G-PAC 4: Contrast de dues mostres i contrast de variàncies. Turisme

PID_00141418

Universitat Oberta de Catalunya
www.uoc.edu

Índex

1. Introducció

En aquesta quarta prova d'avaluació continuada estendrem l'anàlisi feta a la PAC anterior per al cas que vulguem establir inferències entre dos grups, mesurar diferències sobre diversos aspectes d'un mateix grup, és a dir, quan treballem amb mostres aparellades o bé quan el nostre objectiu sigui fer inferència estadística sobre la variància d'una o de dues mostres.

A més, aquesta prova d'avaluació continuada està basada en l'aplicació d'aquestes tècniques estadístiques a dades de l'entorn del turisme en general. La comprensió correcta d'aquesta aplicació ens permetrà ser capaços d'elaborar informes sobre tot allò que estigui relacionat amb l'estudi del sector turístic.

La PAC es compon de tres exercicis amb diversos apartats. En el primer exercici s'aborden qüestions generals sobre com fer inferències entre dos grups sobre la diferència de mitjanes d'un mateix grup o de diferents grups tant si la fem amb intervals de confiança com amb contrastos d'hipòtesis. En el segon exercici es tracten aspectes més específics sobre la diferència entre dues proporcions de diferents grups. Finalment, el tercer exercici se centra en la variància de la població i es treballa la inferència estadística sobre la variància d'una o de dues poblacions.

2. Objectius i competències

Els **objectius** que es volen assolir amb aquesta quarta PAC són els següents:

1) Saber què és la inferència estadística de dues mostres i quan cal aplicar-la.

2) Construir intervals de confiança i realitzar contrasts d'hipòtesi per a diferències de mitjanes i proporcions; i interpretar els resultats.

3) Definir i conèixer les propietats de la distribució khi-quadrat i la distribució *F* d'Snedecor.

4) Construir intervals de confiança i fer contrasts d'hipòtesi per a la variància poblacional; i interpretar els resultats.

5) Construir intervals de confiança i fer contrastos d'hipòtesis per a comparar les variàncies de dues poblacions normals a partir de dades mostrals; i interpretar els resultats.

Addicionalment, en aquesta PAC 4 es treballaran les **competències** següents:

- Capacitat per a generar coneixement econòmic rellevant a partir de dades, aplicant els instruments tècnics pertinents.

- Capacitat per a generar coneixement rellevant per a la gestió d'organitzacions turístiques, i espais i destinacions turístiques a partir de dades, aplicant els instruments tècnics pertinents.

- Capacitat per a valorar críticament situacions empresarials concretes i establir possibles evolucions d'empreses i de mercats.

- Capacitat per a utilitzar i aplicar les tecnologies de la informació i la comunicació en els àmbits acadèmic i professional.

3. Guia de continguts

3.1. Introducció: Mostres aparellades i mostres no aparellades

Quan treballem amb dos grups d'observacions mostrals diferents, de vegades ens trobarem amb dues mostres aparellades o dependents i de vegades amb dues mostres no aparellades o independents.

El fet que dues mostres siguin dependents o no està determinat per les fonts (persones o objectes) que ens aporten les observacions. Si en l'obtenció de totes mostres s'han utilitzat les mateixes fonts o fonts associades, tenim dues *mostres aparellades* o *dependents*. Al contrari, si s'han utilitzat fonts completament diferents parlarem de *mostres no aparellades* o *independents*.

Exemple

Suposem que, en iniciar el semestre, seleccionem a l'atzar 30 estudiants matriculats en *Estadística I* i els passem un test de coneixements previs. Al final del semestre, seleccionem 30 estudiants més a l'atzar i els passem un test de coneixements adquirits durant el curs. En aquest cas, consideraríem ambdues mostres com independents.

Al contrari, si el test de coneixements adquirits es passés als mateixos 30 estudiants que van fer el test inicial, aleshores parlaríem de mostres dependents.

