Esercitazione 4 - Statistica (parte II)

Davide Passaretti

27/2/2017

Contents

1 Verifica delle ipotesi

Nella verifica delle ipotesi, si vuole far inferenza su uno o più parametri della popolazione utilizzando i dati campionari a disposizione. La procedura è caratterizzata da due ipotesi discordanti (e spesso complementari):

- H_0, l'ipotesi nulla, ovvero l'ipotesi di partenza riguardo il parametro d'interesse;
- H_1, l'ipotesi alternativa, ossia l'ipotesi che dovremmo prendere in considerazione qualora i nostri dati ci diano una marcata evidenza della non plausibiltà di H_0.

Dal momento che si tratta di decidere se "restare" con H_0 oppure virare su H_1, possiamo commettere due tipi di errore cui sono associate due rispettive probabilità di sbagliare:

- α, il livello di significatività del test, è la probabilità di $\overbrace{\text{rifiutare } H_0 \text{ se } H_0 \text{ è vera}}^{\text{errore di I tipo}}$. Di solito è la probabilità che controlliamo e che impostiamo ad un valore basso (valori consueti sono 1%, 5%, 10%), in quanto è quella associata all'errore "più pericoloso";
- β è la probabilità di $\overbrace{\text{non rifiutare } H_0 \text{ se } H_0 \text{ è falsa}}^{\text{errore di II tipo}}$. È la probabilità di un *miss*, in quanto *manchiamo* l'obiettivo di riconoscere che l'ipotesi di partenza non è vera.

La probabilità, invece, di identificare l'ipotesi alternativa come vera (nel caso in cui H_1 sia effettivamente vera) è ovviamente $1-\beta$: in questo caso il nostro test è stato abbastanza *potente* da farci giustamente cambiare idea riguardo l'ipotesi di partenza. Per questo motivo, $1-\beta$ è la **potenza del test**.

1.1 I 5 passi per la verifica delle ipotesi

Nell'esecuzione di qualunque test, bisogna rispettare i seguenti passi (crono)logici.

1. Formulare le due ipotesi. Dato un parametro θ, abbiamo:

- $H_0 : \theta = \theta_0$
- H_1 :
  - **test bidirezionale** $\rightarrow \theta \neq \theta_0$
  - **test unidirezionale destro** $\rightarrow \theta > \theta_0$
  - **test unidirezionale sinistro** $\rightarrow \theta < \theta_0$.

2. Fissare α.

3. Determinare la statistica test T_n e la sua distribuzione campionaria.

4. Determinare la regola di decisione, cioè il percentile soglia (test unidirezionale) o i percentili soglia (test bidirezionale) della distribuzione campionaria di T_n.

5. Prendere una decisione in seguito all'esperimento campionario.

1.2 Legame tra intervallo di confidenza e verifica delle ipotesi

L'intervallo di confidenza $\mathbf{IC}_{1-\alpha}(\theta)$ è la regione di tutti i possibili valori $\Theta_0 = \{\theta_{0_1}, \theta_{0_2}, \ldots, \theta_{0_i}, \ldots\}$ per i quali l'ipotesi nulla $H_0 : \theta = \theta_{0_i}$ non è rifiutata ad un livello di significatività α. Per esempio, supponiamo che un intervallo di confidenza del 95% per μ si estenda da -2 a 5. Prendiamo ora $\theta_0 = 4$. Siccome $4 \in [-2, 5]$, risulta che $H_0 : \theta = 4$ non è rifiutata ad un livello α del 5%. Se avessimo preso un qualunque $\theta_0 \notin [-2, 5]$, allora l'ipotesi nulla sarebbe stata rifiutata.

2 Test su un campione

Attraverso i seguenti test su un campione, si fa inferenza su un parametro della popolazione (μ, π, σ^2) tramite un campione estratto da tale popolazione.

2.1 Test sulla media

Il test su μ di una popolazione Normale $X \sim \mathbf{N}(\mu, \sigma^2)$ ha come parametro incognito μ su cui vogliamo fare inferenza. Vogliamo testare l'ipotesi nulla $H_0 : \mu = \mu_0$.

