Introduzione all'Inferenza Statistica

Fabrizio Cipollini
Dipartimento di Statistica, Informatica, Applicazioni (DiSIA) *G. Parenti*
Università di Firenze

Firenze, 3 Febbraio 2015

Casi di studio

Velocità della luce

- Nel 1880, Simon Newcomb la valuta in base al tempo che la luce impiega per fare $\simeq$ 7400m.
- Qual è il suo valore?
- Campione di 66 misure.

Default nel mondo del credito

- Una società finanziaria presta denaro a un cliente (prestito personale).
- Qual è la probabilità che il cliente vada in *default* (cioè non ripaghi il prestito)?
- Campione di 5766 casi.

Casi di studio

Velocità della luce

- Nel 1880, Simon Newcomb la valuta in base al tempo che la luce impiega per fare $\simeq$ 7400m.
- Qual è il suo valore?
- Campione di 66 misure.

Default nel mondo del credito

- Una società finanziaria presta denaro a un cliente (prestito personale).
- Qual è la probabilità che il cliente vada in default (cioè non ripaghi il prestito)?
- Campione di 5960 casi.

Casi di studio

Velocità della luce

- Nel 1880, Simon Newcomb la valuta in base al tempo che la luce impiega per fare $\simeq$ 7400m.
- Qual è il suo valore?
- Campione di 66 misure.

Default nel mondo del credito

- Una società finanziaria presta denaro a un cliente (prestito personale).
- Qual è la probabilità che l'aliante vada in default (cioè non ripaghi il prestito)?
- Campione di 5960 casi.

Casi di studio

Velocità della luce

- Nel 1880, Simon Newcomb la valuta in base al tempo che la luce impiega per fare $\simeq$ 7400m.
- Qual è il suo valore?
- Campione di 66 misure.

Default nel mondo del credito

- Una società finanziaria presta denaro a un cliente (prestito personale).
- Qual è la probabilità che il cliente vada in *default* (cioè non ripaghi il prestito)?
- Campione di 5960 casi.

Casi di studio

Velocità della luce

Nel 1880, Simon Newcomb ha valutato in base al tempo che la luce impiega per fare ≈ 7400m.

- Qual è il suo valore?
- Campione di 66 misure.

Default nel mondo del credito

- Una società finanziaria presta denaro a un cliente (prestito personale).
- Qual è la probabilità che il cliente vada in *default* (cioè non ripaghi il prestito)?
- Campione di 5756 casi.

Casi di studio

Velocità della luce

Nel 1880, Simon Newcomb ha valutato in base al tempo che la luce impiegò per fare ≈ 7400m.

Qual è il suo valore?

- Campione di 66 misure.

Default nel mondo del credito

- Una società finanziaria presta denaro a un cliente (prestito personale).
- Qual è la probabilità che il cliente vada in *default* (cioè non ripaghi il prestito)?
- Campione di 5756 casi.

Casi di studio

Default nel mondo del credito

- Una società finanziaria presta denaro a un cliente (prestito personale).
- Qual è la probabilità che il cliente vada in *default* (cioè non ripaghi il prestito)?
- Campione di 5756 casi.

Situazione pratica

- **Variabile**: '*misura* della velocità della luce'

$$X$$

- **Dati**: $n = 66$ misurazioni espresse in km/s,

$$(x_1 = 264286, x_2 = 336364, \ldots, x_{66} = 321739)$$

- **Obiettivo**: stimare la velocità della luce.

Situazione pratica

- **Variabile**: '*misura* della velocità della luce'

$$X$$

- **Dati**: $n = 66$ misurazioni espresse in km/s,

$$(x_1 = 264286, x_2 = 336364, \ldots, x_{66} = 321739)$$

- **Obiettivo**: stimare la velocità della luce.

Situazione pratica

- **Variabile**: '*misura* della velocità della luce'

$$X$$

- **Dati**: $n = 66$ misurazioni espresse in km/s,

$$(x_1 = 264286, x_2 = 336364, \ldots, x_{66} = 321739) = \underline{x}$$

- **Obiettivo**: stimare la velocità della luce.

Situazione pratica

- **Variabile**: '*misura* della velocità della luce'

$$X$$

- **Dati**: $n = 66$ misurazioni espresse in km/s,

$$(x_1 = 264286, x_2 = 336364, \ldots, x_{66} = 321739) = \underline{x}$$

- **Obiettivo**: stimare la velocità della luce.

Statistica esplorativa

- ***Data cleaning***: due valori negativi $\rightarrow$ eliminare.
- Calcolo di **statistiche descrittive**:

n	min	max	mean	median	sd
64	185000	462500	276234.3	269179.9	55311.5

- **Grafici** (istogrammi):

Statistica esplorativa

- ***Data cleaning***: due valori negativi $\rightarrow$ eliminare.
- Calcolo di **statistiche descrittive**:

n	min	max	mean	median	sd
64	185000	462500	276234.3	269179.9	55311.5

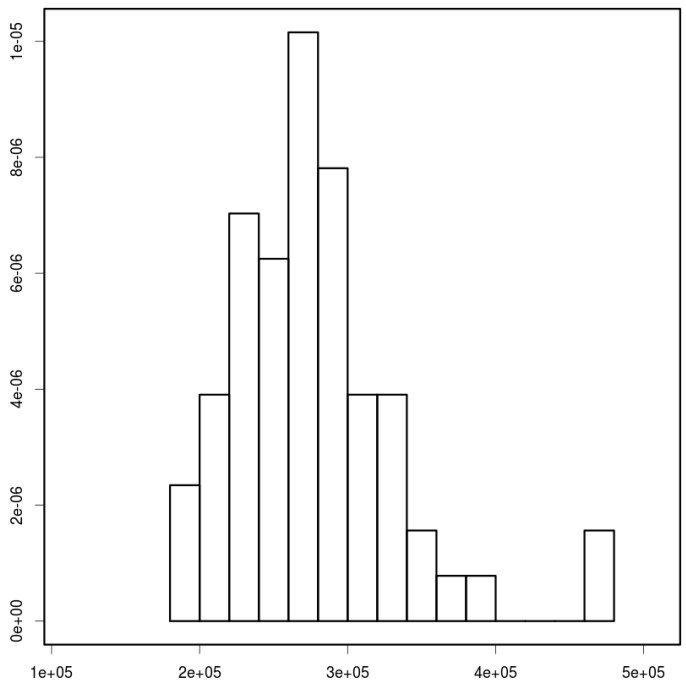
- **Grafici** (istogrammi):

Statistica esplorativa

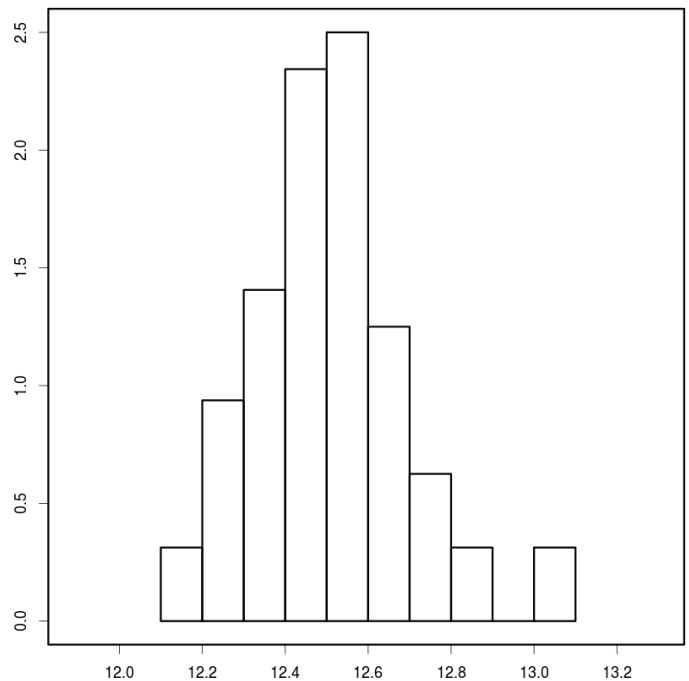
- ***Data cleaning***: due valori negativi → eliminare.
- Calcolo di **statistiche descrittive**:

n	min	max	mean	median	sd
64	185000	462500	276234.3	269179.9	55311.5

- **Grafici** (istogrammi):



