BLOC 1=TEMA 1: INFERÈNCIA ESTADÍSTICA: INTRODUCCIÓ I CONCEPTES BÀSICS

LLIÇONS

1.1 Introducció, objectius i programa de l'assignatura (transparències IntroPlaDocent1516.pdf)

1.2 Estudi d'un cas real. Fil conductor (transparències Introduccio1516.pdf)

1.3 Context i objectius de la inferència estadística.

1.4 Població i mostra. Mostreig aleatori simple.

1.5 Estadístics i distribució en el mostreig.

1.6 Distribucions mostrals

1.3 Context i objectius de la inferència estadística

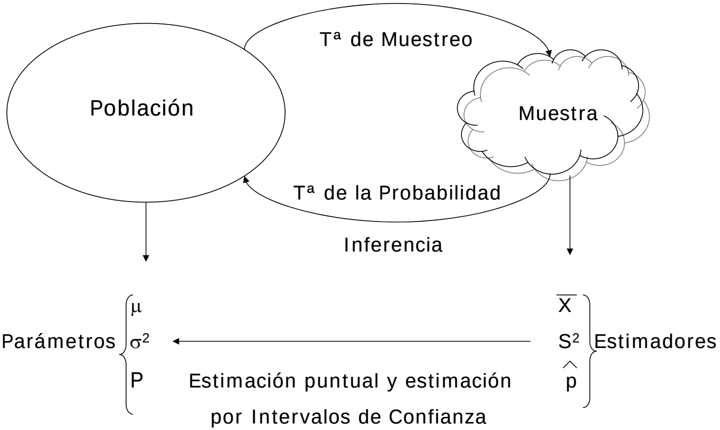
El análisis de datos,
LA RECOPILACIÓN, ORGANIZACIÓN Y RESUMEN DE LOS DATOS;

La probabilidad,
LAS LEYES DEL AZAR DENTRO Y FUERA DEL CASINO;

La inferencia estadística,
LA CIENCIA QUE EXTRAE CONCLUSIONES ESTADÍSTICAS A PARTIR DE DATOS CONCRETOS BASÁNDOSE EN EL CÁLCULO DE PROBABILIDADES.

Raonament Deductiu vs Raonament Inductiu

Fixeu-vos: població i paràmetres, mostra i estimadors



Raonament Deductiu: El llenguatge de la Probabilitat

- Moneda M
- Sabem que està trucada i que probabilitat de cara p=0.6.
- L'enlairem 10 cops i comptem el nombre de cares.
- $X =$ nombre total de cares en 10 llençaments.
- Llei de X: $X \sim$ Binomial amb $n = 10$ i $p = 0.6$.
- Quina és la probabilitat que obtinguem exactament 7 cares?
- $\mathrm{Prob}(X = 7; p = 0.6) = \binom{10}{7}(0.6)^7(0.4)^3 = 120 * 0,0018 = 0.21$
- **DEDUÏM** que, aproximadament, 21 de cada 100 cops que repetissim **AQUEST EXPERIMENT ALEATORI de (10 llençaments de moneda)** obtindriem **EXACTAMENT** 7 cares.

Establim hipòtesis sobre el mecanisme que han generat les dades i a partir d'aquest es dedueixen les probabilitats dels valors possibles.

Raonament Inductiu: El llenguatge de l'Inferència Estadística

- Moneda M
- No coneixem p=Prob(cara), no sabem si està trucada
- L'enlairem 10 cops i comptem el nombre de cares (li diem X).
- OBSERVEM que han surtit 7 cares $\Rightarrow x_1 = 7$
- Sortiràn sempre 7 cares?