2.1.1 Varianza nota

I seguenti passi per il test su μ assumono (abbastanza inverosimilmente) che la varianza σ^2 della popolazione sia nota:

Passi per la verifica dell'ipotesi
$H_0 : \mu = \mu_0$ test bidirezionale: $H_1 : \mu \neq \mu_0$ test unidirezionale destro: $H_1 : \mu > \mu_0$ test unidirezionale sinistro: $H_1 : \mu < \mu_0$
Fissare il livello di significatività α
Distribuzione della statistica: $\frac{\bar{X}-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim \mathbf{N}(0,1)$
$\mathcal{A}_b$ (regione di accettazione test bidirezionale): $[\mathcal{Z}_{\frac{\alpha}{2}}, \mathcal{Z}_{1-\frac{\alpha}{2}}]$ $\mathcal{A}_d$ (regione di accettazione test unidirezionale destro): $[-\infty, \mathcal{Z}_{1-\alpha}]$ $\mathcal{A}_s$ (regione di accettazione test unidirezionale sinistro): $[\mathcal{Z}_{\alpha}, +\infty]$
Se $\mathcal{Z}_{oss}$ è fuori dalla regione di accettazione, allora rifiuto H_0

Prendiamo in esame il campione di 9 imprese dell'Esercitazione 3, esercizio 2.1.1. Riassumiamo: la media campionaria è 39726.96 mentre la varianza è nota ed uguale a 302500.
Da precedenti studi si ha che la media μ_0 è uguale a 40000 €. Vogliamo testare utilizzando il campione appena raccolto se con la crisi la media dei ricavi mensili si sia abbassata, ad una significatività $\alpha = 1\%$.

Passi per la verifica dell'ipotesi
$H_0 : \mu = 40000$ test unidirezionale sinistro: $H_1 : \mu < 40000$
$\alpha = 0.01$
Distribuzione della statistica: $\frac{\bar{X}-\mu_0}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim \mathbf{N}(0,1)$
$\mathcal{A}_s$: $[\mathcal{Z}_{0.01}, +\infty] = [-2.326, +\infty]$
$\mathcal{Z}_{oss} = \frac{39726.96-40000}{\left(\frac{\sqrt{302500}}{\sqrt{9}}\right)} = -1.49 \in \mathcal{A}_s \Longrightarrow$ non rifiuto H_0

Calcolare il p-value.

Il p-value è la probabilità di osservare sotto H_0 un valore della statistica più estremo di quello osservato. In pratica, tale valore coincide con:

- test bidirezionale: $\mathbf{P}(\mathcal{Z} > |\mathcal{Z}_{oss}|) \times 2$;
- test unidirezionale: $\mathbf{P}(\mathcal{Z} > |\mathcal{Z}_{oss}|)$.

Nel nostro caso, trattandosi di test unidirezionale, il p-value è:

$$p = \mathbf{P}(\mathcal{Z} > |-1.49|) = \mathbf{P}(\mathcal{Z} > 1.49) = 1 - \Phi(1.49) = 1 - 0.932 = 0.068$$

Visto che $p > \alpha$, non rifiutiamo H_0.

Calcolare la potenza del test nel caso in cui il vero valore di μ sia $\mu_1 = 39500$ €.

Per fare ciò, dobbiamo:

1. Riportare gli estremi della regione di accettazione sulla scala originaria della X sotto H_0. Dato che abbiamo $\mathcal{A}_s$, ci basta riscalare $\mathcal{Z}_{0.01} = -2.326$ per ottenere $\mathcal{A}_s^X = [x_0, +\infty]$:

$$x_0 = X_{0.01}^{H_0} = \mu_0 + \mathcal{Z}_{0.01}\,\frac{\sigma}{\sqrt{n}} = 40000 - 2.326\,\frac{\sqrt{302500}}{\sqrt{9}} = 39573.5$$

2. Calcolare l'area sotto H_1 che non interseca $\mathcal{A}_s^X$:

$$\mathbf{P}(X^{H_1} \leq x_0) = \mathbf{P}\left(\frac{x_0 - \mu_1}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right) = \mathbf{P}\left(\frac{39573.5 - 39500}{\left(\frac{302500}{\sqrt{9}}\right)}\right) = \Phi(0.4) = 0.655$$

Nel caso in cui $\mu = 39500$, il test rifiuterebbe giustamente H_0 nel 65.5% dei campioni.

