



Departament d'Estadística
i Investigació Operativa

UNIVERSITAT POLITÈCNICA DE CATALUNYA



Models de VA i simulació

Annex

Bloc B – Probabilitat i Estadística

2024

Índex

Mostra aleatòria simple

Descriptiva (gràfica)

Simulació de valors aleatoris amb R

Anàlisi gràfica de la normalitat

Models derivats de la Normal

Probabilitats i quantils amb models de VA amb R

Mostra Aleatòria Simple (MAS)

Sigui la VA: $X: \Omega \rightarrow \mathbb{R}$
 $\omega_i \rightarrow X(\omega_i) = x_i$

Direm que M.A.S. de grandària n de la v.a. X
 és una funció vectorial $M = (X_1, X_2, \dots, X_n)$ tal que
 $M: \Omega^n \rightarrow \mathbb{R}^n$
 $\omega = (\omega_1, \omega_2, \dots, \omega_n) \rightarrow M(\omega) = (X_1, X_2, \dots, X_n)$

Direm que és una MAS si i només si es compleixen les dues condicions següents:

- (1) **Tots els elements** de la població tenen la **mateixa probabilitat** de pertànyer a la mostra.
- (2) **Qualsevol combinació** de n elements té la **mateixa probabilitat** de pertànyer a la mostra.

La informació aportada per les diferents unitats ha de ser **independent** entre sí:
 - les X_i han de ser VA independents i idènticament distribuïdes (i.i.d.)

R té funcions per generar mostres aleatòries simples seguint un model. Per exemple,
R: rbinom(), rpois(), rexp(), rnorm()

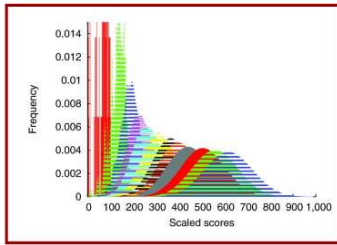
Descriptiva (gràfica)

L'**Estadística Descriptiva** permet resumir dades gràficament (i també numèricament tal com es veurà relacionant-ho amb inferència estadística)

En la següent taula hi ha algunes funcions (bàsiques) en R per **Estadística Descriptiva gràfica** en dades discretes i contínues de forma univariant o bivariant:

	UNIVARIANT VAC	UNIVARIANT VAD	BIVARIANT
GRÀFIQUES	hist() boxplot()	barplot(table())	plot(,)

(més funcions gràfiques en R: <https://www.r-graph-gallery.com/>)



(consulteu el material disponible a la pàgina web, instaleu-vos R a l'ordinador i practiqueu)