Raonament Inductiu: El llenguatge de l'Inferència Estadística

- Moneda M
- No coneixem p=Prob(cara), no sabem si està trucada
- L'enlairem 10 cops i comptem el nombre de cares (li diem X).
- OBSERVEM que han surtit 7 cares $\Rightarrow x_1 = 7$
- Sortiràn sempre 7 cares?
- La tornem a enlairar 10 cops i OBSERVEM 5 cares $\Rightarrow x_2 = 5$
- Quins valors pot prendre X?
- Pot prendre valors v entre entre 0 i 10 amb probabilitat $\mathrm{Prob}(X = v; p) = \binom{10}{v} p^v (1-p)^{10-v}$
- $X \sim$ Binomial amb $n = 10$ i paràmetre p **DESCONEGUT**
- Basant-se en moltes repeticions $x_1, x_2, \cdots, x_l$, l'Inferència Estadística dona mètodes per aproximar (estimar) el valor de p=Prob(cara)

Donades les freqüències (o altres mesures) observades d'una variable, es tracta d'inferir el model probabilístic que les ha generat.

Raonament Inductiu: Estimació puntual

- Què podríem dir de la probabilitat de cara p basant-nos en l'evidència que de 10 llençaments 7 han resultat en cara?
- Podem **INFERIR** que la probabilitat de cara és $\hat{p} = \frac{7}{10} = 0.7$

Raonament Inductiu: Estimació puntual

- Què podríem dir de la probabilitat de cara p basant-nos en l'evidència que de 10 llençaments 7 han resultat en cara?
- Podem **INFERIR** que la probabilitat de cara és $\hat{p} = \frac{7}{10} = 0.7$
- Tenim una mica d'evidència per creure que la moneda **ESTÀ TRUCADA**
- Podem saber si realment el verdader valor de p és 0.7?
- Quantes vegades hauriem de realitzar l'experiment per tenir una **ESTIMACIÓ** de p *FIABLE (Acurada, precisa)*?
- Si realment p fos 0.7, quina és la probabilitat d'obtenir exactament 7 cares ?
- $\mathrm{Prob}(X = 7; \hat{p} = 0.7) = \binom{10}{7}(0.7)^7(0.3)^3 = 120 * 0,0022 = 0.27$ Creieu que és probable?

Com dona les respostes l'Inferència Estadística?

1. Què sabem sobre la distribució d'on venen les dades?
   - Coneixem la llei $\Rightarrow$ *Estadística Paramètrica* . Les dades provenen d'una llei
     - discreta (Binomial, Poisson, ...)
     - continua (Normal, exponencial, ...).

     Treballem amb el llenguatge de la probabilitat adient per aquestes lleis.
   - Desconeixem la llei $\Rightarrow$ *Estadística No Paramètrica*
     Segurament sabem quelcom sobre la forma de la distribució: Simètrica, amb cues pessades, ...
2. En funció de la qüestió plantejada, es decideix com es presenta la resposta
   - *Estimació puntual*: Utilitza les observacions per obtenir una única aproximació numèrica
   - *Estimació per intervals*: Obté un rang de valors possibles per un paràmetre
   - *Prova d'Hipòtesis*: Decideix si s'accepta o rebutja una afirmació sobre la distribució estudiada.

LLEGIU Peña. 7.1 pàgines 257–259

POBLACIÓ: conjunt dels elements (individus, peces, animals,monedes,...) pels que estudiem una característica donada.

- La població dels nens pels que estudiem el $X =$ BMI: Index de massa corporal
- La població d'infectats amb VIH prenent una teràpia antiretroviral (HAART) pels que estudiem el $T =$ Temps que triguen en reduir la càrrega viral
- La població de cotxes pels que estudiem si són o no CONTAMINANTS (Diòxid de carboni $< 120 gr/Km$) $Y = 1$ (SI), $Y = 0$ (NO)
- La població d'estudiants universitaris i pels que volem conèixer el seu $R =$ Rendiment acadèmic en el primer any

Podem accedir a tota la població? INE

CENS: estudi exhastiu de tota la població

Per primer cop l'INE `http://www.ine.es/` combina l'us de registres administratius amb el treball de camp, que inclou una gran enquesta per a conèixer les característíques de persones i vivendes

El Censo de Población y Viviendas 2011 se ha planteado como una operación basada en la combinación de los siguientes elementos:

Un fichero precensal realizado a partir de los registros administrativos disponibles, tomando al Padrón como elemento básico de su estructura.

Un trabajo de campo que incluye dos grandes operaciones: Un Censo de Edificios exhaustivo que permite la georreferenciación de todos los edificios.

