

TRACCE PROVA SCRITTA

PROVA 1

STATISTICA DESCRITTIVA/INFERENZIALE

1) La moda:

- è il valore che si manifesta con maggior frequenza in una serie di osservazioni
- è la sommatoria di tutti i valori diviso il numero delle osservazioni
- corrisponde al 25° percentile
- è una misura di tendenza centrale

2) Quando si utilizza il test del chi quadrato:

- per confrontare due medie
- per confrontare due proporzioni
- per calcolare i limiti di confidenza di una media
- per calcolare l'errore standard di una media
- nessuna delle altre risposte è corretta

3) Supponiamo che in una partita di calcio la probabilità che vinca la squadra di casa sia 0.5 e la probabilità che vinca la squadra ospite sia 0.2. Qual è la probabilità di pareggio?

STATISTICA METODOLOGICA/INFERENZIALE

Il/la candidato illustri in generale in quali casi può essere utile una Cluster Analysis e ne descriva una possibile tecnica (algoritmo top down o bottom up) applicata ad una situazione concreta a sua scelta (se possibile riferita al contesto universitario). Ci si focalizzi sulla problematica di origine e sui contenuti che l'analisi elaborata può apportare.

LINGUAGGI ED ELABORAZIONE DATI

Si consideri il seguente estratto della tabella *STUDENTI* con owner *CARRIERA*, in cui CDS identifica il corso di studi di iscrizione, TIPOLOGIA_LAUREA distingue fra lauree triennali (LT) e magistrali (LM), ID è l'identificativo della persona, mentre MATRICOLA identifica la carriera.

ID	MATRICOLA	COGNOME	NOME	TIPOLOGIA_LAUREA	CdS
1	697352	Rossi	Mario	LT	C1
2	967362	Rossi	Maria	LM	C2
2	697326	Rossi	Maria	LT	C1
3	693752	Rossi	Marta	LT	C1
1	967352	Rossi	Mario	LM	C6
4	697353	Rossi	Marzia	LT	C1
2	697326	Rossi	Maria	LT	C3
1	697352	Rossi	Mario	LT	C4
4	697353	Rossi	Marzia	LT	C3
4	967353	Rossi	Marzia	LM	C2
3	963752	Rossi	Marta	LM	C5
1	967352	Rossi	Mario	LM	C5

Tabella 1: Studenti

Nel sistema è presente anche la tabella *ESAMI* che contiene tutti gli esami sostenuti dagli studenti ed ha la struttura riportata in Tabella 2. Anche in questo caso l'immagine non è esaustiva e riporta solo le prime righe della tabella. Lo stato dell'esame può essere Superato (S) o Rimandato (R).

MATRICOLA	CdS	CODICE_ESAME	STATO	DATA	VOTO
697352	C1	E1	R	05/02/2020	
697352	C1	E2	S	09/09/2020	30
697326	C1	E1	S	05/02/2020	28
697326	C1	E2	S	09/09/2020	29
697352	C1	E1	S	07/09/2020	23
693752	C1	E1	R	07/09/2020	

Tabella 2: Esami

1.1 Selezionare la query SQL che conta quanti studenti sono presenti in ogni tipologia di laurea e corso di studi:

SELECT CDS, TIPOLOGIA_LAUREA, COUNT(*) NR
FROM CARRIERA.STUDENTI
GROUP BY CDS, TIPOLOGIA_LAUREA
HAVING COUNT (MATRICOLA) > 1);

SELECT MATRICOLA, CDS, TIPOLOGIA_LAUREA,
COUNT(*)
FROM CARRIERA.STUDENTI
GROUP BY CDS, TIPOLOGIA_LAUREA;

SELECT CDS, TIPOLOGIA_LAUREA, COUNT(*) NR
FROM CARRIERA.STUDENTI
GROUP BY CDS, TIPOLOGIA_LAUREA;

SELECT MATRICOLA, CDS, TIPOLOGIA_LAUREA,
COUNT(*) NR
FROM CARRIERA.STUDENTI
GROUP BY CDS, TIPOLOGIA_LAUREA
HAVING COUNT (MATRICOLA) > 1);

1.2 Selezionare la query SQL per l'estrazione dei codici identificativi degli esami rimandati e delle relative matricole degli studenti che almeno una volta non hanno superato l'esame.

