

PRÀCTICA 1

Exercici exploratori

1. OBJECTIUS

L'objectiu d'aquesta pràctica és consolidar les eines que hem après durant les primeres setmanes de curs. A més se us demana que personalitzeu el vostre gràfic (exemples: títol, nom dels eixos, llegendes, diferents *scales* i/o ajustaments de posició) fent us d'alguna de les comandes que se us han anat introduint al llarg dels seminaris i resumit a les transparències.

2. PROCÈS i AVALUACIÓ

Per fer aquests exercicis:

1. primer heu de familiaritzar-vos amb les dades,
2. després heu de pensar què voleu mostrar i quin gràfic és més adient segons el tipus de dades que teniu.
3. En cas de necessitar-ho apliqueu comandes de *data massage*.
4. Finalment mostreu la gràfica, personalitzeu-la i vam extraieu-ne conclusions.

- **QUÈ SE US DEMANA?:** Se us demanen uns exercicis a fer en grups de quatre-sis persones. **Penseu cada exercici com una comanda (tasca) que us fa un client.** El client vol posar aquestes gràfiques en una revista, per tant li **interessa: i) Que cada gràfica representi perfectament el que us està demanant, ii) contingui un peu de figura on es descrigui el que es veu en la gràfica, iii) contingui una explicació resumida de les conclusions/informació que us dona la gràfica, iv) que cada gràfica estigui ben editada i tingui un toc personal dels creadors/es.** Això és el que us **pagarà** el vostre client.

- **COM S'AVALUA?:** En aquest cas fictici (si entregueu la tasca a temps), se us avaluarà la tasca completa amb una nota de la part grupal (**PG**) entre 0-0,75 punts. A més, *el dia de l'examen parcial hi haurà un exercici referit a aquesta mateixa tasca on podreu treure fins 0,25 punts de la nota referent a la part individual (PI).* La puntuació que traieu en total (**Pract=PG+PI**) se us sumarà al primer examen parcial de forma extra (sempre i quan traieu en aquest 4 o més punts). En cap cas es guardarà pels exàmens de recuperació

En el cas de la nota de la part grupal (PG), el valor dels exercicis està ponderat com s'especifica en l'enunciat corresponent (veure punt 4 - 'ENUNCIATS').

Recordeu el que us demana el vostre client per obtenir la màxima nota.

La tasca (com diu en la guia docent) es compta que es pot fer en menys de 4h, 2h de les quals són tutelades a l'aula. Inclús si es treballa eficientment en grup, 2h haurien de ser suficients. Tot i així, es podrà pujar en un fitxer format .pdf (en el

grup creat durant aquesta sessió) fins diumenge dia 23 de març a les 23h59 (fora d'aquest límit horari no serà puntuada).

- COM HO HEU D'ENTREGAR?: El fitxer .pdf que pugeu, ha de contenir:

1. Una primera plana amb un títol i autors.
2. Tot i que no cal re-copiar els enunciats, s'haurà d'especificar de manera clara la part de la pràctica (Part 1/Part 2), el número i l'apartat d'exercici a la que correspon la resposta que s'està donant. Per cada exercici, s'inclouran els gràfics i/o respostes (seguint els criteris descrits a dalt).
3. A més, al final del treball s'inclourà en apèndix el codi. Per això es recomana que completeu el fitxer PRT1.R que se us proporciona, i que l'enganxeu en l'apèndix un cop completat.

Nota: No respectar aquest format tindrà una penalització en la nota PG.

La tasca se us retornarà comentada i puntuada.

3. DADES

En aquesta pràctica utilitzarem un nou conjunt de dades d'un [estudi](#) de [l'Institut de mètriques i avaluació de la Salut de US](#), prèviament processades. Aquestes dades estimen els patrons de consum d'alcohol mesurats per cada comtat de US durant els anys 2002-2012. Més concretament, per a cada comtat, **les dades representen la prevalença**:

$$Prevalença = \frac{\text{bevedors}}{\text{població adulta}}$$

Les dues mesures d'interès són:

1. *Any drinking* (qualsevol beguda): El consum d'almenys una beguda alcohòlica en els darrers 30 dies
2. *Binge drinking* (consum excessiu d'alcohol): El consum excessiu en una ocasió única com a mínim una vegada en els darrers 30 dies (El consum excessiu en US correspon al consum de més de quatre begudes per a dones o cinc begudes per a homes -dades relacionades amb el pes d'ambdós gèneres)