Estadística descriptiva amb R

Univariant i bivariant

Eines gràfiques



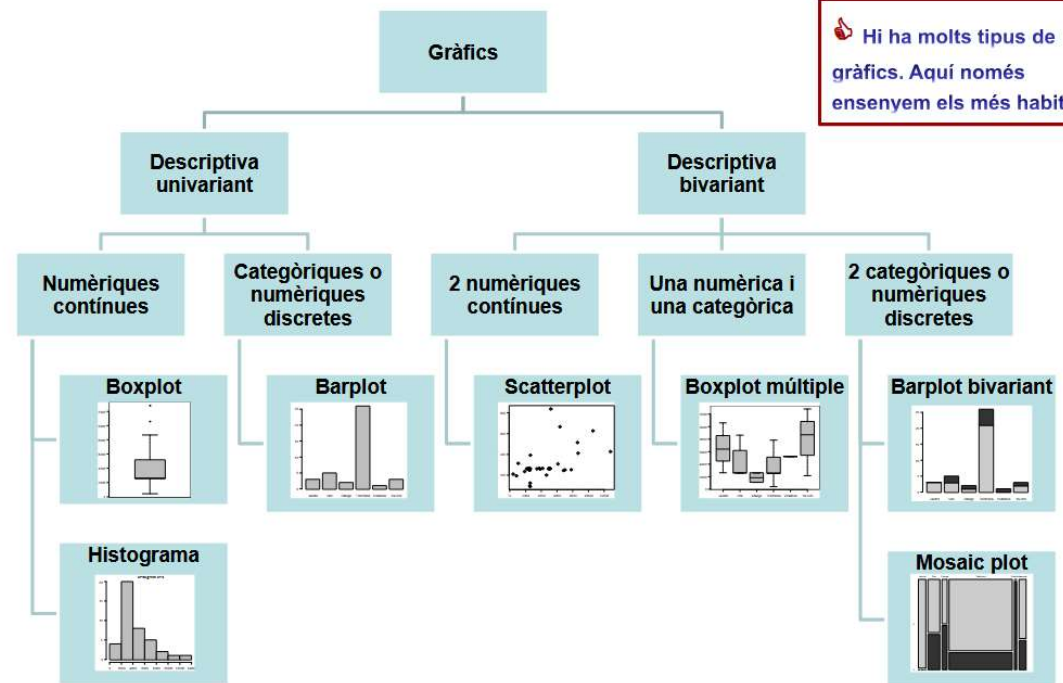
Software R i exploració de dades

Materials d'interès

- Descarregar R.
- Video introductori.
- Guia ràpida de referència
- Tutorial introductori.
- Curs breu: [càlcul amb R](#).
- Introducció al [software R](#).
- Utilitza R al mòbil
- Transparències: [Estadística Descriptiva](#).
- Activitat: Guió Estadística Descriptiva.
- Activitat: Instruccions bàsiques de R.



Hi ha molts tipus de gràfics. Aquí només ensenyem els més habituals



Simulació de VA amb R

R té diverses funcions per a generar valors aleatòriament:

`sample(x, n, replace=y)` **n** valors del conjunt **x**, amb reemplaçament (**y=T**) o no (**y=F**). També podem assignar probabilitats particulars a cada element de **x**.

Exemple: `sample(1:10, 25, replace=TRUE)`

Generalment, volem reproduir valors d'un model probabilístic particular: Poisson, Normal, uniforme, ... Tots els models vistos (tots els que té implementats R, en realitat, apart de les funcions **d...**, **p...** i **q...** tenen la funció **r...** per simular aquests models.

Exemples:

`rnorm(1)` : 1 valor a l'atzar d'un model $N(0,1)$

`rpois(5, 2.5)` : 5 valors a l'atzar d'un model $P(\lambda=2.5)$

`rbinom(50, 30, 0.1)` : 50 valors a l'atzar d'un model $B(n=30, p=0.1)$

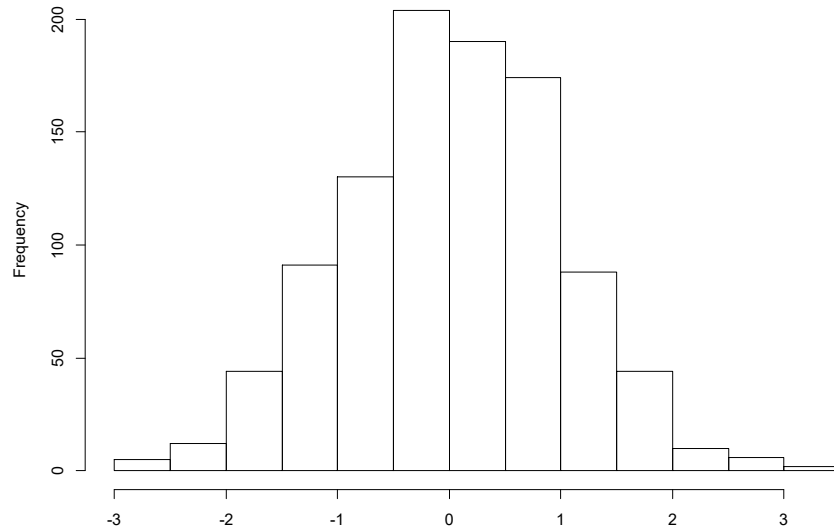
`runif(3, -1, 1)` : 3 valors a l'atzar d'un model $U(-1,1)$

