

Activitat 2: Anàlisi descriptiva i inferencial

Enunciat

Semestre 2019.2

Índex

1	Anàlisi descriptiva	3
1.1	Lectura del fitxer	3
1.2	Anàlisi descriptiva visual	3
1.3	Correlació	3
2	Pes mitjà en néixer	3
2.1	Interval de confiança	3
2.1.1	Càlculs	3
2.1.2	Interpretació	3
2.2	Contrast del valor mitjà amb 3.5kg	3
2.2.1	Escriviu la hipòtesi nul·la i alternativa.	4
2.2.2	Mètode	4
2.2.3	Càlculs	4
2.2.4	Conclusió	4
2.2.5	Relació amb l'interval de confiança.	4
3	Contrast de pes mitjà entre nens i nenes	4
3.1	Hipòtesi	4
3.2	Assumpció de normalitat	4
3.3	Mètode	4
3.4	Càlculs	4
3.5	Interpreteu els resultats	4
3.6	Interval de confiança	5
4	Proporció de nens i nenes	5
4.1	Hipòtesi	5
4.2	Mètode	5
4.3	Càlculs	5
4.4	Interpretació	5
5	Relació entre baix pes i mare fumadora	5
5.1	Anàlisi sobre els casos de baix pes	5
5.1.1	Hipòtesi	6
5.1.2	Càlculs	6
5.1.3	Conclusió del test	6
5.1.4	Interpretació	6
5.2	Anàlisi dels casos de baix pes (amb gestació >36)	6
5.2.1	Hipòtesi	6
5.2.2	Càlculs	6
5.2.3	Conclusió	6

En aquesta activitat usarem el conjunt de dades sobre el pes en néixer que s'ha preprocessat en l'activitat anterior. Una vegada el fitxer estigui preparat, es realitzarà una anàlisi descriptiva i inferencial sobre les dades. Us proporcionem el fitxer **BWprocessed.csv** perquè tots treballem amb el mateix fitxer de dades, independentment del resultat obtingut en l'activitat 1.

Recordem que el fitxer de dades conté el pes i grandària dels nens i nenes nascuts a Espanya durant l'any 2019. L'arxiu conté 300 registres i 11 variables.

Aquestes variables són: ID, HP, City, Time, Day, BW, BPD, AD, Sex, Ge, Sm, on:

- ID: identificador numèric.
- HP: nom de l'hospital.
- City: nom de la ciutat on s'ha produït el naixement.
- Time: hora del naixement (valor entre 0 i 24).
- Day: dia de la setmana (valor entre 1 i 7).
- Sex: sexe del nadó.
- BW: pes al naixement.
- BPD: diàmetre biparietal (en mm), determinat per ultrasons, mesurat abans de néixer.
- AD: diàmetre abdominal (en mm), determinat per ultrasons, mesurat abans de néixer.
- Ge: setmanes de gestació.
- Sm: si la mare és fumadora ('S' en cas afirmatiu, 'N' en cas negatiu).

L'anàlisi es desenvoluparà sobre la base de les preguntes de recerca següents:

1. Interval de confiança al 95% del pes en néixer
2. ¿El pes mitjà en néixer difereix significativament de 3.5kg?
3. ¿El pes en néixer de les nenes és diferent al dels nens?
4. ¿La proporció de naixements de nens és inferior als naixements de nenes?
5. ¿Existeix una relació entre baix pes en néixer i el fet que la mare sigui fumadora?

Notes importants a tenir en compte per lliurar l'activitat:

- És necessari lliurar el fitxer Rmd i el fitxer de sortida (PDF o html). El fitxer de sortida ha d'incloure: el codi i el resultat de l'execució del mateix (pas a pas).
 - S'ha de respectar la mateixa numeració dels apartats que l'enunciat.
 - No es poden realitzar llistats complets de les dades en la solució. El motiu és que es generen fitxers de sortida amb centenars de pàgines que són molt difícils de traçar i corregir. Per comprovar les funcionalitats del codi, podeu usar **head** i **tail** que només mostren unes línies del fitxer.
 - El nivell de confiança per defecte és del 95%, tret que s'indiqui un altre valor.
-

1 Anàlisi descriptiva

1.1 Lectura del fitxer

Llegir el fitxer **BWprocessed.csv**. Validar que els tipus de dades llegides són correctes. Si no és així, realitzar les conversions oportunes.