3.2. Intervals de confiança i contrastos d'hipòtesis per a mostres aparellades

1) Intervals de confiança per a la diferència de mitjanes de mostres aparellades

Donades dues mostres dependents, X_A i X_B, cadascuna d'elles de mida n, considerarem la variable aleatòria que resulti de calcular la seva diferència: $d = X_A - X_B$. Denotarem per $\mu_d = \mu_A - \mu_B$ la seva mitjana i per σ_d la seva desviació estàndard. Per tant, fer inferències sobre la diferència de les dues mitjanes mostrals dependents serà equivalent a fer-les sobre μ_d.

- **Supòsit**: X_A i X_B segueixen una distribució normal o les mostres són suficientment grans per a poder aplicar el teorema central del límit.

Nota

Si:
$X_A \sim N(\mu_A,\sigma_A)$ i $X_B \sim N(\mu_B,\sigma_B)$,
aleshores:
$d = X_A - X_B \approx N(\mu_A-\mu_B,\sigma_A-\sigma_B)$

L'interval de confiança al nivell de confiança $1 - \alpha$ per a $\mu_d = \mu_A - \mu_B$ ve donat per l'expressió:

$$\bar{d} \pm t_{n-1,\alpha/2} \cdot \frac{S_d}{\sqrt{n}}$$

on $t_{n-1,\alpha/2}$ és el valor que, en una t d'Student amb n – 1 graus de llibertat, deixa a la seva dreta una àrea de $\alpha/2$, i s_d és la desviació estàndard mostral de la variable aleatòria d.

2) Contrastos d'hipòtesis sobre la diferència de mitjanes per a mostres aparellades

Per a fer un contrast sobre diferència de mitjanes amb mostres aparellades, considerant que es compleixen els supòsits establerts, cal seguir els passos següents:

Pas 1. Definir les hipòtesis que s'han de contrastar:

$$\begin{cases} H_0 : \mu_d = \mu_0 \\ H_1 : \mu_d \neq \mu_0 \end{cases} \text{ o bé } \begin{cases} H_0 : \mu_d = \mu_0 \\ H_1 : \mu_d > \mu_0 \end{cases} \text{ o bé } \begin{cases} H_0 : \mu_d = \mu_0 \\ H_1 : \mu_d < \mu_0 \end{cases}$$

Pas 2. Calcular l'estadístic de contrast:

$$t^* = \frac{\bar{d}}{s_d / \sqrt{n}}$$

Pas 3. Aplicar el criteri de decisió. Descartarem H_0 si p-valor $\leq \alpha$ (normalment $\alpha = 0{,}05$). O equivalentment, descartarem H_0 si $EC > VC$ (cas unilateral cua a la dreta), si $EC < VC$ (cas unilateral cua a l'esquerra) o bé $EC > VC$ o $EC < -VC$ (cas bilateral).

Nota

El valor crític (VC) correspon al valor de la distribució t d'Student amb n – 1 graus de llibertat.

3.3. Intervals de confiança i contrastos d'hipòtesis per a mostres no aparellades

1) Intervals de confiança per a la diferència de mitjanes de mostres no aparellades

Treballarem ara amb dues mostres independents, X_A i X_B, de mides n_A i n_B respectivament.

- **Supòsit 1**: X_A i X_B segueixen una distribució normal o les mostres són prou grans per a poder aplicar el teorema central del límit.

- **Supòsit 2**: Les desviacions estàndard poblacionals són idèntiques en les dues poblacions, és a dir, $\sigma_1 = \sigma_2 = \sigma$. Aleshores podem estimar σ a partir de l'estimació de la desviació estàndard comuna:

Observació

Si:

$X_A \sim N(\mu_A,\sigma_A)$ i $X_B \sim N(\mu_B,\sigma_B)$, aleshores:

$$\bar{X}_A - \bar{X}_B \approx N\left(\mu_A - \mu_B, \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}\right)$$

$$s = \sqrt{\frac{(n_A - 1).\ s_A^2 + (n_B - 1).\ s_B^2}{n_A + n_B - 2}}$$

L'interval de confiança al nivell de confiança $1 - \alpha$ per a $\mu_A - \mu_B$ ve donat per l'expressió:

$$(\bar{x}_A - \bar{x}_B) \pm t_{n_A + n_B - 2,\ \alpha/2} \cdot S\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

on

$$t_{n_A + n_B - 2,\ \alpha/2}$$

és el valor que, en una t d'Student amb $n_A + n_B - 2$ graus de llibertat, deixa a la seva dreta una àrea de α /2.

