

Traccia n. 1

Il presente testo fa riferimento ad un ambiente virtuale a cui connettersi mediante "Desktop remoto" all'indirizzo 131.114.22.7 utilizzando le credenziali sorteggiate. Nell'ambiente sono disponibili 4 file .csv 1 file .xlsx e un file SQL a cui faranno riferimento i seguenti quesiti. Si ricorda che:

- **NON** va comunicato ai membri della commissione lo username dell'utente usato per collegarsi all'ambiente
- L'ambiente è usato solo per consultazione, si possono però modificare i file al fine di esplorare le informazioni necessarie a rispondere ad alcuni quesiti
- **NON devono essere lasciate informazioni nell'account, qualora eventuali informazioni permettano di identificare l'autore il concorrente sarà escluso dalla procedura**

Quesito 1

Si consideri il contenuto del file UStat.xlsx presente nell'ambiente remoto:

- Si fornisca una breve descrizione delle entità coinvolte e delle relazioni indicando possibili chiavi nei dati
- Interagendo con i dati nel foglio "Contribuzione 2023 per aree" relativi alla contribuzione studentesca quali affermazioni si possono sostenere relativamente alla tasso media nel panorama universitario nazionale?
- Spiegare brevemente la configurazione e il funzionamento della tabella con grafico Pivot e come questi combinino i dati presenti nei fogli della cartella di lavoro

Quesito 2

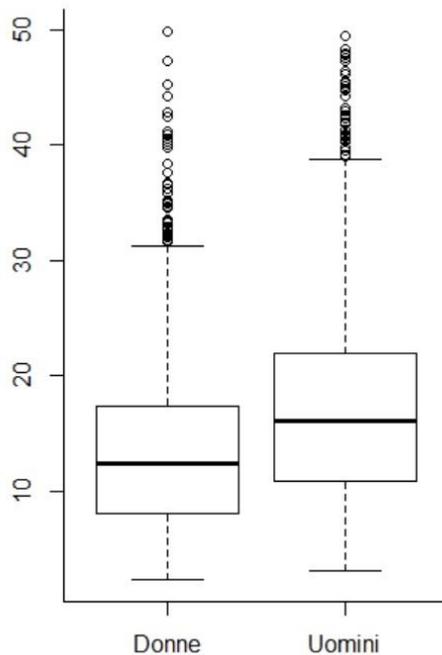
Come è collegato il file UStat.xlsx ai file .csv presenti nella cartella? Se si cambiano i dati nei file di input cosa è necessario fare per aggiornare i dati nelle tabelle Pivot? I file potrebbero essere memorizzati in rete?

Quesito 3

Come analizzeresti e quali metodologie utilizzeresti per valutare se esistono differenze significative nelle contribuzioni universitarie medie, sia per gli studenti paganti che per l'intera popolazione studentesca, tra le diverse aree geografiche d'Italia (SUD, CENTRO, NORD-EST, NORD-OVEST, ISOLE)?

Inoltre, come procederesti per un confronto specifico tra l'Università di Pisa e l'Università degli studi di Firenze?

Quesito 4



Analizza il boxplot fornito e discuti le differenze nelle distribuzioni delle retribuzioni tra uomini e donne. In particolare, considera:

- La posizione della mediana per entrambi i gruppi.
- La simmetria o asimmetria delle distribuzioni.
- La presenza e la distribuzione di eventuali outlier.
- Cosa suggerisce il boxplot riguardo alle differenze nelle retribuzioni tra i due gruppi.

Quesito 5

Considerando la query SQL nel file Query.sql dell'ambiente remoto:

- Nella query, è utilizzato un LEFT JOIN con P05_AUTOCERTIFICAZIONI. Per quale motivo si è scelto un LEFT JOIN anziché un INNER JOIN? Cosa cambierebbe nei risultati?
- Come è correlata la subquery per il campo "**entrato_in_anagrafe_ANS**" con la query principale? In che modo i risultati della subquery si basano sui dati della query principale?
- Se nella query viene aggiunta la condizione **CDS.tipo_corso_cod = 'LS'**, quale valore assumerà il campo titolo_univ_prec_flg? Spiega perché.

Quesito 6

Descrivi il processo generale di un test di ipotesi in statistica inferenziale. In particolare, illustra i seguenti concetti:

- Ipotesi nulla (H_0) e ipotesi alternativa (H_1).
- Livello di significatività (α).



- P-value.
- Errori di Tipo I e Tipo II.