SELECT ES.CODICE_ESAME, ES.MATRICOLA
FROM CARRIERA.ESAMI ES
WHERE ES.STATO ='R';

SELECT CODICE_ESAME, MATRICOLA
FROM CARRIERA.ESAMI
HAVING STATO ='R';

SELECT ES1.CODICE_ESAME, ES1.MATRICOLA
FROM CARRIERA.ESAMI ES1
NOT HAVING
(SELECT 1
FROM ESAMI ES2
WHERE ES1.MATRICOLA = ES2.MATRICOLA
AND ES2.STATO ='S');

SELECT ES1.CODICE_ESAME, ES1.MATRICOLA
FROM CARRIERA.ESAMI ES1
WHERE NOT EXISTS
(SELECT 1
FROM ESAMI ES2
WHERE ES1.MATRICOLA = ES2.MATRICOLA
AND ES2.STATO ='S');

1.3 Scrivere la query SQL per l'estrazione dei codici esami, delle matricole e dei nominativi (cognome e nome) degli studenti che hanno preso 30.

LINGUAGGIO R (selezionare 1 a scelta multipla)

Si ipotizzi di ricevere l'estrazione totale delle due tabelle precedenti (Tabella 1 e Tabella 2) rispettivamente nei due file: *studenti.csv* ed *esami.csv*.

2.1 Supponendo di avere acquisito i due dataset *df_studenti* e *df_esami*, scrivere il codice R per aggiungere a *df_esami* il solo identificativo dello studente (ID).

- ```
df_esami <- merge(df_esami, df_studenti, by = c('Matricola', 'CdS')
 , all = FALSE
 , no.dups = TRUE)
```
- ```
df_esami <- merge(df_esami, df_studenti[,c('Matricola', 'CdS', 'ID')]
                  , by = c('Matricola', 'CdS')
                  , all = FALSE
                  , no.dups = TRUE)
```
- ```
df_esami <- merge(df_esami, df_studenti, by = c('Matricola')
 , all = FALSE
 , no.dups = TRUE)
```
- ```
df_esami <- merge(df_esami, df_studenti, by = c('Matricola', 'CdS')
                  , all = TRUE
                  , no.dups = TRUE)
```

2.2 Scrivere il codice R per l'importazione della tabella *df_studenti* [o eventualmente descrivere i passi svolti nell'ambiente di sviluppo].

2.3 Calcolare la media esami della matricola 697326.

PROVA 2 (prova estratta)

STATISTICA DESCRITTIVA/INFERENZIALE

- 1) Quale delle seguenti è considerata una misura di dispersione?
- media
 - deviazione standard
 - moda
 - mediana
 - 50° percentile
- 2) Se la probabilità di ammalarsi è di 0,2 in ciascuna delle due persone in osservazione, qual è la probabilità che si ammalino entrambe?
- 0,04
 - 0,064
 - 0,2
 - 0,4
 - 0,64
- 3) Il candidato ordini i seguenti eventi da quello meno probabile a quello più probabile
- A. estrarre un asso da un mazzo di 40 carte.
 - B. estrarre un multiplo di 5 nel gioco della tombola con 90 palline.
 - C. esca 10 lanciando due dadi e sommando i risultati delle due facce
 - D. escano due croci lanciando due monete

STATISTICA METODOLOGICA/INFERENZIALE

Il/la candidato/a descriva i test di significatività statistica più comuni a sua conoscenza e illustri in quali casi possono essere utilizzati anche con esempi concreti a sua scelta.

LINGUAGGI ED ELABORAZIONE DATI

Si consideri il seguente estratto della tabella *STUDENTI* con owner *CARRIERA*, in cui CDS identifica il corso di studi di iscrizione, TIPOLOGIA_LAUREA distingue fra lauree triennali (LT) e magistrali (LM), ID è l'identificativo della persona, mentre MATRICOLA identifica la carriera.