Una gran encuesta por muestreo para conocer las características de las personas y las viviendas.

Las cifras de población se han obtenido mediante el recuento de los registros contenidos en el fichero precensal, ponderados cuando así se requiere con unos factores de recuento obtenidos a partir de la encuesta.

Podem accedir a tota la població? IDESCAT

Vegeu al web de l'IDESCAT la informació sobre la població a Catalunya. Aneu a Padró Continu
http://www.idescat.cat/cat/poblacio/poblrecomptes.html

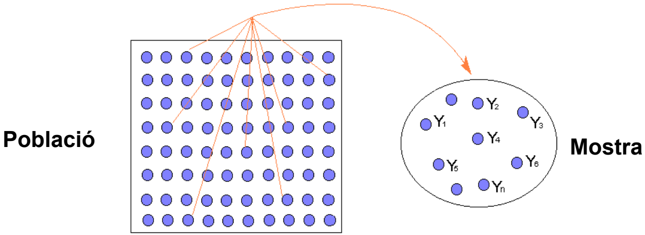
És sovint difícil o impossible accedir a TOTA la població ja que un estudi de tota una població

- pot ser molt costós i per tant inviable
- pot ser molt llarg en el temps
- pot ser impossible
- pot implicar la destrucció de l'element

MOSTRA versus POBLACiÓ

MOSTRA: Subconjunt REPRESENTATIU de la població.
PUNT CLAU: La mostra ha d'estar **BEN** escollida per representar a tota la població sota estudi.

Concepte intuitiu de mostra aleatòria simple (m.a.s.)



m.a.s.: Tot element de la població té la mateixa probabilitat de ser escollit per formar part de la mostra

↕

Y_1, Y_2, ..., Y_n són INDEPENDENTS

MOSTRA ALEATÒRIA SIMPLE

Una mostra és aleatòria simple quan

- Tots els subconjunts (mostres) possibles de grandària n de la població tenen la mateixa probabilitat de ser escollits
- Cada element de la població té la mateixa probabilitat de ser escollit $\Rightarrow$ per garantir la REPRESENTATIVAT
- La selecció d'un element de la població no ha d'influir en la selecció d'un altre $\Rightarrow$ per garantir l'INDEPENDÈNCIA

MOSTRA ALEATÒRIA SIMPLE

Mostra aleatòria simple amb una mica de notació

A l'estudi **Healthy for Life**: tenim molts nens i nenes i per cadascun mesurem moltes coses: edat, pes, alçada, etc.

- Indiquem els n nens de l'estudi per
$$w_1, w_2, \ldots, w_n$$
- Quan parlem d'edat ens refererirem com E. Per exemple $E(w_1) = 3$ vol dir que el nen 1 de l'estudi té 3 anys i $E(w_2) = 5$ el nen 2 té 5 anys. Sovint indicarem $E_1 = E(w_1)$ i $E_2 = E(w_2)$ i quan ja coneixem el valor fem servir $e_1 = 3$ i $e_2 = 5$
- Si P representa el pes, $P_1 = P(w_1)$, $p_1 = 14$ i $P_2 = P(w_2)$, $p_2 = 20$ vol dir que el nen 1 pesa 14 Kg i el 2 pesa 20 Kg
- Si A és alçada $A_1 = A(w_1)$ i $A_2 = A(w_2)$. Els nens 1 i 2 fan $a_1 = 96$cm i $a_2 = 110$cm
- Si els mesurem en metres definim $A^* = A/100$ i per tant, $a_1^* = 0.96$ i $a_2^* = 1.1$
- Si B representa el BMI (Body Mass Index), $B = P/A_*^2$. $B_1 = B(w_1)$, $B_2 = B(w_2)$. $b_1 = 14/(0.96)^2 = 15.2$ i $b_2 = 20/(1.1)^2 = 16.52$

Mostra aleatòria simple més FORMALMENT

- P=pes, A=alçada, B=BMI són variables aleatòries, de forma genèrica fem servir les lletres X, Y, Z
- Distribució F de X (llei Binomial, Normal, etc.)
- Individus de l'estudi per $w_1, w_2, \ldots, w_n$

