

Document de Treball 1

Objectiu

L'objectiu d'aquesta sessió és:

1. Repassar la diferència entre mostres emparellades i independents.
2. Repassar mostres emparellades.
3. Exemple.

1. Diferència entre mostres emparellades i independents

Material de l'aulavirtual:

Teoria	Activitats
Secció: Unitat 3. Anàlisi estadística de dues o més variables. Pdf: Tema 5. Comparació de dues mitjanes poblacionals. Pàgines: 1 a 7	Document: Exercicis unitat 3. Dades: Dades unitat 3. Exercicis: 2 i 5

Unitat 3. Anàlisi estadística de dues o més variables

-  Tema 5 Comparació de dues mitjanes poblacionals
Pujat 27/01/2020 10:18
-  Tema 6 Comparació de k mitjanes poblacionals
Pujat 27/01/2020 10:18
-  Tema 8 Comparació de variables categòriques
Pujat 27/01/2020 10:19
-  Exercicis unitat 3
Pujat 27/01/2020 10:19
-  Dades unitat 3
Pujat 27/01/2020 10:20
-  Exercici
Pujat 13/03/2020 13:20

Destresa:

Entendre la diferència entre mostres relacionades i mostres independents de l'explicació i els exemples.

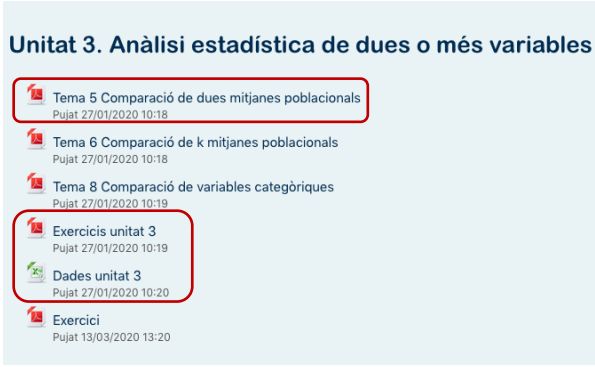
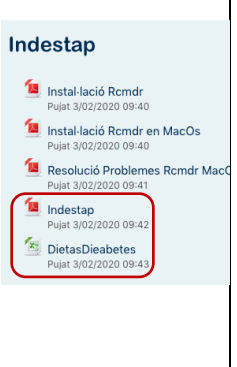
Activitat:

Llegir els exercicis 2 i 5 del document "Exercicis unitat 3", diferenciar quin exercici és per a mostres emparellades i quin per a mostres independents.

Solució:
Exercici 2: Mostres independents
Exercici 5: Mostres emparellades

2. Mostres emparellades (o aparellades)

Material de l'aulavirtual:

Teoria	Activitats amb R	Activitats sense R	Indestap
Secció: Unitat 3. Anàlisi estadística de dues o més variables. Pdf: Tema 5. Comparació de dues mitjanes poblacionals. Pàgines: 8 a 24	Document: Exercicis unitat 3. Dades: Dades unitat 3. Exercicis: 1 al 6 (exercicis tant per a mostres emparellades, com per a independents)	Secció: Problemes Carpeta: Dues mostres. (exercicis tant per a mostres emparellades, com per a independents)	Document: Indestap Secció del document: 2.4 Preguntes: 2.4a i 2.4b
			

Destresa:

- Saber quan s'han d'utilitzar els test paramètrics i quan els test no paramètrics.
- Saber realitzar contrast d'hipòtesi per a la igualtat de mitjanes poblacionals (cas paramètric).
- Saber calcular els intervals de confiança per a la diferència de mitjanes poblacionals (cas paramètric).
- Saber realitzar contrast d'igualtat de distribucions poblacionals (en el cas no paramètric).
- Saber realitzar contrastos unidireccionals (anàleg a una mostra).

Destresa	Pàgina del document de teoria on es troba la resposta	Exemples en el document
a)	10	11 a 14
b)	16	17
c)	18	20 a 22
d)	23	24

Activitat:

Es poden realitzar totes les activitats de dues mostres emparellades, en l'apartat "material de l'aulavirtual" teniu quines són les activitats que podeu realitzar. Haureu de llegir els problemes i saber identificar quins són mostres emparellades i quins són mostres independents. Dintre d'un mateix exercici pot haver un apartat per a mostres emparellades i un de mostres independents. Recordeu, que en el document d'indestap teniu un exemple amb els passos a seguir amb el R. Per tots aquells que no teniu R, teniu activitats amb les eixides.