ID	MATRICOLA	COGNOME	NOME	TIPOLOGIA_LAUREA	CDS
1	697352	Rossi	Mario	LT	C1
2	967362	Rossi	Maria	LM	C2
2	697326	Rossi	Maria	LT	C1
3	693752	Rossi	Marta	LT	C1
1	967352	Rossi	Mario	LM	C6
4	697353	Rossi	Marzia	LT	C1
2	697326	Rossi	Maria	LT	C3
1	697352	Rossi	Mario	LT	C4
4	697353	Rossi	Marzia	LT	C3
4	967353	Rossi	Marzia	LM	C2
3	963752	Rossi	Marta	LM	C5
1	967352	Rossi	Mario	LM	C5

Tabella 1: Studenti

Nel sistema è presente anche la tabella *ESAMI* che contiene tutti gli esami sostenuti dagli studenti ed ha la struttura riportata in Tabella 2. Anche in questo caso l'immagine non è esaustiva e riporta solo le prime righe della tabella. Lo stato dell'esame può essere Superato (S) o Rimandato (R).

MATRICOLA	CDS	CODICE_ESAME	STATO	DATA	VOTO
697352	C1	E1	R	05/02/2020	
697352	C1	E2	S	09/09/2020	30
697326	C1	E1	S	05/02/2020	28
697326	C1	E2	S	09/09/2020	29
697352	C1	E1	S	07/09/2020	23
693752	C1	E1	R	07/09/2020	

Tabella 2: Esami

1.1 Selezionare la query SQL per rimuovere tutti gli studenti che sono iscritti a più corsi di laurea.

DROP FROM STUDENTI
WHERE MATRICOLA IN (SELECT MATRICOLA
FROM CARRIERA.STUDENTI
GROUP BY MATRICOLA
HAVING COUNT (MATRICOLA) > 1);

DELETE FROM STUDENTI
WHERE MATRICOLA IN (SELECT MATRICOLA
FROM CARRIERA.STUDENTI
GROUP BY MATRICOLA
HAVING COUNT (MATRICOLA) > 1);

DELETE FROM STUDENTI
WHERE ID IN (SELECT ID
FROM CARRIERA.STUDENTI
GROUP BY ID
HAVING COUNT (ID) > 1);

DROP FROM STUDENTI
WHERE ID IN (SELECT ID
FROM CARRIERA.STUDENTI
GROUP BY ID
HAVING COUNT (ID) > 1);

1.2 Selezionare la query SQL per l'estrazione dei codici esami e delle matricole degli studenti che hanno superato l'esame E1 il 05/02/2020.

SELECT CODICE_ESAME, MATRICOLA
FROM CARRIERA.ESAMI
WHERE 1=1
AND DATA = TO_DATE('05/02/2020', 'DD/MM/YYYY');

SELECT CODICE_ESAME, MATRICOLA
FROM CARRIERA.ESAMI
WHERE 1=1
AND DATA = DATE('05/02/2020', 'DD/MM/YYYY');

SELECT CODICE_ESAME, MATRICOLA
FROM CARRIERA.ESAMI
WHERE 1=1
AND CODICE_ESAME='E1'
AND DATA = TO_DATE('05/02/2020', 'DD/MM/YYYY');

SELECT CODICE_ESAME, MATRICOLA
FROM CARRIERA.ESAMI
WHERE 1=1
AND CODICE_ESAME='E1'
AND DATA = DATE('05/02/2020', 'DD/MM/YYYY');

1.3 Scrivere la query SQL per l'estrazione dei codici esami, delle matricole e dei nominativi (cognome e nome) degli studenti che hanno svolto l'esame E2.

LINGUAGGIO R

Si ipotizzi di ricevere l'estrazione totale delle due tabelle precedenti (Tabella 1 e Tabella 2) rispettivamente nei due file: *studenti.csv* ed *esami.csv*.

