# La distribuzione $t$

Federico Plazzi

3 Novembre 2015

# Popolazione e campioni

# Popolazione e campioni

## Definizioni ed assunzioni di partenza

- **Campione**: l'insieme di individui che abbiamo potuto osservare.

# Popolazione e campioni

## Definizioni ed assunzioni di partenza

- **Campione**: l'insieme di individui che abbiamo potuto osservare.
- **Popolazione**: l'insieme, più vasto, di individui, da cui il campione è tratto, che è generalmente l'obiettivo del nostro lavoro.

# Popolazione e campioni

## Definizioni ed assunzioni di partenza

- **Campione**: l'insieme di individui che abbiamo potuto osservare.
- **Popolazione**: l'insieme, più vasto, di individui, da cui il campione è tratto, che è generalmente l'obiettivo del nostro lavoro.
- **Supponiamo innanzitutto che i valori della nostra variabile considerati su tutta la popolazione siano distribuiti in modo normale.**

# Popolazione e campioni

## Definizioni ed assunzioni di partenza

- **Campione**: l'insieme di individui che abbiamo potuto osservare.
- **Popolazione**: l'insieme, più vasto, di individui, da cui il campione è tratto, che è generalmente l'obiettivo del nostro lavoro.
- **Supponiamo innanzitutto che i valori della nostra variabile considerati su tutta la popolazione siano distribuiti in modo normale.**
- **Diciamo anche che il nostro campioni è estratto dalla popolazione in modo casuale.**

# Popolazione e campioni

## Definizioni ed assunzioni di partenza

- **Campione**: l'insieme di individui che abbiamo potuto osservare.
- **Popolazione**: l'insieme, più vasto, di individui, da cui il campione è tratto, che è generalmente l'obiettivo del nostro lavoro.
- **Supponiamo innanzitutto che i valori della nostra variabile considerati su tutta la popolazione siano distribuiti in modo normale.**
- **Diciamo anche che il nostro campioni è estratto dalla popolazione in modo casuale.**
- Chiamiamo $\mu_{pop}$, $\sigma^2_{pop}$ e $\sigma_{pop}$ rispettivamente media, varianza e deviazione standard *di tutta la popolazione*

# Popolazione e campioni

## Definizioni ed assunzioni di partenza

- **Campione**: l'insieme di individui che abbiamo potuto osservare.
- **Popolazione**: l'insieme, più vasto, di individui, da cui il campione è tratto, che è generalmente l'obiettivo del nostro lavoro.
- **Supponiamo innanzitutto che i valori della nostra variabile considerati su tutta la popolazione siano distribuiti in modo normale.**
- **Diciamo anche che il nostro campioni è estratto dalla popolazione in modo casuale.**
- Chiamiamo $\mu_{pop}$, $\sigma^2_{pop}$ e $\sigma_{pop}$ rispettivamente media, varianza e deviazione standard *di tutta la popolazione*
- Chiamiamo infine $N$ la *taglia del campione*.

# La grande domanda della statistica inferenziale

**Cosa posso dire della popolazione a partire dal campione che osservo?**

# Cosa posso dire della popolazione a partire dal campione che osservo?

Per rispondere, dobbiamo per prima cosa capire quali sono i legami tra campioni e popolazione...

# Legami tra campioni e popolazione

## Come procedere?

Estraiamo molti (diciamo 5000, che è il limite di R per eseguire un test di Shapiro-Wilk) campioni di taglia $N$.

# Legami tra campioni e popolazione

### Come procedere?

Estraiamo molti (diciamo 5000, che è il limite di R per eseguire un test di Shapiro-Wilk) campioni di taglia $N$.

### Risultato fondamentale

Le medie campionarie ($\mu_{\overline{X}}$) di un elevato numero di campioni estratti dalla nostra popolazione si distribuiscono in modo **normale**!