2.1.2 Varianza incognita

Passi per la verifica dell'ipotesi
$H_0 : \mu = \mu_0$ test bidirezionale: $H_1 : \mu \neq \mu_0$ test unidirezionale destro: $H_1 : \mu > \mu_0$ test unidirezionale sinistro: $H_1 : \mu < \mu_0$
Fissare il livello di significatività α
Distribuzione della statistica: $\frac{\bar{X}-\mu_0}{\left(\frac{s}{\sqrt{n}}\right)} \sim t_{n-1}$
$\mathcal{A}_b$ (regione di accettazione test bidirezionale): $[t_{\frac{\alpha}{2},\,n-1}, t_{1-\frac{\alpha}{2},\,n-1}]$ $\mathcal{A}_d$ (regione di accettazione test unidirezionale destro): $[-\infty, t_{1-\alpha,\,n-1}]$ $\mathcal{A}_s$ (regione di accettazione test unidirezionale sinistro): $[t_{\alpha,\,n-1}, +\infty]$
Se t_{oss} è fuori dalla regione di accettazione, allora rifiuto H_0

Poniamo ora il caso di non conoscere la varianza σ^2. Dobbiamo allora stimarla dal campione. Avevamo già calcolato la stima corretta s^2 nell'esercizio 2.1.2 dell'Esercitazione 3 e risultava uguale a 225846.1.
Testare la stessa ipotesi dell'esercizio precedente utilizzando lo stesso livello α.

Passi per la verifica dell'ipotesi
$H_0 : \mu = 40000$ test unidirezionale sinistro: $H_1 : \mu < 40000$
$\alpha = 0.01$
Distribuzione della statistica: $\frac{\bar{X}-\mu_0}{\left(\frac{s}{\sqrt{n}}\right)} \sim t_{n-1}$
$\mathcal{A}_s$: $[t_{0.01,\,8}, +\infty] = [-2.9, +\infty]$
$t_{oss} = \frac{39726.96-40000}{\left(\frac{\sqrt{225846.1}}{\sqrt{9}}\right)} = -1.72 \in \mathcal{A}_s \implies$ non rifiuto H_0

2.2 Test sulla proporzione

Il test su π testa l'ipotesi di partenza che π sia uguale ad un valore π_0. Ricordiamo che la varianza della Bernoulli è funzione del parametro π oggetto di inferenza, perciò non è conoscibile: per stimarla, utilizziamo $\pi_0(1-\pi_0)$. Si noti che $\mathscr{F}_n \xrightarrow{\textbf{TLC}} \mathbf{N}\left(\pi_0, \frac{\pi_0(1-\pi_0)}{n}\right)$ sotto H_0, dunque si possono sfruttare i percentili della $\mathcal{Z}$.

Passi per la verifica dell'ipotesi
$H_0 : \pi = \pi_0$ test bidirezionale: $H_1 : \pi \neq \pi_0$ test unidirezionale destro: $H_1 : \pi > \pi_0$ test unidirezionale sinistro: $H_1 : \pi < \pi_0$
Fissare il livello di significatività α
Distribuzione della statistica: $\frac{\hat{\pi}-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \xrightarrow{\textbf{TLC}} \mathbf{N}(0,1)$
$\mathcal{A}_b$ (regione di accettazione test bidirezionale): $[\mathcal{Z}_{\frac{\alpha}{2}}, \mathcal{Z}_{1-\frac{\alpha}{2}}]$ $\mathcal{A}_d$ (regione di accettazione test unidirezionale destro): $[-\infty, \mathcal{Z}_{1-\alpha}]$ $\mathcal{A}_s$ (regione di accettazione test unidirezionale sinistro): $[\mathcal{Z}_{\alpha}, +\infty]$
Se $\mathcal{Z}_{oss}$ è fuori dalla regione di accettazione, allora rifiuto H_0

Riguardo una certa riforma, risulta che metà della popolazione è favorevole. In un recente sondaggio somministrato a sole 10 persone, solo 2 persone hanno espresso un parere favorevole.
Verificare l'ipotesi che la percentuale di consensi sia cambiata (test bidirezionale) utilizzando un livello di significatività del 5%.

