# PROBLEMA 1

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.

1. Riporta la stima puntuale della media della popolazione.
2. Trova e interpreta l’errore standard della media campionaria.
3. Trova l’intervallo di confidenza al 95% e interpretalo.
4. 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 **x̅ = 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 **x ̅** (distribuzione campionaria della media).

# c)

Un intervallo di confidenza al 95% si trova con la formula: **I.C. = x̅ ± 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 **x̅ = 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).

# PROBLEMA 2

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.

1. Mostra che l’errore standard della media campionaria è 0.089.
2. 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. = x̅ ± 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.

# PROBLEMA 3

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:

x ̅ = 2.76; σ = 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. = x̅ ± 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 zscore 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).**

**PROBLEMA 4**

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.

1. Trova un intervallo di confidenza al 95% per la media della popolazione delle donne e dei maschi.
2. 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. = x̅ ± 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 zscore 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.

# PROBLEMA 5

In un sondaggio condotto negli USA, è stato chiesto agli intervistati se essi fossero favorevoli alle unioni civili. Dei 2003 adulti intervistati, il 54% ha detto SI, il 42% NO e il 4% non ha espresso opinioni. Trova l’errore standard della stima per la proporzione campionaria di chi risponde SI. Fornisci un’interpretazione.

## Soluzione

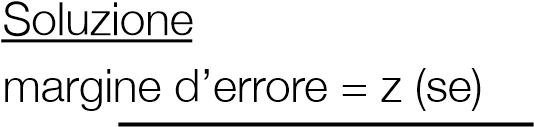
se = √ (0.54 x 0.46) / 2003 = 0.01

l’errore standard della stima per la proporzione campionaria di chi risponde SI fornisce una previsione dell’errore che si commette nell’utilizzare la proporzione campionaria di chi risponde SI per stimare la proporzione della popolazione che risponderebbe SI

# PROBLEMA 6

Quando in un sondaggio nel 2006 è stato chiesto se la Turchia dovesse entrare a far parte della UE, la percentuale di chi ha risposto SI è stata pari al 51% in Danimarca (n = 1008) e al 42% nel Regno Unito (n = 1312). In riferimento al risultato per la Danimarca, il report prodotto indicava che il margine d’errore è più o meno 3.1%. Spiega come è stato ottenuto questo risultato.

## Soluzione



se = √ (0.51 x 0.49) / 1008 = 0.016

il margine d’errore è un multiplo (z\*) dell’errore standard della stima: 1.96 (0.016) = 0.031 = 3.1%

# PROBLEMA 7

In un sondaggio condotto negli USA, una domanda ha chiesto “Ritieni che debba essere responsabilità del governo ridurre le differenze tra ricchi e poveri?”. Coloro che hanno risposto SI comprendevano 90 dei 142 soggetti che si autodefinivano “Democratici” e 26 dei 102 autodefinitesi “Repubblicani”.

1. Trova la stima puntuale della proporzione della popolazione che dovrebbe rispondere SI in ciascun gruppo;
2. Trova l’intervallo di confidenza al 95% per la proporzione della popolazione che risponde SI tra i

Democratici;

1. Trova l’intervallo di confidenza al 99% per la proporzione della popolazione che risponde SI tra i

Democratici;

1. Trova l’intervallo di confidenza al 95% per la proporzione della popolazione che risponde SI tra i Repubblicani;
2. Trova l’intervallo di confidenza al 99% per la proporzione della popolazione che risponde SI tra i

Repubblicani;

1. Spiega come interpretare tali intervalli.

## Soluzione

1. p(“Si”|D) = 90/142 = 0.63

p(“Si”|R) = 26/102 = 0.25

1. se = 0.04, 95% IC per π(“Si”|D) = 0.63 ± 1.96 (0.04) = 0.63 ± 0.08 [0.55, 0.71]
2. 99% IC per π(“Si”|D) = 0.63 ± 0.10 [0.53, 0.73]
3. se = 0.04, 95% IC per π(“Si”|R) = 0.25 ± 0.08 [0.17, 0.33]
4. 99% IC per π(“Si”|R) = 0.25 ± 0.10 [0.15, 0.35]
5. L’intervallo di confidenza è un intervallo di valori entro cui ricade il parametro π; la probabilità associata al fatto che l’intervallo contenga il parametro è denominata livello di fiducia.

# PROBLEMA 8

In un sondaggio è stato chiesto se si è o meno d’accordo con la seguente affermazione: “è meglio per tutti se l’uomo lavora e la donna si prende cura della casa e della famiglia”. La proporzione campionaria di chi era d’accordo è stata 0.66 nel 1977 e 0.36 nel 2004 (n=883). a) mostra che l’errore standard stimato nel 2004 è stato 0.016

1. mostra che il margine di errore per un intervallo di confidenza al 95% utilizzato nella stima del

2004 era 0.03

1. costruisci l’intervallo di confidenza al 95% per il 2004 e interpreta

## Soluzione

1. se = √ (0.36 x 0.64) / 883 = 0.016
2. margine d’errore = z (se) = 1.96 (0.016) = 0.031
3. 95% IC per π(“Si”|2004) = 0.36 ± 0.03 [0.33, 0.39] questo intervallo conterrà il parametro π(“Si”|2004) nel 95% dei campioni casuali di ampiezza 883 estratti dalla popolazione.