`rnbinom(1, 5, 0.7)` : 1 valor a l'atzar d'un model $BN(r=5, p=0.7)$

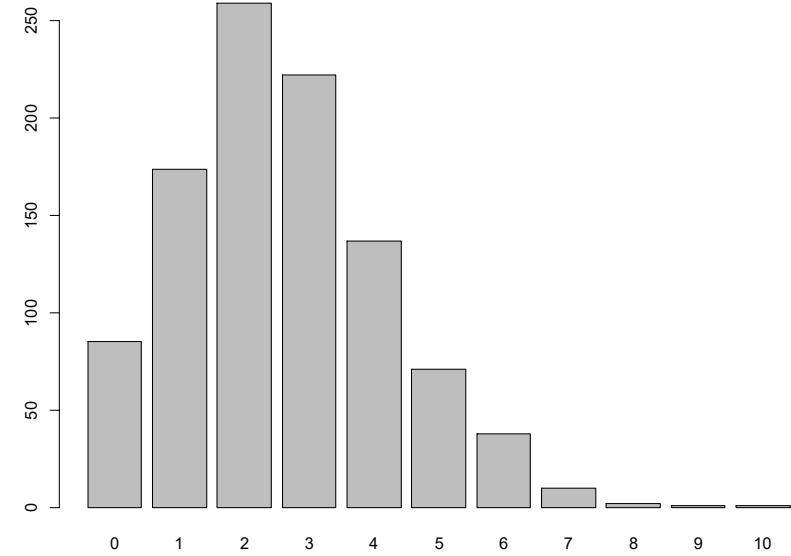
compte: R contempla el nombre de fracassos abans de **r** èxits, no el nombre total d'intents

Galeria

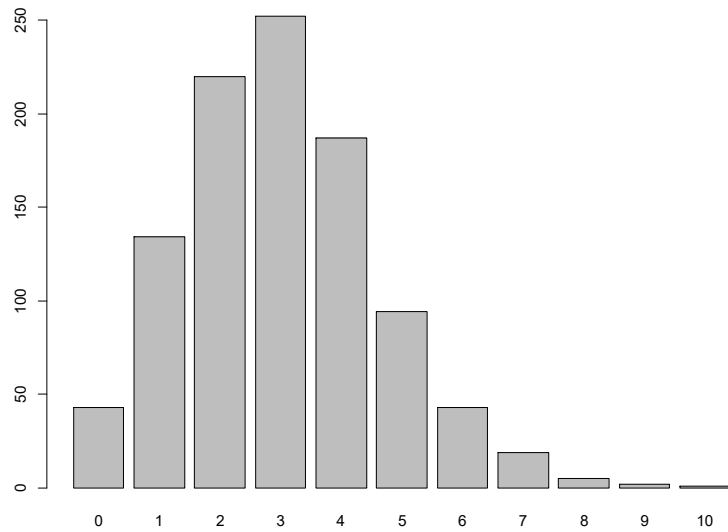
Histogram of rnorm(1000)



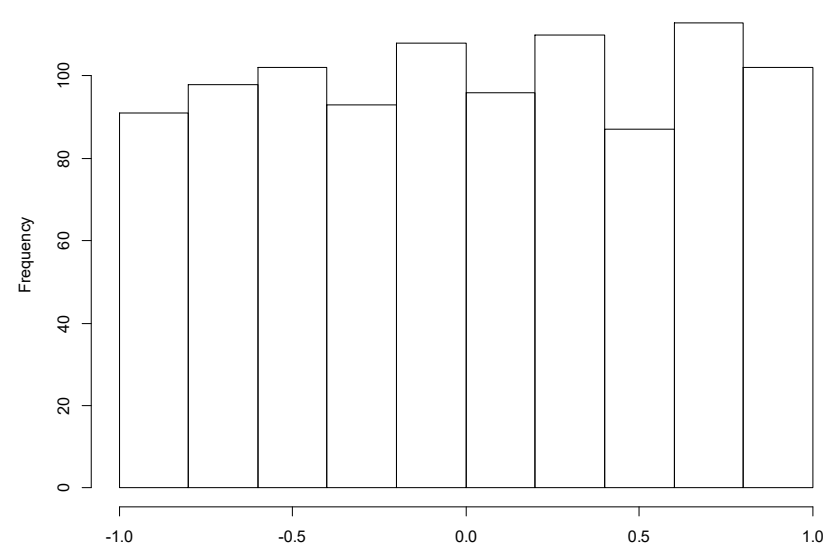
rpois(1000, 2.5)



rbinom(1000, 30, 0.1)



