

Un esempio di Report

Emanuele Meier

2021-12-05

Contents

1	Premessa	5
2	Il progetto	7
3	Analisi dei risultati delle prove	11
3.1	Informazioni generali	11
3.2	Correlazioni tra le dimensioni valutate con la prova	14
4	Il contesto e la scuola	17
4.1	Ispettorati	17
5	Applications	25
5.1	Example one	25
5.2	Example two	25
6	Final Words	27

Chapter 1

Premessa

Questo è un esempio di *report* informatizzato che si potrebbe pensare di presentare al cantone. Il documento è stato creato utilizzando bookdown, un pacchetto o libreria comunemente utilizzata per generare dei documenti libri, tramite R direttamente. Ve ne sono diversi esempi che possono essere consultati (<https://bookdown.org>).

Chapter 2

Il progetto

Il presente rapporto fa seguito a due rapporti analoghi prodotti il primo sulla base dei risultati alle prove di matematica in IV elementare (Cirse, 2014) e il secondo alle prove di italiano in III elementare (Cirse, 2016).

Per quanto riguarda questo progetto nel suo insieme si è svolto negli anni scolastici dal 2013 al 2015.

L'insieme delle valutazioni tramite prove standardizzate ha il duplice obiettivo di fornire delle informazioni di monitoraggio del sistema educativo (oggetto specifico di questo rapporto) e di fornire ai docenti, direttori e ispettori delle informazioni di dettaglio relative all'andamento delle specifiche classi.

Lo svolgimento del progetto di Matematica e la struttura del rapporto riprendono quelli degli altri rapporti. La presentazione e le dimensioni teoriche sono molto simili e si rimanda per gli approfondimenti del caso ai documenti già esistenti; alcune parti che si riteneva potessero giovare alla lettura ancorché già esistenti sono state messe in appendice.

I primi mesi del progetto sono stati impiegati per il consolidamento della rete di collaborazione. Questa è stata composta dal gruppo di esperti del territorio (in questo caso il termine non fa riferimento esclusivo agli esperti disciplinari della Scuola Media bensì a persone che fossero portatori di un sapere e di una conoscenza utile alla riflessione su questo oggetto), individuati grazie alla collaborazione con l'Ufficio Scuole Comunali (USC) e con il Dipartimento Formazione e Apprendimento (DFA).

La prima fase del progetto, sino ad agosto 2013, è stata dedicata alla definizione dei contenuti da testare e all'identificazione delle persone che avrebbero potuto sviluppare gli item. La scelta dei settori da investigare è fondamentale in quanto sarebbe estremamente difficoltoso valutare contemporaneamente tutte le competenze presenti in una disciplina in modo esatto. Il modello di analisi che è stato utilizzato è legato alla Item Response Theory (IRT) ed è costruito al

fine di avere delle misure precise di costrutti ben definiti e quanto più possibili unitari, in appendice è presente una spiegazione maggiormente dettagliata (vedi Appendice 2 “Competenze fondamentali e prove”).

Nell’elaborazione della prova sono stati presi in considerazione, attingendoli al Piano di studio della scuola dell’obbligo (Divisione della scuola, 2015), sei Traguardi di competenza: Grandezze e misure – Eseguire e applicare, Grandezze e misure – Sapere, riconoscere e descrivere, Numeri e calcolo – Matematizzare e modellizzare, Grandezze e misure – Matematizzare e modellizzare, Numeri e calcolo – Sapere, riconoscere e descrivere, Numeri e calcolo – Eseguire e applicare. L’insieme delle risposte alle domande ha costituito una misura generale chiamata Matematica Generale.

Successivamente, sono stati sviluppati gli item per testare ognuna delle parti stabilite. Nel costruire gli item si è dovuto tener conto di alcune indicazioni. In primo luogo ogni item doveva essere quanto più possibile mono dimensionale. Il lavoro in questa fase è stato svolto in stretta collaborazione con gli esperti di Didattica della Matematica operanti nel DECS e nel DFA. Per poter misurare la capacità di discriminazione dell’item (si parla di capacità discriminativa relativamente al fatto che l’item riceva risposte corrette dagli allievi più abili e scorrette da quelli meno abili) e anche la sua coerenza con la dimensione che si desiderava valutare è infatti necessario che ogni item sia attinente a una e una sola dimensione. Questa caratteristica rende questi item in sé differenti da quelli che normalmente sono utilizzati dai docenti durante la loro attività professionale. In secondo luogo si è dovuto creare un numero di item sovrabbondante rispetto all’uso finale. Si è dovuto infatti prevedere che successivamente alla prova campione sarebbe stato eliminato almeno il 30% degli item prodotti.