2) Contrastos d'hipòtesis sobre la diferència de mitjanes per a mostres no aparellades

Per a fer un contrast sobre diferència de mitjanes amb mostres no aparellades, considerant que es compleixen els supòsits establerts, cal seguir els passos següents:

Pas 1. Definir les hipòtesis que es volen contrastar:

$$\begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A \neq \mu_B \end{cases} \text{ o bé } \begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A > \mu_B \end{cases} \text{ o bé } \begin{cases} H_0 : \mu_A = \mu_B \\ H_1 : \mu_A < \mu_B \end{cases}$$

Pas 2. Calcular l'estadístic de contrast:

$$t^* = \frac{\bar{X}_A - \bar{X}_B}{s \cdot \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$$

Pas 3. Aplicar el criteri de decisió. Descartarem H_0 si p-valor $\leq \alpha$ (normalment $\alpha = 0{,}05$). O equivalentment, descartarem H_0 si $EC > VC$ (cas unilateral cua a la dreta), si $EC < VC$ (cas unilateral cua a l'esquerra) o bé $EC > VC$ o $EC < -VC$ (cas bilateral).

Nota

El valor crític (*VC*) correspon al valor de la distribució t d'Student amb $n_A + n_B - 2$ graus de llibertat.

3) Intervals de confiança per a la diferència de dues proporcions

Treballarem ara amb dues mostres independents, X_A i X_B, de grandàries n_A i n_B respectivament.

- **Supòsit**: Les mostres són prou grans per a aplicar l'aproximació i fer una distribució normal.

L'interval de confiança al nivell de confiança 1 – α per a $p_A - p_B$ és determinat per l'expressió:

$$(p_A - p_B) \pm z_{\alpha/2} \cdot \sqrt{p(1-p)\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

en què $z_{\alpha/2}$ és el valor que, en una normal estàndard, deixa a la seva dreta una àrea de α/2 i en què $p = \dfrac{p_A n_A + p_B n_B}{n_A + n_B}$ és l'estimació de la proporció comuna.

4) Contrastos d'hipòtesis sobre la diferència de proporcions

Per a fer un contrast sobre diferència de proporcions, considerant que es compleixen els supòsits establerts, cal seguir els passos següents:

Pas 1. Definir les hipòtesis que es volen contrastar:

$$\begin{cases} H_0 : \pi_A = \pi_B \\ H_1 : \pi_A \neq \pi_B \end{cases} \text{ o bé } \begin{cases} H_0 : \pi_A = \pi_B \\ H_1 : \pi_A > \pi_B \end{cases} \text{ o bé } \begin{cases} H_0 : \pi_A = \pi_B \\ H_1 : \pi_A < \pi_B \end{cases}$$

Pas 2. Calcular l'estadístic de contrast:

$$z^* = \frac{p_A - p_B}{\sqrt{p(1-p)\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

Pas 3. Aplicar el criteri de decisió. Descartarem H_0 si p-valor ≤ α (normalment α = 0,05). O equivalentment, descartarem H_0 si $EC > VC$ (cas unilateral cua a la dreta), si $EC < VC$ (cas unilateral cua a l'esquerra) o bé $EC > VC$ o $EC < -VC$ (cas bilateral).

Nota

El valor crític (*VC*) correspon al valor de la distribució normal.

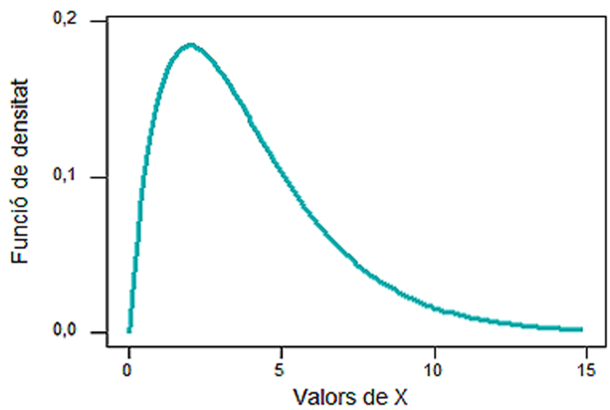
3.4. La distribució khi-quadrat

Una variable aleatòria amb distribució khi-quadrat (χ^2) amb *n* graus de llibertat es defineix com la suma dels quadrats de *n* distribucions normals, independents i estandarditzades.