Histogram of runif(1000, -1, 1)



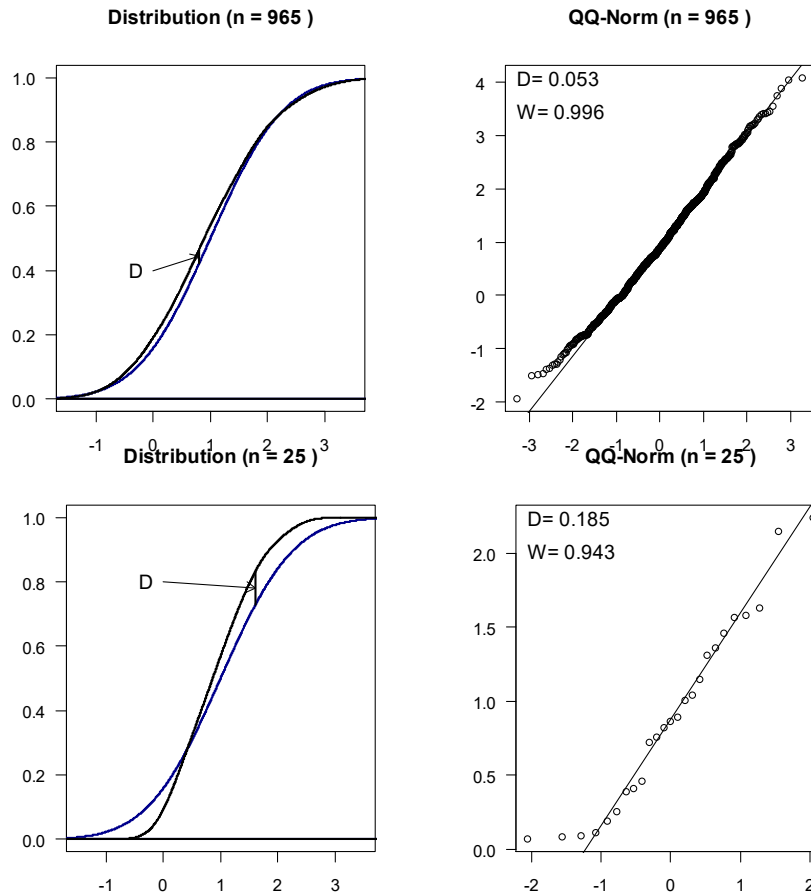
Anàlisi gràfica de Normalitat

- Hi ha dues mesures que ajuden a valorar el grau d'ajustament, afinitat o similitud a una certa distribució de referència.

Kolmogorov-Smirnov (Estadístic D)	Shapiro-Wilk (Estadístic W)
<p>Distància màxima entre la funció de distribució empírica i la teòrica.</p> <p>Valors elevats indiquen No Normalitat</p> <p>Entre 0 i 1 (usualment, prop de 0).</p> <p>Valors alts indiquen des-ajustament</p>	<p>Mesura la correlació entre els quantils observats i els teòrics.</p> <p>Valors elevats indiquen Normalitat</p> <p>Entre 0 i 1 (usualment, prop de 1).</p> <p>Valors alts indiquen bon ajustament</p>

- Totes dues mesures fluctuen a les mostres i han de interpretar-se amb prudència. En farem només una anàlisi descriptiva i visual (qqplot), especialment respecte la distribució Normal (QQnorm)
- A continuació mostrem alguns exemples generats a partir de distribucions conegudes i amb diferents mides mostrals.