Passi per la verifica dell'ipotesi
$H_0 : \pi = 0.5$ test bidirezionale: $H_1 : \pi \neq 0.5$
$\alpha = 0.05$
Distribuzione della statistica: $\frac{\hat{\pi}-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \xrightarrow{\textbf{TLC}} \mathbf{N}(0,1)$
$\mathcal{A}_b$ (regione di accettazione test bidirezionale): $[\mathcal{Z}_{0.025}, \mathcal{Z}_{0.975}] = [-1.96, 1.96]$
$\mathcal{Z}_{oss} = \frac{0.2-0.5}{\sqrt{\frac{0.5(1-0.5)}{10}}} = -1.9 \in \mathcal{A}_s \Longrightarrow$ non rifiuto H_0

Calcolare il p-value.

Trattandosi di test bidirezionale, il p-value è:

$$p = 2 \times \mathbf{P}(\mathcal{Z} > |-1.9|) = 2 \times \mathbf{P}(\mathcal{Z} > 1.9) = 2(1 - \Phi(1.49)) = 2(1 - 0.971) = 0.057$$

Visto che $p > \alpha$, non rifiutiamo H_0.

Avrebbe senso verificare l'ipotesi $H_0 : \pi = 0$?

No, perché da una popolazione con zero probabilità di consensi non possono essere estratte persone favorevoli.

Avrebbe senso verificare l'ipotesi $H_0 : \pi = 1$?

No, perché da una popolazione con il 100% di consensi non possono essere estratte persone non favorevoli.

2.3 Test sulla varianza

Il test sulla varianza σ^2 verifica l'ipotesi nulla che una popolazione Normale abbia varianza σ_0^2.

Passi per la verifica dell'ipotesi
$H_0 : \sigma^2 = \sigma_0^2$ test bidirezionale: $H_1 : \sigma^2 \neq \sigma_0^2$ test unidirezionale destro: $H_1 : \sigma^2 > \sigma_0^2$ test unidirezionale sinistro: $H_1 : \sigma^2 < \sigma_0^2$
Fissare il livello di significatività α
Distribuzione della statistica: $\frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$
$\mathcal{A}_b$ (regione di accettazione test bidirezionale): $[\chi^2_{1-\frac{\alpha}{2},\, n-1}, \chi^2_{\frac{\alpha}{2},\, n-1}]$ $\mathcal{A}_d$ (regione di accettazione test unidirezionale destro): $[0,\, \chi^2_{\alpha,\, n-1}]$ $\mathcal{A}_s$ (regione di accettazione test unidirezionale sinistro): $[\chi^2_{1-\alpha,\, n-1}, +\infty]$
Se χ^2_{oss} è fuori dalla regione di accettazione, allora rifiuto H_0

Riguardo le solite 9 imprese, supponiamo di non conoscere la varianza come nell'esercizio 2.1.2. Vogliamo verificare $H_0 : \sigma^2 = 100000$ contro l'ipotesi alternativa che la varianza sia aumentata. Dal campione risulta $s^2 = 225846.1$. **Verificare H_0 usando $\alpha = 5\%$.**

Passi per la verifica dell'ipotesi
$H_0 : \sigma^2 = \sigma_0^2$ test unidirezionale destro: $H_1 : \sigma^2 > \sigma_0^2$
$\alpha = 0.05$
Distribuzione della statistica: $\frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$
$\mathcal{A}_d : [0,\, \chi^2_{0.05,\, 8}] = [0, 15.51]$
$\chi^2_{oss} = \frac{8 \times 225846.1}{100000} = 18.07 \notin \mathcal{A}_d \Longrightarrow$ rifiuto H_0

3 Test su due campioni

I test su due campioni hanno l'obiettivo di confrontare due popolazioni A e B partendo dall'ipotesi nulla che una funzione dei parametri appartenenti alle due popolazioni $f(\theta_A, \theta_B)$ sia uguale ad un valore k_0. Per esempio, se si vuole verificare l'ipotesi nulla che la differenza tra le medie sia uguale a 5, allora $f(\mu_A, \mu_B) = \mu_A - \mu_B = 5$. I due campioni hanno rispettivamente ampiezze n_A e n_B.

