

# Differenze di genere nella scuola primaria e secondaria nella provincia di Rimini: evidenze dal progetto RIR

Beatrice Franzolini

email: [franzolini@pm.me](mailto:franzolini@pm.me)

24 agosto 2021

## Sommario

L'analisi dei dati di RiminiInRete per gli anni scolastici dal 2013-2014 al 2019-2020, ha evidenziato che in media le ragazze tendono a ottenere valutazioni più elevate, anche se in maniera contenuta, in tutti gli insegnamenti. Restringendo l'analisi al 5% migliore degli studenti, i ragazzi ottengono risultati superiori in Matematica e Scienze, mentre il vantaggio delle ragazze diventa non significativo. Considerando il 5% peggiore degli studenti non si registra nessuna differenza di genere statisticamente significativa. Tali risultati appaiono in linea con la letteratura scientifica relativa al gender gap di voto nelle materie non-scientifiche e scientifiche.

L'andamento delle differenze di voto durante il percorso dalle elementari alle superiori non presenta tendenze particolarmente significative: in generale le differenze medie permangono in maniera approssimativamente costante durante tutto il percorso. L'analisi delle diverse coorti non evidenzia nessun effetto coorte sul gender gap. Gli studenti stranieri ottengono voti più bassi rispetto agli studenti italiani, con differenze medie fino a quasi un voto in Italiano. Tuttavia tali differenze non subiscono variazioni in base al genere dello studente.

## 1 Popolazione di riferimento e obiettivi dell'indagine

La popolazione di riferimento è composta dagli studenti che hanno ricevuto almeno una valutazione di fine anno presso una scuola primaria o secondaria di primo grado nella provincia Rimini negli anni scolastici dal 2013-2014 al 2019-2020. Gli obiettivi dell'indagine sono:

1. identificare le eventuali differenze di genere nelle valutazioni medie di fine anno e nelle diverse materie. (*Es. I maschi ottengono valutazioni medie significativamente diverse dalle femmine? I maschi ottengono valutazioni in italiano significativamente diverse dalle femmine?*)
2. Indagare l'evoluzione delle differenze di genere durante il percorso scolastico dello studenti dalle elementari, alle medie, fino alle superiori. (*Es. Le differenze di genere sono più ampie durante certi cicli di studio rispetto ad altri?*)
3. Indagare l'andamento temporale delle differenze di genere negli anni considerati/nelle diverse coorti di studenti. (*Es. Le differenze di genere sono state più rilevanti per alcune coorti di studenti?*)
4. Indagare l'interazione tra il genere e la nazionalità degli studenti rispetto alle valutazioni ottenute. (*Es. Le differenze di genere sono analoghe tra studenti italiani e studenti stranieri?*)

## 2 Dataset iniziale, valori mancanti e duplicati (nota tecnica)

Il dataset iniziale si riferisce a 3111 unità di rilevazione (i.e. studenti), osservati fino ad un massimo di 7 periodi temporali (i.e. anni scolastici).

Le unità per cui non è riportato il sesso anagrafico sono 54: per 3 di queste è stato possibile ricostruire il valore della variabile attraverso i dati anagrafici aggiuntivi, mentre le restanti 51 sono state escluse dall'indagine (per queste 51 non è disponibile nessun dato anagrafico).

Grazie al confronto dei dati anagrafici e del percorso scolastico sono state individuate 11 duplicazioni nel dataset, per le quali lo stesso studente risulta registrato come due unità di rilevazione disgiunte: per le 11 unità duplicate è stato possibile ricostruire tutte le variabili. Il numero effettivo di unità statistiche per cui è disponibile la variabile *Sesso* è quindi di 3049 unità, di cui 1609 maschi e 1440 femmine.

Le valutazioni risultanti pari a 0 nel dataset iniziale, vengono considerate come valori mancanti.

Infine, vengono escluse 122 unità di rilevazione per le quali non è riportata nessuna delle variabili di voto per nessun anno scolastico.

Il numero di studenti considerati è quindi 2927, di cui 1540 maschi e 1387 femmine.

## 3 Descrizione dei dati

I dati analizzati corrispondono a 2927 studenti e sono di tipo longitudinale, in quanto per ogni studente sono disponibili informazioni ripetute relative a più anni scolastici, fino ad un massimo di 7. Ogni studente entra nello studio quando riceve la prima valutazione presso una delle scuole primarie o secondarie di primo grado nella provincia di Rimini. La fine del periodo di follow-up<sup>1</sup> può invece essere determinata da uno dei seguenti eventi:

1. Lo studente completa il secondo ciclo di studi.
2. Lo studente interrompe gli studi.
3. Lo studente si iscrive ad una scuola al di fuori della provincia di Rimini (dato censurato).
4. Lo studente non completa il secondo ciclo di studi entro l'anno scolastico 2019-2020 (dato censurato).

Il periodo medio di follow-up per studente, determinato in base alla presenza/assenza nel dataset della valutazione media di fine anno, è di 3.94 anni.

Poichè (ragionevolmente) l'iscrizione ad una scuola al di fuori della provincia e il non completamento degli studi entro l'anno 2019-2020 non sono associati a un particolare andamento delle valutazioni dello studente, non si ritiene che i dati censurati possano determinare disorsioni nei risultati.

Le covariate considerate sono descritte sinteticamente nella Tabella 1. Oltre a 6 variabili di voto, sono presenti anche due variabili anagrafiche e tre variabili relative al percorso scolastico dello studente. La variabile anagrafica denominata *Sesso* coincide con il sesso riportato nei documenti anagrafici e assume due valori *F* per le femmine e *M* per i maschi.

Ad ogni osservazione è associata una coorte: gli studenti corrispondono a 14 diverse coorti. Per gli studenti che non hanno mai ripetuto anni scolastici la coorte è determinata dall'anno in cui hanno iniziato il ciclo di studi (es. Coorte=2007 significa che lo studente ha frequentato la prima elementare nel 2007). Quando uno studente non supera un anno scolastico esce dalla sua coorte e entra in quella successiva. Il dataset contiene informazioni relative a 14 diverse coorti, il cui percorso è descritto nella Tabella 2 e che sono rappresentate nell'Appendice B.

Nell'Appendice A sono riportate alcune statistiche descrittive e grafici dei dati a disposizione.

---

<sup>1</sup>Il periodo di follow-up è il periodo di tempo in cui uno studente rimane all'interno dello studio.

<b>Variabile</b>	<b>Descrizione</b>	<b>Nome variabile</b>	<b>Valori</b>
Sesso	Sesso riportato sui documenti anagrafici	<b>Sesso</b>	<i>F,M</i>
StrIta	Lo studente è classificato come “straniero” se nato in uno stato diverso dall’Italia o se la sua cittadinanza alla nascita era non-italiana	<b>StrIta</b>	<i>ITA,STR</i>
Classe	Classe frequentata dallo studente	<b>Classe</b>	<i>P1, P2, ..., SS5</i>
Coorte	Coorte di appartenenza	<b>Coorte</b>	<i>2006,...,2019</i>
Scuola	Scuola frequentata dallo studente	<b>Scuola</b>	<i>RN...</i>
Voto	Media di fine anno	<b>Voto</b>	[1, 10]
Mat	Voto di matematica di fine anno	<b>Mat</b>	[1, 10]
MatScie	Media del voto in matematica e scienze	<b>MatScie</b>	[1, 10]
Ita	Voto di italiano di fine anno	<b>Ita</b>	[1, 10]
Ingl	Voto di inglese di fine anno	<b>Ingl</b>	[1, 10]
Comp	Voto di comportamento di fine anno	<b>Comp</b>	[1, 10]

Tabella 1: Variabili utilizzate nell’analisi

	Classe frequentata nell’anno scolastico...						
Coorte	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19	2019/20
2006	SP3	SS1	SS2	SS3	SS4	SS5	
2007	SP2	SP3	SS1	SS2	SS3	SS4	SS5
2008	SP1	SP2	SP3	SS1	SS2	SS3	SS4
2009	P5	SP1	SP2	SP3	SS1	SS2	SS3
2010	P4	P5	SP1	SP2	SP3	SS1	SS2
2011	P3	P4	P5	SP1	SP2	SP3	SS1
2012	P2	P3	P4	P5	SP1	SP2	SP3
2013	P1	P2	P3	P4	P5	SP1	SP2
2014		P1	P2	P3	P4	P5	SP1
2015			P1	P2	P3	P4	P5
2016				P1	P2	P3	P4
2017					P1	P2	P3
2018						P1	P2
2019							P1

Tabella 2: Coorti

## 4 Risultati

I risultati riportati in questa sezione sono ottenuti attraverso l'impiego di modello lineari a effetti misti per ulteriori dettagli si veda la Sezione 5.