Una volta prodotti gli item e sottoposti a verifica di contenuto con l’assistenza degli esperti, si è proceduto alla preparazione della prova campione. Questa prova aveva lo scopo di valutare la pertinenza degli item e di individuare quelli più efficaci a misurare e a discriminare. Ordinandoli per difficoltà crescente si dovrebbe trovare un numero inizialmente molto elevato di allievi che risponderà correttamente e questo numero dovrebbe ridursi al crescere della difficoltà. Se un item ad esempio, riceverà, un numero di risposte corrette elevato ma solo dagli allievi meno abili questo sarà scartato, altrimenti saranno eliminati gli item non discriminanti, quelli cioè ai quali tutti o nessuno avranno risposto. Queste procedure hanno infatti lo scopo di costruire delle scale valide non in termini astratti ma all’interno delle popolazioni reali. Gli item costruiti sono infatti coerenti con i contenuti presentati nella scuola ticinese e la loro difficoltà è valutata rispetto agli allievi della stessa scuola. Concretamente, sono stati realizzati 10 fascicoli differenti, ciascuno di essi richiedeva un tempo di risposta di 45 minuti. Questa distribuzione apparentemente complessa era necessaria per garantire che ogni item fosse testato su almeno 300 allievi e ogni allievo venisse confrontato con 2 fascicoli, l’uno a distanza di una settimana dall’altro. La distanza di una settimana è stata ritenuta quella minima per poter ritenere l’effetto di apprendimento residuale. Ogni allievo ha quindi ricevuto due fascicoli

diversi assegnati casualmente e le classi sono state estratte in modo da essere rappresentative della popolazione degli studenti ticinesi. Si è scelto di utilizzare un campionamento basato sulle classi scolastiche, in quanto meno invasivo e per altro accettato all'interno delle principali ricerche nazionali e internazionali. Sono stati testati circa 1600 allievi pari al 50% della popolazione complessiva.

La somministrazione è stata curata da personale esterno alla scuola che si è occupato di portare le prove nelle singole classi, far eseguire il lavoro agli allievi e recuperare poi i materiali distribuiti. Quest'ultima fase è estremamente rilevante in quanto gli esercizi proposti in questa fase dei lavori non hanno ancora subito alcun processo di validazione e non possono essere ritenuti efficaci alla valutazione delle competenze specifiche, non vi sono infatti valori che ne indichino l'efficacia o la difficoltà in alcun modo. Si deve anche sottolineare come un esercizio diffuso in maniera non corretta (ad esempio tramite fotocopie del materiale) potrebbe rendere l'attività di valutazione non valida in quanto introdurrebbe una condizione di non equità di fronte alla prova. Un problema più ampio legato alla disponibilità degli esercizi è quello definito in letteratura "teaching for testing" (Flukiger, 2004) del quale si discuterà più oltre.

Tutti i questionari sono stati quindi raccolti e le risposte sono state inserite in un archivio al fine di poter valutare la bontà metrica degli esercizi. I singoli esercizi, le scale e l'insieme degli esercizi sono stati quindi valutati utilizzando il modello di Rasch, al fine di capire come costruire le successive prove e quali esercizi conservare. Queste analisi hanno permesso di identificare 90 esercizi con buone capacità metriche divisi nei 6 settori. Di seguito vedremo un sunto delle analisi effettuate.

Dagli esercizi è stato possibile realizzare due fascicoli, la prova complessivamente comprendeva 90 item, 15 per ogni traguardo di competenza. La finestra di tempo nella quale questa prova doveva avvenire è stata di due settimane.

Per evitare inoltre gli effetti visti nelle medesime ricerche si è stabilito di far svolgere le attività di somministrazione esclusivamente a personale formato specificatamente, ciò anche per non portare un aggravio di lavoro al personale insegnante.

Una volta raccolte le prove gli item sono stati nuovamente sottoposti ad una analisi relativa alla identificazione della difficoltà rispetto alla popolazione degli allievi, sulla base di questi valori sono state successivamente svolte le analisi. Nel periodo estivo sono stati elaborati tre tipi di rapporti. Il primo tipo è stato consegnato ad ogni docente che avesse avuto una classe testata, il secondo ad ogni direttore nel cui istituto fosse stata testata una classe e il terzo ad ogni ispettore. Ogni docente ha ricevuto un rapporto relativo alla sua classe nel quale si mostrava per la classe e per ogni allievo relativamente ad ogni singolo settore il punteggio medio rispetto al circondario e all'insieme della popolazione testata. Analogamente ai direttori e agli ispettori sono stati inviati rapporti, relativamente, all'istituto e al circondario di competenza. Il direttore dell'Ufficio Scuole Comunali ha ricevuto l'insieme dei rapporti assieme ad una

sintesi gene-rale. Nel capitolo relativo alle analisi verranno letti i dati raccolti secondo tre punti di osservazione differenti: il territorio, le classi e gli allievi.

