

## PiE 2: Entregable 1 (Q1 23-24)

### Instruccions:

- L'entregable s'ha de entregar a través d'Atenea. La data límit és el **diumenge 22 d'octubre a les 23:59h**.
- Pels exercicis 1 i 2, hem generat 5 conjunts de dades diferents. Els arxius acaben amb un nombre  $X \in \{1, 2, 3, 4, 5\}$ . A l'arxiu `parelles.pdf` hi trobareu els conjunts de dades que heu d'analitzar, identificats pels valors de  $X$ . Per a llegir les dades amb **R**, recomanem la funció `read.csv`.
- Heu d'entregar un informe en format **pdf** que inclogui la resolució dels exercicis i el codi que heu fet servir (integreu el codi amb les vostres explicacions; no bolqueu tot el codi en un apèndix). Recomanem que feu servir Rmarkdown, però no és obligatori. Si us hi animeu, el tutorial creat per Rstudio està prou bé.
- Si teniu cap dubte, ens podeu preguntar en persona o per correu electrònic (`marta.perez@upc.edu` o `victor.pena.pizarro@upc.edu`).
- Els exercicis els podeu trobar a la segona pàgina d'aquest document. Que tingueu sort.

**Exercici 1:** Un grup d'investigadors en ciències de l'esport vol saber els efectes d'un programa d'entrenament de tres mesos en el pes mig de dones entre 18 i 25 anys. Per evitar discrepàncies grans entre individus, només consideren dones amb alçades entre 155cm i 165cm. Els investigadors van pesar les participants abans i després del programa d'entrenament. Podeu trobar les dades a l'arxiu `entrenamentX.csv`, on  $X \in \{1, 2, 3, 4, 5\}$ . Els pesos estan en kg.

- Feu un gràfic que resumeixi les dades i interpreteu-lo.
- Hi ha evidències a nivell  $\alpha = 0.05$  que el pes esperat canvia després de la rutina d'entrenament? Si és així, hi ha evidència de pèrdua o guany de pes? Contesteu la pregunta fent servir i) valors crítics, ii) intervals de confiança i iii)  $p$ -valors.
- Trobeu un interval de confiança al 95% per a la proporció poblacional de dones que perden pes amb la rutina d'entrenament. Comenteu els resultats i compareu-los amb els de l'apartat anterior.
- Hi ha evidències per a concloure (a nivell  $\alpha = 0.05$ ) que el percentatge de dones que perden pes amb el programa és més gran del 70%? Justifiqueu la resposta.

**Exercici 2:** Una farmacèutica vol provar que un nou medicament és més efectiu que el que es fa servir habitualment per a curar una malaltia. A l'arxiu `medicamentX.csv`, on  $X \in \{1, 2, 3, 4, 5\}$ , hi trobareu els resultats. Els individus que s'han recuperat de la malaltia estan marcats amb un 1 i els que no amb un 0. Contesteu les següents preguntes.

- Trobeu les proporcions mostrals de curats amb el nou medicament i el que es fa servir normalment per separat. Feu estimacions puntuals i per interval (a nivell de confiança del 95%), també per separat. Feu un gràfic per a comparar els dos grups i interpreteu-lo.
- Feu una prova d'hipòtesi a nivell  $\alpha = 0.01$  l'objectiu del qual és provar que el nou medicament és més efectiu. Escriviu amb cura les hipòtesis i expliqueu per què heu triat aquestes i no unes altres. Resoleu la prova amb i) valors crítics i ii)  $p$ -valors. Quina conclusió n'extraieu?
- Feu un interval de confiança del 99% per a la diferència poblacional entre la proporció de curats amb el nou medicament i el que es fa servir habitualment. L'interval conté el zero o no? Compareu el resultat amb el que heu trobat a l'apartat anterior.
- Si les mides mostrals dels grups són les mateixes, quina és la mida mostral mínima per assegurar que l'interval de l'apartat anterior té una llargària inferior a 0.1? Justifiqueu la resposta.

**Exercici 3:** Un grup de científics vol fer la prova d'hipòtesi per a mitjanes normals  $H_0 : \mu \leq 0$  contra  $H_1 : \mu > 0$  a nivell  $\alpha = 0.05$ . La variància poblacional és coneguda:  $\sigma^2 = 1$ . El seu objectiu és entendre com la probabilitat de rebutjar  $H_0$  quan és falsa (és a dir, la potència de la prova) depèn del veritable valor de  $\mu$  i la mida mostral  $n$ .

- Grafiqueu la probabilitat de rebutjar  $H_0$  per a  $\mu$  entre 0 i 5 pels casos  $n \in \{5, 10, 15, 50\}$ . Grafiqueu les quatre corbes en un sol gràfic. Comenteu els resultats. [Per a fer el gràfic, us pot ser útil el paquet `library(ggplot2)`.]
- Quina és la mínima mida mostral que necessitem per tal que la probabilitat de rebutjar  $H_0$  quan  $\mu = 1$  sigui com a mínim 0.8? Quina seria la resposta si, en canvi,  $\mu = 0.5$ ? Comenteu els resultats.