Una *mostra aleatòria simple* de grandària n (*size* en anglès) de X és una col.lecció de n variables aleatòrias tal que:

1. $X_1, X_2, \ldots, X_n$ són **independents**. Per exemple, l'elecció del nen 1 no influeix en l'elecció del nen 2 i per tant els seus pesos $P_1, P_2, \ldots,$ alçades $A_1, A_2, \ldots,$ etc. són independents
2. $X_1, X_2, \ldots, X_n$ tenen la **mateixa distribució**. Per exemple, els pesos es distribueixen segons una llei normal de mitjana μ i variància σ^2, és a dir, $P_1 \sim N(\mu, \sigma^2)$, $P_2 \sim N(\mu, \sigma^2)$, etc.
3. Ho indiquem com *i.i.d.*=independent idènticament distribuït

$$(X_1, X_2, \ldots, X_n) \overset{i.i.d.}{\sim} X \sim N(\mu, \sigma^2)$$

El conjunt $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ d'observacions concretes de $(X_1, X_2, \ldots, X_n)$ es denomina *realització de la mostra*.

Tipus de mostreig

- Si el nombre d'elements de la població N és petit, en relació a la grandària de la mostra n, les observacions es realitzen amb reemplaçament de manera que es garanteix que la població és idèntica en totes les extraccions
- Si $n/N > 0.1$ cal fer correccions $\Rightarrow$ *Mostreig per poblacions finites*

Quan els elements de la població no són homogenis respecte a la característica que estudiem es poden fer altres tipus de mostreig que garanteixin que la mostra té una composició analoga a la població. Per exemple,

- Per garantir que a la mostra hi hagi la mateixa proporció de nois i noies, de persones de determinades edats $\Rightarrow$ Mostreig estratificat
- de persones de cada comarca $\Rightarrow$ Mostreig per conglomerats
- Quan els elements de la població estan ordenats en llistes $\Rightarrow$ Mostreig sistemàtic

LLEGIR PEÑA 7.2 pàgines 260–264

RECOLLINT DADES A L'AULA

GRUPS (5 persones)	ALÇADA CM	GENERE 1=NOIA, 0=NOI	NOMBRE DE GERMANS
1			
1			
1			
1			
1			
2			
2			
2			
2			
2			
3			
3			
3			
3			
3			

Donada una mostra aleatòria simple $X_1, X_2, \ldots, X_n$ podem definir molts estadístics diferents:

$$
\begin{aligned}
T_1(X_1, \ldots, X_n) &= \sum_{i=1}^{n} X_i \\
T_2(X_1, \ldots, X_n) &= \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X} \\
T_3(X_1, \ldots, X_n) &= \min\{X_1, \ldots, X_n\} \\
T_4(X_1, \ldots, X_n) &= \max\{X_1, \ldots, X_n\} \\
T_5(X_1, \ldots, X_n) &= \sum_{i=1}^{n} (X_i - \overline{X})^2 \\
T_6(X_1, \ldots, X_n) &= \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} = S^2
\end{aligned}
$$

Com que $T_2(X_1, \ldots, X_n) = \overline{X}$ estima la mitjana de la població, μ, direm que és un estimador de μ, com que $T_6(X_1, \ldots, X_n) = S^2$ estima la variança de la població, σ^2, direm que és un estimador de σ^2.

Estadístics i Estimadors en Healthy for Life

- $X =$ pes podem suposar $\sim \mathrm{Normal}(\mu, \sigma^2)$
- $\mu =$ mitjana dels pes de TOTS els nens de la població. Ens caldria pesar a tots els nens per saber quan val μ exactament
- σ^2 variància dels pesos de TOTS els nens de la població
- $X_1, \cdots, X_n$ pes dels n nens de la mostra. Totes les components $X_1, X_2, \ldots, X_n$ són $\mathrm{Normal}(\mu, \sigma^2)$
- L'estadístic $\sum_{i=1}^{n} X_i$ proporciona la suma de tots els pesos dels nens de la mostra
- L'estadístic $\frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$ proporciona la mitjana del pes dels nens de la mostra i és un estimador de la mitjana dels pesos de tots els nens, és un estimador de μ.
- L'estadístic $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$ estima la variança σ^2 dels pesos dels nens de la població

Definicions matemàtiques

Definició

Una variable aleatòria (v.a) és una representació "abstracte" d'un resultat mesurable d'un experiment. Una v.a. no és el resultat d'un sol experiment: conceptualment és la representació dels possibles resultats de la repetició en condicions idèntiques de molts experiments. És molt més que un número o un conjunt de números, és la portadora de les propietats probabilístiques del que s'està mesurant.