3.1 Test sul confronto tra medie

Di solito, vogliamo verificare l'ipotesi nulla che due popolazioni Normali abbiano media uguale, cioè $\mu_A - \mu_B = 0$. Ci concentreremo su questo caso negli esercizi.

3.1.1 Varianze note

Passi per la verifica dell'ipotesi
$H_0 : \mu_A - \mu_B = k_0$ test bidirezionale: $H_1 : \mu_A - \mu_B \neq k_0$ test unidirezionale destro: $H_1 : \mu_A - \mu_B > k_0$ test unidirezionale sinistro: $H_1 : \mu_A - \mu_B < k_0$
Fissare il livello di significatività α
Distribuzione della statistica: $\frac{(\bar{X}_A - \bar{X}_B) - \mu_0}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim \mathbf{N}(0,1)$
$\mathcal{A}_b$ (regione di accettazione test bidirezionale): $[\mathcal{Z}_{\frac{\alpha}{2}}, \mathcal{Z}_{1-\frac{\alpha}{2}}]$ $\mathcal{A}_d$ (regione di accettazione test unidirezionale destro): $[-\infty, \mathcal{Z}_{1-\alpha}]$ $\mathcal{A}_s$ (regione di accettazione test unidirezionale sinistro): $[\mathcal{Z}_{\alpha}, +\infty]$
Se $\mathcal{Z}_{oss}$ è fuori dalla regione di accettazione, allora rifiuto H_0

Riprendiamo l'esercizio 3.1.1 dell'Esercitazione 3. Abbiamo due supermercati A e B e vogliamo confrontare le loro vendite medie giornaliere. Uno studio precedente afferma che i due supermercati incassano in media la stessa somma. Sappiamo che $\sigma_A^2 = 3600$ e $\sigma_B^2 = 3900$, mentre le medie sono stimate dai campioni: ricordiamo che $\bar{x}_A = 500$ è calcolata su un campione di 90 unità, mentre $\bar{x}_B = 600$ è calcolata su 30 unità.
Vogliamo verificare che i risultati dello studio precedente siano ancora validi utilizzando un livello di significatività del 10%.

Passi per la verifica dell'ipotesi
$H_0 : \mu_A - \mu_B = 0$ test bidirezionale: $H_1 : \mu_A - \mu_B \neq 0$
$\alpha = 0.1$
Distribuzione della statistica: $\frac{(\bar{X}_A - \bar{X}_B) - \mu_0}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim \mathbf{N}(0,1)$
$\mathcal{A}_b$: $[\mathcal{Z}_{0.05}, \mathcal{Z}_{0.95}] = [-1.645, 1.645]$
$\mathcal{Z}_{oss} = \frac{500-600}{\left(\sqrt{\frac{3600}{90} + \frac{3900}{30}}\right)} \sim \mathbf{N}(0,1) = -7.67 \notin \mathcal{A}_b \Longrightarrow$ rifiuto H_0

Se prendiamo l'intervallo di confidenza del 90% calcolato nell'Esercitazione 3 (esercizio 3.1.1):
$\mathbf{IC}_{0.9}(\mu_A - \mu_B) = [-121.45, -78.45]$, possiamo notare che esso non contiene $k_0 = 0$, dunque rifiutiamo H_0.

Calcolare il *p*-value.

Trattandosi di test bidirezionale, il p-value è:

$$p = 2 \times \mathbf{P}(\mathcal{Z} > 7.67) \approx 2 \times 0 = 0$$

3.1.2 Varianze incognite ma supposte uguali

Per la discussione sul perché assumere che le varianze delle due popolazioni siano uguali qualora ci siano incognite, si veda l'Esercitazione 3, paragrafo 3.1.2. In tale paragrafo, abbiamo anche determinato uno stimatore *pooled* s_p^2 per la varianza che sarà usato a seguire.

