

**Analisi statistiche:** Dopo aver importato nel dataframe **totale** i dati “**inquinamento.txt**” relativi a 1000 giornate estive in cui si sono registrati il quantitativo di **plastica** (in kg) recuperata in mare e il numero di **pesci** ritrovati morti per aver ingerito materiale plastico, svolgere i seguenti punti:

1. estrarre da **totale** un **campione** di 31 giornate con relativi valori di **plastica** e **pesci**. Con i dati di **campione** calcolare per **pesci** I e III quartile, mediana, media aritmetica e varianza campionaria della distribuzione unitaria, per **plastica** gli stessi indici ma per una distribuzione nelle classi:  
[4 , 9] , [9 , 12] , [12 , 18] ;
2. ipotizzando per la variabile **plastica** una distribuzione normale mentre per **pesci** una di Poisson, stimare i rispettivi parametri con il metodo dei momenti basandosi sui valori presenti in **campione** e tramite essi calcolare le probabilità di osservare una giornata con raccolta plastica superiore al 95° percentile di **totale\$plastica** e una con numero di pesci trovati morti per essa inferiore del 5° percentile di **totale\$pesci**;
3. utilizzando i dati per **plastica** in **campione**, condurre un test di adattamento delle frequenze registrate nella distribuzione in classi determinata al punto 1 con la distribuzione normale stimata al punto precedente;
4. Condurre un test d'indipendenza, con significatività del 5%, tra **plastica** e **pesci** usando i dati in **campione**;
5. Usando i dati di **campione\$pesci** e ipotizzando la loro provenienza da una distribuzione normale, condurre un test, con significatività dell' 1%, sull'ipotesi alternativa che la varianza della popolazione sia diverso dal valore  $\sigma_0=4.19$ ;
6. Prevedere, attraverso un modello di regressione lineare semplice basato su **campione**, il numero di pesci trovati intossicati in una giornata in cui si sono raccolti 15 Kg di plastica e commentare se il modello lineare sia buono o meno.

**Problema di probabilità:**

7. Se il tasso medio di accessi a un certo sito web è di 5 per minuto, si calcoli:
  - la probabilità che nei prossimi 3 minuti ci siano esattamente 15 accessi;
  - la probabilità che in mezz'ora non si superino i 100 accessi;
  - il numero massimo di accessi in un ora che si registrano nel 95% dei casi.

**Domande teoriche:**

8. L'indice chi-quadro d'indipendenza è più vicino a 0 quanto più le coppie di dati sono vicine dall'essere perfettamente dipendenti ☐ V ☐ F, perché...
9. L'indice di bontà d'adattamento  $R^2$  può assumere valori tra [-1, 1] e in particolare -1 quando c'è perfetta dipendenza lineare negativa e 1 quando c'è perfetta dipendenza lineare positiva ☐ V ☐ F, perché...
10. Un evento di probabilità nulla è certamente impossibile ☐ V ☐ F, perché...
11. In un test d'ipotesi, il campione si dice significativo quando porta a confermare l'ipotesi nulla  $H_0$  ☐ V ☐ F, perché...
12. Lo stimatore è un numero aleatorio mentre la stima un valore numerico ☐ V ☐ F, perché...

**ATTENZIONE: VISUALIZZARE a console E COMMENTARE I RISULTATI**  
(# dopo i comandi )

SALVARE I DATI DI **campione** NEL FILE **campione\_tuocognome.txt**  
TRAMITE COMANDO  
**write.table(campione, "campione\_tuocognome .txt", row.names=FALSE)**

SALVARE IL FILE COMANDI COME **tuocognome.R**  
e consegnare i **DUE** file tramite unistudium.