(a) Misurazione



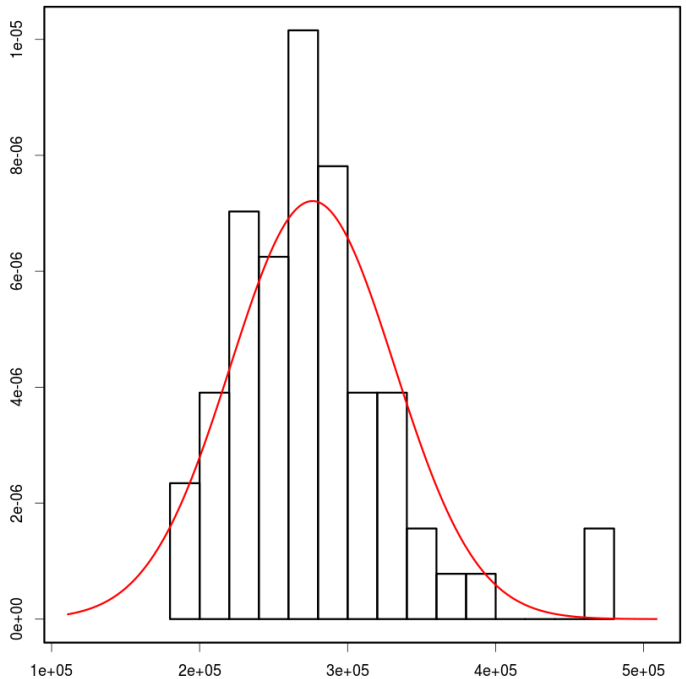
(b) ln(Misurazione)

Statistica esplorativa

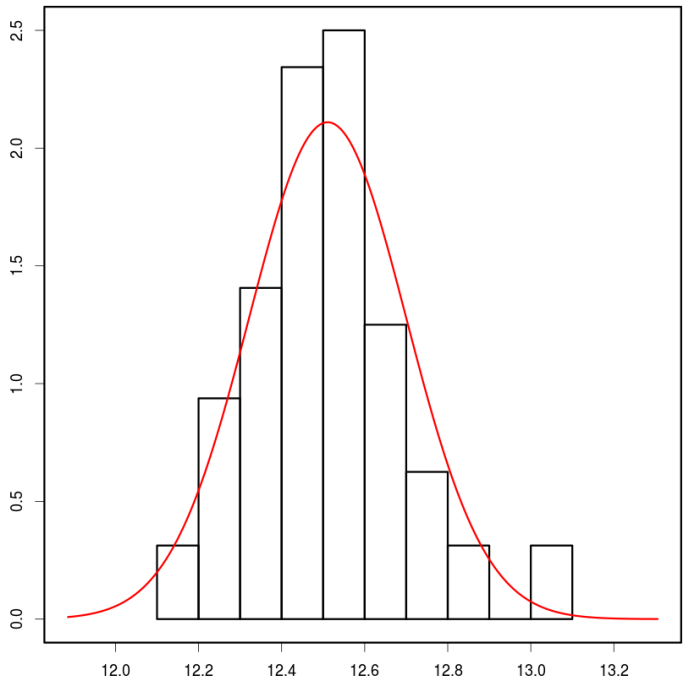
- ***Data cleaning***: due valori negativi → eliminare.
- Calcolo di **statistiche descrittive**:

n	min	max	mean	median	sd
64	185000	462500	276234.3	269179.9	55311.5

- **Grafici** (istogrammi):



(a) Misurazione



(b) ln(Misurazione)

Dall'esplorazione all'inferenza

- Ogni misura può dare un valore diverso.

$$X = \text{`}misura\text{ della velocità della luce'}$$

è una variabile casuale.

- Ragionevole pensare che X si distribuisca intorno ad una media μ con una certa standard deviation σ,

$$X \sim \left(\mu, \sigma\right)$$

  - μ = velocità della luce (se l'esperimento è ben progettato!)
  - σ = grado di imprecisione dell'esperimento
  - Valore/i di μ e σ non dati/campionati
  - Potremmo anche assumere che X abbia una distribuzione Normale (vedi pratica)

Dall'esplorazione all'inferenza

- Ogni misura può dare un valore diverso.

$$X = \text{`}misura\text{ della velocità della luce'}$$

è una variabile casuale.

- Ragionevole pensare che X si distribuisca intorno ad una media μ con una certa standard deviation σ,

$$X \sim \left(\mu, \sigma\right)$$

  - μ = velocità della luce (se l'esperimento è ben progettato!)
  - σ = grado di imprecisione dell'esperimento
  - Media e sd di X, non del campione!
  - Potremmo anche assumere che X abbia una distribuzione Normale (vedi grafico).

Dall'esplorazione all'inferenza

- Ogni misura può dare un valore diverso.

$$X = \text{`}misura\text{ della velocità della luce'}$$

  è una variabile casuale.
- Ragionevole pensare che X si distribuisca intorno ad una media μ con una certa standard deviation σ,

$$X \sim \left(\mu, \sigma\right)$$

  - μ = velocità della luce (se l'esperimento è ben progettato!)
  - σ = grado di imprecisione dell'esperimento
  - Media e sd di X, non del campione!
  - Potremmo anche assumere che X abbia una distribuzione Normale (vedi grafico).

Dall'esplorazione all'inferenza

- Ogni misura può dare un valore diverso.

$$X = \text{`}misura\text{ della velocità della luce'}$$

  è una variabile casuale.
- Ragionevole pensare che X si distribuisca intorno ad una media μ con una certa standard deviation σ,

$$X \sim \left(\mu, \sigma\right)$$

  - μ = velocità della luce (se l'esperimento è ben progettato!)
  - σ = grado di imprecisione dell'esperimento
  - Media e sd di X, non del campione!
  - Potremmo anche assumere che X abbia una distribuzione Normale (vedi grafico).

Dall'esplorazione all'inferenza

- Ogni misura può dare un valore diverso.

$$X = \text{`}misura\text{ della velocità della luce'}$$

è una variabile casuale.

- Ragionevole pensare che X si distribuisca intorno ad una media μ con una certa standard deviation σ,

$$X \sim \left(\mu, \sigma\right)$$

  - μ = velocità della luce (se l'esperimento è ben progettato!)
  - σ = grado di imprecisione dell'esperimento
  - Media e sd di X, non del campione!
  - Potremmo anche assumere che X abbia una distribuzione Normale (vedi grafico).

Dall'esplorazione all'inferenza

- Ogni misura può dare un valore diverso.

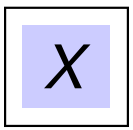
$$X = \text{`}misura\text{ della velocità della luce'}$$

  è una variabile casuale.
- Ragionevole pensare che X si distribuisca intorno ad una media μ con una certa standard deviation σ,

$$X \sim N(\mu, \sigma)$$

  - μ = velocità della luce (se l'esperimento è ben progettato!)
  - σ = grado di imprecisione dell'esperimento
  - Media e sd di X, non del campione!
  - Potremmo anche assumere che X abbia una distribuzione Normale (vedi grafico).

Gli ingredienti dell'inferenza



- **Variabile casuale**: X = '*misura* della velocità della luce'
- **Modello** per X: meccanismo aleatorio che genera le osservazioni. Nel caso in esame (μ, σ)
- **Parametri**: μ parametro d'interesse; σ parametro di disturbo. Si fa inferenza sui parametri.
- **Campione**: $\underline{x}$, fatto da n osservazioni $(x_1, \ldots, x_n)$. È l'informazione per fare inferenza sui parametri.

Gli ingredienti dell'inferenza

$$X \sim (\mu, \sigma)$$

- **Variabile casuale**: X = '*misura* della velocità della luce'
- **Modello** per X: meccanismo aleatorio che genera le osservazioni. Nel caso in esame (μ, σ)
- **Parametri**: μ parametro d'interesse; σ parametro di disturbo. Si fa inferenza sui parametri.
- **Campione**: $\underline{x}$, fatto da n osservazioni $(x_1, \ldots, x_n)$. È l'informazione per fare inferenza sui parametri.

Gli ingredienti dell'inferenza

$$X \sim N(\mu, \sigma)$$

- **Variabile casuale**: X = '*misura* della velocità della luce'
- **Modello** per X: meccanismo aleatorio che genera le osservazioni. Nel caso in esame (μ, σ) oppure $N(\mu, \sigma)$.
- **Parametri**: μ parametro d'interesse; σ parametro di disturbo. Si fa inferenza sui parametri.
- **Campione**: $\underline{x}$, fatto da n osservazioni $(x_1, \ldots, x_n)$. È l'informazione per fare inferenza sui parametri.

Gli ingredienti dell'inferenza

$$X \sim (\mu, \sigma)$$

- **Variabile casuale**: X = '*misura* della velocità della luce'
- **Modello** per X: meccanismo aleatorio che genera le osservazioni. Nel caso in esame (μ, σ) oppure $N(\mu, \sigma)$.
- **Parametri**: μ parametro d'interesse; σ parametro di disturbo. Si fa inferenza sui parametri.
- **Campione**: $\underline{x}$, fatto da n osservazioni $(x_1, \ldots, x_n)$.
  È l'informazione per fare inferenza sui parametri.