# Distribuzione delle medie campionarie

# Distribuzione delle medie campionarie

## Media delle medie campionarie

$$\mu_{\overline{X}} = \mu_{pop} \tag{1}$$

# Distribuzione delle medie campionarie

## Media delle medie campionarie

$$\mu_{\overline{X}} = \mu_{pop} \tag{1}$$

## Varianza delle medie campionarie

$$\sigma^2_{\overline{X}} = \frac{\sigma^2_{pop}}{N} \tag{2}$$

# Distribuzione delle medie campionarie

Media delle medie campionarie

$$\mu_{\overline{X}} = \mu_{pop} \tag{1}$$

Varianza delle medie campionarie

$$\sigma^2_{\overline{X}} = \frac{\sigma^2_{pop}}{N} \tag{2}$$

Deviazione standard delle medie campionarie (*errore standard*)

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2_{pop}}{N}} = \frac{\sigma_{pop}}{\sqrt{N}} \tag{3}$$

# Torniamo alla deviata normale

Se le medie campionarie si distribuiscono in modo normale, possiamo usare la deviata normale ($Z$), che già conosciamo, per verificare se la media del nostro campione somiglia a quella di tutti gli altri concepibili campioni o se si tratta di un valore estremo.

# Torniamo alla deviata normale

Se le medie campionarie si distribuiscono in modo normale, possiamo usare la deviata normale ($Z$), che già conosciamo, per verificare se la media del nostro campione somiglia a quella di tutti gli altri concepibili campioni o se si tratta di un valore estremo.

In altre parole, possiamo usare $Z$ per rispondere alla domanda:
*il nostro campione appartiene o no ad una certa popolazione?*

# Torniamo alla deviata normale

Se le medie campionarie si distribuiscono in modo normale, possiamo usare la deviata normale ($Z$), che già conosciamo, per verificare se la media del nostro campione somiglia a quella di tutti gli altri concepibili campioni o se si tratta di un valore estremo.

In altre parole, possiamo usare $Z$ per rispondere alla domanda:
*il nostro campione appartiene o no ad una certa popolazione?*

$$Z = \frac{X - \mu}{\sigma} \tag{4}$$

# Costruzione e stima di $Z$

- In questo caso, la nostra $X$ sarà la nostra media campionaria, $\overline{X}$.

# Costruzione e stima di $Z$

- In questo caso, la nostra $X$ sarà la nostra media campionaria, $\overline{X}$.
- La media della distribuzione è la media della distribuzione delle medie campionarie, ossia, per la (1), la media di popolazione, $\mu_{pop}$.

# Costruzione e stima di $Z$

- In questo caso, la nostra $X$ sarà la nostra media campionaria, $\overline{X}$.
- La media della distribuzione è la media della distribuzione delle medie campionarie, ossia, per la (1), la media di popolazione, $\mu_{pop}$.
- Cosa possiamo dire dell'errore standard della popolazione a partire dal campione?

# Stima della varianza di popolazione

- Con l'esame di un elevato numero di campioni, si vede che

$$\sigma^2 = \sigma_{pop}^2 \cdot \frac{N-1}{N} \tag{5}$$

# Stima della varianza di popolazione

- Con l'esame di un elevato numero di campioni, si vede che

$$\sigma^2 = \sigma^2_{pop} \cdot \frac{N-1}{N} \tag{5}$$

- Perciò

$$\hat{\sigma}^2_{pop} = \sigma^2 \cdot \frac{N}{N-1} \tag{6}$$

# Stima della varianza di popolazione

- Con l'esame di un elevato numero di campioni, si vede che

$$\sigma^2 = \sigma^2_{pop} \cdot \frac{N-1}{N} \tag{5}$$

- Perciò

$$\hat{\sigma}^2_{pop} = \sigma^2 \cdot \frac{N}{N-1} \tag{6}$$

- Sostituendo con la definizione di varianza,

$$\hat{\sigma}^2_{pop} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N} \cdot \frac{N}{N-1} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N-1} \tag{7}$$

# Il $t$ di Student

- Abbiamo così terminato la costruzione del nostro $Z$, che però, visti i diversi "rimaneggiamenti" che si sono resi necessari, non chiamiamo più $Z$, ma $t$:

$$t = \frac{\overline{X} - \mu_{pop}}{\frac{\hat{\sigma}_{pop}}{\sqrt{N}}} = \frac{\overline{X} - \mu_{pop}}{\frac{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}}{\sqrt{N-1}\sqrt{N}}} \tag{8}$$