Definició

Un estadístic és una VARIABLE ALEATÒRIA que és funció de la mostra (de les dades) i no depèn del valor del paràmetre.

Definició

Donada una mostra aleatòria simple $X_1, X_2, \ldots, X_n$ i una funció $T : \mathbb{R}^n \longrightarrow \mathbb{R}$, un estadístic $T(X_1, X_2, \ldots, X_n)$ és una variable aleatòria que no depèn dels paràmetres de la distribució de X.

Que és un estimador?

Definició

Els estimadors són funcions de les dades (estadístics) que aproximen el valor del paràmetre.
TOTS els estimadors són ESTADÍSTICS. A l'inrevès no!!!
Un estimador d'un paràmetre $\theta \in \Theta$ és un estadístic, que usualment denotem per $\hat{\theta}$ (notació dels barrets), tal que els possibles valors que pot prendre coincideixen amb els valors que el paràmetre θ representa.

$$\begin{array}{cccc} \hat{\theta}: & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & (x_1, x_2, \ldots, x_n) & \longrightarrow & T(x_1, x_2, \ldots, x_n) = t \quad \in \Theta \subset \mathbb{R} \end{array}$$

- $X_1, X_2, \ldots, X_n$ mostra aleatòria de X
- Estimadors possibles de $\mu = \mathrm{E}(X)$:
  $\hat{\mu}_1 = \frac{\sum_{i=1}^{n} X_i}{n}$, $\hat{\mu}_2 = \frac{\sum_{i=1}^{n-3} X_i}{n-3}$, ...
- Estimadors possibles de $\sigma^2 = \mathrm{Var}(X)$:
  $\hat{\sigma}_1^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$, $\hat{\sigma}_2^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$, ...

Definición

La distribució de probabilitat de l'estadístic (RECORDEM: Un estadístic és una variable aleatòria, per tant li correspon una distribució de probabilitat) $T(X_1, X_2, \ldots, X_n)$ s'anomena *distribució mostral* o *distribució en el mostreig de T*.

Treballant les alçades dels estudiants a classe el 17/02

POBLACIÓ: A=alçades (en cm)

$A \sim N(\mu = 174,54cm, \sigma^2 = 72,14cm^2)$

ESTADÍSTIC: $\bar{A}_n = \frac{\sum_{i=1}^{n} A_i}{n}$.

La distribució (mostral) de $\bar{A}_n$ depèn de:

- La població base, $A \sim N(174,54, 72.14)$
- La grandària mostral n
- $\bar{A}_n \sim N(\mu = 174,54, \sigma^2/n = 72.14/n)$
- Si $n = 5$, $\bar{A}_5 \sim N(174,54, 72.14/5 = 14,43)$
- Si $n = 10$, $\bar{A}_{10} \sim N(174,54, 72.14/10 = 7,21)$

LLEGIR PEÑA 7.4.1 pàgines 270–271 i EXERCICI VOLUNTARI 1

Distribució de l'alçada dels nois de la UB

X = alçada en cm dels nois que estudien a la UB.
Agafem mostres de 25 nois ($n = 25$), com es distribueix la mitjana mostral $\bar{X}_{25}$?