Gli ingredienti dell'inferenza

$$X \sim (\mu, \sigma)$$

- **Variabile casuale**: X = '*misura* della velocità della luce'
- **Modello** per X: meccanismo aleatorio che genera le osservazioni. Nel caso in esame (μ, σ) oppure $N(\mu, \sigma)$.
- **Parametri**: μ parametro d'interesse; σ parametro di disturbo. Si fa inferenza sui parametri.
- **Campione**: $\underline{x}$, fatto da n osservazioni $(x_1, \ldots, x_n)$.
  È l'informazione per fare inferenza sui parametri.

Tipi di inferenza

- Stima puntuale: Uso $\underline{x}$ per (cercare di) indovinare il vero valore del parametro
  - Es: Cerco di indovinare μ (velocità della luce).
- Stima per intervallo: Uso $\underline{x}$ per ricavare un intervallo che, con alta probabilità, include il vero valore del parametro
  - Es: Calcolo l'intervallo che, col 95% di probabilità, include μ (velocità della luce).
- Test delle ipotesi: Uso $\underline{x}$ per accettare/rifiutare un'ipotesi su valore del parametro (entro un certo margine di incertezza)
  - Es: Confermo o respingo una data ipotesi sul valore di μ (velocità della luce) fatta precedentemente da un altro studioso.

Tipi di inferenza

- Stima puntuale: Uso $\underline{x}$ per (cercare di) indovinare il vero valore del parametro
  - Es: Cerco di indovinare μ (velocità della luce).
- Stima per intervallo: Uso $\underline{x}$ per ricavare un intervallo che, con alta probabilità, include il vero valore del parametro
  - Es: Calcolo l'intervallo che, col 95% di probabilità, include μ (velocità della luce).
- Test delle ipotesi: Uso $\underline{x}$ per accettare/rifiutare un'ipotesi su valore del parametro (entro un certo margine di incertezza)
  - Es: Confermo o respingo una data ipotesi sul valore di μ (velocità della luce) fatta precedentemente da un altro scienziato.

Tipi di inferenza

- Stima puntuale: Uso $\underline{x}$ per (cercare di) indovinare il vero valore del parametro
  - Es: Cerco di indovinare μ (velocità della luce).
- Stima per intervallo: Uso $\underline{x}$ per ricavare un intervallo che, con alta probabilità, include il vero valore del parametro
  - Es: Calcolo l'intervallo che, col 95% di probabilità, include μ (velocità della luce).
- Test delle ipotesi: Uso $\underline{x}$ per accettare/rifiutare un'ipotesi su valore del parametro (entro un certo margine di incertezza)
  - Es: Confermo o respingo una data ipotesi sul valore di μ (velocità della luce) fatta precedentemente da un altro studioso.

Tipi di inferenza

- Stima puntuale: Uso $\underline{x}$ per (cercare di) indovinare il vero valore del parametro
  - Es: Cerco di indovinare μ (velocità della luce).
- Stima per intervallo: Uso $\underline{x}$ per ricavare un intervallo che, con alta probabilità, include il vero valore del parametro
  - Es: Calcolo l'intervallo che, col 95% di probabilità, include μ (velocità della luce).
- Test delle ipotesi: Uso $\underline{x}$ per accettare/rifiutare un'ipotesi su valore del parametro (entro un certo margine di incertezza)
  - Es: Confermo o respingo una certa ipotesi sul valore di μ (velocità della luce) fatta precedente da un altro studioso.

Tipi di inferenza

- Stima puntuale: Uso $\underline{x}$ per (cercare di) indovinare il vero valore del parametro
  - Es: Cerco di indovinare μ (velocità della luce).
- Stima per intervallo: Uso $\underline{x}$ per ricavare un intervallo che, con alta probabilità, include il vero valore del parametro
  - Es: Calcolo l'intervallo che, col 95% di probabilità, include μ (velocità della luce).
- Test delle ipotesi: Uso $\underline{x}$ per accettare/rifiutare un'ipotesi su valore del parametro (entro un certo margine di incertezza)
  - Es: Confermo o respingo una certa ipotesi sul valore di μ (velocità della luce) fatta precedente da un altro studioso.

Tipi di inferenza

- Stima puntuale: Uso $\underline{x}$ per (cercare di) indovinare il vero valore del parametro
  - Es: Cerco di indovinare μ (velocità della luce).
- Stima per intervallo: Uso $\underline{x}$ per ricavare un intervallo che, con alta probabilità, include il vero valore del parametro
  - Es: Calcolo l'intervallo che, con 95% di probabilità, include μ (velocità della luce).
- Test delle ipotesi: Uso $\underline{x}$ per accettare/rifiutare un'ipotesi su valore del parametro (entro un certo margine di incertezza)
  - Es: Confermo o respingo una certa ipotesi sul valore di μ (velocità della luce) fatta precedente da un altro studioso.

Statistica

- Un parametro è una quantità scalare; il campione $\underline{x}$ è fatto da tante osservazioni $\rightarrow$ sintesi.
- Ogni sintesi di $\underline{x}$ è chiamata statistica.
  Es: min, max, media, mediana, sd, etc.

Statistica

- Un parametro è una quantità scalare; il campione $\underline{x}$ è fatto da tante osservazioni $\rightarrow$ sintesi.
- Ogni sintesi di $\underline{x}$ è chiamata statistica.
  Es: min, max, media, mediana, sd, etc.

Azione

- (...) è ragionevole ritenere che $\boxed{\overline{x} = \text{media}(\underline{x})}$ sia una buona statistica per stimare μ (la media di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\boxed{\widehat{\mu} = \overline{x} = 276234.3 \text{ km/s}}.$$

- Commento (nel 2015): Sottostima dell'8.5%.
- (...) è ragionevole ritenere che $\boxed{s = \text{sd}(\underline{x})}$ sia una buona statistica per stimare σ (la sd di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\boxed{\widehat{\sigma} = s = 55311.5 \text{ km/s}}.$$

- Commento: $\widehat{\sigma}/\widehat{\mu} \simeq 20\%$ dà un'idea del grado di imprecisione (come errore relativo) dell'esperimento.

Azione

- (...) è ragionevole ritenere che $\overline{x} = \text{media}(\underline{x})$ sia una buona statistica per stimare μ (la media di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\mu} = \overline{x} = 276234.3 \text{ km/s}.$$

- Commento (nel 2015): Sottostima dell'8.5%.
- (...) è ragionevole ritenere che $s = \text{sd}(\underline{x})$ sia una buona statistica per stimare σ (la sd di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\sigma} = s = 55311.5 \text{ km/s}.$$

- Commento: $\widehat{\sigma}/\widehat{\mu} \simeq 20\%$ dà un'idea del grado di imprecisione (come errore relativo) dell'esperimento.

Azione

- (...) è ragionevole ritenere che $\overline{x} = \text{media}(\underline{x})$ sia una buona statistica per stimare μ (la media di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\mu} = \overline{x} = 276234.3 \text{ km/s}.$$

- Commento (nel 2015): Sottostima dell'8.5%.
- (...) è ragionevole ritenere che $s = \text{sd}(\underline{x})$ sia una buona statistica per stimare σ (la sd di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\sigma} = s = 55311.5 \text{ km/s}.$$

- Commento: $\widehat{\sigma}/\widehat{\mu} \simeq 20\%$ dà un'idea del grado di imprecisione (come errore relativo) dell'esperimento.

Azione

- (...) è ragionevole ritenere che $\overline{x} = \text{media}(\underline{x})$ sia una buona statistica per stimare μ (la media di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\mu} = \overline{x} = 276234.3 \text{ km/s}.$$

- Commento (nel 2015): Sottostima dell'8.5%.
- (...) è ragionevole ritenere che $s = \text{sd}(\underline{x})$ sia una buona statistica per stimare σ (la sd di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\sigma} = s = 55311.5 \text{ km/s}.$$

- Commento: $\widehat{\sigma}/\widehat{\mu} \simeq 20\%$ dà un'idea del grado di imprecisione (come errore relativo) dell'esperimento.

Azione

- (...) è ragionevole ritenere che $\overline{x} = \text{media}(\underline{x})$ sia una buona statistica per stimare μ (la media di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\mu} = \overline{x} = 276234.3 \text{ km/s}.$$

- Commento (nel 2015): Sottostima dell'8.5%.
- (...) è ragionevole ritenere che $s = \text{sd}(\underline{x})$ sia una buona statistica per stimare σ (la sd di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\sigma} = s = 55311.5 \text{ km/s}.$$

- Commento: $\widehat{\sigma}/\widehat{\mu} \simeq 20\%$ dà un'idea del grado di imprecisione (come errore relativo) dell'esperimento.