Come possono essere rilevanti ai fini di analisi dati in ambito Universitario?

Traccia n. 2

Il presente testo fa riferimento ad un ambiente virtuale a cui connettersi mediante “Desktop remoto” all’indirizzo 131.114.22.7 utilizzando le credenziali sorteggiate. Nell’ambiente sono disponibili 4 file .csv 1 file .xlsx e un file SQL a cui faranno riferimento i seguenti quesiti. Si ricorda che:

- **NON** va comunicato ai membri della commissione lo username dell’utente usato per collegarsi all’ambiente
- L’ambiente è usato solo per consultazione, si possono però modificare i file al fine di esplorare le informazioni necessarie a rispondere ad alcuni quesiti
- **NON devono essere lasciate informazioni nell’account, qualora eventuali informazioni permettano di identificare l’autore il concorrente sarà escluso dalla procedura**

Quesito 1

Si consideri il contenuto del file UStat.xlsx presente nell’ambiente remoto:

- Si fornisca una breve descrizione delle entità coinvolte e delle relazioni indicando possibili chiavi nei dati
- Interagendo con i dati relativi al foglio “Analisi bilancio di genere” quali affermazioni si possono sostenere relativamente ai cambiamenti nel bilancio di genere nel quinquennio 2015-2019?
- Spiegare brevemente la configurazione e il funzionamento della tabella con grafico Pivot e come questi combinino i dati presenti nei fogli della cartella di lavoro. Quali vantaggi possono derivare da una rappresentazione tridimensionale del dato?

Quesito 2

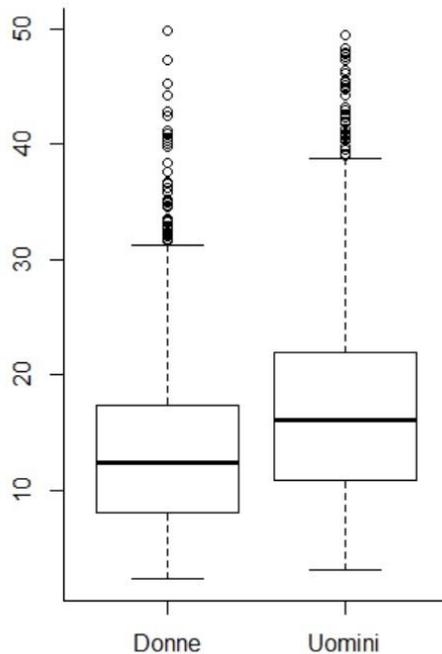
Come è collegato il file UStat.xlsx ai file .csv presenti nella cartella? Se si cambiano i dati nei file di input cosa è necessario fare per aggiornare i dati nelle tabelle Pivot? Come cambierebbe la struttura di caricamento se la sorgente dati fosse un Web service REST con bearer token?

Quesito 3

Come analizzeresti e quali metodologie utilizzeresti per valutare se esistono differenze significative nelle contribuzioni universitarie medie, sia per gli studenti paganti che per l'intera popolazione studentesca, tra le diverse regioni d'Italia?

Inoltre, come procederesti per un confronto specifico tra l'Università di Pisa e l'Università degli studi di Firenze?

Quesito 4



Descrivi i principali componenti di un boxplot e la loro interpretazione. Utilizzando il grafico fornito come esempio, spiega cosa rappresentano:

- La linea all'interno della scatola.
- I confini della scatola.
- Le estremità delle "whiskers" (i baffi).
- Gli outlier rappresentati come cerchi.

Quesito 5

Considerando la query SQL nel file Query.sql dell'ambiente remoto:

- Come cambierebbero i risultati della query se il LEFT JOIN con P05_AUTOCERTIFICAZIONI fosse sostituito da un INNER JOIN? Quali differenze ci sarebbero nei dati restituiti
- Come è correlata la subquery per il campo "data_iscr" con la query principale? In che modo i risultati della subquery si basano sui dati della query principale?
- Se nella query viene aggiunta la condizione CDS.tipo_corso_cod = 'LM', quale valore assumerà il campo titolo_univ_prec_flg? Spiega perché

Quesito 6

In un contesto in cui si conduce un test di ipotesi, spiega come deve essere interpretato il p-value. Discuti il significato pratico del p-value, in che modo esso influisce sulla decisione di rifiutare o meno l'ipotesi nulla, e quali possono essere le limitazioni dell'uso del p-value nella pratica statistica.