### 4.1 Differenze di genere valori medi di tutti gli anni considerati

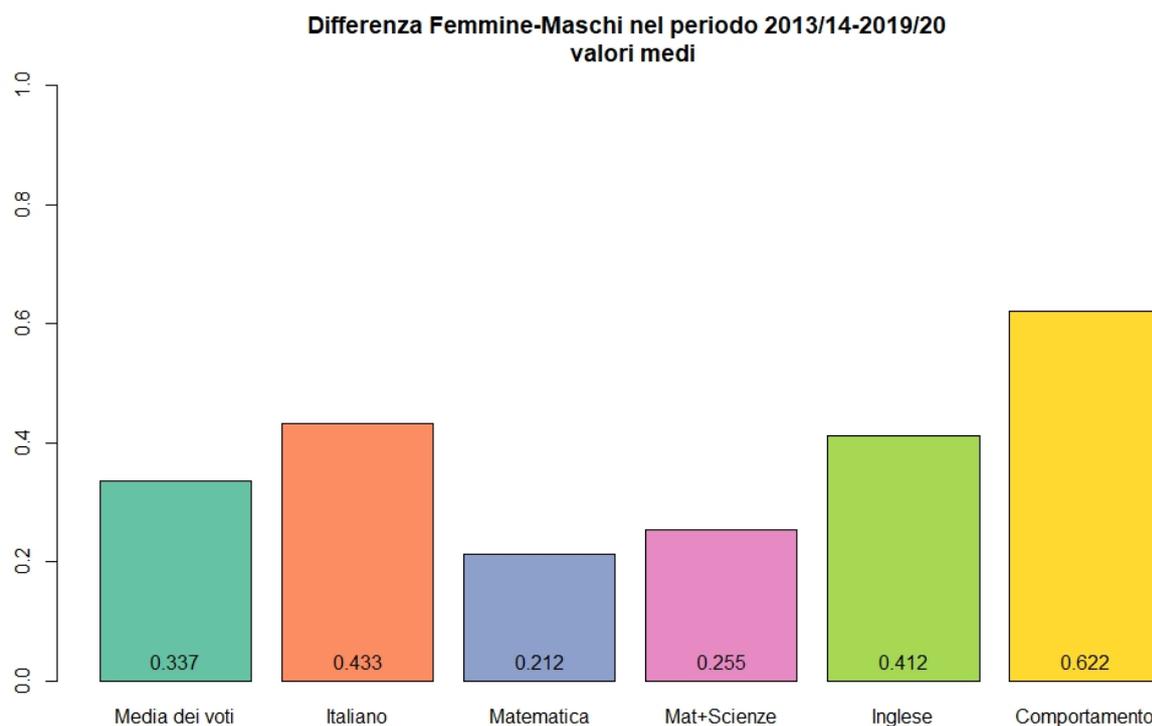


Figura 1: Variazione media tra femmine e maschi per i diversi insegnamenti - valore calcolato attraverso un modello lineare a effetti misti

In Figura 1 sono rappresentate le differenze di genere (F - M) in media per le 6 variabili di voto. Le variazioni evidenziano un vantaggio delle ragazze, che in media performano meglio in tutti gli ambiti, con un vantaggio più rilevante nelle materie non-STEM (come Italiano e Inglese). In particolare, la media dei voti è in media più alta di 0.34 per le ragazze rispetto ai ragazzi e il voto in Italiano è in media più alto di 0.43 per le ragazze rispetto ai ragazzi. Mentre le differenze medie in Matematica e Matematica e Scienze sono di 0.21 e 0.25, sempre a favore delle ragazze. Il valore per le valutazioni in Inglese è 0.41. Per il Comportamento il differenziale tra femmine e maschi è pari a 0.62. Tutte e sei le differenze risultano statisticamente significative.

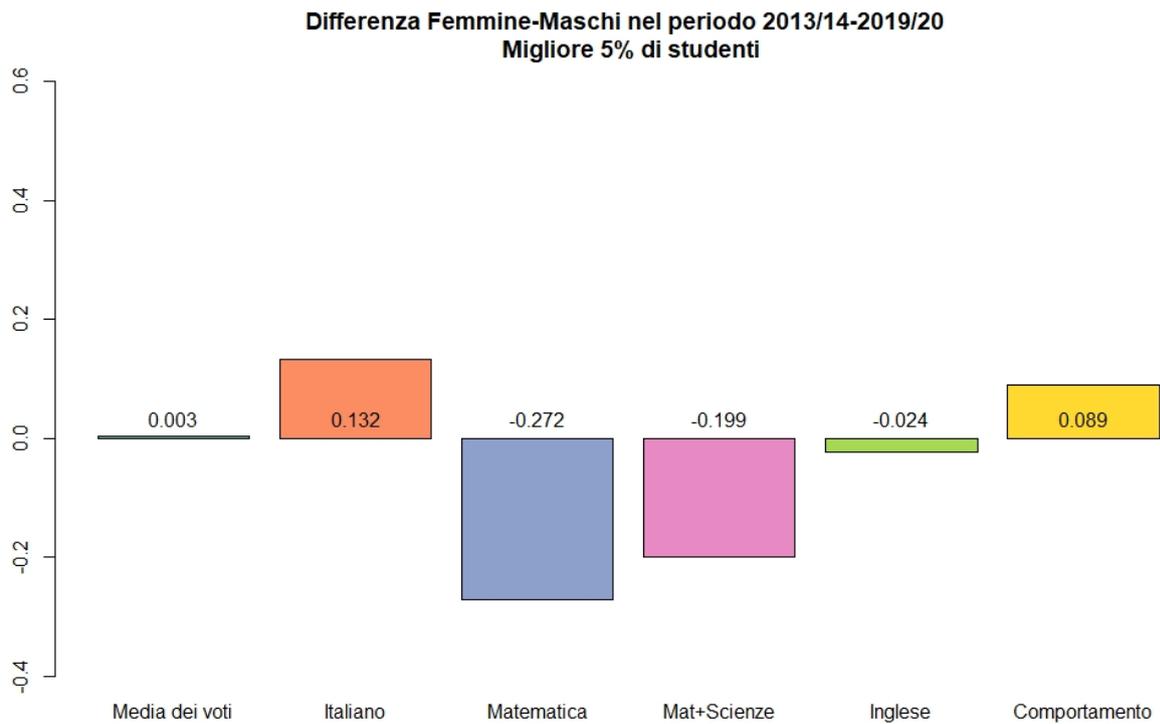


Figura 2: Variazione media tra femmine e maschi considerando solo il 5% migliore degli studenti per i diversi insegnamenti - valore calcolato attraverso un modello lineare a effetti misti

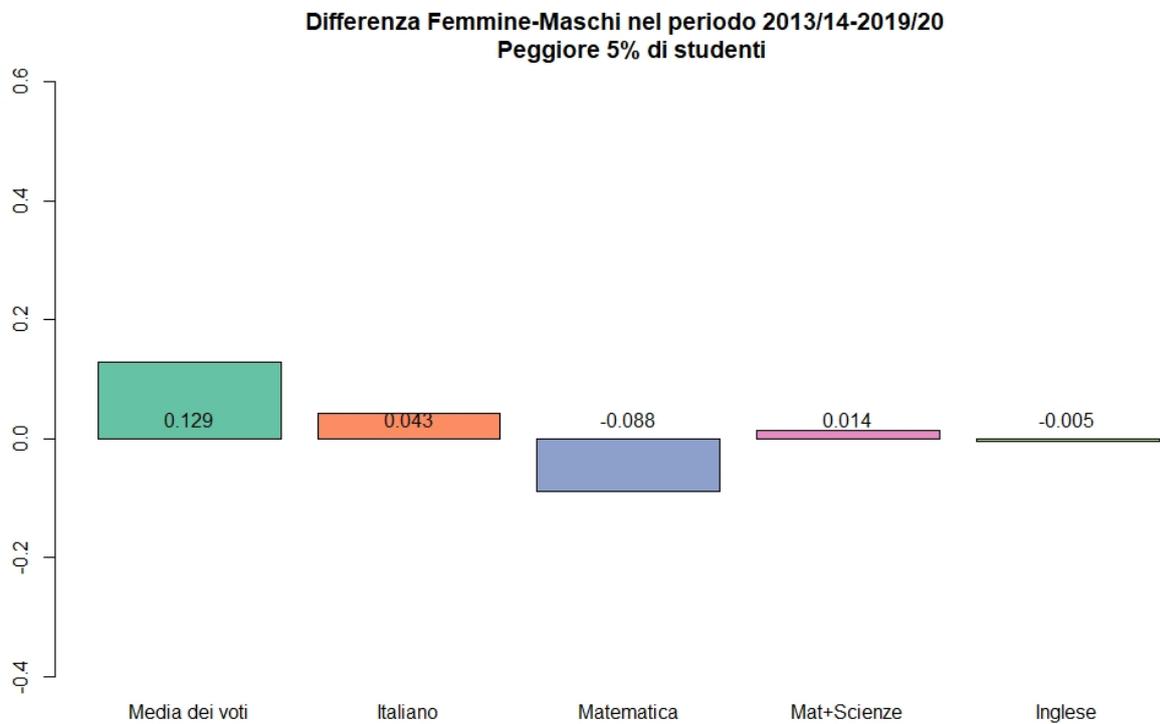


Figura 3: Variazione media tra femmine e maschi considerando solo il 5% peggiore degli studenti per i diversi insegnamenti - valore calcolato attraverso un modello lineare a effetti misti