Azione

- (...) è ragionevole ritenere che $\overline{x} = \text{media}(\underline{x})$ sia una buona statistica per stimare μ (la media di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\mu} = \overline{x} = 276234.3 \text{ km/s}.$$

- Commento (nel 2015): Sottostima dell'8.5%.

- (...) è ragionevole ritenere che $s = \text{sd}(\underline{x})$ sia una buona statistica per stimare σ (la sd di X).
- Dal campione a disposizione (vedi statistiche descrittive)

$$\widehat{\sigma} = s = 55311.5 \text{ km/s}.$$

- Commento: $\widehat{\sigma}/\widehat{\mu} \simeq 20\%$ dà un'idea del grado di imprecisione (come errore relativo) dell'esperimento.

Commenti

- L'inferenza è affetta da errore.
- Se si cambia il campione si ottengono stime diverse.
- Se si cambia statistica (esempio si usa mediana($\underline{x}$) per stimare μ) si ottengono stime diverse. Meglio la media, la mediana o un'altra statistica?
- Come valutare il grado di incertezza/imprecisione della stima effettuata a prescindere dal vero valore del parametro (che di solito non si conosce)?

Commenti

- L'inferenza è affetta da errore.
- Se si cambia il campione si ottengono stime diverse.
- Se si cambia statistica (esempio si usa mediana($\underline{x}$) per stimare μ) si ottengono stime diverse. Meglio la media, la mediana o un'altra statistica?
- Come valutare il grado di incertezza/imprecisione della stima effettuata a prescindere dal vero valore del parametro (che di solito non si conosce)?

Commenti

- L'inferenza è affetta da errore.
- Se si cambia il campione si ottengono stime diverse.
- Se si cambia statistica (esempio si usa mediana(x) per stimare μ) si ottengono stime diverse. Meglio la media, la mediana o un'altra statistica?
- Come valutare il grado di incertezza/imprecisione della stima effettuata a prescindere dal vero valore del parametro (che di solito non si conosce)?

Commenti

- L'inferenza è affetta da errore.
- Se si cambia il campione si ottengono stime diverse.
- Se si cambia statistica (esempio si usa mediana(x) per stimare μ) si ottengono stime diverse. Meglio la media, la mediana o un'altra statistica?
- Come valutare il grado di incertezza/imprecisione della stima effettuata a prescindere dal vero valore del parametro (che di solito non si conosce)?

Commenti

- L'inferenza è affetta da errore.
- Se si cambia il campione si ottengono stime diverse.
- Se si cambia statistica (esempio si usa mediana($\underline{x}$) per stimare μ) si ottengono stime diverse. Meglio la media, la mediana o un'altra statistica?
- Come valutare il grado di incertezza/imprecisione della stima effettuata a prescindere dal vero valore del parametro (che di solito non si conosce)?

distribuzione campionaria

Distribuzione campionaria di una statistica

- Definizione:

Distribuzione della statistica
considerando **tutti** i possibili campioni

- Capire il concetto (procedimento costruttivo):
  1. Prendo una scatola, metto dentro tanti biglietti numerati (es. 200) secondo la distribuzione di X che preferisco (es. $N(\mu = 5, \sigma = 8)$).
  2. Fisso la dimensione del campione (es. $n = 10$)
  3. Prendo il primo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\bar{x} = 7.81$)
  4. Prendo il secondo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\bar{x} = 9.26$)
  5. Continuo finché non mi stanco (non troppo presto)
  6. Faccio l'istogramma → ecco qua è (una approssimazione della) la distribuzione campionaria.
  7. Potrei fare tutto con un calcolatore (molto più facile)

Distribuzione campionaria di una statistica

- Definizione:

> **Distribuzione della statistica**
> considerando **tutti** i possibili campioni

- Capire il concetto (procedimento costruttivo):
  1. Prendo una scatola, metto dentro tanti bigliettini numerati (es. 200) secondo la distribuzione di X che preferisco (es. $N(\mu = 5, \sigma = 8)$).
  2. Fisso la dimensione del campione (es. $n = 10$)
  3. Prendo il primo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 7.91$).
  4. Prendo il secondo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 3.72$).
  5. Continuo finché non mi stanco (non troppo presto!).
  6. Faccio l'istogramma... *et voilà*: quella è (una buonissima approssimazione del-) la distribuzione campionaria.
  7. Potrei fare tutto con un calcolatore (numeri casuali).

Distribuzione campionaria di una statistica

- Definizione:

Distribuzione della statistica
considerando **tutti** i possibili campioni

- Capire il concetto (procedimento costruttivo):
  1. Prendo una scatola, metto dentro tanti bigliettini numerati (es. 200) secondo la distribuzione di X che preferisco (es. $N(\mu = 5, \sigma = 8)$).
  2. Fisso la dimensione del campione (es. $n = 10$)
  3. Prendo il primo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 7.91$).
  4. Prendo il secondo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 3.72$).
  5. Continuo finché non mi stanco (non troppo presto!).
  6. Faccio l'istogramma... *et voilà*: quella è (una buonissima approssimazione del-) la distribuzione campionaria.
  7. Potrei fare tutto con un calcolatore (numeri casuali).

Distribuzione campionaria di una statistica

- Definizione:

Distribuzione della statistica
considerando **tutti** i possibili campioni

- Capire il concetto (procedimento costruttivo):
  1. Prendo una scatola, metto dentro tanti bigliettini numerati (es. 200) secondo la distribuzione di X che preferisco (es. $N(\mu = 5, \sigma = 8)$).
  2. Fisso la dimensione del campione (es. $n = 10$)
  3. Prendo il primo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 7.91$).
  4. Prendo il secondo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 3.72$).
  5. Continuo finché non mi stanco (non troppo presto!).
  6. Faccio l'istogramma... *et voilà*: quella è (una buonissima approssimazione del-) la distribuzione campionaria.
  7. Potrei fare tutto con un calcolatore (numeri casuali).

Distribuzione campionaria di una statistica

- Definizione:

Distribuzione della statistica
considerando **tutti** i possibili campioni

- Capire il concetto (procedimento costruttivo):
  1. Prendo una scatola, metto dentro tanti bigliettini numerati (es. 200) secondo la distribuzione di X che preferisco (es. $N(\mu = 5, \sigma = 8)$).
  2. Fisso la dimensione del campione (es. $n = 10$)
  3. Prendo il primo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 7.91$).
  4. Prendo il secondo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 3.72$).
  5. Continuo finché non mi stanco (non troppo presto!).
  6. Faccio l'istogramma... *et voilà*: quella è (una buonissima approssimazione del-) la distribuzione campionaria.
  7. Potrei fare tutto con un calcolatore (numeri casuali).

Distribuzione campionaria di una statistica

- Definizione:

> **Distribuzione della statistica**
> considerando **tutti** i possibili campioni

- Capire il concetto (procedimento costruttivo):
  1. Prendo una scatola, metto dentro tanti bigliettini numerati (es. 200) secondo la distribuzione di X che preferisco (es. $N(\mu = 5, \sigma = 8)$).
  2. Fisso la dimensione del campione (es. $n = 10$)
  3. Prendo il primo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 7.91$).
  4. Prendo il secondo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 3.72$).
  5. Continuo finché non mi stanco (non troppo presto!).
  6. Faccio l'istogramma... *et voilà*: quella è (una buonissima approssimazione del-) la distribuzione campionaria.
  7. Potrei fare tutto con un calcolatore (numeri casuali).

Distribuzione campionaria di una statistica

- Definizione:

Distribuzione della statistica
considerando **tutti** i possibili campioni

- Capire il concetto (procedimento costruttivo):
  1. Prendo una scatola, metto dentro tanti bigliettini numerati (es. 200) secondo la distribuzione di X che preferisco (es. $N(\mu = 5, \sigma = 8)$).
  2. Fisso la dimensione del campione (es. $n = 10$)
  3. Prendo il primo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 7.91$).
  4. Prendo il secondo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 3.72$).
  5. Continuo finché non mi stanco (non troppo presto!).
  6. Faccio l'istogramma... *et voilà*: quella è (una buonissima approssimazione del-) la distribuzione campionaria.
  7. Potrei fare tutto con un calcolatore (numeri casuali).

Distribuzione campionaria di una statistica

- Definizione:

Distribuzione della statistica
considerando **tutti** i possibili campioni

- Capire il concetto (procedimento costruttivo):
  1. Prendo una scatola, metto dentro tanti bigliettini numerati (es. 200) secondo la distribuzione di X che preferisco (es. $N(\mu = 5, \sigma = 8)$).
  2. Fisso la dimensione del campione (es. $n = 10$)
  3. Prendo il primo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 7.91$).
  4. Prendo il secondo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 3.72$).
  5. Continuo finché non mi stanco (non troppo presto!).
  6. Faccio l'istogramma... *et voilà*: quella è (una buonissima approssimazione del-) la distribuzione campionaria.
  7. Potrei fare tutto con un calcolatore (numeri casuali).