Chapter 3

Analisi dei risultati delle prove

Per organizzare i contenuti del rapporto si procederà con una logica a imbuto. Inizialmente saranno presentate le analisi più generali poi si scenderà progressivamente nel dettaglio. Da un punto di vista di informazioni rispetto al sistema dopo quelle generali saranno considerati i circondari, quindi alcune dimensioni territoriali e poi le classi. Nel capitolo successivo saranno approfondite le dimensioni relative alle caratteristiche degli allievi.

Il livello di circondario è stato scelto anche per la funzione di vigilanza e di organizzazione che viene svolta dagli ispettorati e dal collegio degli ispettori. Il Cantone infatti ha compiti di istituzione e direzione della scuola svolti in collaborazione con i comuni. Il collegio degli ispettori e l'ufficio scuole comunali (USCO) sono, per le scuole comunali, gli organi di raccordo tra le scuole del territorio e il DECS (Legge della scuola, 1990).

```
joined <- read.csv2("Data/joined.csv") # join tramite concatena.. di modo da avere solo dati com
```

3.1 Informazioni generali

Per avere una visione di insieme sono stati inizialmente analizzati i punteggi ottenuti nei singoli traguardi e la loro media. Questo permette un rapido raffronto tra le differenti parti che compongono la prova. Il punteggio nel traguardo è stato calcolato definendo per ogni item la difficoltà sul traguardo specifico e quindi sommandone i valori. Al fine di ottenere un valore che tenesse conto della difficoltà dei singoli item rispetto alla prova nel suo insieme è stato quindi fatto il medesimo calcolo tenendo però conto non più solo dei singoli traguardi ma della difficoltà complessiva questo valore è stato definito “Matematica generale”.

```
joined %>% select(Gagi,GEO_CA:Matematica_Generale) %>%
  datatable()
```

Show entries Search:

	Gagi	GEO_CA	GEO_MM	GEO_SRD	GM_EA	GM_SRD	NC_MT	Matematica_Generale
1	1101283270	62.86	31.96	35.26	46.55	88.58	47.11	50.0335894103599
2	1127450612	0.3	15.81	42.5	51.81	51.25	50.6	38.2402526358585
3	1016277651	59.29	36.38	51.25	53.13	55.21	61.76	56.5418655831486
4	970596913	18.22	16.28	42.15	0	23.53	26.29	29.0413353831264
5	805754094	37.18	25.04	35.26	27.46	54.89	53.89	41.2798558740657
6	856577273	43.76	31.49	36.15	50.13	58.56	45.68	46.9318990721707
7	1501768731	26.15	21.99	3.85	0.8	29.95	38.9	26.1515694842026
8	856577348	24.34	31.49	44.04	4.27	33.42	29.24	30.3668218372904
9	1100308229	61.22	31.34	42.5	66.07	51.25	30.71	49.1227256903479
10	959135484	48.51	21.99	33.01	37.77	53.62	44.53	40.1073540446619

Showing 1 to 10 of 2,506 entries Previous 2 3 4 5 ... 251 Next

```
joined %>% select(Gagi,GEO_CA:Matematica_Generale) %>%
  summarise(across(GEO_CA:Matematica_Generale, mean, na.rm=TRUE)) %>%
  t() %>%
  datatable()
```

Show entries Search:

	V1
GEO_CA	45.420538707103
GEO_MM	36.5836312849162
GEO_SRD	52.6319353551476
GM_EA	44.6091101356744
GM_SRD	59.4451636073424
NC_MT	50.8103112529928
Matematica_Generale	51.6836107573348