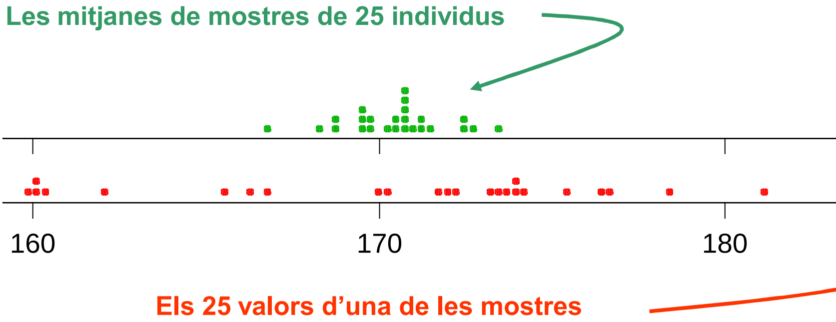
Muestra 1	Muestra 2	
183,7	171,9	
178,5	188,5	
160,0	168,3	
171,9	176,4	
176,6	173,0	
160,0	171,0	
174,3	161,8	
160,0	169,2	
170,1	169,0	
170,1	177,6	
160,3	165,5	
165,5	171,7	
172,3	181,9	...
176,5	173,7	
175,5	185,5	
173,8	169,1	
181,1	184,0	
166,7	163,0	
166,3	153,8	
174,0	186,5	
173,4	167,9	
171,6	176,0	
174,0	171,0	
173,2	177,5	
162,2	165,7	

Mitjana = 170,9

Mitjana = 172,8

Medias muestrales
170,9
172,8
169,5
168,8
171,1
171,1
172,5
170,4
170,5
170,7
172,4
169,6
169,7
166,7
170,6
169,8
171,2
169,4
170,4
168,7
170,7
173,6
168,2
171,5
170,7

Distribució de l'alçada dels nois de la UB



Hi ha més dispersió en els valors individuals que en les mitjanes mostrals

Estudiem FORMALMENT la Distribució de la mitjana mostral

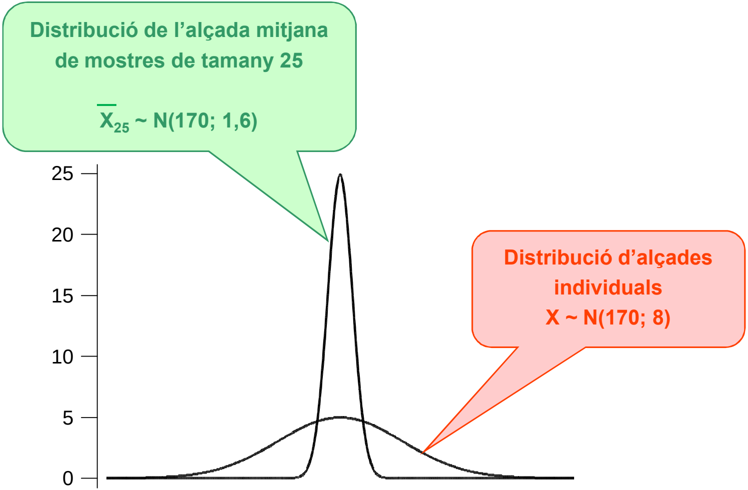
- X denota la població
- Denotem per μ la mitjana de la població, $E(X) = \mu$
- Denotem per σ^2 la variància de la població,
  $0 < \mathrm{Var}(X) = \sigma^2 < \infty$
  HEU DE DISTINGIR X (població) de $\overline{X}$ (mitjana mostral)
- Prenem una mostra aleatòria simple de X, és dir, $X_1, \ldots, X_n$
- Considerem la Mitjana mostral: $\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$

COM ES COMPORTA LA MITJANA MOSTRAL ?

- $\mathrm{E}(\overline{X}_n) = \mathrm{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n \mathrm{E}(X_i)}{n} = \frac{n\mu}{n} = \mu$
- $\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\mathrm{Var}(\sum_{i=1}^n X_i)}{n^2} = \frac{\sum_{i=1}^n \mathrm{Var}(X_i)}{n^2} = \frac{\sigma^2}{n}$
  ja que $X_1, \ldots, X_n$ són independents, $\mathrm{E}(X_i) = \mu$ i $\mathrm{Var}(X_i) = \sigma^2$.
- La llei de $\overline{X}_n$ és més simètrica que la de X i amb menys dispersió.

Distribució de la població d'alçades i Distribució de les mitjanes de les alçades dels nois de la UB

Si ens diguessin que el model es comporta com
$X \sim N(\mu = 170, \sigma = 8)$.