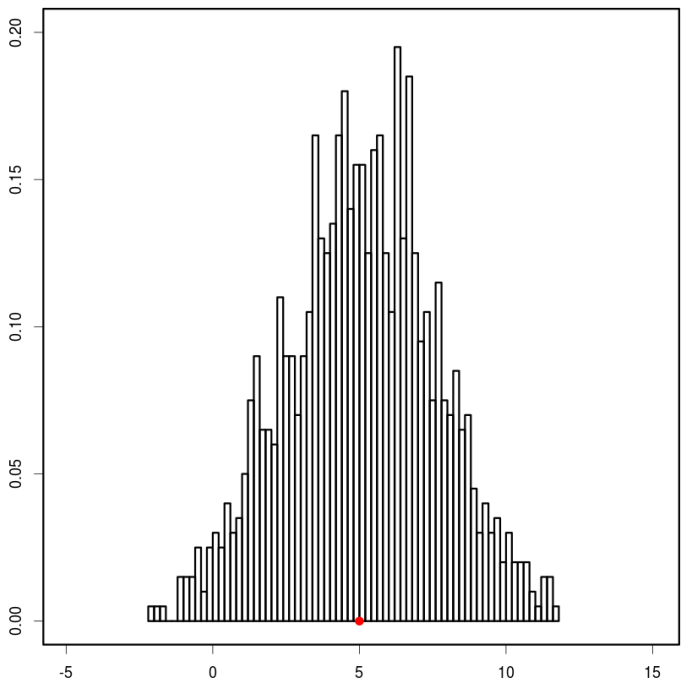
Distribuzione campionaria di una statistica

- Definizione:

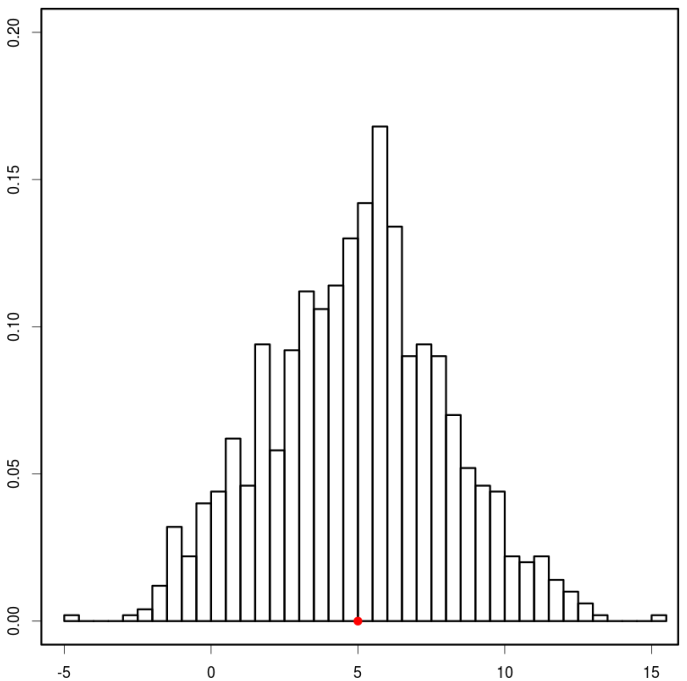
Distribuzione della statistica
considerando **tutti** i possibili campioni

- Capire il concetto (procedimento costruttivo):
  1. Prendo una scatola, metto dentro tanti bigliettini numerati (es. 200) secondo la distribuzione di X che preferisco (es. $N(\mu = 5, \sigma = 8)$).
  2. Fisso la dimensione del campione (es. $n = 10$)
  3. Prendo il primo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 7.91$).
  4. Prendo il secondo campione (con rimbussolamento), ne calcolo la statistica, segno il risultato (es. $\overline{x} = 3.72$).
  5. Continuo finché non mi stanco (non troppo presto!).
  6. Faccio l'istogramma... *et voilà*: quella è (una buonissima approssimazione del-) la distribuzione campionaria.
  7. Potrei fare tutto con un calcolatore (numeri casuali).