Les seves **característiques** més importants són:

- Tots els valors de χ^2 són positius.
- És una corba amb asimetria cap a la dreta.
- La mitjana de la distribució és igual als seus graus de llibertat *n*.
- En realitat, la χ^2 no és una distribució, sinó una família completa de distribucions. Per a cada valor dels graus de llibertat hi ha una distribució diferent.
- A mida que augmenten els graus de llibertat, la corba es va fent més simètrica, fins que, en el límit, s'identifica amb una normal.

En el gràfic següent, que correspon a una χ^2 amb 4 graus de llibertat, es poden apreciar aquestes propietats:



Fent servir les taules estadístiques es podran calcular les probabilitats associades a aquesta distribució.

3.5. Contrastos d'hipòtesis sobre la variància

Treballem amb una mostra aleatòria de mida *n* i variància mostral s^2 obtinguda d'una població *X* (que segueix una distribució qualsevol), amb mitjana μ i desviació estàndard σ desconegudes.

- **Supòsit**: La mostra és suficientment gran per a poder aplicar l'aproximació per a fer a una distribució normal.

Per a fer un contrast sobre la variància, cal seguir els passos següents:

Pas 1. Definir les hipòtesis que es volen contrastar:

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases} \text{ o bé } \begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases} \text{ o bé } \begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$$

Pas 2. Calcular l'estadístic de contrast:

$$\chi^* = \frac{(n-1) \cdot s^2}{\sigma_0^2}$$

Pas 3. Aplicar el criteri de decisió. Descartarem H_0 si p-valor ≤ α (normalment α = 0,05). O equivalentment, descartarem H_0 si $EC > VC$ (cas unilateral cua a la dreta), si $EC < VC$ (cas unilateral cua a l'esquerra) o bé $EC > VC$ (cas bilateral cua a la dreta) o $EC < VC$ (cas bilateral cua a l'esquerra).

Nota

El valor crític (*VC*) correspon al valor de la distribució χ^2 amb $n - 1$ graus de llibertat al nivell de significació $1 - \alpha$ (cas unilateral cua a la dreta), α (cas unilateral cua a l'esquerra) o bé $1 - \alpha/2$ (cas bilateral cua a la dreta) i $\alpha/2$ (cas bilateral cua a l'esquerra).

L'interval de confiança al nivell de confiança 1 – α per a la variància és determinar per l'expressió:

$$\left(\frac{(n-1) \cdot s^2}{\chi^2_{1-\alpha/2,\, n-1}}, \frac{(n-1) \cdot s^2}{\chi^2_{\alpha/2,\, n-1}} \right)$$

3.6. La distribució *F* d'Snedecor

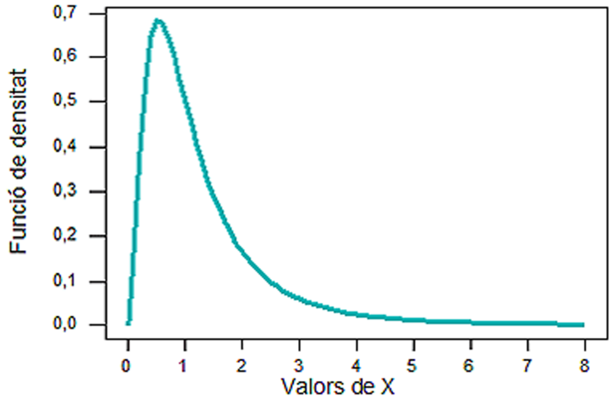
Una variable aleatòria que es distribueix segons una *F* d'Snedecor amb *m* graus de llibertat (al numerador) i *n* graus de llibertat (al denominador), es defineix de la següent a partir de dues distribucions khi-quadrat independents amb *m* i *n* graus de llibertat, respectivament:

$$F_{m,n} = \frac{\chi_2^m / m}{\chi_2^n / n}$$

Les **característiques** principals són:

- Es tracta d'una distribució contínua.
- La distribució *F* sempre pren valors positius.
- La seva funció de densitat presenta asimetria cap a la dreta.
- A mesura que augmenten els valors, la corba s'aproxima a l'eix de les *x*, però mai l'arriba a tocar.
- Existeix tota una família de distribucions *F*. Una en particular es determina per dos paràmetres: els graus de llibertat del numerador i els del denominador.