# Il $t$ di Student

- Abbiamo così terminato la costruzione del nostro $Z$, che però, visti i diversi "rimaneggiamenti" che si sono resi necessari, non chiamiamo più $Z$, ma $t$:

$$t = \frac{\overline{X} - \mu_{pop}}{\frac{\hat{\sigma}_{pop}}{\sqrt{N}}} = \frac{\overline{X} - \mu_{pop}}{\frac{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}}{\sqrt{N-1}\sqrt{N}}} \tag{8}$$

- **Sembra incredibile, ma la distribuzione di $t$ si può studiare, esattamente come la distribuzione normale di $Z$!**

# Il $t$ di Student

- Abbiamo così terminato la costruzione del nostro $Z$, che però, visti i diversi "rimaneggiamenti" che si sono resi necessari, non chiamiamo più $Z$, ma $t$:

$$t = \frac{\overline{X} - \mu_{pop}}{\frac{\hat{\sigma}_{pop}}{\sqrt{N}}} = \frac{\overline{X} - \mu_{pop}}{\frac{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}}{\sqrt{N-1}\sqrt{N}}} \tag{8}$$

- **Sembra incredibile, ma la distribuzione di $t$ si può studiare, esattamente come la distribuzione normale di $Z$!**
- La distribuzione di $t$ dipende anche dai *gradi di libertà*, pari ad $N - 1$.

# Il $t$ di Student

- Abbiamo così terminato la costruzione del nostro $Z$, che però, visti i diversi "rimaneggiamenti" che si sono resi necessari, non chiamiamo più $Z$, ma $t$:

$$t = \frac{\overline{X} - \mu_{pop}}{\frac{\hat{\sigma}_{pop}}{\sqrt{N}}} = \frac{\overline{X} - \mu_{pop}}{\frac{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}}{\sqrt{N-1}\sqrt{N}}} \tag{8}$$

- **Sembra incredibile, ma la distribuzione di $t$ si può studiare, esattamente come la distribuzione normale di $Z$!**
- La distribuzione di $t$ dipende anche dai *gradi di libertà*, pari ad $N - 1$.
- $H_0$ : il campione *appartiene* alla popolazione $\Rightarrow t = 0$

# Il $t$ di Student

- Abbiamo così terminato la costruzione del nostro $Z$, che però, visti i diversi "rimaneggiamenti" che si sono resi necessari, non chiamiamo più $Z$, ma $t$:

$$t = \frac{\overline{X} - \mu_{pop}}{\frac{\hat{\sigma}_{pop}}{\sqrt{N}}} = \frac{\overline{X} - \mu_{pop}}{\frac{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}}{\sqrt{N-1}\sqrt{N}}} \tag{8}$$

- **Sembra incredibile, ma la distribuzione di $t$ si può studiare, esattamente come la distribuzione normale di $Z$!**
- La distribuzione di $t$ dipende anche dai *gradi di libertà*, pari ad $N - 1$.
- $H_0$ : il campione *appartiene* alla popolazione $\Rightarrow t = 0$
- $H_1$ : il campione *non appartiene* alla popolazione $\Rightarrow t \neq 0$