Esempi

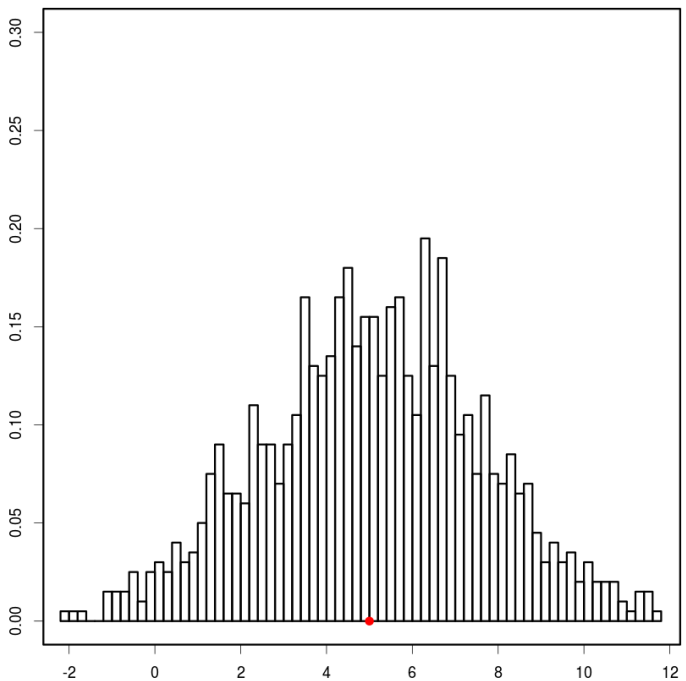


(a) $\overline{X}$ per $X \sim N(5, 8)$

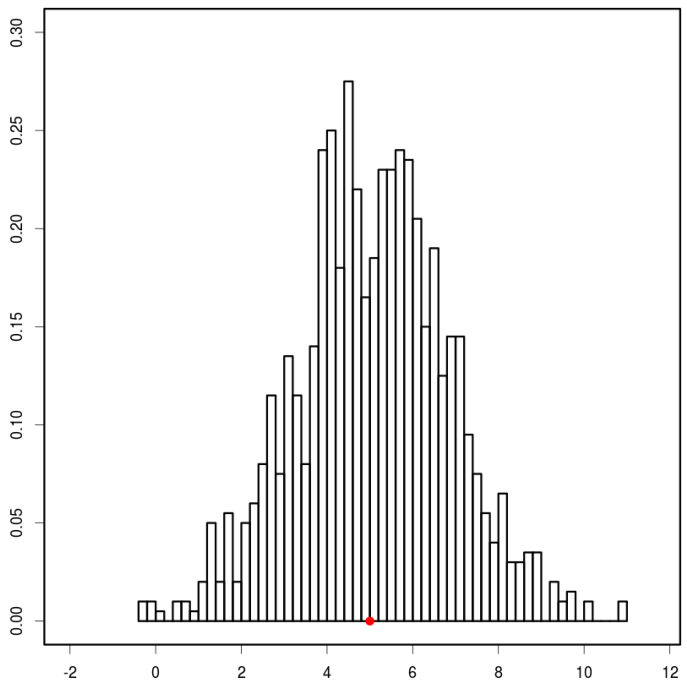


(b) Mediana per $X \sim N(5, 8)$

Esempi

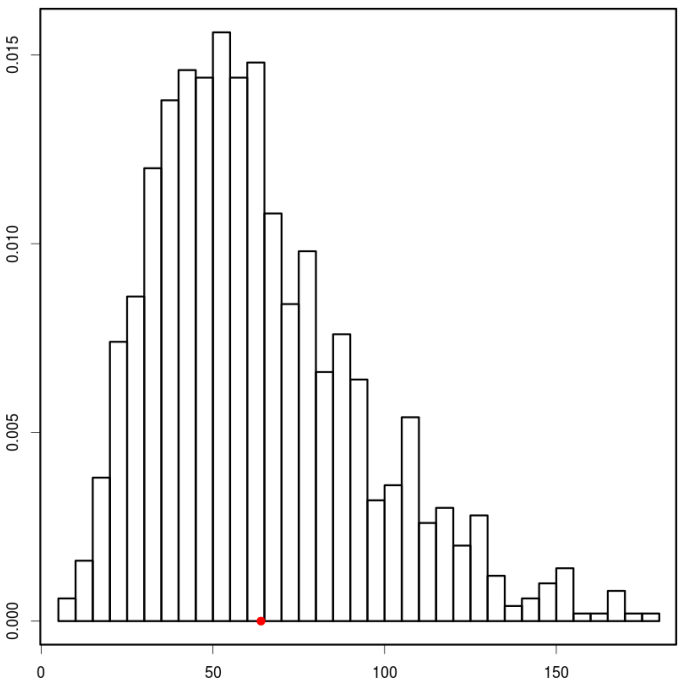


(a) $\overline{X}_{10}$ per $X \sim N(5, 8)$

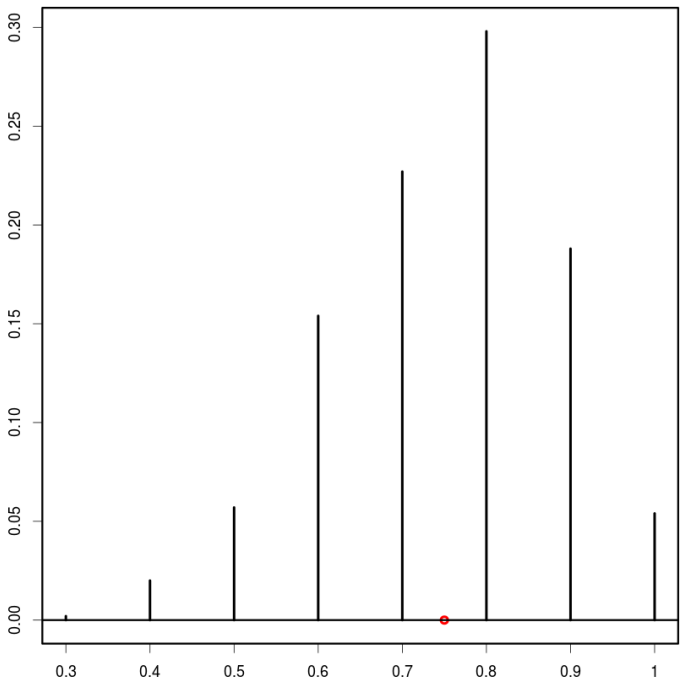


(b) $\overline{X}_{20}$ per $X \sim N(5, 8)$

Esempi



(a) S^2 per $X \sim N(5, 8)$



(b) $\overline{X}$ per $X \sim Be(p = 0.75)$

Perchè è utile

- Indica la forma della distribuzione della statistica.
- Posizione (è centrata sul vero valore del parametro?) e variabilità (meno è, meglio è) della distribuzione della statistica.
- Comparando la distribuzione campionaria di statistiche diverse posso capire qual è migliore.
- Indispensabile per valutare l'incertezza associata alla stima effettuata:

  standard error = stima della sd della statistica
- Indispensabile per stima per intervallo e test delle ipotesi.

Perchè è utile

- Indica la forma della distribuzione della statistica.
- Posizione (è centrata sul vero valore del parametro?) e variabilità (meno è, meglio è) della distribuzione della statistica.
- Comparando la distribuzione campionaria di statistiche diverse posso capire qual è migliore.
- Indispensabile per valutare l'incertezza associata alla stima effettuata:

  standard error = stima della sd della statistica
- Indispensabile per stima per intervallo e test delle ipotesi.

Perchè è utile

- Indica la forma della distribuzione della statistica.
- Posizione (è centrata sul vero valore del parametro?) e variabilità (meno è, meglio è) della distribuzione della statistica.
- Comparando la distribuzione campionaria di statistiche diverse posso capire qual è migliore.
- Indispensabile per valutare l'incertezza associata alla stima effettuata:

  standard error = stima della sd della statistica
- Indispensabile per stima per intervallo e test delle ipotesi.

Perchè è utile

- Indica la forma della distribuzione della statistica.
- Posizione (è centrata sul vero valore del parametro?) e variabilità (meno è, meglio è) della distribuzione della statistica.
- Comparando la distribuzione campionaria di statistiche diverse posso capire qual è migliore.
- Indispensabile per valutare l'incertezza associata alla stima effettuata:

standard error = stima della sd della statistica

- Indispensabile per stima per intervallo e test delle ipotesi.

Perchè è utile

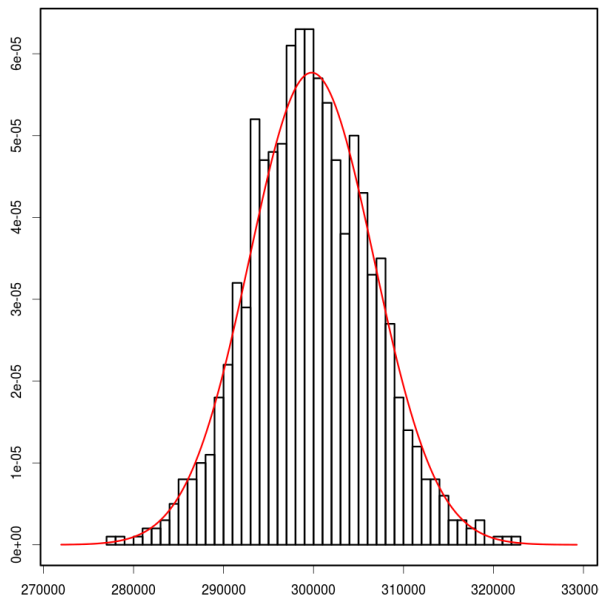
- Indica la forma della distribuzione della statistica.
- Posizione (è centrata sul vero valore del parametro?) e variabilità (meno è, meglio è) della distribuzione della statistica.
- Comparando la distribuzione campionaria di statistiche diverse posso capire qual è migliore.
- Indispensabile per valutare l'incertezza associata alla stima effettuata:

  standard error = stima della sd della statistica

- Indispensabile per stima per intervallo e test delle ipotesi.

Come si ricava

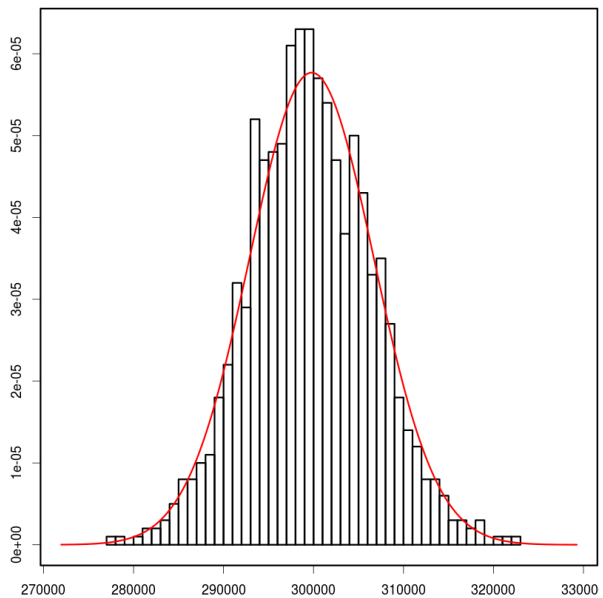
- Analiticamente in modo esatto, valido per n qualsiasi (pochi casi fortunati)
- Analiticamente in modo approssimato, valido per $n \to +\infty$ (quasi sempre; approccio molto generale)
- Numericamente, mediante simulazione al computer
- Esempio: Distribuzione campionaria di $\overline{X}$



$\overline{X}$ per $X \sim N(299792.5, 55000)$

Come si ricava

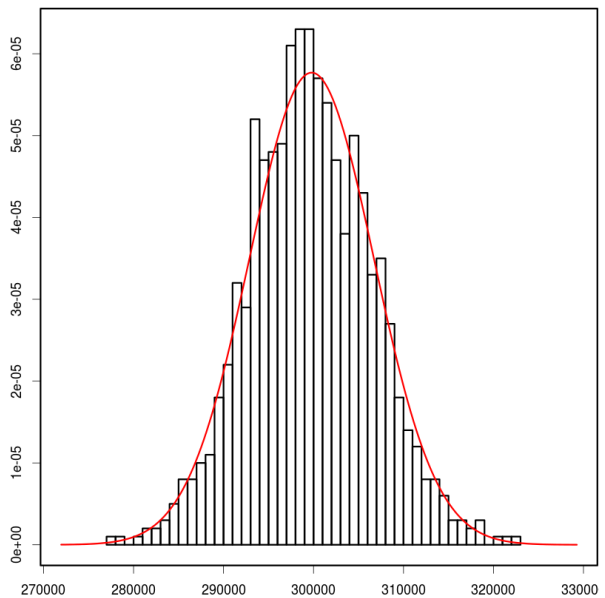
- Analiticamente in modo esatto, valido per n qualsiasi (pochi casi fortunati)
- Analiticamente in modo approssimato, valido per $n \to +\infty$ (quasi sempre; approccio molto generale)
- Numericamente, mediante simulazione al computer
- Esempio: Distribuzione campionaria di $\overline{X}$