## 4.2 Evoluzione delle differenze di genere nel percorso scolastico degli studenti

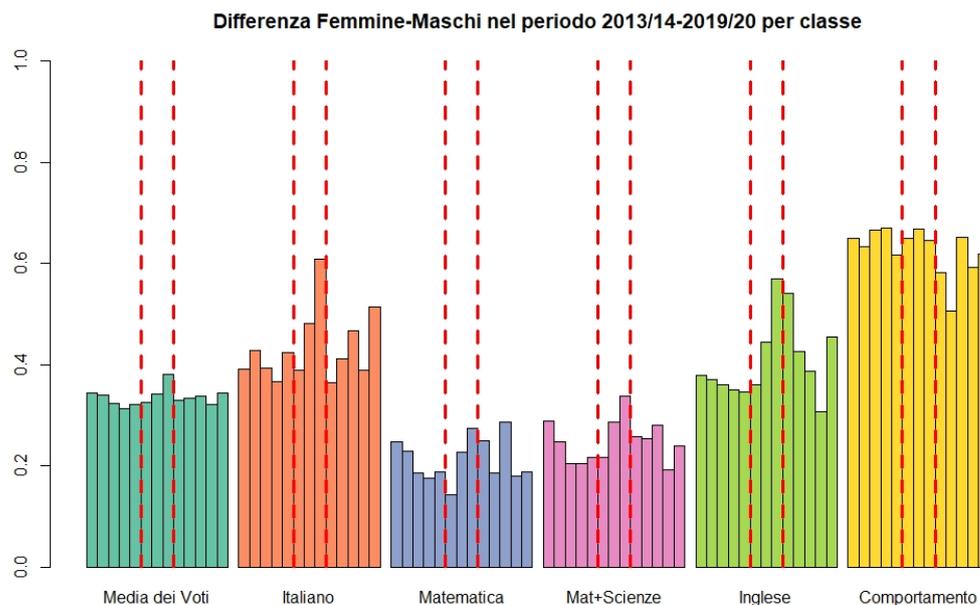


Figura 4: Ogni barra rappresenta una delle 13 classi, le linee rosse verticali dividono la scuola primaria dalla secondaria di primo grado e quest'ultima della secondaria di secondo grado.

Per la media dei voti, Matematica e Matematica+Scienze non c'è evidenza statistica di un trend nelle differenze di genere medie durante il percorso scolastico. Mentre per Italiano e Inglese il trend è debolmente significativo.

## 4.3 Andamento nelle diverse coorti di studenti

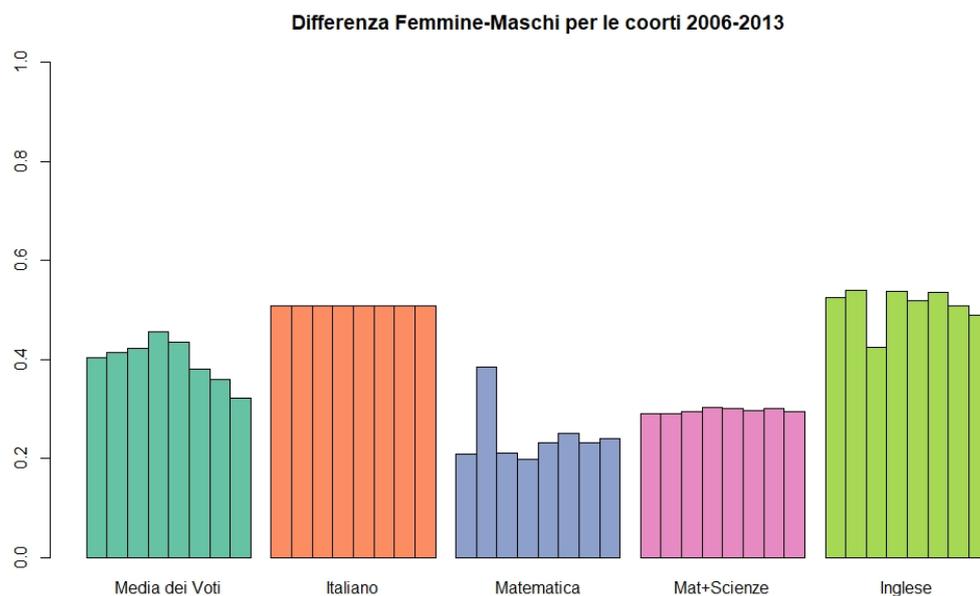


Figura 5: Ogni barra rappresenta una coorte di studenti.

Non c'è nessun effetto significativo delle diverse coorti sulle differenze di genere: non si sono registrate differenze nelle gender gap tra le diverse coorti considerate che non siano attribuibili alla naturale variabilità tra gli studenti.

#### 4.4 Differenze di genere e "nazionalità"

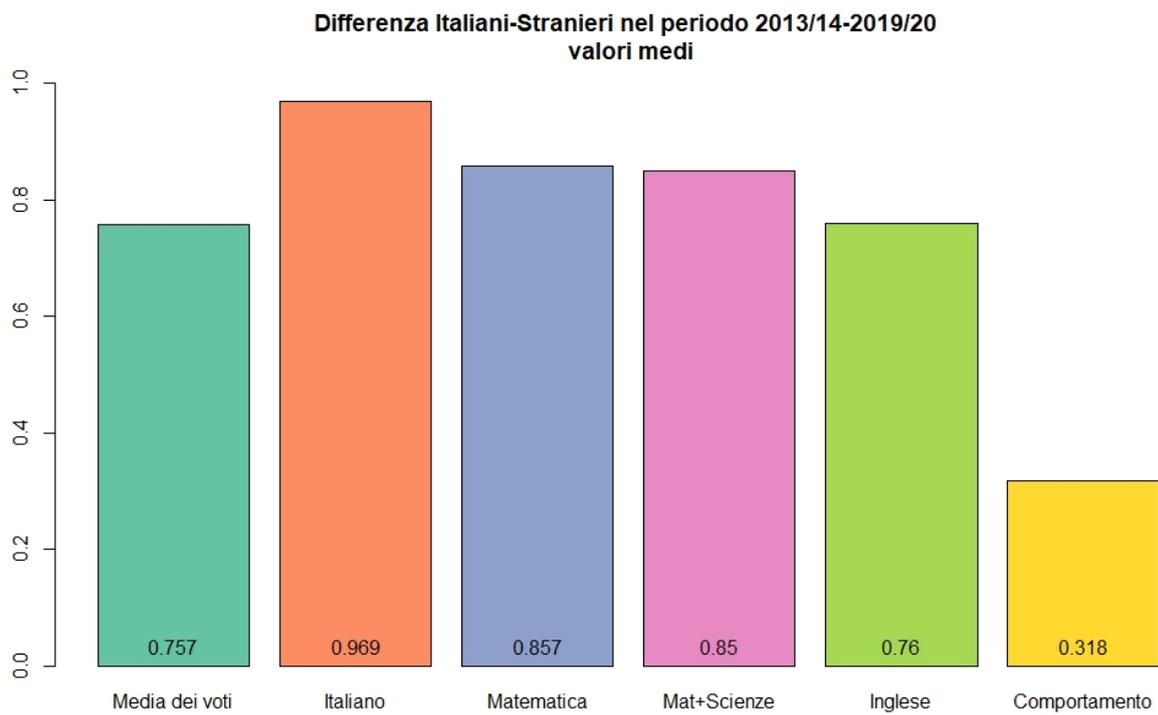


Figura 6: Variazione media tra italiani e stranieri per i diversi insegnamenti. Tutte le differenze sono statisticamente significative.