**TABLE A.2**

***t* Distribution: Critical Values of *t***

| | | *Significance level* | | | | | |
|---|---|---|---|---|---|---|---|
| *Degrees of freedom* | *Two-tailed test:* | 10% | 5% | 2% | 1% | 0.2% | 0.1% |
| | *One-tailed test:* | 5% | 2.5% | 1% | 0.5% | 0.1% | 0.05% |
| **1** | | 6.314 | 12.706 | 31.821 | 63.657 | 318.309 | 636.619 |
| **2** | | 2.920 | 4.303 | 6.965 | 9.925 | 22.327 | 31.599 |
| **3** | | 2.353 | 3.182 | 4.541 | 5.841 | 10.215 | 12.924 |
| **4** | | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| **5** | | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| **6** | | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| **7** | | 1.894 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| **8** | | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| **9** | | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| **10** | | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| **11** | | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| **12** | | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| **13** | | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| **14** | | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |
| **15** | | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| **16** | | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015 |
| **17** | | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |
| **18** | | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |
| **19** | | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| **20** | | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850 |
| **21** | | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 | 3.819 |
| **22** | | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792 |
| **23** | | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 | 3.768 |
| **24** | | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745 |
| **25** | | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 | 3.725 |
| **26** | | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707 |
| **27** | | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.690 |
| **28** | | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674 |
| **29** | | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 | 3.659 |
| **30** | | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 | 3.646 |
| **32** | | 1.694 | 2.037 | 2.449 | 2.738 | 3.365 | 3.622 |
| **34** | | 1.691 | 2.032 | 2.441 | 2.728 | 3.348 | 3.601 |
| **36** | | 1.688 | 2.028 | 2.434 | 2.719 | 3.333 | 3.582 |
| **38** | | 1.686 | 2.024 | 2.429 | 2.712 | 3.319 | 3.566 |
| **40** | | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 | 3.551 |
| **42** | | 1.682 | 2.018 | 2.418 | 2.698 | 3.296 | 3.538 |
| **44** | | 1.680 | 2.015 | 2.414 | 2.692 | 3.286 | 3.526 |
| **46** | | 1.679 | 2.013 | 2.410 | 2.687 | 3.277 | 3.515 |
| **48** | | 1.677 | 2.011 | 2.407 | 2.682 | 3.269 | 3.505 |
| **50** | | 1.676 | 2.009 | 2.403 | 2.678 | 3.261 | 3.496 |
| **60** | | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 | 3.460 |
| **70** | | 1.667 | 1.994 | 2.381 | 2.648 | 3.211 | 3.435 |
| **80** | | 1.664 | 1.990 | 2.374 | 2.639 | 3.195 | 3.416 |
| **90** | | 1.662 | 1.987 | 2.368 | 2.632 | 3.183 | 3.402 |
| **100** | | 1.660 | 1.984 | 2.364 | 2.626 | 3.174 | 3.390 |
| **120** | | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 | 3.373 |
| **150** | | 1.655 | 1.976 | 2.351 | 2.609 | 3.145 | 3.357 |
| **200** | | 1.653 | 1.972 | 2.345 | 2.601 | 3.131 | 3.340 |
| **300** | | 1.650 | 1.968 | 2.339 | 2.592 | 3.118 | 3.323 |
| **400** | | 1.649 | 1.966 | 2.336 | 2.588 | 3.111 | 3.315 |
| **500** | | 1.648 | 1.965 | 2.334 | 2.586 | 3.107 | 3.310 |
| **600** | | 1.647 | 1.964 | 2.333 | 2.584 | 3.104 | 3.307 |
| $\infty$ | | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291 |

.

# Coppie di campioni

- Supponiamo ora di non estrarre un campione alla volta, bensì *coppie* di campioni e di confrontarli a due a due.

# Coppie di campioni

- Supponiamo ora di non estrarre un campione alla volta, bensì *coppie* di campioni e di confrontarli a due a due.
- Chiamiamo $N_a$ la dimensione del primo campione ed $N_b$ la dimensione del secondo.

# Coppie di campioni

- Supponiamo ora di non estrarre un campione alla volta, bensì *coppie* di campioni e di confrontarli a due a due.
- Chiamiamo $N_a$ la dimensione del primo campione ed $N_b$ la dimensione del secondo.
- Chiamiamo $\Delta$ la differenza tra le medie di due campioni. Possiamo studiare la distribuzione di $\Delta$ ed avremo quindi $\mu_\Delta$, $\sigma^2_\Delta$ e $\sigma_\Delta$.

# Legami tra differenze di coppie di campioni e popolazione

# Legami tra differenze di coppie di campioni e popolazione

## Media delle differenze

$$\mu_\Delta = 0 \tag{9}$$

# Legami tra differenze di coppie di campioni e popolazione

Media delle differenze

$$\mu_\Delta = 0 \tag{9}$$

Varianza delle differenze

$$\sigma^2_\Delta = \frac{\sigma^2_{pop}}{N_a} + \frac{\sigma^2_{pop}}{N_b} \tag{10}$$

# Legami tra differenze di coppie di campioni e popolazione

Media delle differenze

$$\mu_\Delta = 0 \tag{9}$$

Varianza delle differenze

$$\sigma_\Delta^2 = \frac{\sigma_{pop}^2}{N_a} + \frac{\sigma_{pop}^2}{N_b} \tag{10}$$