$\overline{X}$ per $X \sim N(299792.5, 55000)$

Come si ricava

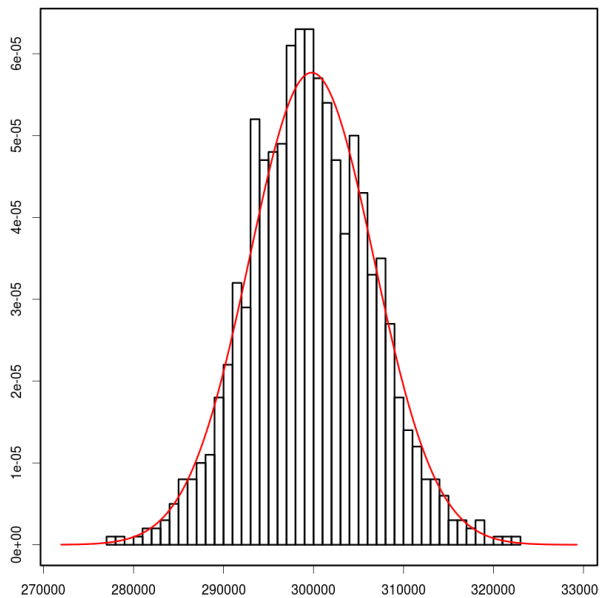
- Analiticamente in modo esatto, valido per n qualsiasi (pochi casi fortunati)
- Analiticamente in modo approssimato, valido per $n \to +\infty$ (quasi sempre; approccio molto generale)
- Numericamente, mediante simulazione al computer
- Esempio: Distribuzione campionaria di $\overline{X}$



$\overline{X}$ per $X \sim N(299792.5, 55000)$

Come si ricava

- Analiticamente in modo esatto, valido per n qualsiasi (pochi casi fortunati)
- Analiticamente in modo approssimato, valido per $n \to +\infty$ (quasi sempre; approccio molto generale)
- Numericamente, mediante simulazione al computer
- Esempio: Distribuzione campionaria di $\overline{X}$



$\overline{X}$ per $X \sim N(299792.5, 55000)$

Distribuzione campionaria di $\overline{X}$

Si dimostra che

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \overset{n \text{ 'grande'}}{\approx} N\left(\mu, se = \frac{\widehat{\sigma}}{\sqrt{n}}\right)$$

- La distribuzione di $\overline{X}$ è:
  - esatta se ;
  - approssimata (valida per n grande) se ;
  - per n 'grande' non cambia se si usa $\widehat{\sigma}$ (linea rossa nel grafico) al posto di σ (linea blu nel grafico).
- La distribuzione di $\overline{X}$ è centrata su μ.
- La sua sd e la sua stima (lo *standard error*) vanno a zero per $n \to \infty$.

Distribuzione campionaria di $\overline{X}$

Si dimostra che

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \overset{n \text{ 'grande'}}{\approx} N\left(\mu, se = \frac{\widehat{\sigma}}{\sqrt{n}}\right)$$

- La distribuzione di $\overline{X}$ è:
  - esatta se ;
  - approssimata (valida per n 'grande') se ;
  - per n 'grande' non cambia se si usa $\widehat{\sigma}$ (linea rossa nel grafico) al posto di σ (linea blu nel grafico).
- La distribuzione di $\overline{X}$ è centrata su μ.
- La sua sd e la sua stima (lo *standard error*) vanno a zero per $n \to \infty$.

Distribuzione campionaria di $\overline{X}$

Si dimostra che

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \overset{n \text{ 'grande'}}{\approx} N\left(\mu, se = \frac{\widehat{\sigma}}{\sqrt{n}}\right)$$

- La distribuzione di $\overline{X}$ è:
  - esatta se $X \sim N(\mu, \sigma)$;
  - approssimata (valida per n 'grande') se ;
  - per n 'grande' non cambia se si usa $\widehat{\sigma}$ (linea rossa nel grafico) al posto di σ (linea blu nel grafico).
- La distribuzione di $\overline{X}$ è centrata su μ.
- La sua sd e la sua stima (lo *standard error*) vanno a zero per $n \to \infty$.

Distribuzione campionaria di $\overline{X}$

Si dimostra che

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \overset{n \text{ 'grande'}}{\approx} N\left(\mu, se = \frac{\widehat{\sigma}}{\sqrt{n}}\right)$$

- La distribuzione di $\overline{X}$ è:
  - esatta se $X \sim N(\mu, \sigma)$;
  - approssimata (valida per n 'grande') se $X \sim (\mu, \sigma)$;
  - per n 'grande' non cambia se si usa $\widehat{\sigma}$ (linea rossa nel grafico) al posto di σ (linea blu nel grafico).
- La distribuzione di $\overline{X}$ è centrata su μ.
- La sua sd e la sua stima (lo *standard error*) vanno a zero per $n \to \infty$.

Distribuzione campionaria di $\overline{X}$

Si dimostra che

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \overset{n \text{ 'grande'}}{\approx} N\left(\mu, se = \frac{\widehat{\sigma}}{\sqrt{n}}\right)$$

- La distribuzione di $\overline{X}$ è:
  - esatta se $X \sim N(\mu, \sigma)$;
  - approssimata (valida per n 'grande') se $X \sim (\mu, \sigma)$;
  - per n 'grande' non cambia se si usa $\widehat{\sigma}$ (linea rossa nel grafico) al posto di σ (linea blu nel grafico).
- La distribuzione di $\overline{X}$ è centrata su μ.
- La sua sd e la sua stima (lo *standard error*) vanno a zero per $n \to \infty$.

Distribuzione campionaria di $\overline{X}$

Si dimostra che

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \overset{n \text{ 'grande'}}{\approx} N\left(\mu, se = \frac{\widehat{\sigma}}{\sqrt{n}}\right)$$

- La distribuzione di $\overline{X}$ è:
  - esatta se $X \sim N(\mu, \sigma)$;
  - approssimata (valida per n 'grande') se $X \sim (\mu, \sigma)$;
  - per n 'grande' non cambia se si usa $\widehat{\sigma}$ (linea rossa nel grafico) al posto di σ (linea blu nel grafico).
- La distribuzione di $\overline{X}$ è centrata su μ.
- La sua sd e la sua stima (lo *standard error*) vanno a zero per $n \to \infty$.

Distribuzione campionaria di $\overline{X}$

Si dimostra che

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}},\right) \overset{n \text{ 'grande'}}{\approx} N\left(\mu, se = \frac{\widehat{\sigma}}{\sqrt{n}}\right)$$

- La distribuzione di $\overline{X}$ è:
  - esatta se $X \sim N(\mu, \sigma)$;
  - approssimata (valida per n 'grande') se $X \sim (\mu, \sigma)$;
  - per n 'grande' non cambia se si usa $\widehat{\sigma}$ (linea rossa nel grafico) al posto di σ (linea blu nel grafico).
- La distribuzione di $\overline{X}$ è centrata su μ.
- La sua sd e la sua stima (lo *standard error*) vanno a zero per $n \to \infty$.

Stima per intervallo

- Poiché $\overline{X} \approx N\left(\mu, se = \frac{\widehat{\sigma}}{\sqrt{n}}\right)$, in base alle caratteristiche della normale, si ha

$$95\% = P\left(-1.96 \leq \frac{\overline{X} - \mu}{se} \leq 1.96\right)$$
$$= P\left(\overline{X} - 1.96 \cdot se \leq \mu \leq \overline{X} + 1.96 \cdot se\right)$$

- La corrispondente stima per intervallo (al 95%) è

$$[\overline{x} - 1.96 \cdot se, \overline{x} + 1.96 \cdot se] = [262683.3, 289785.4]$$

- Commento (nel 2015): Non include il vero valore di μ (299792.5 km/s). Verosimilmente, l'esperimento non era ben progettato.

Stima per intervallo

- Poiché $\overline{X} \approx N\left(\mu, se = \frac{\widehat{\sigma}}{\sqrt{n}}\right)$, in base alle caratteristiche della normale, si ha

$$95\% = P\left(-1.96 \leq \frac{\overline{X} - \mu}{se} \leq 1.96\right)$$
$$= P\left(\overline{X} - 1.96 \cdot se \leq \mu \leq \overline{X} + 1.96 \cdot se\right)$$

- La corrispondente stima per intervallo (al 95%) è

$$\boxed{[\overline{x} - 1.96 \cdot se, \overline{x} + 1.96 \cdot se] = [262683.3, 289785.4]}$$

- Commento (nel 2015): Non include il vero valore di μ (299792.5 km/s). Verosimilmente, l'esperimento non era ben progettato.