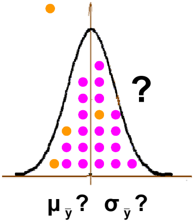


Seguim parlant de la distribució de la mitjana mostral

Estudiem $\bar{X}_n$. Considerem mostres de grandària n, i per cadascuna calculem la mitjana $\bar{x}_n^j$

$$\begin{array}{lcccccl}
\text{Mostra 1} \Rightarrow & x_{11} & x_{12} & x_{13} & \dots & x_{1n} & \Rightarrow \bar{x}_n^1 (\textit{mitjana Mostra } 1) \\
\text{Mostra 2} \Rightarrow & x_{21} & x_{22} & x_{23} & \dots & x_{2n} & \Rightarrow \bar{x}_n^2 (\textit{mitjana Mostra } 2) \\
& \vdots & \vdots & \vdots & & & \qquad \vdots \\
\text{Mostra s} \Rightarrow & x_{s1} & x_{s2} & x_{s3} & \dots & x_{sn} & \Rightarrow \bar{x}_n^s (\textit{mitjana Mostra } s)
\end{array}$$

Fem un histograma amb els valors $\bar{x}_n^1, \bar{x}_n^2, \bar{x}_n^3, \cdots, \bar{x}_n^{s-1}, \bar{x}_n^s$ i "ajustem" una corba. Aquesta corba es correspon amb la densitat de $\bar{X}_n$ i s'anomena Distribució mostral de la mitjana.



Distribució de la mitjana mostral quan n és gran

- **Quan n és gran ($n > 30$), la distribució de $\overline{X}_n$ és aproximadament Normal de mitjana μ i variància $\frac{\sigma^2}{n}$.**
- La mitjana estandaritzada

$$\frac{\overline{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \sim \mathrm{N}(0, 1)$$

- X normal $\Rightarrow \overline{X}_n$ normal (combinació lineal de normals).
- Quan n és petita la distribució exacte de $\overline{X}_n$ depèn de la població d'on prové X.

LLEGIR PEÑA 7.4.3 pàgines 272–273

Mostreig en poblacions normals. Distribucions derivades de la normal

AQUESTES LLEIS I EL TEOREMA DE FISHER, ES FARAN SERVIR MÉS ENDAVANT I S'ANIRÀN ASSIMILANT MICA A MICA

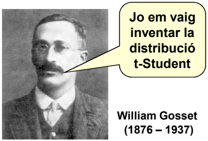
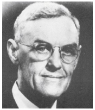


Figure: χ^2 khi-quadrat de Pearson i t de Student (Gosset)



George W. Snedecor
(1882 - 1974)

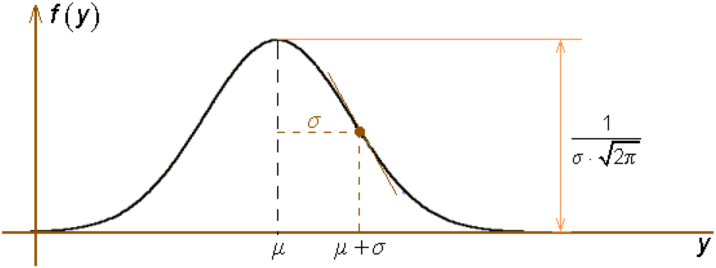
Ronald A. Fisher
(1890 – 1962)

Figure: F de Fisher-Snedecor

Funció de densitat de la llei normal

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\frac{(y-\mu)^2}{\sigma^2}} \qquad \text{amb :} \quad \begin{cases} -\infty < \mu < \infty \\ \sigma > 0 \end{cases}$$

Geometria de la llei normal



La distribució χ^2 quadrat

$X_1, X_2, \ldots, X_n$ v.a. independents, $X_i \sim N(0,1)$

Definició

La distribució de la v.a. $Y = X_1^2 + X_2^2 + \cdots + X_\nu^2$ s'anomena khi-quadrat amb ν graus de llibertat. $Y \sim \chi^2_\nu$