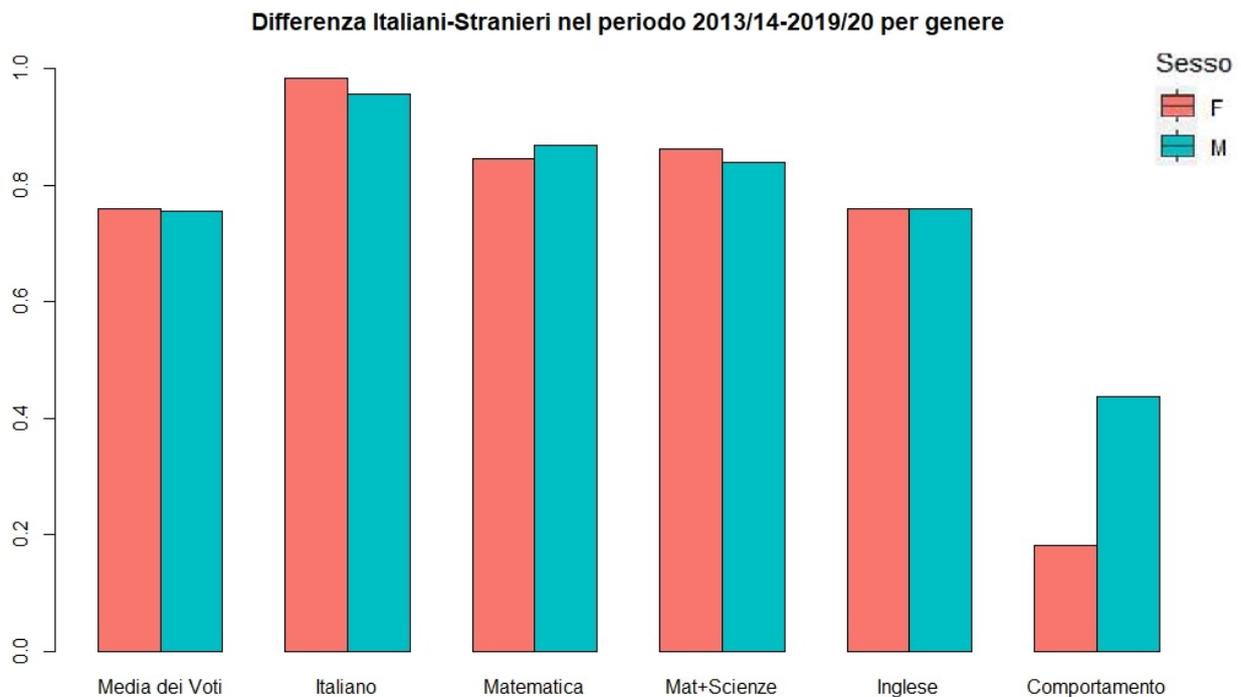


Figura 7: Variazione media tra ragazze italiane e straniere e tra ragazzi italiani e stranieri.

## 5 Nota metodologica sintetica

I risultati riportati nella sezione precedente sono ottenuti attraverso modelli lineari a effetti misti che (rispetto alle medie semplici condizionate) permettono di correggere rispetto a eventuali sbilanciamenti tra il numero di maschi e il numero di femmine e potenziali fattori confondenti. Tuttavia, essendo femmine e maschi approssimativamente equamente distribuiti nelle diverse coorti, nelle diverse classi e tra italiani e stranieri, in generale i risultati non divergono in maniera decisiva rispetto al calcolo delle medie semplici.

Modello per la Figura 1 e Figura 6:

$$\text{Voto}_{i,j} = \alpha + \beta \cdot \text{Sesso}_i + \gamma \cdot \text{Classe}_{i,j} + \xi \cdot \text{StrIta}_i + \delta_{\text{Coorte}(i,j)} + \eta_i + \epsilon_{i,j} \quad (1)$$

Il modello è stato stimato per ogni variabile di voto e in tutti i casi evidenzia che tutti gli effetti considerati sono statisticamente significativi, i.e. la valutazione differisce significativamente in base al sesso, la classe frequentata e tra stranieri e italiani.

Per la costruzione dei grafici nelle Figure 2 e 3 si è utilizzata una stima a due step. Ad ogni studente è stato associato un singolo voto che rappresentasse la performance dello studente depurata dall'“effetto classe” e dall'“effetto nazionalità”. Tale voto è definito pari all'effetto random  $\eta_i$  nel modello

$$\text{Voto}_{i,j} = \alpha + \gamma \cdot \text{Classe}_{i,j} + \xi \cdot \text{StrIta}_i + \delta_{\text{Coorte}(i,j)} + \eta_i + \epsilon_{i,j} \quad (2)$$

In base al valore di  $\eta_i$  sono stati identificati i due gruppi di studenti formati rispettivamente dal 5% migliore degli studenti e dal 5% peggiore degli studenti. I differenziali rappresentati nelle figure sono poi ottenuti attraverso la stima modello in (1) applicato ai due gruppi.

I valori rappresentati nelle Figure 4 e 5 corrispondono ai coefficienti random  $\beta$  delle variabili Classe e Coorte rispettivamente nei modelli

$$\text{Voto}_{i,j} = \alpha + \beta_{\text{Classe}_{i,j}} \cdot \text{Sesso}_i + \gamma \cdot \text{Classe}_{i,j} + \xi \cdot \text{StrIta}_i + \delta_{\text{Coorte}(i,j)} + \eta_i + \epsilon_{i,j} \quad (3)$$

$$\text{Voto}_{i,j} = \alpha + \beta_{\text{Coorte}_{i,j}} \cdot \text{Sesso}_i + \gamma \cdot \text{Classe}_{i,j} + \xi \cdot \text{StrIta}_i + \delta_{\text{Coorte}(i,j)} + \eta_i + \epsilon_{i,j} \quad (4)$$

Infine la Figura 7 è ottenuta attraverso il modello

$$\text{Voto}_{i,j} = \alpha + \beta \cdot (\text{Sesso}_i : \text{StrIta}_i) + \gamma \cdot \text{Classe}_{i,j} + \xi \cdot \text{StrIta}_i + \delta_{\text{Coorte}(i,j)} + \eta_i + \epsilon_{i,j} \quad (5)$$

## 6 Confronto con la letteratura scientifica esistente

I risultati ottenuti sono in linea con le evidenze scientifiche relative al gender gap di voto, per le quali le ragazze tendono ad avere valutazioni in media superiori e meno variabili rispetto ai ragazzi, con un vantaggio medio più contenuto nelle materie STEM. Mentre, tra i migliori studenti, i ragazzi tendono a sovraperformare nelle materie STEM. Si veda, ad esempio, O'Dea et al. (2018) and Voyer & Voyer (2014). La Figura 8 cerca di rappresentare sommariamente le distribuzioni di voto attese dall'evidenze scientifiche attuali.

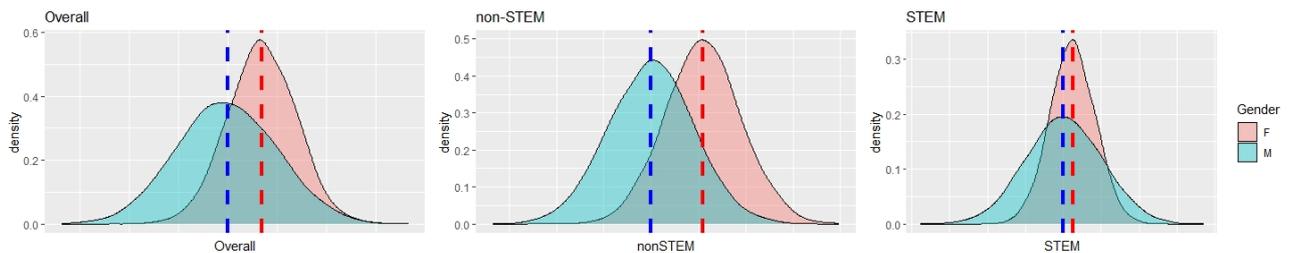
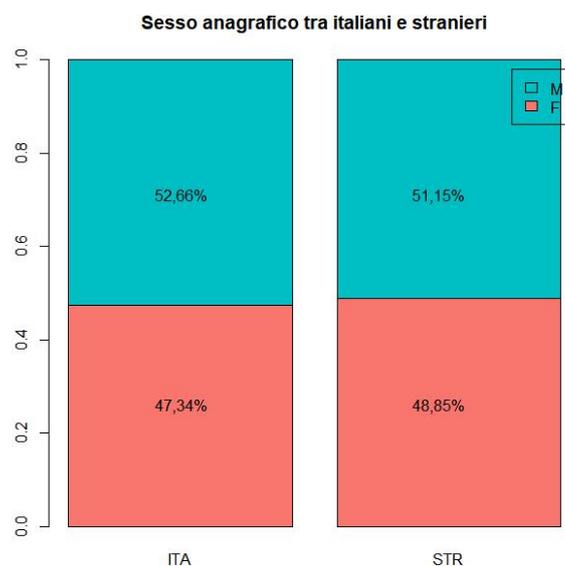


Figura 8: Differenze attese dalla letteratura scientifica nelle distribuzioni dei voti per genere

## Appendice A - Statistiche descrittive

### A.1 Variabili anagrafiche: Sesso e StrIta

Sesso	Conteggio	Percentuale	StrIta	Conteggio	Percentuale
M	1540	52.61%	ITA	2180	74.48%
F	1387	47.39%	STR	747	25.52%
Tot	2927	100%	Tot	2927	100%



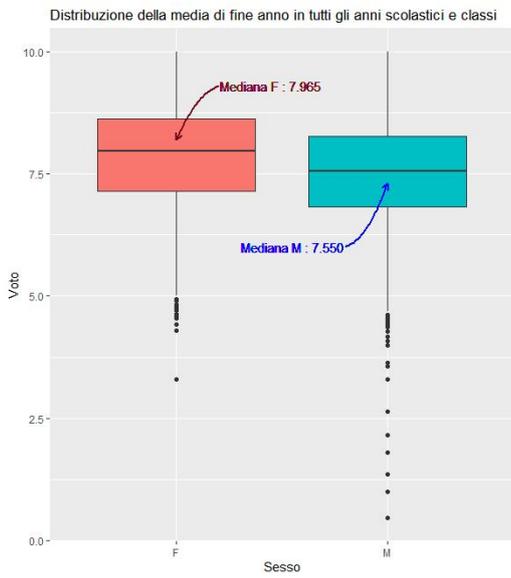
Circa il 25% degli studenti è stato classificato come "straniero".

