ESERCIZIO 5.22 (svolto da me)

In una GSS si è domandato, "Quale ritieni debba essere il numero ideale di figli per una famiglia?" La distribuzione delle risposte date dalle 497 donne intervistate presenta una media pari a 3.02. La deviazione standard della popolazione è conosciuta ed è pari a 1.81.

a) Riporta la stima puntuale della media della popolazione.

b) Trova e interpreta l'errore standard della media campionaria.

c) Trova l'intervallo di confidenza al 95% e interpretalo.

d) È plausibile che la popolazione abbia media = 2.0? Fornisci una spiegazione.

RISOLUZIONE:

a)

La stima puntuale della media della popolazione è il valore della media campionaria. Quindi per i dati del problema **$\bar{x}$ = 3.02.**

b)

L'errore standard della media campionaria si calcola con la formula **se = σ/√n** dove **σ** è la deviazione standard della popolazione e **n** è la numerosità del campione.

se = 1.81/√497 = 0.0812

L'errore standard è un indicatore della variabilità delle osservazioni intorno alla media campionaria, rappresenta la deviazione standard della distribuzione che ha come centro **$\bar{x}$** (distribuzione campionaria della media).

c)

Un intervallo di confidenza al 95% si trova con la formula: **I.C. = $\bar{x}$ ± z*se**. Dove **z** è lo z-score corrispondente a un livello di fiducia del 95%. Ovvero quello z-score che nella distribuzione normale standardizzata lascia aldilà della coda destra e sinistra una proporzione totale del 0.05 (quindi 0.025 per coda).

Per un livello di fiducia al 95% **z = 1.96** (ricavabile dalla tavola a pag.286 del libro)

Ricavo l'intervallo: **I.C. = 3.02 ± 1.96*0.0812 = 3.02 ± 0.16 = (2.86, 3.18)**

Quindi l'intervallo di fiducia al 95% è **(2.86, 3.18).**

Ciò mi dice che: ripetendo il processo di campionamento innumerevoli volte, mi aspetto, con una fiducia al 95%, che il vero valore medio **μ** della popolazione sia contenuto nell'intervallo **(2.86, 3.18)**.

d)

Non è plausibile che il valore effettivo della media **μ** della popolazione sia **2.0** poiché, tenendo in considerazione il punto **c)**, **2.0** è un valore estremamente distante sia dal valore dalla stima puntuale della media **$\bar{x}$ = 3.02** sia dall'intervallo di confidenza **(2.9, 3.2)**

(Se consideriamo la distribuzione come una normale con media **3.02** e deviazione standard pari all'errore standard infatti si trova che il valore **2.0** dista circa **12 errori standard** dalla media, con una probabilità estremamente bassa).

ESERCIZIO 5.23

In riferimento al Problema 5.22, per i 397 maschi del campione, la media era pari a 2.89 e la deviazione standard della popolazione era pari a 1.77.

a) Mostra che l'errore standard della media campionaria è 0.089.

b) Trova l'intervallo di confidenza al 95% per la media della popolazione e spiega cosa significa "fiducia al 95%".

RISOLUZIONE:

a)

L'errore standard della media campionaria si calcola con la formula **se = σ/√n** dove **σ** è la deviazione standard della popolazione e **n** è la numerosità del campione.

se = 1.77/√397 = 0.0888, approssimabile a **0.089**.

b)

Un intervallo di confidenza al 95% si trova con la formula: **I.C. = $\bar{x}$ ± z*se**. Dove **z** è lo z-score corrispondente a un livello di fiducia del 95%. Ovvero quello z-score che nella distribuzione normale standardizzata lascia aldilà della coda destra e sinistra una proporzione totale del 0.05 (quindi 0.025 per coda).

Per un livello di fiducia al 95% **z = 1.96** (ricavabile dalla tavola a pag.286 del libro).

Ricavo l'intervallo: **I.C. = 2.89 ± 1.96*0.089 = 2.89 ± 0.17 = (2.72, 3.06)**

Quindi l'intervallo di fiducia al 95% è **(2.72, 3.06).**

Una **fiducia al 95%** significa che: su infiniti campioni estratti nello stesso modo e con stessa dimensione campionaria, mi aspetto che il 95% di essi avranno un intervallo di confidenza che contiene l'effettivo valore medio **μ** della popolazione mentre circa il 5% avranno un intervallo che non lo conterrà.