NOTES:

- ❖ Trobareu tres fitxers per començar:
 - PRT1.R on heu d'escriure el script corresponent a cada part que enganxareu després en l'apèndix com s'indica en el punt 3 de l'apartat 2 descrit més amunt.
 - Trobareu **dos fitxers .csv separats, un per 'any drinking' i l'altre per 'binge drinking'**. Ambdós fitxers tenen el mateix format (tenen 3179 observacions i 35 variables).
 - Si inspeccioneu els fitxers de dades amb les comandes que hem anat veient durant els seminaris, haureu de tenir en compte que inclouen estimacions al comtat, a l'estat, i a nivells nacionals tots barrejats en un sol fitxer.

Procediu a la seva lectura i assigneu les dades a una variable.

Si voleu tenir una millor visió de les dades veieu la visualització interactiva “alcohol use” en el següent [enllaç](#).

4. ENUNCIATS

- **PART 1 (70% PG) - Any drinking in 2012**

1. Extraieu el dataframe amb columnes ‘estat (State)’ i ‘ubicació (location)’ que contingui tota la informació referent a l’any 2012. Quantes observacions i variables té el nou dataframe? Feu una captura de pantalla de les primeres línies i les columnes corresponents i incloeu-la amb el seu peu de taula (descripció) corresponent. (10% PG).
2. Afegiu una columna anomenada ‘DIF_SEXE_2012’ al vostre dataframe creat en l’exercici anterior, de manera que contingui el *valor de la diferència* entre la prevalença corresponent al gènere masculí i la prevalença corresponent al gènere femení del 2012. Assigneu al nou dataframe un nom perquè pugueu accedir-hi més endavant. El dataframe final NO ha de contenir valors NA’s. Responen:
 - i) Quantes observacions i variables té el nou dataframe? Feu una captura de pantalla de les primeres línies i les columnes corresponents i incloeu-la amb el seu peu de figura corresponent.
NOTA: Per obtenir la màxima nota d’aquest apartat utilitzeu pipes combinant l’exercici 1 i 2 (12% PG).
 - ii) Dibuixeu la distribució d’aquesta nova variable ‘DIF_SEXE_2012’ (15%PG). Extraieu-ne conclusions. **NOTA:** No us hauria de sortir cap warning.
3. Ara que teniu les dades preparades i us heu familiaritzat amb la distribució de la nova variable, responeu els tres apartats següents:
 - i) Hi ha ubicacions (‘locations’) on les dones beguin més que els homes? Raoneu la vostra resposta (Exemple: Les dones bevien més que els homes en X ubicacions). (6% PG)
 - ii) Feu una gràfica que compari les diferències de la prevalença entre sexes (‘DIF_SEXE_2012’) en les respectives localitzacions. Anomeneu cinc localitzacions on els homes veuen bastant més que les dones, per exemple, amb una diferència superior a 24. (12% PG)
 - iii) Quines són les ubicacions on la prevalença dels homes i dones és molt semblant? Feu un gràfic amb el que es pugui comparar les diferències de la prevalença entre sexes en aquestes respectives ubicacions. Raoneu la vostra resposta. (15% PG). **NOTA:** Per raonar la vostra elecció de “bastant similar” podeu fer us del gràfic que heu fet en l’exercici 2.ii. i raonar com heu triat en el vostre cas el valor “bastant similar”.

- **PART 2 (30% NG) – *Binge drinking***

1. Extraieu un dataframe que només tingui observacions a nivell de les ubicacions que duen per nom: 'NOM **county**' (a nivell comtat) i assigneu-lo a una variable ('County_DATA'). *NOTA: Confirmeu que heu extret les dades correctes comprovant que el nombre d'observacions és 3002. El nombre de columnes/variables no us hauria de variar.* (8% PG).
2. Ara que les vostres dades estan configurades, podeu començar a respondre algunes preguntes sobre el vostre nou dataframe ('County_DATA'). Responen: quina va ser la prevalença màxima de consumició excessiva a nivell d'estat ('state') el 2012 per a tots dos sexes (both_sexes_2012)? Doneu els noms dels estats amb les 5 màximes més altes? (10% NG)
3. Compareu en un mateix gràfic la prevalença al llarg dels anys (comparació temporal) per "both_sexes", dels comtats 'Little River County' i 'Bacon County'. (12% PG)