Gli studenti sono approssimativamente equamente distribuiti tra maschi e femmine, sia sull'intera popolazione sia considerando separatamente gli studenti italiani e gli studenti stranieri.

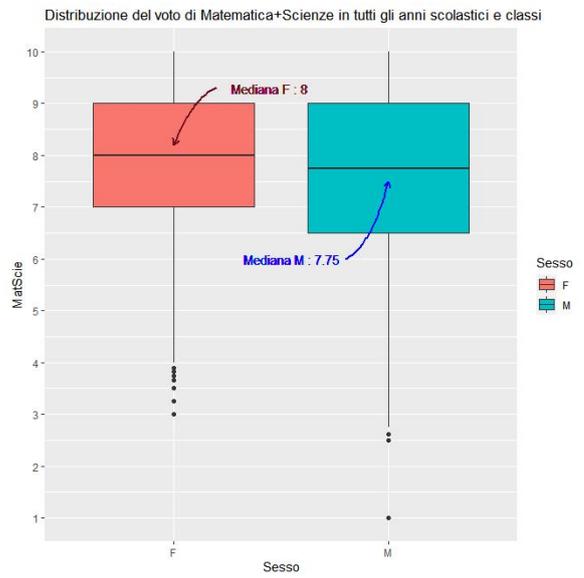
### A.2 Variabili di Voto: Voto, Mat, MatScie, Ita, Ingl, Comp

Si noti che le statistiche descrittive in questa sezione sono calcolate considerando insieme tutte le osservazioni e non distinguendo per le diverse classi o anni scolastici.

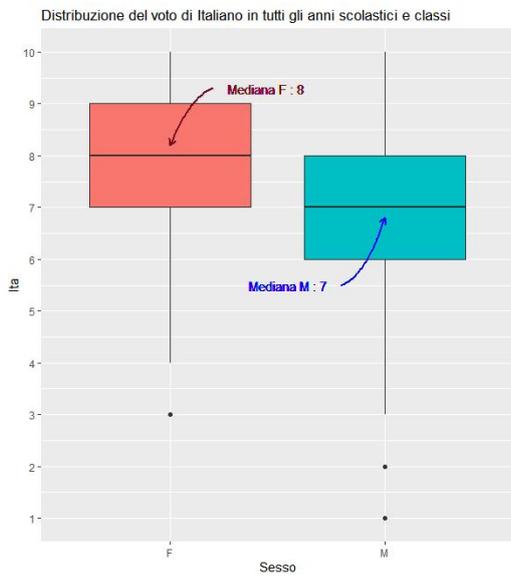
	Voto	Mat	MatScie	Ita	Ingl	Comp
valori disponibili	11543	11519	11474	11528	11525	5358
valori mancanti	0	24	69	15	18	6185
media	7.69	7.68	7.84	7.69	7.78	8.79
deviazione standard	1.02	1.50	1.35	1.32	1.39	1.01
min	0.46	1.0	1.0	1.0	1.0	1.0
25%	6.92	6.0	6.75	7.0	7.0	8.0
50%(mediana)	7.73	8.0	8.0	8.0	8.0	9.0
75%	8.45	9.0	9.0	9.0	9.0	10.0
max	10.0	10.0	10.0	10.0	10.0	10.0