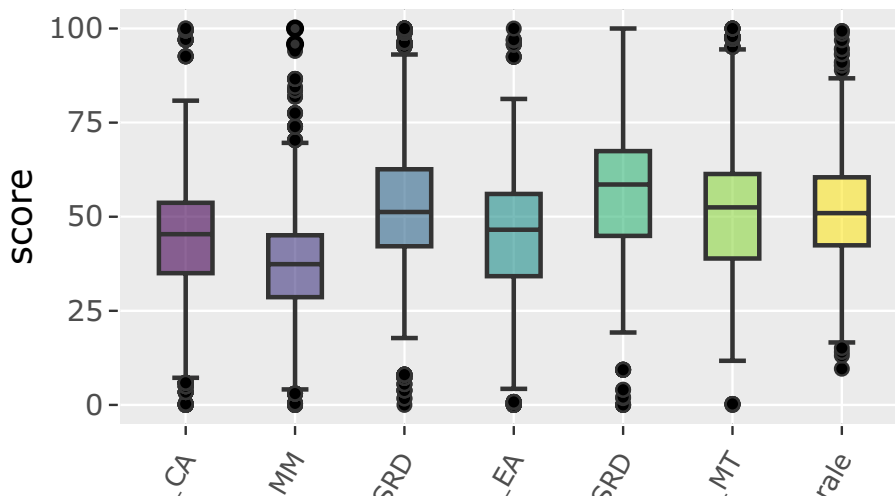
Showing 1 to 7 of 7 entries Previous Next

```
joined_long <- joined %>%
  gather(dimensione, score, GEO_CA:Matematica_Generale, factor_key = TRUE)

# plot risultati matematica
ris_medi <- joined_long %>%
  ggplot(aes(x=dimensione, y=score, fill=dimensione)) +
  geom_boxplot() +
  scale_fill_viridis(discrete = TRUE, alpha = 0.6) +
  theme(
    legend.position="none",
    plot.title = element_text(size=11)
  ) +
```

```
ggtitle("Punteggi medi degli allievi nelle diverse dimensioni") +
  xlab("") +
  theme(axis.text.x = element_text(angle = 60, vjust = 1, hjust=1, size = 8))
ggplotly(ris_medi)
```

Punteggi medi degli allievi nelle diverse dime



Il traguardo che ha raggiunto il valore medio più elevato è “Grandezze e misure Sapere, riconoscere e descrivere” mentre quello con il valore più basso è “Numeri e calcolo Matematizzare e modellizzare”. Si può notare come, all’interno dei due Ambiti di competenza scelti (Grandezze e misure e Numeri e calcolo) gli Aspetti di competenza relativi a Matematizzare e modellizzare risultino quelli appartenentemente più difficili e quelli relativi a Sapere riconoscere e descrivere risultino i due con i risultati più elevati.

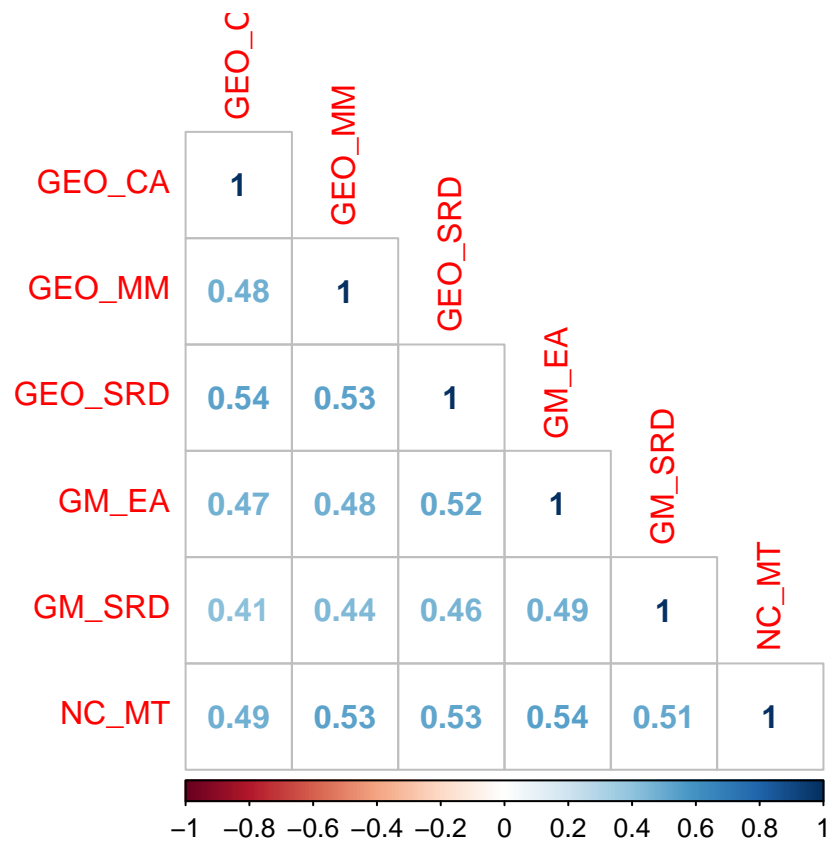
Tutti i punteggi sono stati normalizzati² in modo da assumere valori compresi tra 0 e 100. I punteggi non equivalgono però a percentuali corrispondenti al numero di esercizi svolti correttamente: ottenere 50 in un certo settore non significa infatti aver svolto correttamente il 50% degli item di quel settore. In ciascun settore gli item sono stati infatti ponderati per il rispettivo coefficiente di difficoltà; questo è stato calcolato ponderando la quantità di risposte corrette allo specifico item. In pratica, chi ha svolto correttamente gli esercizi con elevato coefficiente di difficoltà ottiene un punteggio superiore a chi ha svolto un uguale numero di esercizi con coefficiente di difficoltà inferiore. All’interno dei fascicoli gli item utilizzati sono stati proposti in ordine crescente di difficoltà ed è quindi probabile che chi ha risposto correttamente agli item più difficili abbia risposto correttamente anche a quelli precedenti.

