Esame di INFORMATICA E STATISTICA per le Biotecnologie

N. matricola : 985081
COGNOME - NOME: Biscuola Luca

<1> Utilizzando la funzione integrate calcolare l’interale di probabilita’ sotto la curva della distribuzione normale tra i valori 0 e 0.5, dal risulato estrarre il valore calcolato e salvarlo in una variabile x, utilizzando un’unica istruzione R. Suggerimento: leggere il manuale della funzione integrate().
<2> In un campione di 12 persone esposte ad un determinato fattore ambientale e’ stato osservato che 8 di queste si sono ammalate. Data H0 "La proporzione di persone esposte che si ammalano e’ uguale alla proporzione di persone esposte che non si ammalano", HA "La proporzione di persone esposte che si ammalano e’ maggiore rispetto alla proporzione di persone esposte che non si ammalano" e la distribuzione nulla della statistica test la cui distribuzione di probabilita’ e’ riportata di seguito (Pr[0 malati] = 0.000244, Pr[1 malato] = 0.00293, Pr[2 malati] = 0.016113, Pr[3 malati] = 0.053711, Pr[4 malati] = 0.12085, Pr[5 malati] = 0.193359, Pr[6 malati] = 0.225586, Pr[7 malati] = 0.193359, Pr[8 malati] = 0.12085, Pr[9 malati] = 0.053711, Pr[10 malati] = 0.016113, Pr[11 malati] = 0.00293, Pr[12 malati] = 0.000244), calcolare il p-value ed indicare a quale dei seguenti valori corrisponde il p-value corretto e se sia possibile rifiutare l’ipotesi nulla dato un livello di significativita’ alfa = 0.05. "A") 0.387207, rifiuto H0; "B") 0.387207, non rifiuto H0; "C") 0.193848, rifiuto H0; "D") 0.193848, non rifiuto H0.
<3> Effettuare un t test per campioni indipendenti confrontando un campione di 55 valori estratti dalla distribuzione normale con media=30 e deviazione standard 3 ed un secondo campione contenente 46 valori estratti dalla distribuzione normale con media=27 e deviazione standard=4, salvare il p value del test in una variabile x, utilizzando un’unica istruzione R. (assumere uguale la varianza nei due campioni)
<4> L’altezza delle piante di una determinata varieta’ e’ caratterizzata da un certo grado di variabilita’. Si suppone tuttavia che l’altezza media delle piante di tale varieta’ sia di 40.81 cm. Al fine di verificare tale ipotesi sono stati raccolti dati relativi all’altezza di un campione di 7 piante: l’altezza media delle piante appartenenti al campione e’ risultata pari a 42.01 cm con deviazione standard di 1.24 cm. Applicando il test t per un campione e facendo riferimento alla tavola statistica della distribuzione t (tavola_statistica_distribuzione_t.jpg), l’evidenza derivante dai dati e’ sufficientemente forte da poter rifiutare l’ipotesi nulla (H0: "l’altezza media e’ di 40.81 cm"; HA: "l’altezza media non e’ di 40.81 cm") assumendo un livello di significativita’ alfa = 0.05? "A") no; "B") si’.
<5> OGGETTO_013_b contiene le misurazioni di una variabile riferita ad un test ematologico in due gruppi di soggetti portatori e non portatori di una mutazione genetica X. Testare l’ipotesi che la media nelle due popolazioni sia uguale e salvare il valore assoluto della differenza della stima dei valori medi nelle due popolazioni in una variabile x utilizzando un’unica istruzione R. Suggerimento: indagate la struttura dell’oggetto restituito dal test statistico prima di costruire la soluzione da sottomettere.
<6> Il test t per un campione e’ stato applicato al fine di verificare se il valor medio di emoglobina in portatori di una mutazione genetica sia di 16 g/dl (H0: "il valor medio di emoglobina nei portatori della mutazione e’ di 16 g/dl"; HA: "il valor medio di emoglobina nei portatori della mutazione non e’ di 16 g/dl"). Basandosi sul p-value ottenuto (p-value = 0.002), se assumessi un livello di significativita’ alfa = 0.01 incorrerei in errore nel prendere la decisione riguardo H0 sapendo che il valor medio di emoglobina nei portatori della mutazione non e’ di 16 g/dl (realta’: H0 falsa)? "A") Si’; "B") No.
<7> OGGETTO_014_b contiene i conteggi degli individui con una determinata allergia in due gruppi. Effettuare un test di Fisher per comparare la proporzione di individui allergigi nei due gruppi usando un livello di confidenza do 90% e salvare in una variabile x l’estremo superiore dell’intervallo di confidenza calcolato. Suggerimento: indagate la struttura dell’oggetto restituito dalla funzione che realizza il test statistico.
<8> Quale tra i valori di odds ratio stimati su dati raccolti nel contesto di quattro studi sperimentali indipendenti (studio 1: OR = 2.17, studio 2: OR = 1.84, studio 3: OR = 1.02, studio 4: OR = 0.99) indicherebbe evidenza piu’ forte in merito all’efficacia di una terapia innovativa (valori variabile terapia: innovativa, standard) sulla guarigione da una determinata patologia (valori variabile guarigione: guarito, non guarito), considerando come successo l’evento "guarito" (gruppo di trattamento mediante terapia innovativa rispetto al gruppo di trattamento mediante terapia standard)? "A") OR = 2.17 (studio 1); "B") OR = 1.84 (studio 2); "C") OR = 1.02 (studio 3); "D") OR = 0.99 (studio 4).