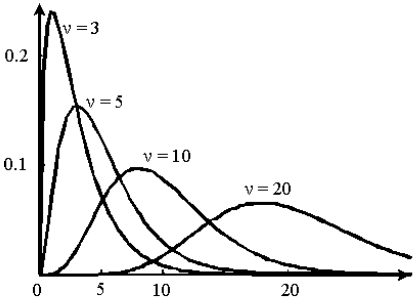


Figure: khi-quadrat de Pearson

La distribució t de Student

Y, Z v.a.'s indep. $Z \sim N(0,1)$ i $Y \sim \chi^2_\nu$

Definició

La distribució de la v.a. $t = \frac{Z}{\sqrt{Y/\nu}}$ s'anomena t de Student amb ν graus de llibertat.

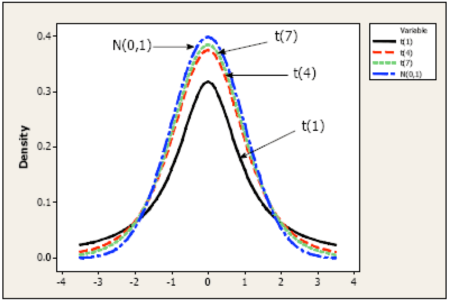


Figure: Student (Gosset)

La distribució F de Fisher

$U \sim \chi^2_m$, $V \sim \chi^2_n$ indep.

Definició

La distribució de la v.a. $F = \frac{U/m}{V/n}$ s'anomena F de Fisher amb m i n graus de llibertat resp.

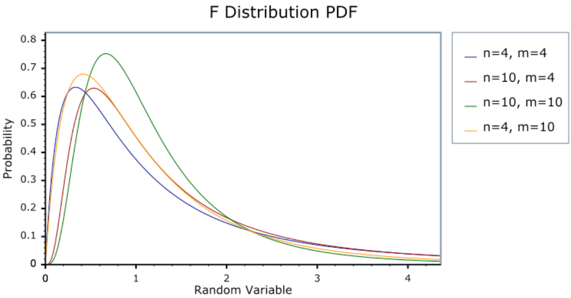


Figure: Fisher i Snedecor

Teorema de Fisher

Comencem amb $X_1, X_2, \ldots, X_n$ una mostra de $N(\mu, \sigma)$.
Considerem

- la mitjana mostral $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$
- la variància mostral $S_n^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

EL TEOREMA DE FISHER ESTABLEIX QUE $\overline{X}_n$ i S_n^2 són independents

Conseqüències del Teorema de Fisher

- L'estadístic
$$\frac{(n-1)S_n^2}{\sigma^2}$$
es distribueix com una χ^2_{n-1} amb $n-1$ graus de llibertat.
- L'estadístic
$$\sqrt{n-1}\frac{\bar{X}_n-\mu}{S_n}$$
és es distribueix com una t de Student amb $n-1$ graus de llibertat.

Donades dues mostres aleatòries simples i independents

$$X_1, X_2, \ldots, X_{n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1) \qquad Y_1, Y_2, \ldots, Y_{n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2)$$

definim $S_1^2 = \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2}{n_1 - 1}$ i $S_2^2 = \frac{\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2}{n_2 - 1}$.
L'estadístic

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

es distribueix com una F de Fisher amb $n_1 - 1, n_2 - 1$ graus de llibertat.
Si $\sigma_1^2 = \sigma_2^2$, l'estadístic $\frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}$

Conseqüències del Teorema de Fisher (III)

Donades dues mostres aleatòries simples, independents entre si,

$$X_1, X_2, \ldots, X_{n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1) \qquad Y_1, Y_2, \ldots, Y_{n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2)$$

l'estadístic

$$\frac{\bar{X}_{n_1} - \bar{Y}_{n_2} - (\mu_1 - \mu_2)}{(n_1 - 1)S_1^2/\sigma_1^2 + (n_2 - 1)S_2^2/\sigma_2^2} \sqrt{\frac{n_1 + n_2 - 2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

segueix una distribució t de Student amb $n_1 + n_2 - 2$ graus de llibertat.

LLEGIR PEÑA 7.4.5 pàgines 276–280.