Il valore “Matematica generale” è quindi quella variabile costituita dalla somma dei valori di difficoltà degli item ai quali l’allievo ha risposto poi normalizzata

rispetto a una scala da 0 a 100. Questo valore è stato calcolato per ogni allievo e può essere considerato come un indicatore sintetico della sua prestazione (Crescentini, Salvisberg, & Zanolla, 2014) alla prova. Il sistema con il quale è stato calcolato il valore di Matematica Generale è, rispetto ai precedenti rapporti, più raffinato e permette di valutare meglio la competenza generale. Non si tratta infatti della media dei valori degli altri traguardi ma è stato ricalcolato rispetto alla difficoltà di tutti gli item, il valore medio per tutta la popolazione è pari a 62.52.

3.2 Correlazioni tra le dimensioni valutate con la prova

```
joined %>% select(GEO_CA:NC_MT) %>%
  cor() %>%
  corrrplot(method="number", type = "lower")
```



3.2. CORRELAZIONI TRA LE DIMENSIONI VALUTATE CON LA PROVA15

La tabella precedente mostra come i risultati nelle diverse dimensioni siano tra loro correlati. Le correlazioni tra loro esistenti sono forti. Questo risultato non è inatteso in quanto si tratta di dimensioni parte del costrutto più generale valutato che corrisponde alla matematica in generale. Questa è una giustificazione aggiuntiva alla costruzione della dimensione Matematica generale che rappresentasse il punteggio globale del singolo allievo alla prova.

Il quadro generale delle analisi di seguito riportate rispecchia in gran parte il percorso di indagine prodotto nei rapporti precedenti (Crescentini et al., 2014) che ha riguardato le prove standardizzate di matematica e di italiano. Analogamente a quanto fatto allora, in questa sede verranno presentate le analisi descrittive generali dei punteggi ottenuti dagli allievi e le analisi di approfondimento riguardanti i circondari, le classi e gli allievi.

Chapter 4

Il contesto e la scuola

4.1 Ispettorati

I circondari sono le unità organizzative territoriali più ampie per quanto riguarda le scuole comunali. Il nostro territorio per quanto riguarda le scuole comunali è diviso in nove aree individuate da un numero crescente (attualmente i circondari sono passati da 9 a 7 a causa di una riorganizzazione del sistema, si è preferito però conservare l'organizzazione precedente per coerenza con la raccolta delle informazioni) procedendo da sud verso nord. La popolazione degli allievi che hanno partecipato al test è composta da 2982 bambini (50.8% maschi e 49.2% femmine). La ricerca ha coinvolto 193 classi, tra pubbliche (95%) e private (5%), di queste 54 sono pluriclassi. Tutti i circondari sono stati testati nel mese di maggio 2014. La distribuzione degli allievi per circondario è rappresentata nel Grafico 3.

```
joined <- read.csv2("Data/joined.csv")

library(ggplot2)
library(viridis)
library(plotly)
library(tidyverse)

table(joined$circondari)
```

