G-PAC 2: Probabilitat i variables aleatòries. Organització

PID_00141416

Índex

1. Introducció

En aquesta PAC 2 treballarem els conceptes de la probabilitat, la combinatòria, les variables aleatòries, el teorema central del límit, les distribucions de variables discretes i les distribucions normal i *t* d'Student.

Per fer-ho, iniciarem el treball amb el plantejament dels principals conceptes i principis de la teoria de la probabilitat i introduirem les nocions essencials que afirma la teoria freqüentista del càlcul de probabilitats: la probabilitat que un succés es doni es pot obtenir quan l'experiment fet és complet i podem assenyalar els èxits obtinguts sobre el total de casos possibles, és a dir:

$$P(A) = \frac{\text{casos favorables}}{\text{casos possibles}}$$

A partir d'aquí, la nostra intenció és comptar els resultats i les possibilitats que hi ha que un succés s'esdevingui, introduint la combinatòria i les diferents opcions que aquesta permet segons l'experiment que estiguem fent (permutacions, variacions, variacions amb repetició i combinacions).

Un segon bloc d'aquesta PAC està orientat a treballar les funcions de distribució de la probabilitat. En aquest sentit, el primer que haurem de fer serà separar les funcions de distribució discretes de les contínues, segons el tipus de dades amb què treballem.

Finalment, aquesta PAC se centrarà a introduir el teorema central del límit. Ens permetrà afirmar que la distribució de la mitjana mostral es pot aproximar a una normal (i per tant, aplicar-ne les propietats), sempre que la mostra sigui prou gran ($n > 30$) i independentment del tipus de distribució de la variable que estiguem considerant.

2. Objectius i competències

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

1) Entendre les propietats principals de la teoria de la probabilitat i de la combinatòria.

2) Observar les funcions de distribució discretes i contínues més importants.

3) Treballar el teorema central del límit i les seves aplicacions.

Addicionalment, en aquesta PAC 2 es treballaran les **competències** següents:

- Capacitat per a generar coneixement econòmic rellevant a partir de dades, aplicant els instruments numèrics pertinents.
- Capacitat per a valorar críticament situacions empresarials concretes i establir possibles evolucions d'empreses i mercats.
- Capacitat per a utilitzar i aplicar tecnologies de la informació i la comunicació en els àmbits acadèmic i professional.

3. Guia de continguts

Aquesta PAC 2 ens permetrà introduir elements importants com els relacionats amb la probabilitat, les variables aleatòries, les distribucions discretes i el teorema central del límit.

1) Fitxa 18: Probabilitat

L'estudi de la probabilitat d'un succés ens permetrà mesurar la tendència que té un succés d'esdevenir-se. Aquesta mesura serà un nombre comprès entre 0 i 1: 0 quan la probabilitat sigui nul·la i 1 quan el succés es doni sempre. Tenint en compte aquest fet, la freqüència relativa d'un succés s'obté dividint el nombre de vegades que el resultat de l'experiment és favorable entre el nombre de vegades total que fem l'experiment.

2) Fitxa 2: Combinatòria

Lligat amb el concepte de probabilitat tenim la combinatòria, que ens permet comptar els resultats d'un succés i els possibles resultats que es poden obtenir d'un experiment.

	Cal agafar-los tots?	Es poden repetir	Importa l'ordre	Nombre total
Variacions	No	No	Sí	$N(N-1)\ldots(N-k+1)$
Variacions amb repetició	No	Sí	Sí	N^k
Permutacions	Sí	No	Sí	$N!$
Combinacions	No	No	No	$\binom{N}{k} = \frac{N!}{k!(N-k)!}$

A partir d'aquí, entendrem una variable aleatòria com la funció que associa un valor numèric a cada element de l'espai mostral. Aquesta podrà ser discreta sempre que prengui un nombre finit de valors i contínua si pot prendre qualsevol valor en un interval dels reals. D'entre les distribucions discretes més comunes trobem les de Bernoulli, la binomial, la distribució geomètrica i la de Poisson. D'entre les contínues, considerarem la normal i la *t* d'Student.

3) Fitxa 7: Distribucions discretes de probabilitat

a) Les distribucions discretes. Les més importants són la de Bernoulli, la binomial, la distribució geomètrica i la de Poisson.

b) La distribució binomial. Comencem l'estudi d'algunes distribucions discretes. En aquest apartat estudiarem quan s'utilitza la distribució binomial, quina és la seva funció de quantia i com es fan servir les taules estadístiques per a calcular probabilitats que hi estan associades.

c) La distribució de Poisson. Veurem com la distribució de Poisson simplifica, sota certes circumstàncies, el càlcul de probabilitats binomials. Estudiarem la funció de quantia poissoniana, i també l'ús que cal fer de les taules estadístiques per a calcular probabilitats d'una manera senzilla.

d) La distribució hipergeomètrica. En aquest epígraf estudiarem quan s'utilitza el model hipergeomètric i quina és la seva funció de quantia de probabilitat. Veurem que, en determinades condicions, la llei binomial i la hipergeomètrica són equivalents.