3. Exemple

Una empresa que fabrica fertilitzants orgànics està desenvolupant dos nous fertilitzants que, per abreviar, cridarem B i S, respectivament. En una de les proves s'estudiaren 50 plantes d'una mateixa espècie que es van plantar en un hivernacle. Es va dividir a l'atzar les 50 plantes en dos grups d'igual grandària, fertilitzant un grup amb el fertilitzant B i l'altre amb l'S. Part de les dades obtingudes després de realitzar l'estudi es van guardar en el fitxer **Plantes.RData***, que conté les següents variables:

<i>abans</i>	altures de les plantes abans de fertilitzar-les
<i>després</i>	altures de les plantes després de fertilitzar-les
<i>fertilitzant</i>	fertilitzant utilitzat { B, S }

Contesta raonadament a les següents preguntes, amb $\alpha = 0.05$:

- Proporciona l'estimació puntual de la diferència de les altures mitjanes de les plantes abans i després. Seria correcta l'estimació per intervals? Si és possible, calcula el del 95%; en un altre cas, justifica la teua resposta.
- Podem afirmar que les altures mitjanes de les plantes abans de fertilitzar-les són iguals?
- Repetiu l'apartat anterior per a les altures després.
- És possible concloure que l'altura de les plantes augmenta amb el fertilitzant B?

*fitxer disponible en la secció "Documents de Treball"

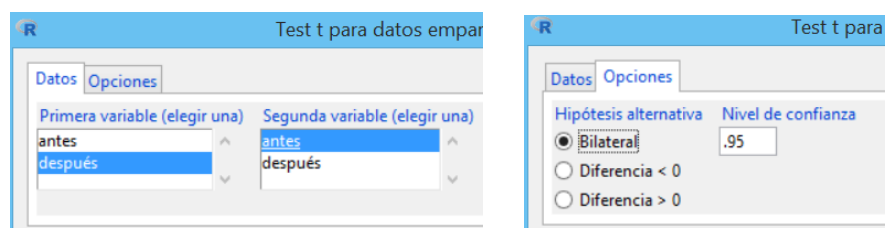
Noteu que si comparem les altures de les mateixes plantes abans i després de rebre el tractament tenim mostres aparellades (o emparellades), però si la comparació la fem en funció del fertilitzant, es tracta de mostres independents.

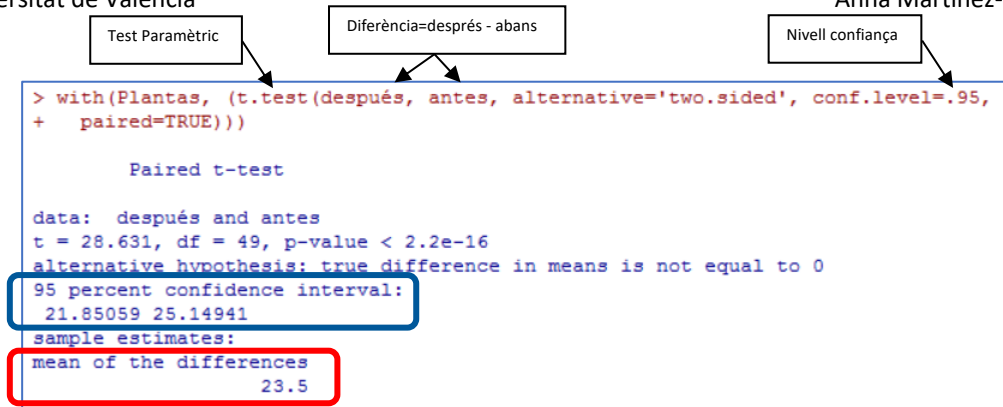
- Proporciona l'estimació puntual de la diferència de les altures mitjanes de les plantes abans i després. Seria correcta l'estimació per intervals? Si és possible, calcula el del 95%; en un altre cas, justifica la teua resposta.

L'estimació puntual de la diferència de mitjanes poblacionals és la diferència de mitjanes mostrals i, sempre és vàlida. Per tant, bastaria amb calcular estadístics per a ambdues mostres per obtenir-la, encara que l'obtindrem d'una altra forma.