A continuació es mostra el gràfic de la funció de densitat d'una *F* d'Snedecor amb 6 graus de llibertat en el numerador i 8 en el denominador:



La distribució *F* compleix la propietat recíproca, que serà de molta utilitat per a trobar valors que determinin probabilitats en la cua esquerra de la distribució. Segons aquesta propietat podem establir que:

$$P(F_{m,n} \geq x) = P\left(\frac{1}{F_{n,m}} < x\right) = P\left(F_{n,m} \geq \frac{1}{x}\right) = \alpha$$

Fent servir les taules estadístiques es podran calcular les probabilitats associades a aquesta distribució.

3.7. Contrastos d'hipòtesis sobre la comparació de dues variàncies

Treballarem ara amb dues mostres independents, X_A i X_B (que segueixen una distribució qualsevol), de mides *m* i *n* amb variàncies mostrals respectives s_1^2 i s_2^2, i mitjanes (μ_1 i μ_2) i desviacions estàndard (σ_1 i σ_2) desconegudes.

Supòsit: Les mostres són suficientment grans per a poder aplicar l'aproximació per a fer una distribució normal.

Per a fer un contrast sobre la comparació de variàncies, i considerant que es compleixen els supòsits establerts, cal seguir els passos següents:

Pas 1. Definir les hipòtesis a contrastar:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases} \text{ o bé } \begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 > \sigma_2^2 \end{cases} \text{ o bé } \begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \end{cases}$$

O de manera equivalent:

$$\begin{cases} H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} \neq 1 \end{cases} \quad \text{o bé} \quad \begin{cases} H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} > 1 \end{cases}$$

$$\text{o bé} \quad \begin{cases} H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} = 1 \\ H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} < 1 \end{cases}$$

Pas 2. Calcular l'estadístic de contrast:

$$f^* = \frac{s_1^2}{s_2^2}$$

Pas 3. Aplicar el criteri de decisió. Descartarem H_0 si *p*-valor ≤ α (normalment α = 0,05). O equivalentment, descartarem H_0 si $EC > VC$ (cas unilateral cua a la dreta), si $EC < VC$ (cas unilateral cua a l'esquerra) o bé $EC > VC$ (cas bilateral cua a la dreta) o $EC < VC$ (cas bilateral cua a l'esquerra).

Nota

El valor crític (*VC*) correspon al valor de la distribució F d'Snedecor amb $m - 1$ graus de llibertat al numerador i $n - 1$ al denominador al nivell de significació $1 - \alpha$ (cas unilateral cua a la dreta), α (cas unilateral cua a l'esquerra) o bé $1 - \alpha/2$ (cas bilateral cua a la dreta) i $\alpha/2$ (cas bilateral cua a l'esquerra).

L'interval de confiança al nivell de confiança 1 – α per a la comparació de variàncies és determinat per l'expressió:

$$\left(\frac{s_1^2}{s_2^2} \frac{1}{f_{1-\alpha/2,n-1,m-1}} , \frac{s_1^2}{s_2^2} f_{1-\alpha/2,m-1,n-1} \right)$$

4. Fonts d'informació

Per a resoldre la PAC 4 haureu de fer servir les bases de dades i llegir la lectura proposada, que trobareu a la unitat didàctica corresponent. També podeu consultar tot un seguit de fonts d'informació que us proposem a continuació. Trobareu arxius amb dades actualitzades a la unitat didàctica corresponent.

1) Webs

- Instituto Nacional de Estadística (INE).

2) Bases de dades

a) *Encuesta de ocupación hotelera*, al web de l'INE. Es faciliten dades agregades i cal seleccionar també els períodes.

3) Publicacions rellevants

a) *Notas de prensa INE*. Informació cojuntural de l'Instituto Nacional de Estadística.

b) *Cifras INE. Boletín informativo del Instituto Nacional de Estadística*.