Situazione pratica

- **Variabile**: X = 'cliente finanziato è andato in *default*?'
- **Dati**: $n = 5756$ casi di clienti precedentemente finanziati,

$$\underline{x} = (x_1 = \text{No}, x_2 = \text{No}, x_3 = \text{Sì}, x_4 = \text{No}, \ldots, x_{5756} = \text{No})$$

$\underline{x}$ include 278 Sì e 5478 No.
- **Obiettivo**: stimare la probabilità che un cliente vada in *default*.

Situazione pratica

- **Variabile**: X = 'cliente finanziato è andato in *default*?'
- **Dati**: $n = 5756$ casi di clienti precedentemente finanziati,

$$\underline{x} = (x_1 = \text{No}, x_2 = \text{No}, x_3 = \text{Sì}, x_4 = \text{No}, \ldots, x_{5756} = \text{No})$$

  $\underline{x}$ include 278 Sì e 5478 No.
- **Obiettivo**: stimare la probabilità che un cliente vada in *default*.

Situazione pratica

- **Variabile**: X = 'cliente finanziato è andato in *default*?'
- **Dati**: $n = 5756$ casi di clienti precedentemente finanziati,

$$\underline{x} = (x_1 = \text{No}, x_2 = \text{No}, x_3 = \text{Sì}, x_4 = \text{No}, \ldots, x_{5756} = \text{No})$$

  $\underline{x}$ include 278 Sì e 5478 No.
- **Obiettivo**: stimare la probabilità che un cliente vada in *default*.

Gli ingredienti

X

- **Variabile casuale**: X = 'cliente finanziato è andato in *default*?' secondo la codifica 1 = Sì, 0 =No.
- **Modello** per X: $Be(p)$, che significa

$$P(X = 1) = p \qquad P(X = 0) = 1 - p$$

- **Parametri**: $p = P(X = 1) = P(default)$.
- **Campione**: di $n = 5756$ osservazioni

$$\underline{x} = (x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 0, \ldots, x_{5756} = 0)$$

In $\underline{x}$ ci sono 278 *default* (1) e 5478 *bonis* (0).

Gli ingredienti

$$X \sim Be(p)$$

- **Variabile casuale**: X = 'cliente finanziato è andato in *default*?' secondo la codifica 1 = Sì, 0 =No.
- **Modello** per X: $Be(p)$, che significa

$$P(X = 1) = p \qquad P(X = 0) = 1 - p$$

- **Parametri**: $p = P(X = 1) = P(default)$.
- **Campione**: di $n = 5756$ osservazioni

$$\underline{x} = (x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 0, \ldots, x_{5756} = 0)$$

In $\underline{x}$ ci sono 278 *default* (1) e 5478 *bonis* (0).

Gli ingredienti

$$X \sim Be(p)$$

- **Variabile casuale**: X = 'cliente finanziato è andato in *default*?' secondo la codifica 1 = Sì, 0 =No.
- **Modello** per X: $Be(p)$, che significa

$$P(X = 1) = p \qquad P(X = 0) = 1 - p$$

- **Parametri**: $p = P(X = 1) = P(default)$.
- **Campione**: di $n = 5756$ osservazioni

$$\underline{x} = (x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 0, \ldots, x_{5756} = 0)$$

In $\underline{x}$ ci sono 278 *default* (1) e 5478 *bonis* (0).

Gli ingredienti

$$\boxed{X \sim Be(p)}$$

- **Variabile casuale**: X = 'cliente finanziato è andato in *default*?' secondo la codifica 1 = Sì, 0 =No.
- **Modello** per X: $Be(p)$, che significa

$$P(X = 1) = p \qquad P(X = 0) = 1 - p$$

- **Parametri**: $p = P(X = 1) = P(default)$.
- **Campione**: di $n = 5756$ osservazioni

$$\underline{x} = (x_1 = 0, x_2 = 0, x_3 = 1, x_4 = 0, \ldots, x_{5756} = 0)$$

In $\underline{x}$ ci sono 278 *default* (1) e 5478 *bonis* (0).

Statistica e distribuzione campionaria

- (. . .) è ragionevole ritenere che la 'proporzione di 1 nel campione', che peraltro coincide con $\boxed{\overline{x} = \text{media}(\underline{x})}$, sia una buona statistica per stimare p.
- In effetti, si dimostra che

$$\overline{X} \approx N\left(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx N\left(p, se = \frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}}\right)$$

ovvero:

- la distribuzione di $\overline{X}$ è centrata su p;
- la sua se (che può servire per il calcolo della [illegible]) va a zero per $n \to \infty$.

Statistica e distribuzione campionaria

- (. . .) è ragionevole ritenere che la 'proporzione di 1 nel campione', che peraltro coincide con $\boxed{\overline{x} = \text{media}(\underline{x})}$, sia una buona statistica per stimare p.
- In effetti, si dimostra che

$$\overline{X} \approx N\left(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx N\left(p, se = \frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}}\right)$$

ovvero:

- la distribuzione di $\overline{X}$ è centrata su p
- la sua sd (che poi serve per il calcolo dello *standard error*) va a zero per $n \to \infty$.

Statistica e distribuzione campionaria

- (. . .) è ragionevole ritenere che la 'proporzione di 1 nel campione', che peraltro coincide con $\boxed{\overline{x} = \text{media}(\underline{x})}$, sia una buona statistica per stimare p.
- In effetti, si dimostra che

$$\overline{X} \approx N\left(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx N\left(p, se = \frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}}\right)$$

ovvero:

  - la distribuzione di $\overline{X}$ è centrata su p
  - la sua sd (che poi serve per il calcolo dello *standard error*) va a zero per $n \to \infty$.

Statistica e distribuzione campionaria

- (. . .) è ragionevole ritenere che la 'proporzione di 1 nel campione', che peraltro coincide con $\overline{x} = \text{media}(\underline{x})$, sia una buona statistica per stimare p.
- In effetti, si dimostra che

$$\overline{X} \approx N\left(p, \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \approx N\left(p, se = \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right)$$

ovvero:

  - la distribuzione di $\overline{X}$ è centrata su p
  - la sua sd (che poi serve per il calcolo dello *standard error*) va a zero per $n \to \infty$.

Stima puntuale e per intervallo

- Dal campione a disposizione (due slides fa)

$$\boxed{\widehat{p} = \overline{x} = 278/5756 = 0.0483 \simeq 4.83\%}.$$

con uno *standard error*

$$\boxed{se = \sqrt{\widehat{p}(1-\widehat{p})}/\sqrt{n} = 0.00283}.$$

- In base alle caratteristiche della normale, si ha

$$95\% = P\left(-1.96 \leq \frac{\overline{X} - p}{se} \leq 1.96\right)$$

$$= P\left(\overline{X} - 1.96 \cdot se \leq p \leq \overline{X} + 1.96 \cdot se\right)$$

- La corrispondente stima per intervallo (al 95%) è

$$[\overline{x} - 1.96 \cdot se, \overline{x} + 1.96 \cdot se] = [0.0428, 0.0538]$$

Stima puntuale e per intervallo

- Dal campione a disposizione (due slides fa)

$$\boxed{\widehat{p} = \overline{x} = 278/5756 = 0.0483 \simeq 4.83\%}.$$

con uno *standard error*

$$\boxed{se = \sqrt{\widehat{p}(1-\widehat{p})}/\sqrt{n} = 0.00283}.$$

- In base alle caratteristiche della normale, si ha

$$\begin{aligned} 95\% &= P\left(-1.96 \leq \frac{\overline{X} - p}{se} \leq 1.96\right) \\ &= P\left(\overline{X} - 1.96 \cdot se \leq p \leq \overline{X} + 1.96 \cdot se\right) \end{aligned}$$

- La corrispondente stima per intervallo (al 95%) è

$$[\overline{x} - 1.96 \cdot se, \overline{x} + 1.96 \cdot se] = [0.0428, 0.0538]$$

Stima puntuale e per intervallo

- Dal campione a disposizione (due slides fa)

$$\boxed{\widehat{p} = \overline{x} = 278/5756 = 0.0483 \simeq 4.83\%}.$$

con uno *standard error*

$$\boxed{se = \sqrt{\widehat{p}(1-\widehat{p})}/\sqrt{n} = 0.00283}.$$

- In base alle caratteristiche della normale, si ha

$$\begin{aligned} 95\% &= P\left(-1.96 \leq \frac{\overline{X} - p}{se} \leq 1.96\right) \\ &= P\left(\overline{X} - 1.96 \cdot se \leq p \leq \overline{X} + 1.96 \cdot se\right) \end{aligned}$$

- La corrispondente stima per intervallo (al 95%) è

$$\boxed{[\overline{x} - 1.96 \cdot se, \overline{x} + 1.96 \cdot se] = [0.0428, 0.0538]}$$


Grazie per l'attenzione!