 \cup Z > 1.96\}$

$z_{oss} = -6.96 < -1.96 \Rightarrow$ rifiuto l'ipotesi nulla (H_0): l'età media al menarca è statisticamente differente tra i due gruppi di donne

Calcolo il p-value (test di significatività statistica):

p-value = $2 * Pr(Z \leq -6.96 \mid H_0$ è vera$) = 2 * Pr(Z \geq 6.96 \mid H_0$ è vera$) < 0.0001$

i dati non supportano l'ipotesi nulla (H_0)

ii) la <u>deviazione standard sia uguale nei due gruppi ma ignota</u> (i precedenti valori della deviazione standard sono quindi statistiche campionarie):

Calcolo la statistica test:

$$s_p = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_0 - 1)\, s_0^2}{(n_1 - 1) + (n_0 - 1)}} = \sqrt{\frac{77 * 1.075^2 + 65 * 1.387^2}{142}} = 1.228$$

$$t_{oss} = \frac{\overline{x}_1 - \overline{x}_0}{s_p \sqrt{1/n_1 + 1/n_0}} = \frac{12.42 - 13.88}{1.228 \sqrt{1/78 + 1/66}} = -7.11$$

Definisco la regione di rifiuto e formulo la decisione (test d'ipotesi):

Gradi di libertà = 142 > 100: la distribuzione campionaria della statistica test è approssimativamente normale ($T_{142} \cong Z$)

Regione di rifiuto di H_0 ($\alpha = 0.05$): $\{T_{142} \cong Z < -1.96 \cup T_{142} \cong Z > 1.96\}$

$t_{oss} = -7.11 < -1.96 \Rightarrow$ rifiuto l'ipotesi nulla (H_0): l'età media al menarca è statisticamente differente tra i due gruppi di donne

Calcolo il p-value (test di significatività statistica):

p-value = $2 * Pr(T_{142} \cong Z \leq -7.11 \mid H_0$ è vera$) < 0.0001$

i dati non supportano l'ipotesi nulla (H_0)

6. Due diversi collutori A e B sono stati somministrati a due gruppi di 10 pazienti ciascuno. E' stato quindi misurato il seguente punteggio clinico di riduzione della quantità di placca dopo 48 ore dalla somministrazione del collutorio:

Punteggio della placca	
Collutorio A	Collutorio B
32	14
60	39
25	24
45	13
65	9
60	3
68	10
83	14
120	1
110	30

(devianza campionaria = 8489.61 nel gruppo trattato con il collutorio A; devianza campionaria = 1284.12 nel gruppo trattato con il collutorio B)

I due collutori hanno in media una diversa efficacia nel ridurre la quantità di placca? Verificate l'ipotesi mediante un <u>test d'ipotesi</u> assumendo che il punteggio della placca abbia distribuzione normale e che la <u>deviazione standard sia uguale nei due gruppi ma ignota</u>.

Media campionaria nel gruppo trattato con il collutorio A: $\bar{x}_1 = 66.8$
Media campionaria nel gruppo trattato con il collutorio B: $\bar{x}_0 = 15.7$

Varianza campionaria nel gruppo trattato con il collutorio A:

$$s_1^2 = 8489.61/9 = 943.29$$

Varianza campionaria nel gruppo trattato con il collutorio B: 15.7

$$s_0^2 = 1284.12/9 = 142.68$$

Definisco il sistema d'ipotesi:

$$\begin{cases} H_0: \mu_0 = \mu_1 \\ H_1: \mu_0 \neq \mu_1 \end{cases}$$

Calcolo la statistica test:

$$s_p = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_0 - 1)\, s_0^2}{(n_1 - 1) + (n_0 - 1)}} = \sqrt{\frac{9 * 943.29 + 9 * 142.68}{18}} = 23.30$$

$$t_{oss} = \frac{\overline{x}_1 - \overline{x}_0}{s_p \sqrt{1/n_1 + 1/n_0}} = \frac{66.8 - 15.7}{23.30 \sqrt{1/10 + 1/10}} = 4.90$$

Definisco la regione di rifiuto e formulo la decisione (test d'ipotesi):

La distribuzione campionaria della statistica test è t di Student con 18 gradi di libertà

Regione di rifiuto di H_0 ($\alpha = 0.05$): $\{T_{18} < -2.101 \cup T_{18} > 2.101\}$

$t_{oss} = 4.90 > 2.101 \Rightarrow$ rifiuto l'ipotesi nulla (H_0): i due collutori hanno in media una diversa efficacia nel ridurre la quantità di placca

7. La "scrapie" è una malattia degli ovini simile al morbo della mucca pazza. In una sperimentazione su cavie è stata utilizzata una sostanza per il trattamento di tale malattia. In un gruppo di 10 cavie infettate e trattate con il farmaco sperimentale, il tempo medio di sopravvivenza era di 116 giorni (deviazione standard = 17.7 giorni). In un gruppo di 10 cavie infettate di controllo, il tempo medio di sopravvivenza era di 85 giorni (deviazione standard = 6.0 giorni).
   C'è una differenza statisticamente significativa nella durata media della sopravvivenza tra i due gruppi? Verificate l'ipotesi mediante un test d'ipotesi assumendo che la durata della sopravvivenza abbia distribuzione normale e che la deviazione standard sia uguale nei due gruppi ma ignota.

Definisco il sistema d'ipotesi:

$$\begin{cases} H_0: \mu_0 = \mu_1 \\ H_1: \mu_0 \neq \mu_1 \end{cases}$$

Calcolo la statistica test:

$$s_p = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_0 - 1)\, s_0^2}{(n_1 - 1) + (n_0 - 1)}} = \sqrt{\frac{9 * 17.7^2 + 9 * 6.0^2}{18}} = 13.22$$

$$t_{oss} = \frac{\overline{x}_1 - \overline{x}_0}{s_p \sqrt{1/n_1 + 1/n_0}} = \frac{116 - 85}{13.22 \sqrt{1/10 + 1/10}} = 5.24$$

Definisco la regione di rifiuto e formulo la decisione (test d'ipotesi):

La distribuzione campionaria della statistica test è t di Student con 18 gradi di libertà

Regione di rifiuto di H_0 ($\alpha = 0.05$): $\{T_{18} < -2.101 \cup T_{18} > 2.101\}$

$t_{oss} = 5.24 > 2.101 \Rightarrow$ rifiuto l'ipotesi nulla (H_0): la durata media della sopravvivenza è statisticamente differente tra i due gruppi