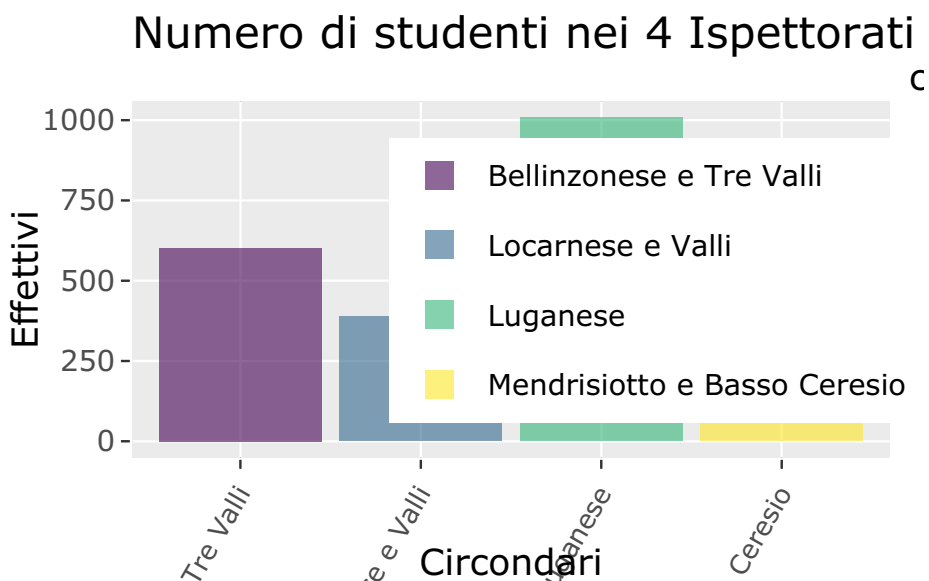
##		
##	Bellinzonese e Tre Valli	Locarnese e Valli
##	602	389
##	Luganese Mendrisiotto e Basso Ceresio	
##	1008	507

```

circPlot <- ggplot(joined, aes(circondari)) +
  geom_bar(aes(fill=circondari)) +
  scale_fill_viridis(discrete = TRUE, alpha = 0.6) +
  ylab("Effettivi") +
  xlab("Circondari") +
  ggtitle("Numero di studenti nei 4 Ispettorati") +
  theme(axis.text.x = element_text(angle = 60, vjust = 1, hjust=1, size = 8))

ggplotly(circPlot)

```



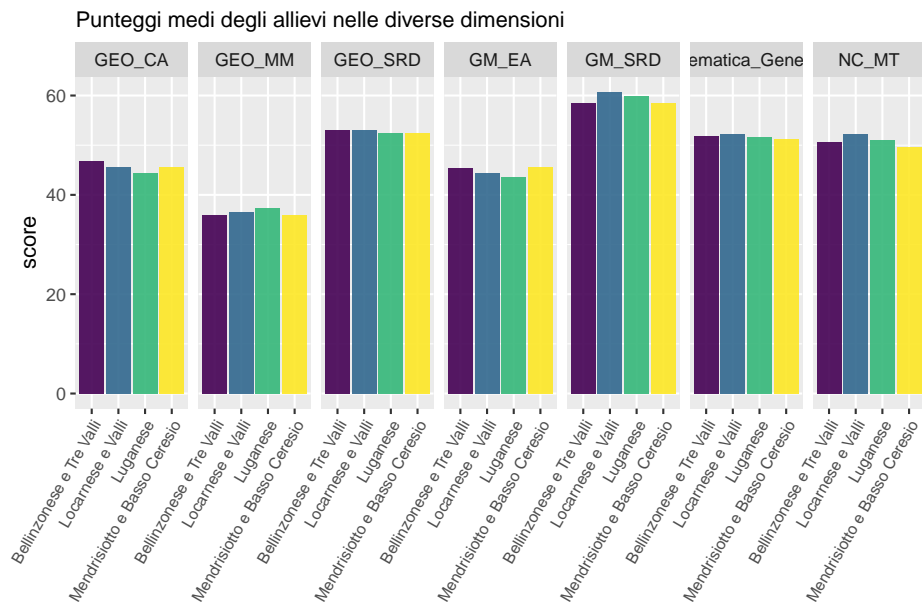
Di seguito (Tabella 2) i punteggi medi nei 9 circondari per la variabile per le diverse sottodimensioni analizzate. Le differenze tra i circondari risultano significative in tutte le dimensioni analizzando l'insieme dei dati tramite l'analisi della varianza. Non è possibile attribuire delle cause specifiche a queste differenze. Vale la pena notare come vi siano delle differenze nelle prestazioni nelle singole dimensioni non omogenee (un singolo circondario può avere una prestazione complessivamente elevata in una dimensione e modesta in un'altra).

```

library(tidyr)
library(DT)

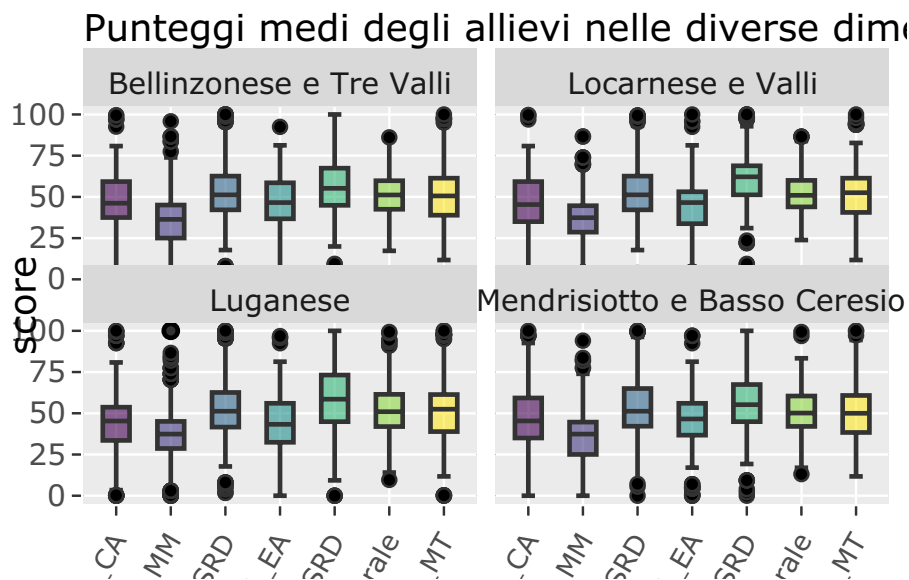
joined %>% select(circondari, GEO_CA:Matematica_Generale) %>%
  group_by(circondari) %>%
  summarise(across(GEO_CA:Matematica_Generale, mean, na.rm=TRUE)) %>%
  mutate(across(GEO_CA:Matematica_Generale, round, 3)) %>%
  datatable()