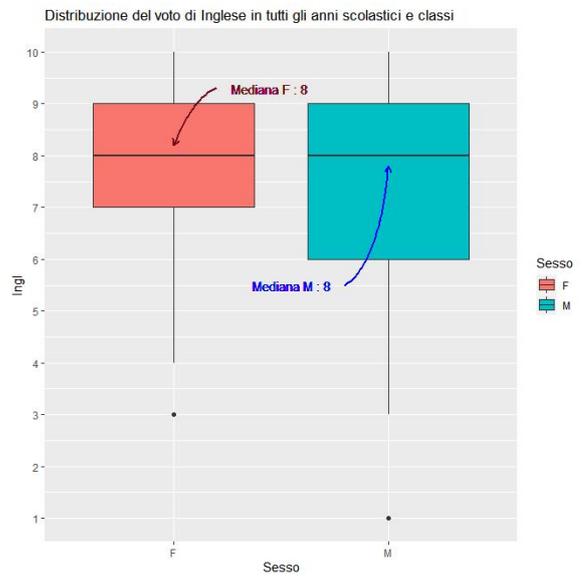
(a) Voto medio



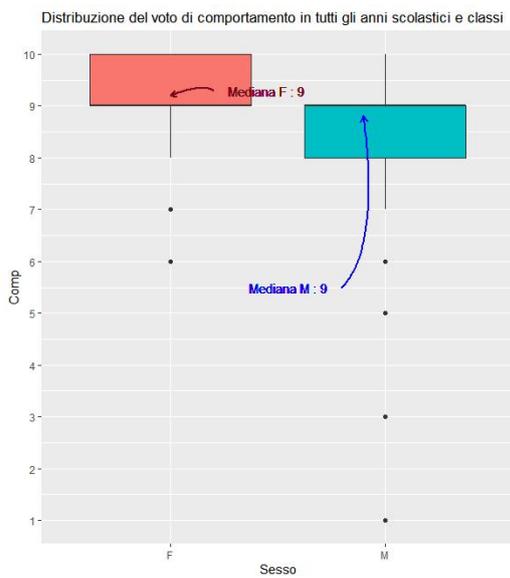
(b) Matematica e Scienze



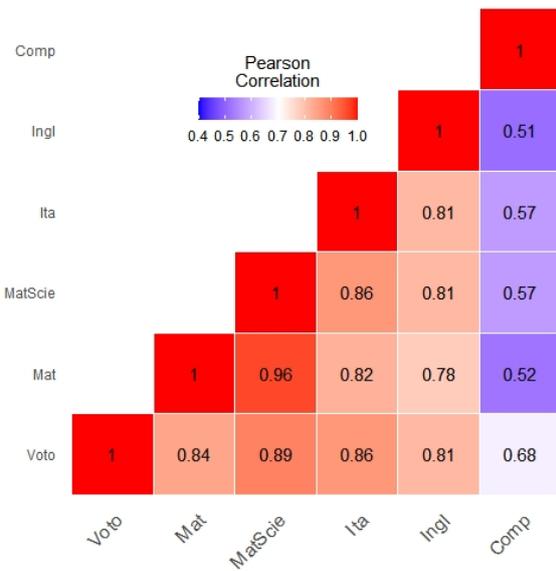
(c) Italiano



(d) Inglese



(e) Comportamento

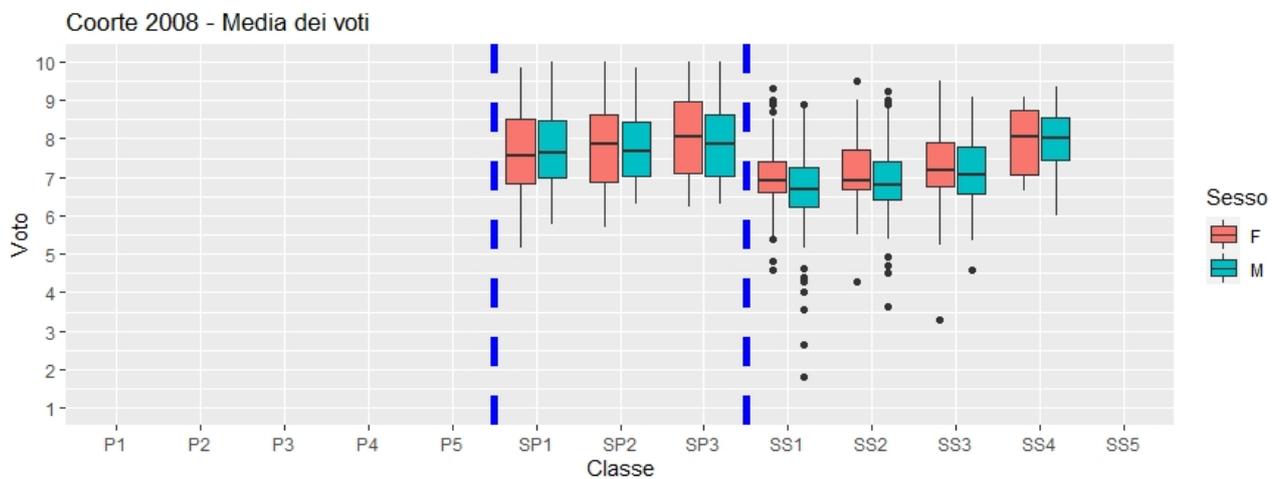
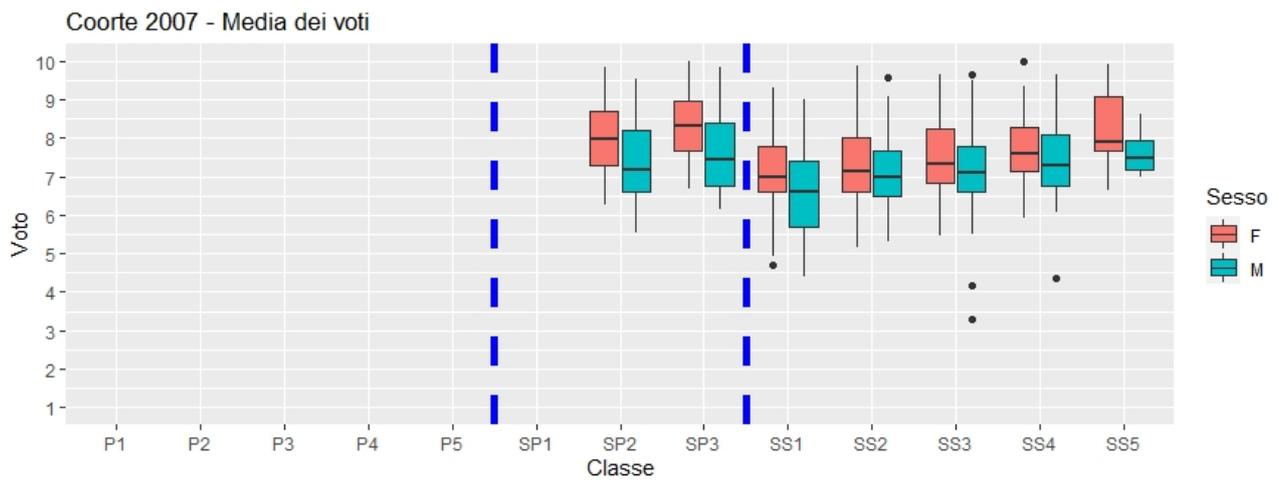
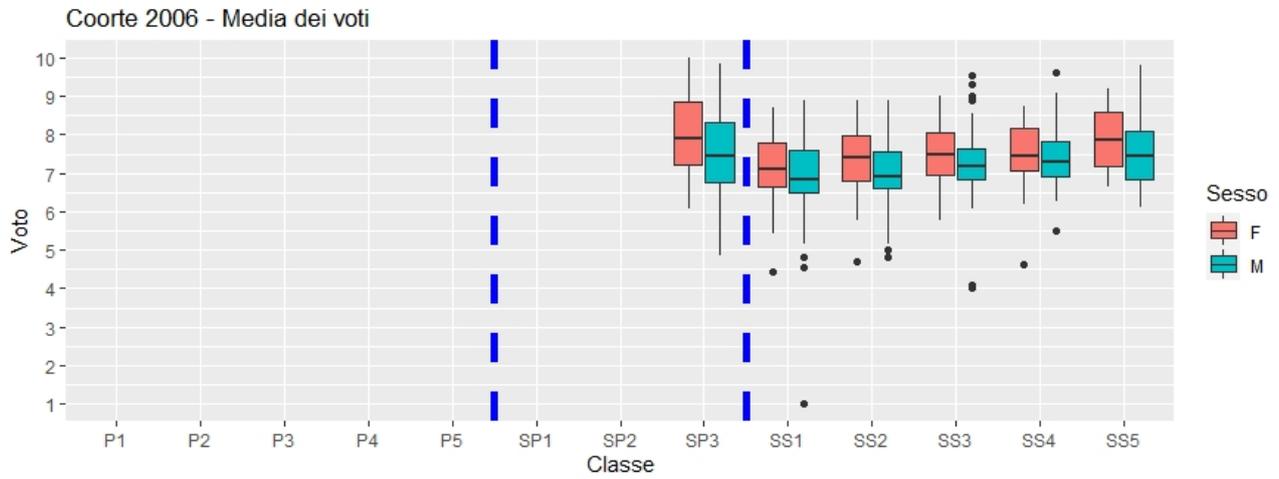


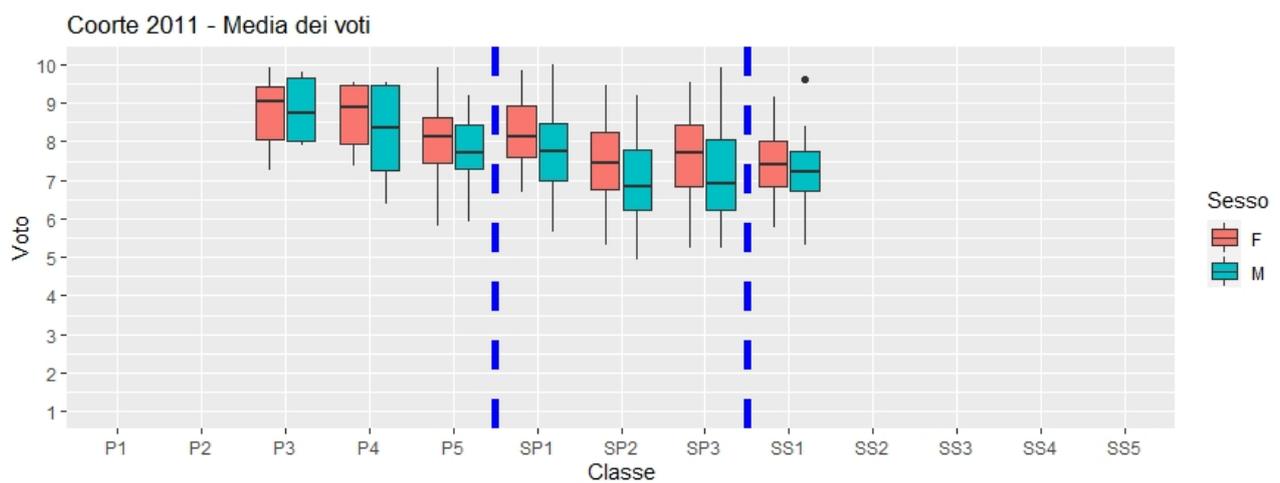
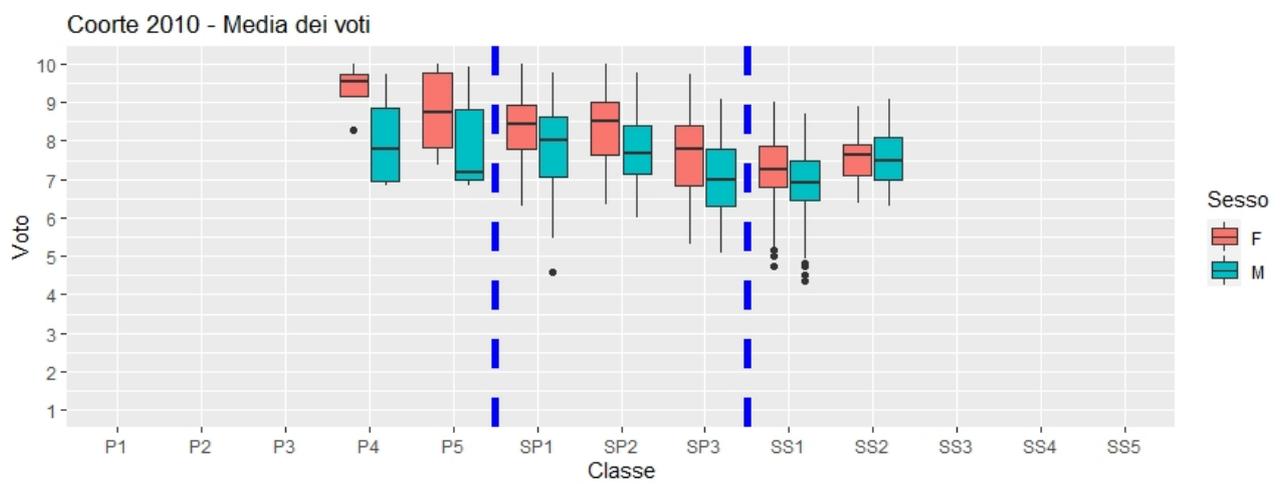
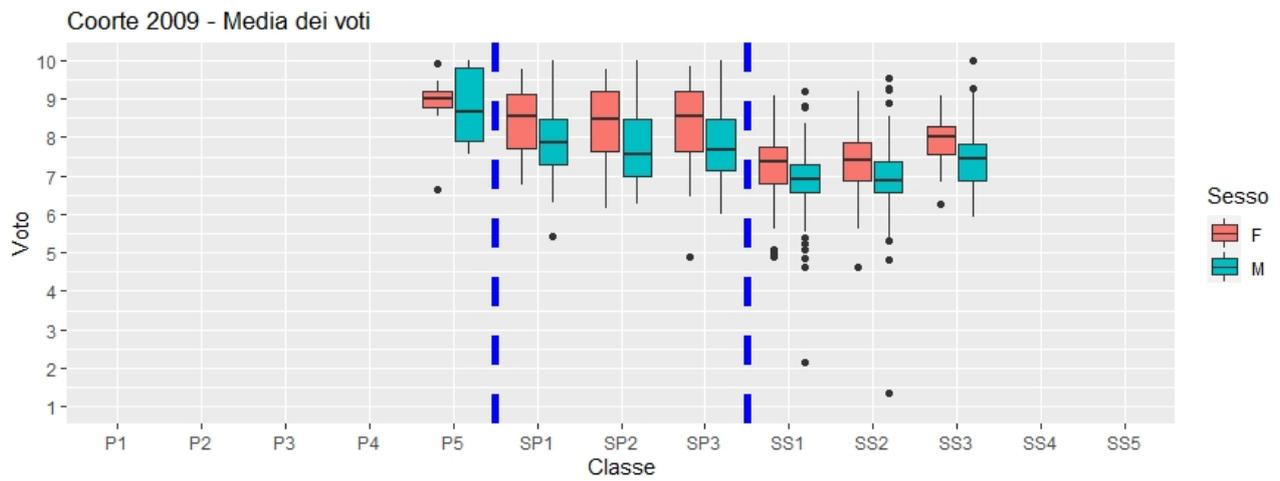
(f) Correlazione

