

Model predictiu d'actualitzacions software

Daniel Massaguer

Jordi Massaguer



A65.cat

Programa Torres Quevedo



Convocatoria
financiada
por el:



FONDO
SOCIAL
EUROPEO

Problema

El cost de mantenir *software appliances* i distribucions Linux depèn de de les actualitzacions dels diferents components.

Quan serà la següent actualització ?



Dificultats

Les actualitzacions depenen de **factors no observables** i que **canvien al llarg del temps** com:

- Dedicació dels autors:
 - Temps complert
 - Temps parcial
 - Temps lliure
- Experiència (software + area)
- Metodologies i eines (e.g., junit, Metodologies àgils)
- Nombre de programadors
- Motivació
- etc

La relació entre actualitzacions i altres **factors observables** es desconeix:

- Quan va ser la última actualització
- Actualització d'altres components
- Maduresa del projecte
- Tipus d'actualització (e.g., de seguretat)
- Tipus de projecte
- etc

$x_i =$ variable aleatòria

$X_i =$ 'temps entre actualitzacions del component i '

Basats en els factors observables, estudiarem 2 opcions de modelatge:
Model paramètric vs model no paramètric

Paramètric:

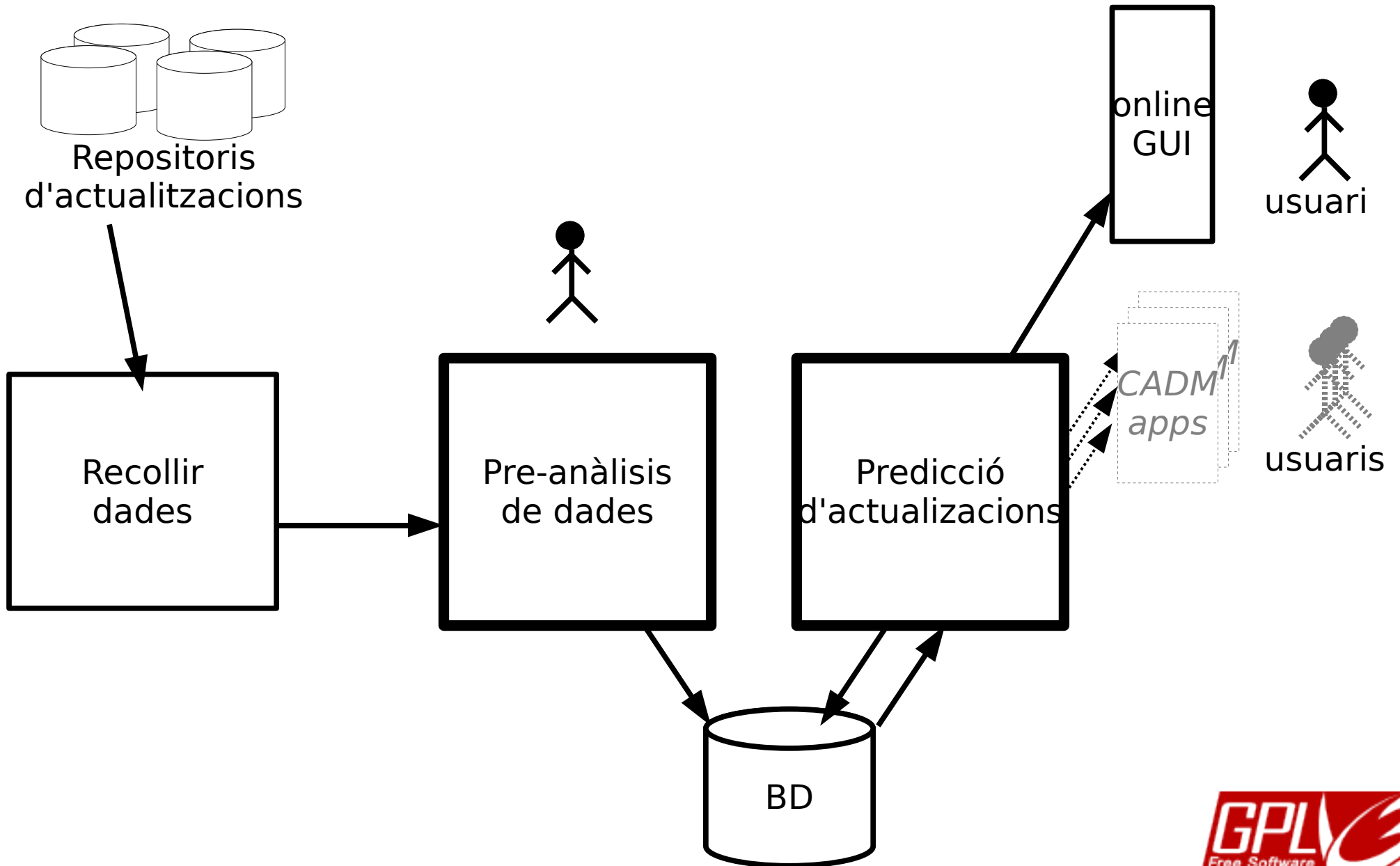
$x_i \sim$ Distribució Estadística

Mètodes estadístics + Machine Learning
per aprendre el tipus de distribució i els paràmetres

No paramètric:

Mètodes estadístics + Machine Learning

Implementació (J2EE + Web Services)



Estudi preliminar: openSUSE 11.0 - i586 - updates rep

Repository:

openSUSE 11.0 updates

(<http://download.opensuse.org/update/11.0/>)

Arquitectura:

i586

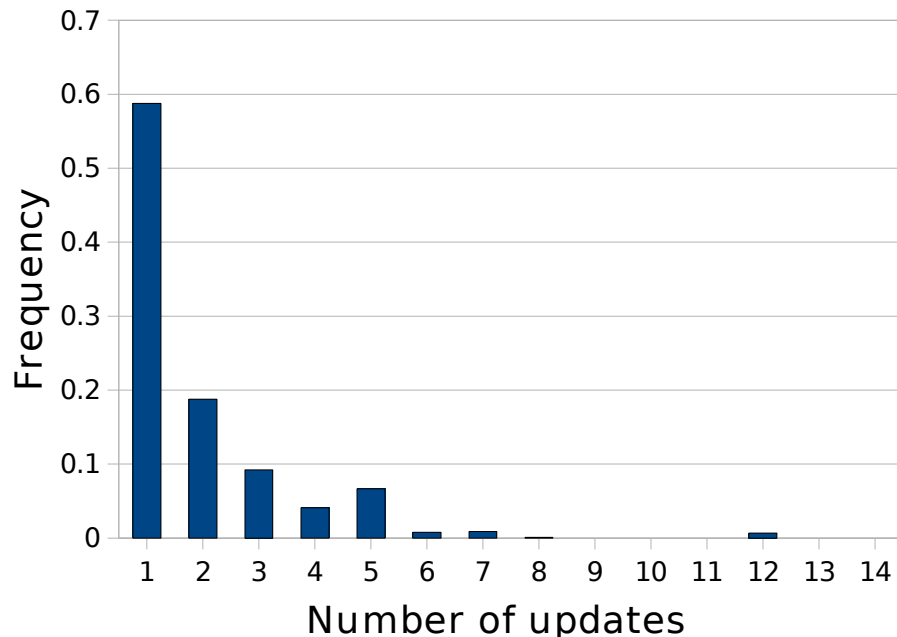
Període d'estudi:

06/06/08 - 06/20/09

(1 any després de la publicació oficial d'openSUSE 11.0)