```

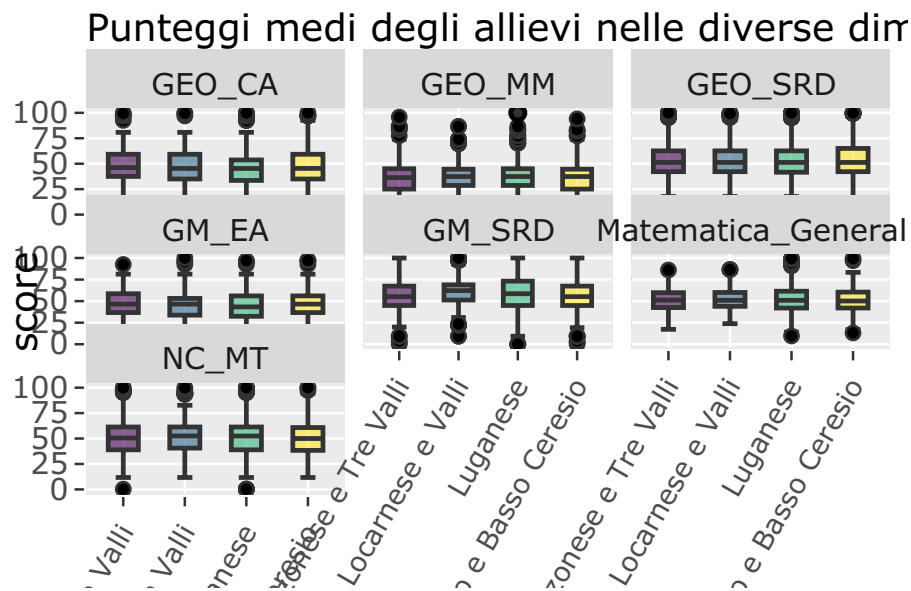



```
ris_medi.circ <- joined_long %>%
  ggplot(aes(x=Dimensioni, y=score, fill=Dimensioni)) +
  geom_boxplot() +
  facet_wrap(~circondari) +
  scale_fill_viridis(discrete = TRUE, alpha = 0.6) +
  theme(
    legend.position="none",
    plot.title = element_text(size=11)
  ) +
  ggtitle("Punteggi medi degli allievi nelle diverse dimensioni") +
  xlab("") +
  theme(axis.text.x = element_text(angle = 60, vjust = 1, hjust=1, size = 8))
ggplotly(ris_medi.circ)
```