1.2 Anàlisi descriptiva visual

Representar de manera visual les variables del conjunt de dades i les distribucions dels seus valors. Trieu la representació més apropiada en cada cas.

1.3 Correlació

Estudiar visualment les possibles correlacions entre:

- les variables pes en néixer i el diàmetre biparietal
- les variables pes i diàmetre abdominal
- les variables pes i setmanes de gestació.

Interpreteu els gràfics. Per avaluar la correlació numèricament, podeu usar la funció **cor**.

2 Pes mitjà en néixer

2.1 Interval de confiança

Es desitja investigar l'interval de confiança del 95% del pes mitjà en néixer a partir de la mostra de dades. S'aconsella definir una funció IC que calculi l'interval de confiança d'una variable donada. Aquesta funció es podrà aprofitar més endavant.

Seguiu els passos que s'especifiquen a continuació.

2.1.1 Càlculs

Nota: heu de realitzar els càlculs manualment. No es poden usar funcions com a **t.test** o similars que ja calculin directament l'interval de confiança. En canvi, sí podeu fer servir funcions com **qnorm**, **pnorm**, **qt**, **pt** per conèixer els valors de la distribució normal o t-Student.

2.1.2 Interpretació

Interpreteu el resultat obtingut i el significat d'interval de confiança.

2.2 Contrast del valor mitjà amb 3.5kg

Independentment del resultat de l'apartat anterior, passem a desenvolupar el procediment per donar resposta a la pregunta següent:

¿El pes mitjà en néixer dels nens i nenes nascuts a Espanya és de 3.5kg?

Responeu a la pregunta amb un 95% de nivell de confiança. Seguiu els passos que s'indiquen a continuació.

Nota: S'ha de calcular a partir de tota la mostra. No es demana el càlcul per separat de nens i nenes, sinó de totes les dades en global.

2.2.1 Escriviu la hipòtesi nul · la i alternativa.

2.2.2 Mètode

Detalleu quin mètode usareu per respondre aquesta pregunta.

2.2.3 Càlculs

Desenvolpeu tots els càlculs. No es poden usar funcions com a **t.test** o similars que ja calculin directament l'interval de confiança. En canvi, sí podeu fer servir funcions com **qnorm**, **pnorm**, **qt**, **pt** per conèixer els valors de la distribució normal o t-Student.

2.2.4 Conclusió

A partir dels valors obtinguts sobre el valor crític, el valor observat i el valor p, responeu la pregunta plantejada.

2.2.5 Relació amb l'interval de confiança.

A partir del resultat sobre l'interval de confiança obtingut en l'apartat anterior, com hauríeu pogut respondre a aquesta mateixa pregunta?

3 Contrast de pes mitjà entre nens i nenes

Es desitja investigar si el pes en néixer de les nenes és diferent al dels nens. Per això, heu de realitzar un contrast sobre la mitjana del pes, considerant dues mostres independents. Seguiu els passos que s'especifiquen a continuació.

3.1 Hipòtesi

Escriviu la hipòtesi nul · la i la hipòtesi alternativa

3.2 Assumpció de normalitat

Justificar si es pot assumir normalitat en aquest cas.

3.3 Mètode

Explicar quin mètode usareu per al contrast a partir de l'anàlisi prèvia sobre l'assumpció de normalitat. Indiqueu també si es tracta d'un contrast unilateral o bilateral.

3.4 Càlculs

Realitzeu tots els càlculs del contrast. No es pot usar funcions del tipus **t.test**. Heu de realitzar els càlculs manualment, seguint les fórmules que s'especifiquen en els materials. Sí que podeu usar funcions del tipus **qnorm**, **pnorm**, **pt**, **qt**, les quals codifiquen els valors de les distribucions de dades normals i t-Student.

3.5 Interpreteu els resultats

Sobre la base del valor p, valor observat i valor crític obtinguts en l'apartat anterior, donar resposta a la pregunta plantejada.

3.6 Interval de confiança

Una altra manera d'avaluar si existeixen diferències del pes mitjà entre nens i nenes és a partir dels respectius intervals de confiança. Calcular aquests intervals de confiança i extreure les conclusions i com aquestes corroboren els resultats del contrast realitzat.

4 Proporció de nens i nenes

A continuació, es planteja la pregunta següent:

¿La proporció de naixements de nens és inferior a la dels naixements de nenes?