4) Fitxa 24: Teorema central del límit

El teorema central del límit ens permet afirmar que la distribució de la mitjana mostral es podrà aproximar a una normal sempre que la mostra sigui prou gran ($n > 30$) i independentment del tipus de distribució de la variable que estiguem considerant. En aquest sentit, es podrà afirmar que la mitjana serà la mateixa que la de la variable d'interès, i la desviació típica de la mitjana mostral serà, aproximadament, l'error estàndard.

5) Les distribucions contínues

a) La distribució normal

Una part important de l'objectiu d'aquest mòdul és que ens familiaritzem amb la distribució normal. Aprendrem què és una corba de densitat normal, com és la gràfica de la distribució normal i les seves propietats més característiques. A més, aprendrem a estandarditzar les variables i a cercar probabilitats o àrees a partir de les taules de la distribució normal.

La **distribució normal** és una de les distribucions de probabilitat contínua més importants, ja que molts fenòmens naturals i socials s'hi ajusten.

Per a indicar que una variable aleatòria contínua segueix una distribució normal de mitjana μ i desviació estàndard σ farem servir l'expressió: $X \sim N(\mu,\sigma)$.

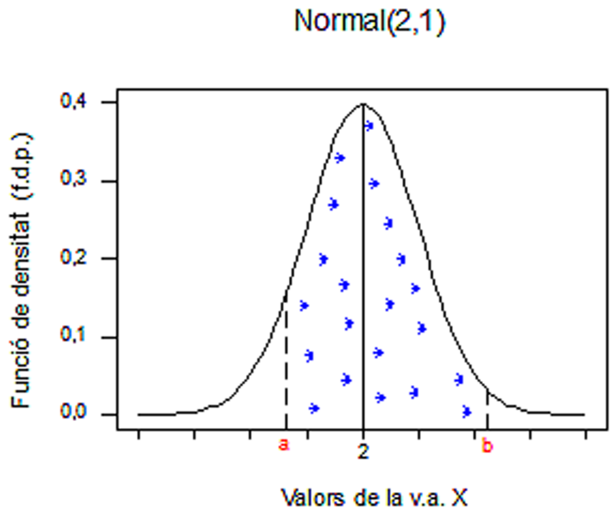
A la mitjana μ i a la desviació estàndard σ se'ls anomena *paràmetres* de la distribució normal.

La funció de densitat de la distribució normal de paràmetres μ i σ és:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

La probabilitat que la variable aleatòria *X* prengui un valor entre dos nombres reals *a* i *b* coincideix amb l'àrea tancada per la funció de densitat entre els punts *a* i *b*, és a dir:

$$P(a \le X \le b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$



Un element molt important de les distribucions normals és la **regla del 68–95–99,7**: En una distribució normal de mitjana μ i desviació estàndard σ:

- El 68% de les observacions es troben entre μ ± σ.
- El 95% de les observacions es troben entre μ ± 2σ.
- El 99,7% de les observacions es troben entre μ ± 3σ.

De les infinites distribucions *N*(μ,σ), té un interès especial la distribució *N*(0,1), és a dir, aquella que té per mitjana 0 i per desviació estàndard 1. Aquesta distribució s'anomena **distribució normal estàndard**.

El gran avantatge que té la distribució normal estàndard és que permet calcular ràpidament les probabilitats associades a aquesta distribució sense recórrer al càlcul integral, ja que els seus valors es troben en unes taules de fàcil maneig.

Per tant, el més aconsellable és transformar la variable *X* que segueix una distribució *N*(μ,σ) en una altra variable *Z* que segueix una distribució *N*(0,1). Aquesta transformació es coneix amb el nom de *tipificació* de la variable i per a fer aquesta transformació cal:

$$z = \frac{x - \mu}{\sigma}$$

b) La distribució *t* d'Student

Imaginem que el valor de la desviació típica de la variable no sigui conegut, llavors aquesta s'haurà d'estimar suposant que sigui una distribució normal.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

En aquest cas, una variable aleatòria que té la distribució *t* d'Student, amb *n* graus de llibertat, es defineix a partir d'un quocient entre dues distribucions independents, una que és una distribució normal estàndard i l'altra una distribució **khi-quadrat de Pearson**, amb *n* graus de llibertat. El quocient que defineix la nova distribució és el següent:

$$t = \frac{Z}{\sqrt{\frac{\chi_n^2}{n}}}$$

Aquesta distribució és força útil a l'hora de fer inferència estadística en els models de regressió lineal per tal de testar la significació individual de paràmetres.

4. Fonts d'informació

En aquest cas, treballarem els diferents conceptes de la probabilitat, la combinatòria, les variables aleatòries i el teorema central del límit, a partir de diverses fonts:

Per a la part de probabilitat i combinatòria usarem bases de dades generades de manera fictícia per a documentar la PAC 2, que trobareu a la unitat didàctica corresponent.

Per a la part del teorema central del límit i la distribució normal treballarem amb les dades de l'*Anuari estadístic* de "la Caixa" sobre municipis i, en concret, mirarem l'evolució de la variació de les activitats industrials amb els anys en percentatge (trobareu les dades actualitzades a la unitat didàctica corresponent).

Bases de dades

a) En aquesta PAC hem treballat les dades que ens resulten de l'*Anuari estadístic* de "La Caixa" sobre dades municipals de tot Espanya.

b) Podem trobar altres bases de dades sobre variables de conjuntura econòmica de Catalunya a l'Idescat.