Anàlisi gràfica de Normalitat



- Aquestes 965 observacions van estar generades seguint una distribució Normal
- D, W i el QQ-Norm mostren que les dades s'ajusten prou bé a una Normal

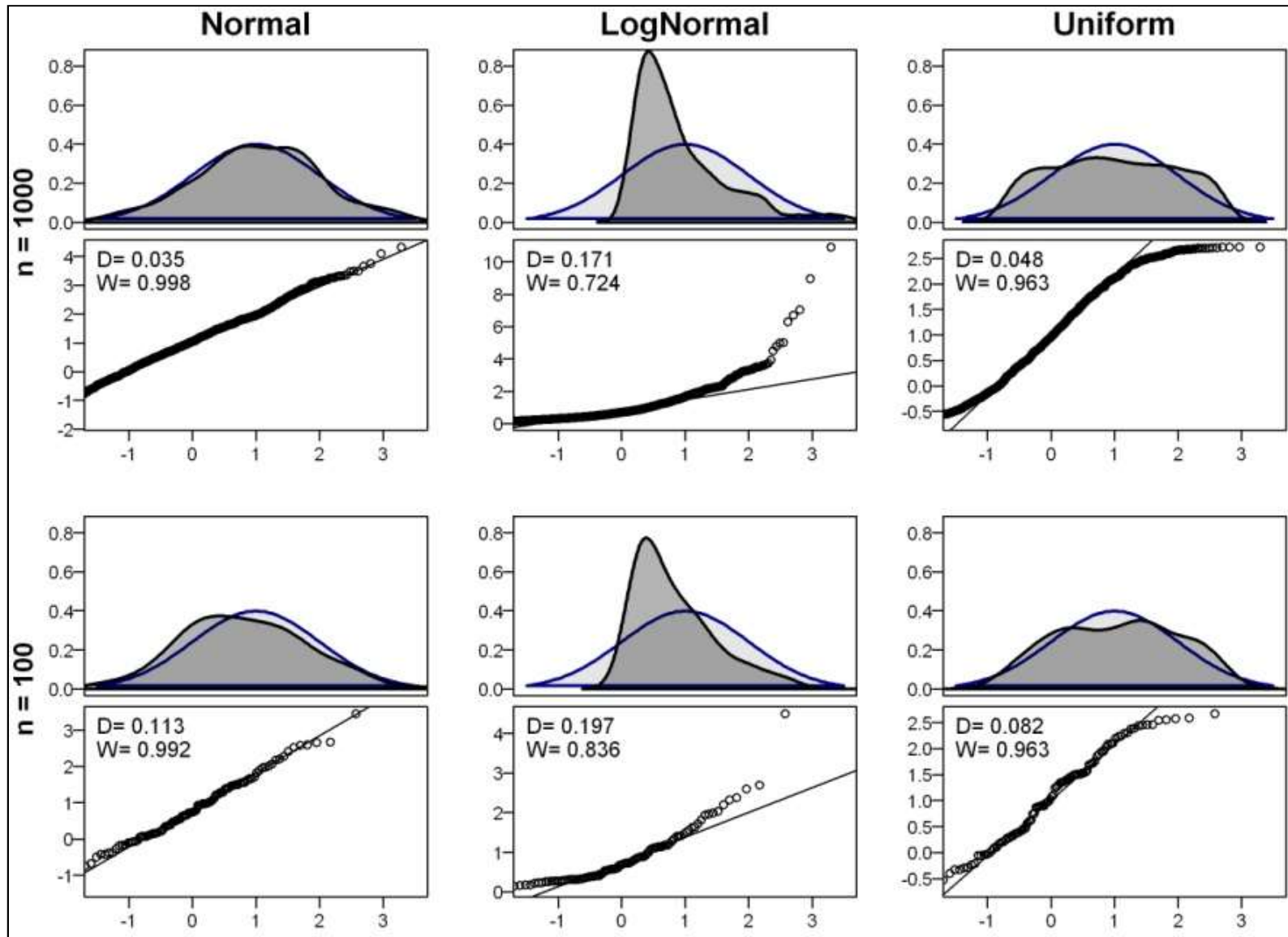
```
x <- rnorm(965)
qqnorm(x)
qqline(x)
```

- Aquestes 25 observacions van estar generades seguint una distribució Exponencial
- F_X i D mostren la distància amb la teòrica, encara que W té una bona correlació.

```
x <- rexp(25)
qqnorm(x)
qqline(x)
```

Noteu la paradoxa: quan la mostra és petita, és important detectar la No Normalitat. En canvi, quan la mostra és gran, és poc important detectar-la (pel bon comportament asimptòtic)

Anàlisi gràfica de Normalitat (n gran)



Anàlisi gràfica de Normalitat (n petita)