La estimació per intervals és correcta si els mètodes paramètrics (basats en la t de Student) són vàlids. Com són **mostres aparellades** (volem comparar l'abans i el després de les mateixes plantes), els mètodes paramètrics són aplicables si les mostra de les diferències és gran ($n \geq 30$) o prové d'una població normal.

En el nostre cas, **la mostra de les diferències** té 50 dades, per la qual cosa la considerem **gran** i, per tant, podem aplicar **mètodes paramètrics**. Notar que, en aquest cas, no és necessari crear la variable diferència, si utilitzem el test t per a mostres aparellades (**Estadístics -> Mitjanes -> Test t per a dades relacionades**):





Per tant, l'estimació puntual de la diferència de mitjanes és 23.5 i l'interval de confiança del 95% és:

$$21.85059 < \mu_{\text{després}} - \mu_{\text{abans}} < 25.14941$$

Alternativament, podríem haver obtingut el mateix resultat creant la variable diferència, $\text{dif} = \text{després} - \text{abans}$ (Dades -> Modificar variables del conjunt de dades actiu -> Crear una nova variable) i utilitzant el test t d'una mostra (Estadístics -> Mitjanes -> Test t per a una mostra).

b) Podem afirmar que les altures mitjanes de les plantes abans de fertilitzar-les són iguals?

Ara volem comparar l'altura abans en els dos grups formats pel fertilitzant. Es tractaria de **mostres independents**, ja que les plantes en tots dos grups no són les mateixes. Ho resoldrem en el Document de Treball 2.

c) Podem afirmar que les altures mitjanes de les plantes després de fertilitzar-les són iguals?

Desitgem comparar l'altura després als dos grups formats pel fertilitzant. De nou es tractaria de **mostres independents**, ja que les plantes en tots dos grups no són les mateixes. Ho resoldrem en el Document de Treball 2.

d) És possible concloure que l'altura de les plantes augmenta amb el fertilitzant B?

Finalment volem comparar l'altura abans i després de les plantes tractades amb el fertilitzant B, per la qual cosa són **mostres aparellades**. De nou, els mètodes paramètrics són aplicables si les mostra de les diferències és gran ($n \geq 30$) o prové d'una població normal.

En el nostre cas, la mostra de les diferències té 25 dades, per la qual cosa hem de fer una prova de normalitat a la diferència d'altures ($\text{dif} = \text{després} - \text{abans}$):

H_0 : la variable dif és normal per al fertilitzant B

H_A : la variable dif no és normal per al fertilitzant B

```

> with(Plantas_B, shapiro.test(dif))

Shapiro-Wilk normality test

data: dif
W = 0.90899, p-value = 0.02895

```

Com el p-valor = 0.02895 < 0.05, hi ha evidència estadística que la variable dif no és normal en el grup del fertilitzant B.

Per tant, no podem aplicar mètodes paramètrics i hem d'utilitzar el test de Wilcoxon per a mostres aparellades (**mètode no paramètric**):

H_0 : els valors de les altures no augmenten amb el fertilitzant B

H_A : els valors de les altures augmenten amb el fertilitzant B

Triem el menú **Estadístics -> Test no paramètrics -> Test de Wilcoxon para muestras pareadas**:

Test de Wilcoxon para datos

Datos Opciones

Primera variable (elegir una) Segunda variable (elegir una)

antes después dif

antes después dif

Hipótesis alternativa Tipo de prueba

☐ Bilateral ☒ Por defecto

☐ Diferencia < 0 ☐ Exacto

☒ Diferencia > 0 ☐ Aproximación normal

☐ Aproximación normal con corrección para la continuidad

Test NO Paramètric

```
> with(Plantas_B, wilcox.test(después, antes, alternative='greater', paired=TRUE))
```

Wilcoxon signed rank test with continuity correction

data: después and antes

V = 325, p-value = 6.459e-06

alternative hypothesis: true location shift is greater than 0

Unilateral: hipòtesi alternativa (>): els valors de les altures augmenten amb el fertilitzant B

Mostres emparellades

Com el p-valor = $6.459 \cdot 10^{-6} < 0.05$, hi ha evidència estadística de que **l'altura augmenta amb el fertilitzant B**.

Alternativament, podríem haver obtingut el mateix resultat utilitzant el test de Wilcoxon per a una mostra (**Estadístics -> Test no paramètrics -> Test de Wilcoxon per a una mostra**) i la variable *dif*.