Per respondre a aquesta pregunta, seguim els passos que s'especifiquen a continuació.

4.1 Hipòtesi

Escriu la hipòtesi nul·la i la hipòtesi alternativa.

4.2 Mètode

Especificar quin mètode usareu per aquest contrast. També indiqueu si el contrast és unilateral o bilateral.

4.3 Càlculs

Realitzeu tots els càlculs manualment. Igual que en els apartats anteriors, només podeu usar les funcions **qnorm**, **pnorm**, **qt**, **pt**.

4.4 Interpretació

A partir del resultat de l'estadístic de contrast observat, el valor crític i el valor p, conclou sobre si es pot afirmar que la proporció de nens que neixen és inferior a la de les nenes.

5 Relació entre baix pes i mare fumadora

A continuació, se'ns planteja la següent pregunta:

¿Existeix relació entre baix pes en néixer i el fet que la mare sigui fumadora?

Per investigar la relació entre baix pes en néixer i ser fill de mare fumadora, es realitzen les anàlisis següents.

5.1 Anàlisi sobre els casos de baix pes

En primer lloc, s'etiqueten els casos de baix pes en néixer com aquells que són inferiors a 2.5kg. Es construeix una taula de contingència on s'avalua la relació entre baix pes en néixer i ser fill de mare fumadora. Una vegada construïda la taula, s'ha d'aplicar el test d'independència de dues variables Chi quadrat (test ChiSquare). Realitzeu tots els càlculs d'aquest test i després, interpreteu si podem observar una relació entre baix pes i ser fill de mare fumadora amb un nivell de confiança del 95%.

Nota: per calcular el test d'independència de dues variables, no podeu usar la funció **chisq.test** per resoldre l'exercici. Heu de realitzar els càlculs manualment. Només podeu usar **chisq.test** per validar els vostres

resultats. Per accedir als valors de la funció de distribució Chi Quadrat, podeu usar les funcions **dchisq**, **pchisq**, **qchisq** de la llibreria **stats**.

Podeu consultar el test d'independència de dues variables ChiSquare en textos d'estadística. Adjuntem alguns enllaços que poden ser útils:

- <https://stattrek.com/chi-square-test/independence.aspx>
- <https://www.spss-tutorials.com/chi-square-independence-test/>

Seguiu els passos que s'especifiquen a continuació.

5.1.1 Hipòtesi

Escriviu la hipòtesi nul·la i alternativa.

5.1.2 Càlculs

Realitzeu els càlculs manualment del test. Per això, es recomana construir una funció **my.chisq** que realitzi aquests càlculs i que pugueu reusar més endavant.

5.1.3 Conclusió del test

Sobre la base dels resultats del test, es pot afirmar que existeix relació entre baix pes i mare fumadora? Expliqueu sobre la base del valor p, valor observat i valor crític.

5.1.4 Interpretació

Analitzeu si existeixen altres variables que puguin influir en l'anàlisi i com afectarien les conclusions extretes del test.

5.2 Anàlisi dels casos de baix pes (amb gestació >36)

Realitzeu la mateixa anàlisi per als naixements a partir de la setmana 36. Per realitzar aquesta anàlisi, es considerarà baix pes en néixer els casos que es trobin en el primer quartil del pes, considerant només els casos amb setmanes de gestació superiors a 36.

Podem concloure que es presenta el mateix patró de relació (d'independència o dependència entre les variables) que en el cas anterior? Interpreteu els resultats.

5.2.1 Hipòtesi

Escriviu la hipòtesi nul·la i alternativa.

5.2.2 Càlculs

Realitzeu els càlculs. Podeu aprofitar la funció desenvolupada anteriorment. Mostreu els resultats dels càlculs.

5.2.3 Conclusió

Interpreteu els resultats obtinguts i responeu a la pregunta plantejada. Justifiqueu les vostres conclusions a partir del valor observat, valor crític i valor p.

6 Conclusions de l'anàlisi

Finalment, a partir de tot l'estudi realitzat, escriviu les conclusions d'aquest estudi, donant resposta ben fonamentada a les preguntes plantejades a l'inici de l'activitat.

7 Puntuació dels apartats

- Apartat 1 (15%)
- Apartat 2 (20%)
- Apartat 3 (15%)
- Apartat 4 (10%)
- Apartat 5 (20%)
- Apartat 6 (10%)
- Qualitat de l'informe dinàmic i del codi R (10%)