Deviazione standard delle differenze

$$\sigma_\Delta = \sqrt{\frac{\sigma_{pop}^2}{N_a} + \frac{\sigma_{pop}^2}{N_b}} \tag{11}$$

# $t$ di Student per due campioni

Il nostro $t$, quindi, sarà in questo caso

$$t = \frac{\Delta - \mu_\Delta}{\sigma_\Delta} = \frac{\Delta - 0}{\sigma_\Delta} = \frac{\Delta}{\sqrt{\frac{\sigma^2_{pop}}{N_a} + \frac{\sigma^2_{pop}}{N_b}}} \tag{12}$$

- Di nuovo, possiamo stimare la varianza di popolazione a partire da quelle campionarie ed usare le nostre stime per il calcolo di $t$.

# $t$ di Student per due campioni

Il nostro $t$, quindi, sarà in questo caso

$$t = \frac{\Delta - \mu_\Delta}{\sigma_\Delta} = \frac{\Delta - 0}{\sigma_\Delta} = \frac{\Delta}{\sqrt{\frac{\sigma^2_{pop}}{N_a} + \frac{\sigma^2_{pop}}{N_b}}} \tag{12}$$

- Di nuovo, possiamo stimare la varianza di popolazione a partire da quelle campionarie ed usare le nostre stime per il calcolo di $t$.
- **Ancora una volta, la distribuzione di $t$ si può studiare, anche se adesso abbiamo due gradi di libertà di cui tenere conto!**

# $t$ di Student per due campioni

Il nostro $t$, quindi, sarà in questo caso

$$t = \frac{\Delta - \mu_\Delta}{\sigma_\Delta} = \frac{\Delta - 0}{\sigma_\Delta} = \frac{\Delta}{\sqrt{\frac{\sigma^2_{pop}}{N_a} + \frac{\sigma^2_{pop}}{N_b}}} \tag{12}$$

- Di nuovo, possiamo stimare la varianza di popolazione a partire da quelle campionarie ed usare le nostre stime per il calcolo di $t$.
- **Ancora una volta, la distribuzione di $t$ si può studiare, anche se adesso abbiamo due gradi di libertà di cui tenere conto!**
- In questo caso, quindi, il test ci dice se due campioni appartengono o no alla stessa popolazione:

# $t$ di Student per due campioni

Il nostro $t$, quindi, sarà in questo caso

$$t = \frac{\Delta - \mu_\Delta}{\sigma_\Delta} = \frac{\Delta - 0}{\sigma_\Delta} = \frac{\Delta}{\sqrt{\frac{\sigma^2_{pop}}{N_a} + \frac{\sigma^2_{pop}}{N_b}}} \tag{12}$$

- Di nuovo, possiamo stimare la varianza di popolazione a partire da quelle campionarie ed usare le nostre stime per il calcolo di $t$.
- **Ancora una volta, la distribuzione di $t$ si può studiare, anche se adesso abbiamo due gradi di libertà di cui tenere conto!**
- In questo caso, quindi, il test ci dice se due campioni appartengono o no alla stessa popolazione:
- $H_0$ : i due campioni *appartengono* alla stessa popolazione $\Rightarrow t = 0$

# $t$ di Student per due campioni

Il nostro $t$, quindi, sarà in questo caso

$$t = \frac{\Delta - \mu_\Delta}{\sigma_\Delta} = \frac{\Delta - 0}{\sigma_\Delta} = \frac{\Delta}{\sqrt{\frac{\sigma^2_{pop}}{N_a} + \frac{\sigma^2_{pop}}{N_b}}} \tag{12}$$

- Di nuovo, possiamo stimare la varianza di popolazione a partire da quelle campionarie ed usare le nostre stime per il calcolo di $t$.
- **Ancora una volta, la distribuzione di $t$ si può studiare, anche se adesso abbiamo due gradi di libertà di cui tenere conto!**
- In questo caso, quindi, il test ci dice se due campioni appartengono o no alla stessa popolazione:
- $H_0$ : i due campioni *appartengono* alla stessa popolazione $\Rightarrow t = 0$
- $H_1$ : i due campioni *non appartengono* alla stessa popolazione $\Rightarrow t \neq 0$