Come possono essere rilevanti ai fini di analisi dati in ambito Universitario?

Traccia n. 3

Il presente testo fa riferimento ad un ambiente virtuale a cui connettersi mediante "Desktop remoto" all'indirizzo 131.114.22.7 utilizzando le credenziali sorteggiate. Nell'ambiente sono disponibili 4 file .csv 1 file .xlsx e un file SQL a cui faranno riferimento i seguenti quesiti. Si ricorda che:

- **NON** va comunicato ai membri della commissione lo username dell'utente usato per collegarsi all'ambiente
- L'ambiente è usato solo per consultazione, si possono però modificare i file al fine di esplorare le informazioni necessarie a rispondere ad alcuni quesiti
- **NON devono essere lasciate informazioni nell'account, qualora eventuali informazioni permettano di identificare l'autore il concorrente sarà escluso dalla procedura**

Quesito 1

Si consideri il contenuto del file UStat.xlsx presente nell'ambiente remoto:

- Si fornisca una breve descrizione delle entità coinvolte e delle relazioni indicando possibili chiavi nei dati
- Interagendo con i dati relativi al foglio "Immatricolati per aree" quali affermazioni si possono sostenere relativamente all'evoluzione storica delle immatricolazioni? Si evidenziano effetti legati alla pandemia?
- Spiegare brevemente la configurazione e il funzionamento della tabella con grafico Pivot e come questi combinino i dati presenti nei fogli della cartella di lavoro. Come si potrebbero confrontare i dati delle Università di Pisa e di Padova?

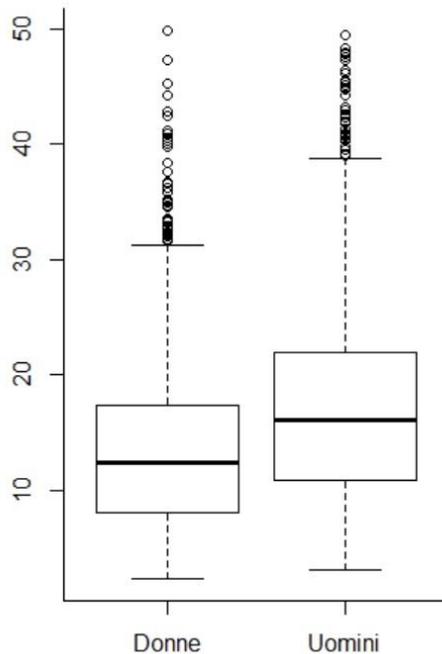
Quesito 2

Come è collegato il file UStat.xlsx ai file .csv presenti nella cartella? Se si cambiano i dati nei file di input cosa è necessario fare per aggiornare i dati nelle tabelle Pivot? Durante il processo di importazione è possibile pulire i dati? È stato fatto per qualcuno dei file?

Quesito 3

Come analizzeresti e quali metodologie utilizzeresti per valutare se esistono differenze significative nelle contribuzioni universitarie medie per l'intera popolazione studentesca tra le università statali e non statali? Inoltre, come procederesti per un confronto specifico tra l'Università di Pisa e Università telematica Pegaso di Napoli?

Quesito 4



Spiega il processo di costruzione di un boxplot a partire da un insieme di dati. Illustra i passaggi necessari per determinare:

- I quartili (Q1, mediana, Q3).
- La definizione e il calcolo degli outlier.
- Il metodo per determinare la lunghezza delle "whiskers".
- Eventuali differenze nei boxplot quando si analizzano dati simmetrici rispetto a dati asimmetrici.

Quesito 5

Considerando la query SQL nel file Query.sql dell'ambiente remoto:

- Se la condizione di LEFT JOIN P05_AUTOCERTIFICAZIONI venisse eliminata, in che modo cambierebbero i risultati della query?
- Come è correlata la subquery per il campo "in_regola_al_3112" con la query principale? In che modo i risultati della subquery si basano sui dati della query principale?
- Se nella query viene aggiunta la condizione STU.aa_imm1=2022, quale valore assumerà il campo "immatricolato"? Spiega perché

Quesito 6

In che modo si sceglie un test statistico appropriato per una data ipotesi? Spiega i principali criteri che devono essere considerati per selezionare il test statistico più adatto, quali:

- La natura delle variabili in esame.
- La distribuzione dei dati.
- La dimensione del campione.
- L'obiettivo dell'analisi.

Come possono essere rilevanti ai fini di analisi dati in ambito Universitario?