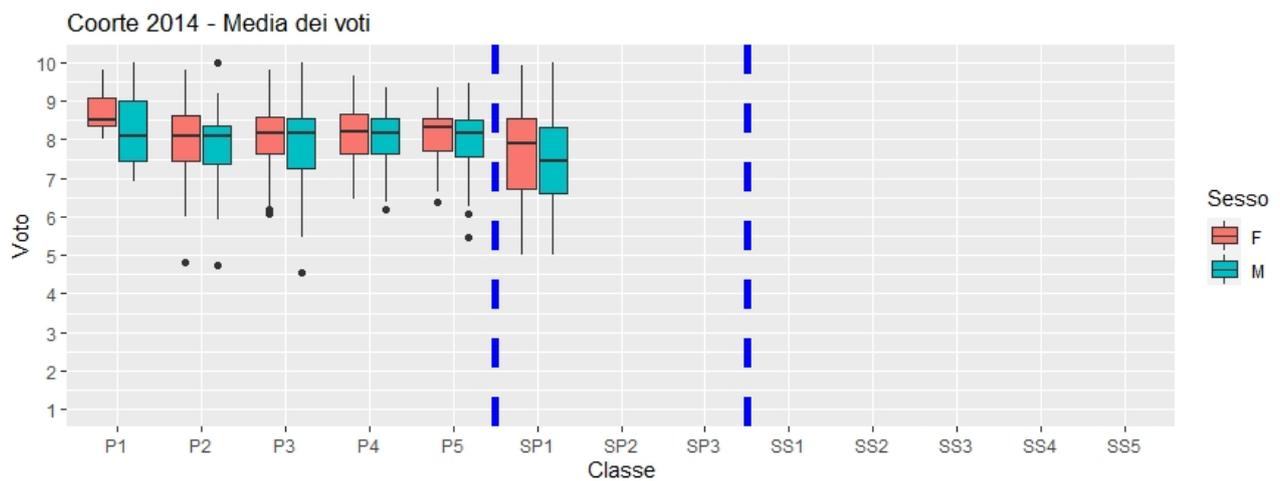
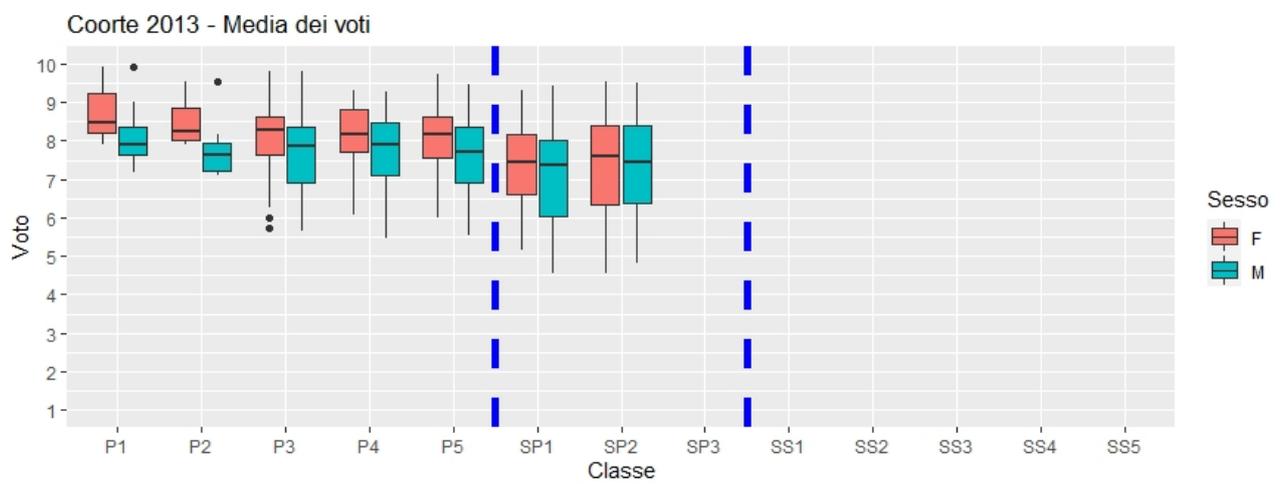
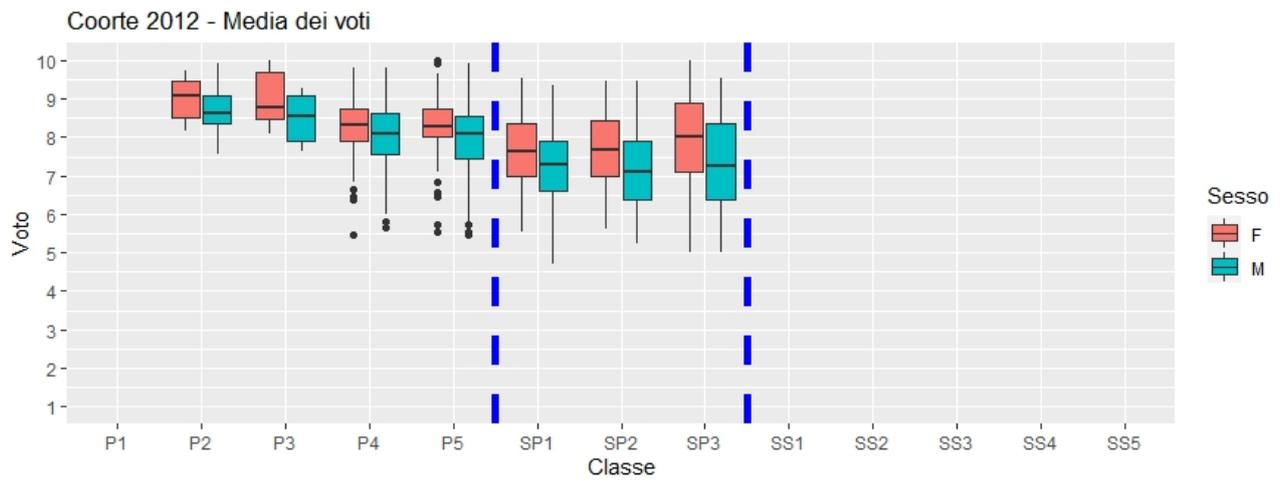
Figura 9: Box-plots complessivi per sesso anagrafico e correlazione tra le valutazioni