2.1 Supponendo di avere acquisito i due dataset *df_studenti* e *df_esami*, scrivere il codice R per aggiungere a *df_esami* solo il cognome e il nome dello studente corrispondente.

- ```
df_esam <- merge(df_esami, df_studenti, by = c('Matricola')
 , all = FALSE
 , no.dups = TRUE)
```
- ```
df_esami <- merge(df_esami, df_studenti[,c('Matricola', 'CdS', 'Cognome', 'Nome')]
                 , by = c('Matricola', 'CdS')
                 , all = FALSE
                 , no.dups = TRUE)
```
- ```
df_esami <- merge(df_esami, df_studenti, by = c('Matricola', 'CdS')
 , all = FALSE
 , no.dups = TRUE)
```
- ```
df_esami <- merge(df_esami, df_studenti, by = c('Matricola', 'CdS')
                 , all = TRUE
                 , no.dups = TRUE)
```

2.2 Scrivere il codice R per l'importazione della tabella *df_esami* [o eventualmente descrivere i passi svolti nell'ambiente di sviluppo].

2.3 Contare il numero di ripetenti per tutti gli appelli dell'esame E1.

PROVA 3

STATISTICA DESCRITTIVA/INFERENZIALE

1) La piramide delle età si utilizza per:

- calcolare i tassi di mortalità per fasce di età
- calcolare i tassi di mortalità per sesso
- calcolare i tassi di mortalità per sesso ed età
- calcolare i tassi di incidenza delle malattie per sesso ed età
- nessuna delle risposte è corretta

2) Se l'incidenza negli esposti ad una qualunque tipologia di rischio è 0,4 e quella nei non esposti è 0,1:

- il rischio assoluto è 1
- il rischio attribuibile è 0,1
- il rischio relativo è 0,4
- il rischio assoluto è 0,1
- il rischio relativo è 4

3) La probabilità di passare l'esame scritto di Statistica è il 50%. La probabilità di passare l'esame orale dato che si è superato lo scritto è il 98%. Il candidato calcoli qual è la probabilità di passare l'esame (è sufficiente scrivere l'impostazione della formula senza il risultato numerico finale).

STATISTICA METODOLOGICA/INFERENZIALE

Il/la candidato/a illustri in generale in quali casi può essere utile una Analisi della varianza e ne descriva una possibile applicazione ad una situazione concreta a sua scelta. Ci si focalizzi sulla problematica di origine e sui contenuti che l'analisi elaborata può apportare.

LINGUAGGI ED ELABORAZIONE DATI

Si consideri il seguente estratto della tabella *STUDENTI* con owner *CARRIERA*, in cui Cds identifica il corso di studi di iscrizione, TIPOLOGIA_LAUREA distingue fra lauree triennali (LT) e magistrali (LM), ID è l'identificativo della persona, mentre MATRICOLA identifica la carriera.

ID	MATRICOLA	COGNOME	NOME	TIPOLOGIA_LAUREA	CdS
1	697352	Rossi	Mario	LT	C1
2	967362	Rossi	Maria	LM	C2
2	697326	Rossi	Maria	LT	C1
3	693752	Rossi	Marta	LT	C1
1	967352	Rossi	Mario	LM	C6
4	697353	Rossi	Marzia	LT	C1
2	697326	Rossi	Maria	LT	C3
1	697352	Rossi	Mario	LT	C4
4	697353	Rossi	Marzia	LT	C3
4	967353	Rossi	Marzia	LM	C2
3	963752	Rossi	Marta	LM	C5
1	967352	Rossi	Mario	LM	C5

Tabella 1: Studenti

Nel sistema è presente anche la tabella *ESAMI* che contiene tutti gli esami sostenuti dagli studenti ed ha la struttura riportata in Tabella 2. Anche in questo caso l'immagine non è esaustiva e riporta solo le prime righe della tabella. Lo stato dell'esame può essere Superato (S) o Rimandato (R).