Passi per la verifica dell'ipotesi
$H_0 : \mu_A - \mu_B = k_0$ test bidirezionale: $H_1 : \mu_A - \mu_B \neq k_0$ test unidirezionale destro: $H_1 : \mu_A - \mu_B > k_0$ test unidirezionale sinistro: $H_1 : \mu_A - \mu_B < k_0$
Fissare il livello di significatività α
Distribuzione della statistica: $\frac{(\bar{X}_A - \bar{X}_B) - k_0}{\sqrt{s_p^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \sim t_{n-1}$
$\mathcal{A}_b$ (regione di accettazione test bidirezionale): $[t_{\frac{\alpha}{2}}, t_{1-\frac{\alpha}{2}}]$ $\mathcal{A}_d$ (regione di accettazione test unidirezionale destro): $[-\infty, t_{1-\alpha}]$ $\mathcal{A}_s$ (regione di accettazione test unidirezionale sinistro): $[t_{\alpha}, +\infty]$
Se t_{oss} è fuori dalla regione di accettazione, allora rifiuto H_0

Supponiamo di non sapere le varianze delle vendite giornaliere dei due supermercati (che supponiamo uguali). Nell'Esercitazione 3, abbiamo calcolato la loro stima congiunta $s_p^2 = 4025.77$.
Verificare la stessa ipotesi nulla dell'esercizio precedente usando $\alpha = 10\%$.

Passi per la verifica dell'ipotesi
$H_0 : \mu_A - \mu_B = 0$ test bidirezionale: $H_1 : \mu_A - \mu_B \neq 0$
$\alpha = 0.1$
Distribuzione della statistica: $\frac{(\bar{X}_A - \bar{X}_B) - k_0}{\sqrt{s_p^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \sim t_{n-1}$
$\mathcal{A}_b$: $[t_{0.05,\,118}, t_{0.95,\,118}] \approx [-1.645, 1.645]$
$t_{oss} = \frac{500-600}{\sqrt{4025.77\left(\frac{1}{90} + \frac{1}{30}\right)}} = -7.48 \notin \mathcal{A}_b \Longrightarrow$ rifiuto H_0

3.2 Test sul confronto tra proporzioni

Di solito, vogliamo verificare l'ipotesi nulla che due popolazioni Bernoulliane abbiano uguale proporzione di successi, cioè $\pi_A - \pi_B = 0$. Ci concentreremo su questo caso nell'esercizio. In questo caso, abbiamo bisogno di uno stimatore *pooled* $\hat{\pi}_p$ per π al fine di stimare la varianza della Bernoulli: si veda il paragrafo 3.2 dell'Esercitazione 3 per calcolarlo.

Passi per la verifica dell'ipotesi
$H_0 : \pi_A - \pi_B = k_0$ test bidirezionale: $H_1 : \pi_A - \pi_B \neq k_0$ test unidirezionale destro: $H_1 : \pi_A - \pi_B > k_0$ test unidirezionale sinistro: $H_1 : \pi_A - \pi_B < k_0$
Fissare il livello di significatività α
Distribuzione della statistica: $\frac{(\hat{\pi}_A - \hat{\pi}_B) - k_0}{\sqrt{\hat{\pi}_p(1-\hat{\pi}_p)\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} \xrightarrow{\textbf{TLC}} \mathbf{N}(0,1)$
$\mathcal{A}_b$ (regione di accettazione test bidirezionale): $[\mathcal{Z}_{\frac{\alpha}{2}}, \mathcal{Z}_{1-\frac{\alpha}{2}}]$ $\mathcal{A}_d$ (regione di accettazione test unidirezionale destro): $[-\infty, \mathcal{Z}_{1-\alpha}]$ $\mathcal{A}_s$ (regione di accettazione test unidirezionale sinistro): $[\mathcal{Z}_{\alpha}, +\infty]$
Se $\mathcal{Z}_{oss}$ è fuori dalla regione di accettazione, allora rifiuto H_0

Riprendiamo l'esercizio 3.2 dell'Esercitazione 3. Abbiamo due campagne sull'inquinamento: per la campagna A osserviamo 220000 aderenti su un milione di persone; per la campagna B osserviamo 34500 aderenti su 150000 persone. Da questi dati campionari abbiamo ottenuto: $n_A = 1000000$, $\hat{\pi}_A = \frac{220000}{1000000} = 0.22$, $n_B = 150000$, $\hat{\pi}_B = \frac{34500}{150000} = 0.23$, da cui $\hat{\pi}_p = \frac{220000+34500}{1000000+150000} = 0.2213$.
Vogliamo verificare l'ipotesi nulla che le proporzioni di consenso nelle due diverse campagne siano uguali, usando $\alpha = 1\%$.