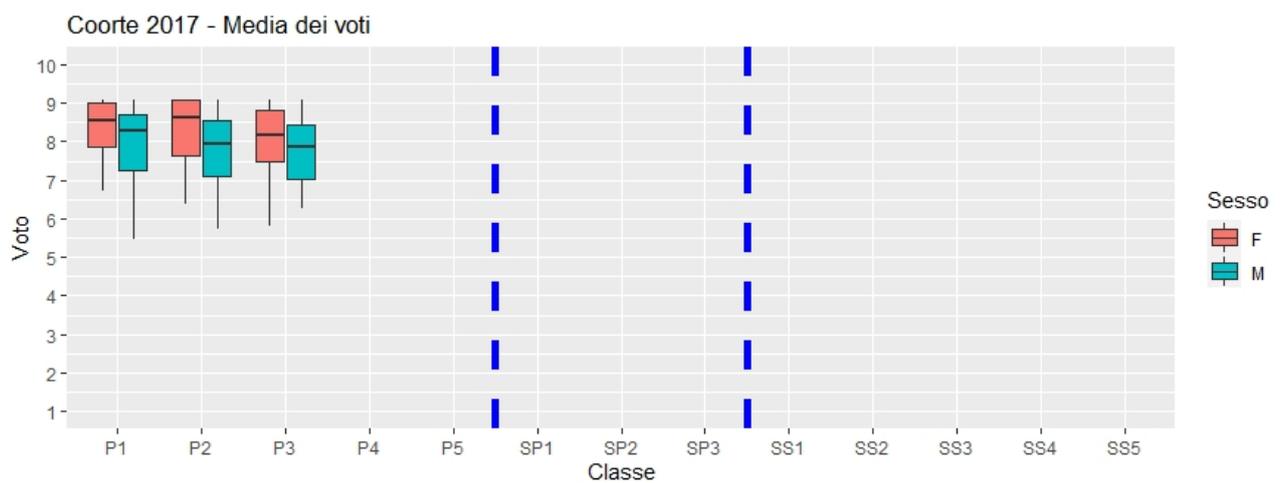
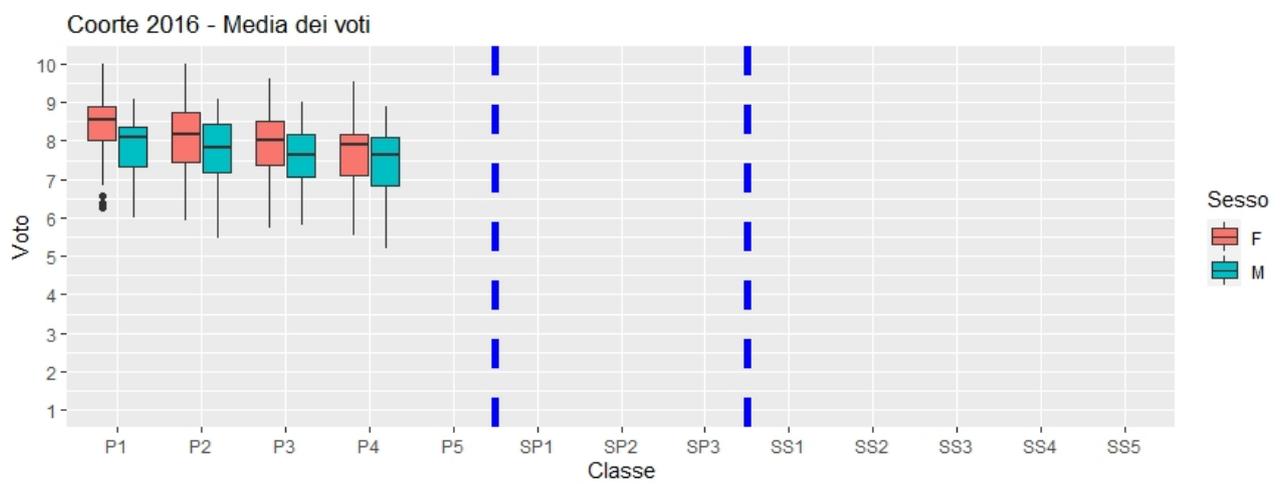
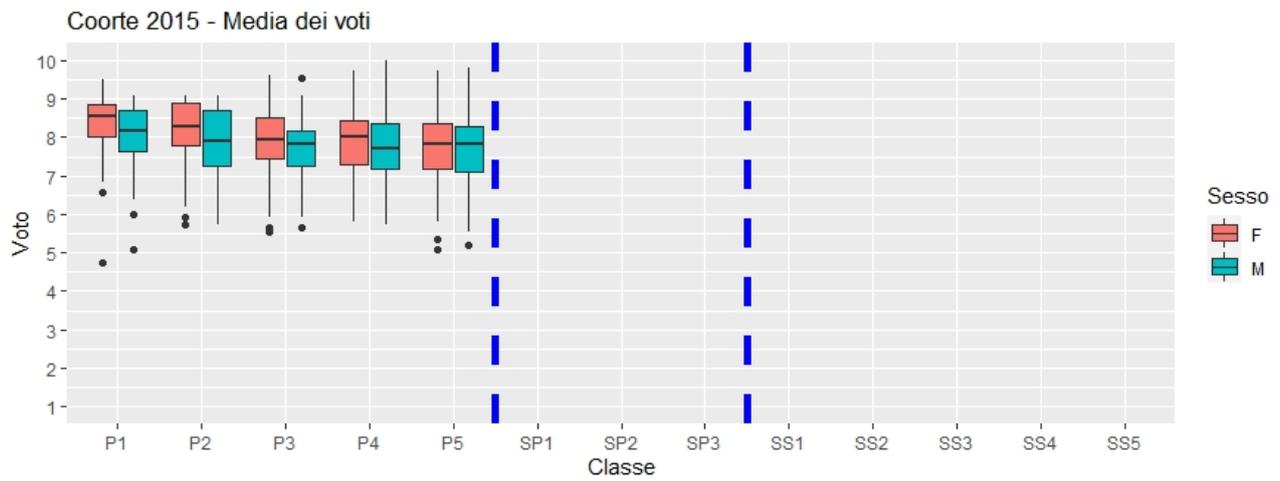
# Appendice B - Andamento temporali delle diverse coorti di studenti per Sesso

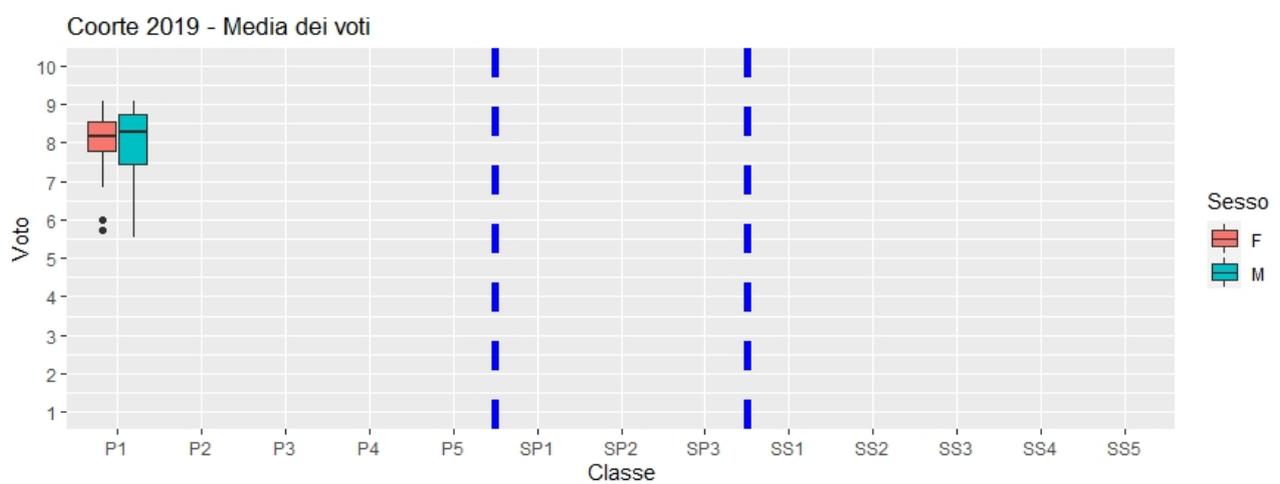
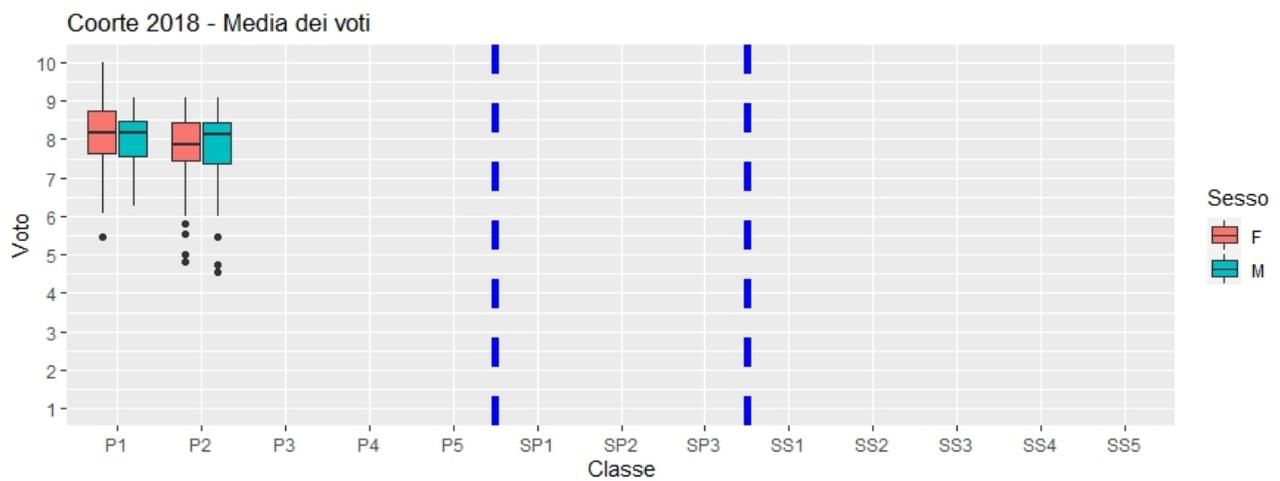
## Andamento coorti: media dei voti











## Riferimenti bibliografici

O'DEA, R. E., LAGISZ, M., JENNIONS, M. D. & NAKAGAWA, S. (2018). Gender differences in individual variation in academic grades fail to fit expected patterns for stem. *Nature communications* **9**, 1–8.

VOYER, D. & VOYER, S. D. (2014). Gender differences in scholastic achievement: a meta-analysis. *Psychological bulletin* **140**, 1174.