MATRICOLA	CdS	CODICE_ESAME	STATO	DATA	VOTO
697352	C1	E1	R	05/02/2020	
697352	C1	E2	S	09/09/2020	30
697326	C1	E1	S	05/02/2020	28
697326	C1	E2	S	09/09/2020	29
697352	C1	E1	S	07/09/2020	23
693752	C1	E1	R	07/09/2020	

Tabella 2: Esami

1.1 Selezionare la query SQL che estrae tutte le matricole degli studenti di laurea magistrale che hanno almeno una carriera triennale.

SELECT S1.MATRICOLA
FROM CARRIERA.STUDENTI S1
WHERE S1.TIPOLOGIA_LAUREA = 'LM'
AND EXISTS (
SELECT NULL
FROM STUDENTI S2
WHERE S1.ID = S2.ID
AND S2.TIPOLOGIA_LAUREA = 'LT');

SELECT S1.MATRICOLA
FROM CARRIERA.STUDENTI S1
WHERE S1.TIPOLOGIA_LAUREA = 'LM'
AND EXISTS (
SELECT 1
FROM STUDENTI S2
WHERE S1.MATRICOLA = S2.MATRICOLA
AND S2.TIPOLOGIA_LAUREA = 'LT');

SELECT S1.MATRICOLA
FROM CARRIERA.STUDENTI S1
WHERE EXISTS (
SELECT 1
FROM STUDENTI S2
WHERE S1.ID = S2.ID
AND S2.TIPOLOGIA_LAUREA = 'LT');

SELECT S1.MATRICOLA
FROM CARRIERA.STUDENTI S1
GROUP BY (S1.ID)
HAVING COUNT(S1.ID)>1;

1.2 Selezionare la query SQL per l'estrazione del codice esame e delle matricole degli studenti che hanno superato l'esame E2 con un voto superiore a 28.

SELECT E1.CODICE_ESAME, E1.MATRICOLA
FROM ESAMI E1
WHERE EXISTS (SELECT 1
FROM ESAMI E2
WHERE E1.MATRICOLA = E2.MATRICOLA
AND E2.VOTO > 28);

SELECT 'E2' CODICE_ESAME, E2.MATRICOLA
FROM ESAMI E2
WHERE EXISTS (SELECT 1
FROM ESAMI E1
WHERE E1.MATRICOLA = E2.MATRICOLA
AND E1.VOTO > 28);

SELECT 'E2' CODICE_ESAME, MATRICOLA
FROM ESAMI
WHERE VOTO > 28;

SELECT CODICE_ESAME, MATRICOLA
FROM ESAMI
WHERE VOTO > 28
AND CODICE_ESAME = 'E2';

1.3 Scrivere la query SQL per l'estrazione dei codici esami, delle matricole e dei nominativi (cognome e nome) degli studenti che sono stati rimandati all'esame E1.

Si ipotizzi di ricevere l'estrazione totale delle due tabelle precedenti (Tabella 1 e Tabella 2) rispettivamente nei due file: *studenti.csv* ed *esami.csv*.

2.1 Supponendo di avere acquisito i due dataset *df_studenti* e *df_esami*, scrivere il codice R per aggiungere a *df_esami* solamente la tipologia di laurea.

- ```
df_esami <- merge(df_esami, df_studenti, by = c('Matricola')
 , all = FALSE
 , no.dups = TRUE)
```
- ```
df_esami <- merge(df_esami, df_studenti, by = c('Matricola', 'CdS')  
                 , all = TRUE  
                 , no.dups = TRUE)
```
- ```
df_esami <- merge(df_esami, df_studenti, by = c('Matricola', 'CdS')
 , all = FALSE
 , no.dups = TRUE)
```
- ```
df_esami <- merge(df_esami, df_studenti[,c('Matricola', 'CdS', 'Tipologia_Laurea')]  
                 , by = c('Matricola', 'CdS')  
                 , all = FALSE  
                 , no.dups = TRUE)
```

2.2 Scrivere il codice R per contare il numero di studenti che hanno superato l'esame E2 con un voto uguale o superiore a 29

2.3 Scrivere il codice R per estrapolare il numero di studenti di ogni corso di studi.