# PROBLEMA 9

Un sondaggio negli USA ha chiesto se le attuali normative ambientali fossero troppo restrittive o meno. Dei 1200 rispondenti, 229 hanno detto che lo sono (“Si”)

Trova:

1. un intervallo di confidenza al 95% per il valore del parametro;
2. un intervallo di confidenza al 99% per il valore del parametro.

## Soluzione

1. P(“Si”) = 229/1200 = 0.19, se = 0.01, 95% IC per π(“Si”) = 0.19 ± 1.96 (0.01) = 0.19 ± 0.02 [0.17, 0.21]
2. 99% IC per π(“Si”) = 0.19 ± 2.58 (0.01) = 0.19 ± 0.03 [0.16, 0.22]

# PROBLEMA 10

Secondo un exit poll su persone che hanno votato all’elezione per governatore, il 40% ha votato per Jones ed il 60% per Smith. Ipotizzando che questo sia un campione casuale di tutti gli elettori:

1. costruisci un intervallo di confidenza al 99% per la proporzione dei voti per Jones nel caso in cui la dimensione campionaria sia 400
2. costruisci un intervallo di confidenza al 99% per la proporzione dei voti per Jones nel caso in cui la dimensione campionaria sia 40
3. saresti disposto a prevedere il vincitore?

## Soluzione

1. N=400

P(Jones) = 0.40

se = 0.02, 99% IC per π(Jones) = 0.40 ± 0.05 [0.35, 0.45]

sono disposto/a a prevedere il vincitore perché tutti i valori contenuti nell’IC sono sotto 0.5 ➝ vincitore Smith

1. N=40

P (Jones) = 0.40

se = 0.08, 99% IC per π(Jones) = 0.40 ± 0.21 [0.19, 0.61]

NON sono disposto/a a prevedere il vincitore perché i valori contenuti nell’IC sono **sia sotto che sopra 0.5** ➝ esito incerto

1. vedi sopra; sulla decisione incide la dimensione campionaria (la stima puntuale è uguale in entrambi i casi, ma al diminuire della dimensione campionaria aumenta SE e quindi l’IC risulta avere una maggiore ampiezza).

# PROBLEMA 11

In un sondaggio è stato chiesto “Quale ritieni debba essere il numero ideale di figli per una famiglia?”. La distribuzione delle risposte date dalle 497 donne intervistate presenta una mediana pari a 2, una media pari a 3.02 ed una deviazione standard (*nel campione*) pari a 1.81.

1. Riporta la stima puntuale della media della popolazione;
2. Trova l’errore standard della media campionaria;
3. Trova l’intervallo di confidenza al 95% e fornisci un’interpretazione;
4. Trova l’intervallo di confidenza al 99% e fornisci un’interpretazione;
5. È plausibile che la popolazione abbia media=2.0? Fornisci una spiegazione.

## Soluzione

1. ȳ = 3.02 ➝ la stima puntuale della media della popolazione coincide con la media campionaria
2. se = s/√n = 1.81/√497 = 0.08
3. 95% IC per µ = 3.02 ± 1.96 (0.08) = 3.02 ± 0.16 [2.86, 3.18] ➝ l’intervallo [2.86, 3.18] conterrà il parametro µ nel 95% dei campioni casuali di ampiezza n estratti dalla popolazione.
4. 99% IC per µ = 3.02 ± 0.21 [2.81, 3.23] ➝ l’intervallo [2.81, 3.23] conterrà il parametro µ nel 99% dei campioni casuali di ampiezza n estratti dalla popolazione
5. Poco probabile visto che tale valore non è compreso nel 99% IC.

# PROBLEMA 12

In riferimento al problema precedente, per i 397 maschi del campione, la media era pari a 2.89 e la deviazione standard (*nel campione*) a 1.77.

1. Mostra che l’errore standard della media campionaria è 0.089;
2. Trova l’intervallo di confidenza al 95% per la media della popolazione e spiega cosa significa “fiducia al 95%”.

## Soluzione

1. se = 1.77 / √397 = 0.089
2. 95% IC per µ = 2.89 ± 0.17 [2.72, 3.06] ho una probabilità pari a 0.95 che il metodo produca un intervallo che contenga il parametro µ

# PROBLEMA 13

In uno studio sono stati registrati i cambiamenti di peso di 17 ragazze con diagnosi di anoressia, sottoposte a terapia cognitivo-comportamentale: 11, 11, 6, 9, 14, -3, 0, 7, 22, -5, -4, 13, 13, 9, 4, 6, 11

1. calcola media e deviazione standard
2. calcola l’errore standard della media campionaria
3. qual’è il valore del *t*-Student per un intervallo di confidenza al 95%?
4. sia µ la media del cambiamento di peso della popolazione per questa terapia. Trova l’intervallo di confidenza al 95% per µ.

## Soluzione

1. ȳ = 7.29, s = 7.18
2. se = 1.74
3. gdl = n - 1 = 17 - 1 = 16 *t*-Student = 2.12 per un 95% IC
4. 95% IC per µ = 7.29 ± 2.12 (1.74) = [3.60, 10.98]