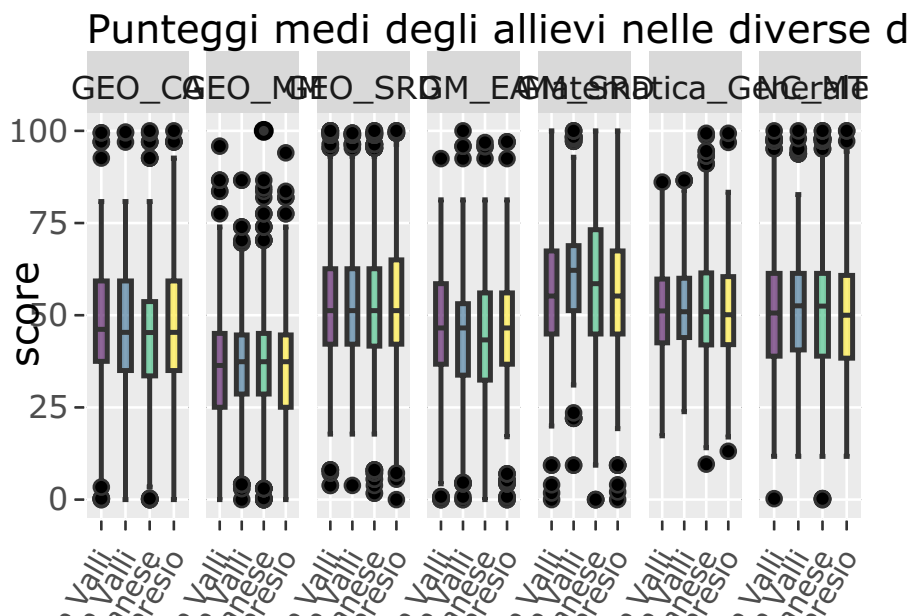
```
## Warning: `group_by()` was deprecated in dplyr 0.7.0.
## Please use `group_by()` instead.
## See vignette('programming') for more help
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```



```
ris_medi.circ2 <- joined_long %>%
  ggplot(aes(x=circondari, y=score, fill=circondari)) +
  geom_boxplot() +
  facet_wrap(~Dimensioni) +
  scale_fill_viridis(discrete = TRUE, alpha = 0.6) +
  theme(
    legend.position="none",
    plot.title = element_text(size=11)
  ) +
  ggtitle("Punteggi medi degli allievi nelle diverse dimensioni") +
  xlab("") +
  theme(axis.text.x = element_text(angle = 60, vjust = 1, hjust=1, size = 8))
ggplotly(ris_medi.circ2)
```



```
ris_medi.circ3 <- joined_long %>%
  ggplot(aes(x=circondari, y=score, fill=circondari)) +
  geom_boxplot() +
  facet_grid(~Dimensioni) +
  scale_fill_viridis(discrete = TRUE, alpha = 0.6) +
  theme(
    legend.position="none",
    plot.title = element_text(size=11)
  ) +
  ggtitle("Punteggi medi degli allievi nelle diverse dimensioni") +
  xlab("") +
  theme(axis.text.x = element_text(angle = 60, vjust = 1, hjust=1, size = 8))
ggplotly(ris_medi.circ3)
```



```
library(apaTables)
isp_aov <- aov(joined$Matematica_Generale ~ joined$circondari, data=joined)

apa.1way.table(iv = circondari, dv = Matematica_Generale, data=joined)

##
##
## Descriptive statistics for Matematica_Generale as a function of circondari.
##
##           circondari      M      SD
## Bellinzonese e Tre Valli 51.75 13.15
##   Locarnese e Valli      52.31 12.24
##   Luganese              51.62 14.26
## Mendrisiotto e Basso Ceresio 51.24 13.57
##
## Note. M and SD represent mean and standard deviation, respectively.
##

apa.aov.table(isp_aov)

##
##
## ANOVA results using joined$Matematica_Generale as the dependent variable
##
```

```
##
##          Predictor          SS    df          MS          F    p partial_eta2
##      (Intercept) 1612359.68     1 1612359.68 8768.75 .000
##   joined$circondari    256.74     3     85.58    0.47 .706          .00
##              Error 460056.67 2502     183.88
##   CI_90_partial_eta2
##
##          [.00, .00]
##
##
## Note: Values in square brackets indicate the bounds of the 90% confidence interval
```

Dalle analisi di varianza effettuate emerge come non vi siano differenze statisticamente significative rispetto ai risultati in matematica generale tra i 4 ispettorati del territorio ticinese.

Chapter 5

Applications

Some *significant* applications are demonstrated in this chapter.

5.1 Example one

5.2 Example two

Chapter 6

Final Words

We have finished a nice book.