Passi per la verifica dell'ipotesi
$H_0 : \pi_A - \pi_B = 0$ test bidirezionale: $H_1 : \pi_A - \pi_B \neq 0$
$\alpha = 0.01$
Distribuzione della statistica: $\frac{(\hat{\pi}_A - \hat{\pi}_B) - k_0}{\sqrt{\hat{\pi}_p(1-\hat{\pi}_p)\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} \xrightarrow{\textbf{TLC}} \mathbf{N}(0,1)$
$\mathcal{A}_b : [\mathcal{Z}_{\frac{\alpha}{2}}, \mathcal{Z}_{1-\frac{\alpha}{2}}] = [-2.58, 2.58]$
$\mathcal{Z}_{oss} = \frac{0.22-0.23}{\sqrt{0.2213\,(1-0.2213)\left(\frac{1}{1000000}+\frac{1}{150000}\right)}} = -8.7 \notin \mathcal{A}_b \Longrightarrow$ rifiuto H_0

Calcolare il *p*-value.

Trattandosi di test bidirezionale, il p-value è:

$$p = 2 \times \mathbf{P}(\mathcal{Z} > 8.7) \approx 2 \times 0 = 0$$

3.3 Test sul confronto tra varianze

Il test per il confronto tra le varianze di due popolazioni Normali (**NB**: l'inferenza sulla varianza è poco robusta alla violazione della Normalità!) verifica l'ipotesi nulla che $\frac{\sigma_A^2}{\sigma_B^2}$ sia uguale a k_0. Di solito, vogliamo testare se le due varianze sono uguali, cioè se $k_0 = 1$. In genere, eseguiamo un test unidirezionale destro (ma un test bidirezionale è altrettanto corretto) in cui si pone la varianza campionaria più alta al numeratore e quella più bassa al denominatore.

Passi per la verifica dell'ipotesi
$H_0 : \frac{\sigma_A^2}{\sigma_B^2} = k_0$ test bidirezionale: $H_1 : \frac{\sigma_A^2}{\sigma_B^2} \neq k_0$ test unidirezionale destro: $H_1 : \frac{\sigma_A^2}{\sigma_B^2} > k_0$
Fissare il livello di significatività α
Distribuzione della statistica: $\frac{\left(\frac{s_A^2}{s_B^2}\right)}{k_0} \sim \mathbf{F}_{n_A-1,\, n_B-1}$
$\mathcal{A}_b$ (regione di accettazione test bidirezionale): $[F_{1-\frac{\alpha}{2},\, n_A-1,\, n_B-1},\ F_{\frac{\alpha}{2},\, n_A-1,\, n_B-1}]$ $\mathcal{A}_d$ (regione di accettazione test unidirezionale destro): $[0,\ F_{\alpha,\, n_A-1,\, n_B-1}]$
Se F_{oss} è fuori dalla regione di accettazione, allora rifiuto H_0

Torniamo all'esercizio 3.3 dell'Esercitazione 3, riguardante i voti di Statistica di due corsi di laurea (EA ed EC). Ci si aspetta che non vi siano differenze sostanziali in variabilità (dunque sotto l'ipotesi nulla, il rapporto tra le varianze deve essere uguale a 1). Per testare tale ipotesi, prendiamo i due campioni a disposizione (rispettivamente di 20 e 15 studenti), da cui abbiamo ottenuto $s^2_{\mathrm{EA}} = 23$ e $s^2_{\mathrm{EC}} = 25$.
Verificare H_0 con un livello di significatività del 5%. utilizzando un test unidirezionale destro.

Passi per la verifica dell'ipotesi
$H_0 : \frac{\sigma_A^2}{\sigma_B^2} = 1$ test unidirezionale destro: $H_1 : \frac{\sigma_A^2}{\sigma_B^2} > 1$
$\alpha = 0.05$
Distribuzione della statistica: $\frac{\left(\frac{s_A^2}{s_B^2}\right)}{k_0} \sim \mathbf{F}_{n_A-1,\, n_B-1}$
$\mathcal{A}_d : [0,\ F_{0.05,\, 14,\, 19}] = [0,\ 2.26]$
$F_{oss} = \frac{25}{23} = 1.087 \in \mathcal{A}_d \Longrightarrow$ non rifiuto H_0