ESERCIZIO 5.25

Nella GSS del 2004 è stato chiesto a un campione di 892 intervistati, "per quante ore in media al giorno guardi la televisione?" I risultati ottenuti sono stati i seguenti:

$\overline{x} = 2.76$; $\sigma = 2.39$

a) Ricava l'intervallo di confidenza al 99%.

RISOLUZIONE:

a)

Un intervallo di confidenza al 99% si trova con la formula: **I.C. = $\overline{x}$ ± z*se**. Dove **se** è l'errore standard della media campionaria e dove **z** è lo z-score corrispondente a un livello di fiducia del 99%. Ovvero quello z-score che nella distribuzione normale standardizzata lascia aldilà della coda destra e sinistra una proporzione totale del 0.01 (quindi 0.005 per coda).

se = σ/√n dove **σ** è la deviazione standard della popolazione e **n** è la numerosità del campione.

se = 2.39/√892 = 0.08

Per un livello di fiducia al 99% **z = 2.58** (ricavabile dalla tavola a pag. 286 del libro)

Ricavo l'intervallo: **I.C. = 2.76 ± 2.58*0.08 = 2.76 ± 0.21 = (2.55, 2.97)**

Quindi l'intervallo di fiducia al 95% è **(2.55, 2.97).**

ESERCIZIO 5.28

In una recente GSS si è chiesto, "in quanti degli ultimi 7 giorni ti sei sentito triste?" Le risposte delle 816 donne hanno avuto media pari a 1.81. La deviazione standard della popolazione delle donne è pari a 1.98. Per i 633 intervistati maschi: media pari a 1.42. La deviazione standard della popolazione dei maschi è pari a 1.83.

a) Trova un intervallo di confidenza al 95% per la media della popolazione delle donne e dei maschi.

b) Spiega perché i valori della media e della deviazione standard suggeriscono che questa variabile non ha una distribuzione normale. Ciò rappresenta un problema per il metodo dell'intervallo di confidenza determinato al punto **a)**? Fornisci una spiegazione.

RISOLUZIONE:

Un intervallo di confidenza al 95% si trova con la formula: **I.C. = $\bar{x}$ ± z*se**. Dove **se** è l'errore standard della media campionaria e dove **z** è lo z-score corrispondente a un livello di fiducia del 95%. Ovvero quello z-score che nella distribuzione normale standardizzata lascia aldilà della coda destra e sinistra una proporzione totale del 0.05 (quindi 0.025 per coda).

se = σ/√n dove **σ** è la deviazione standard della popolazione e **n** è la numerosità del campione.

a) Per le donne:

se = 1.98/√816 = 0.069

Per un livello di fiducia al 95% **z = 1.96** (ricavabile dalla tavola a pag.286 del libro)

Ricavo l'intervallo: **I.C. = 1.81 ± 1.96*0.069 = 1.81 ± 0.14 = (1.67, 1.95)**

Quindi l'intervallo di fiducia al 95% è **(1.67, 1.95).**

Per gli uomini:

se = 1.83/√633 = 0.073

Per un livello di fiducia al 95% **z = 1.96** (ricavabile dalla tavola a pag.286 del libro)

Ricavo l'intervallo: **I.C. = 1.42 ± 1.96*0.073 = 1.42 ± 0.14 = (1.28, 1.56)**

Quindi l'intervallo di fiducia al 95% è **(1.28, 1.56).**

b)

Quei valori mi suggeriscono che la variabile non si distribuisca normalmente perché già a una deviazione standard i valori della coda di sinistra vanno oltre lo 0 diventando negativi (e "impossibili" data la scala di misurazione), sia per le donne che per gli uomini. Infatti: per le donne: 1.81 – 1.98 = -0.17 e per gli uomini: 1.42 – 1.83 = -0.41.

Questo mi suggerisce che la variabile si distribuisca in modo asimmetrico positivo, ovvero che i valori si concentrano fra lo 0 e la media.

Questo però non mi rappresenta un problema per il metodo dell'intervallo di confidenza perché la distribuzione campionaria della media è robusta nei rispetti della violazione del requisito della popolazione normale per grandi numerosità campionarie (come in questo caso). Infatti, come dimostrato dal Teorema del Limite Centrale, per grandi dimensioni campionarie la distribuzione campionaria della media tende a essere normale, aldilà della forma della distribuzione della popolazione.