Number of updates per module

openSUSE 11.0 - i586 - update rep



- **1.742** actualitzacions, **900** mòduls

- **4,53** actualitzacions/dia

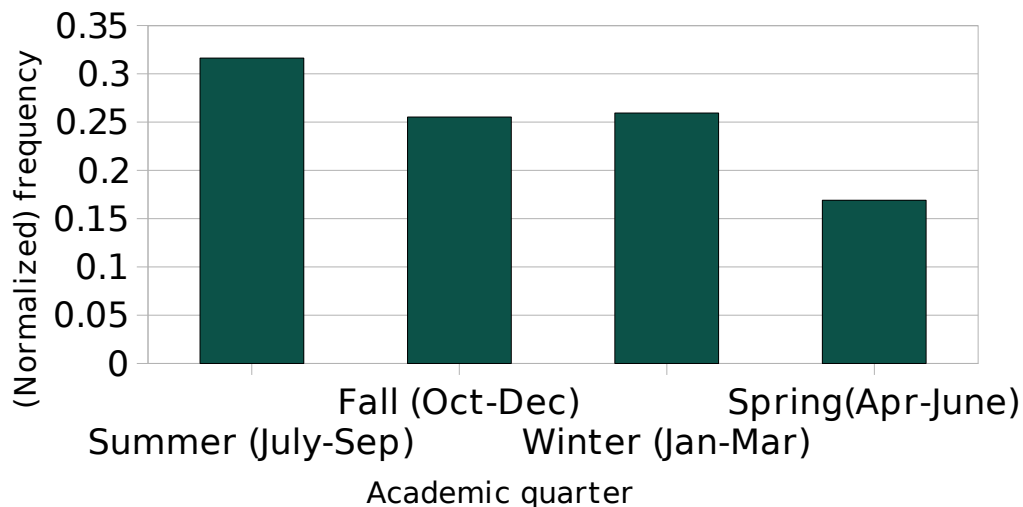
- **1,82** actualitzacions/any per mòdul

- **59%** dels mòduls només han tingut

1 actualització

Estudi preliminar: openSUSE 11.0 - i586 - updates rep

Updates per quarter
i586 - openSUSE 11.0



La majoria d'actualitzacions

(32%) a l'estiu:

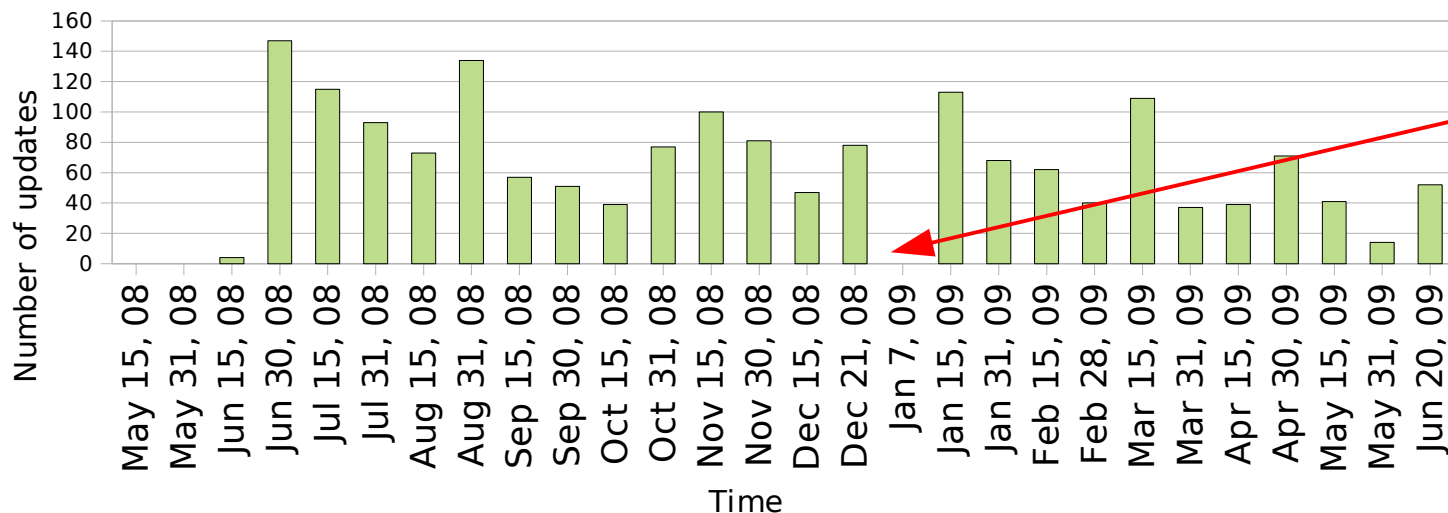
- L'estiu és l'estació més productiva?

o

- Com més aprop de la data de publicació (19 de Juny) més actualitzacions?

Number of updates for 1/2 month

i586 - openSUSE 11.0



0 actualitzacions
durant les vacances
de Nadal
(22/12/08 - 07/01/09)

Estudi preliminar: openSUSE 11.0 - i586 - updates rep

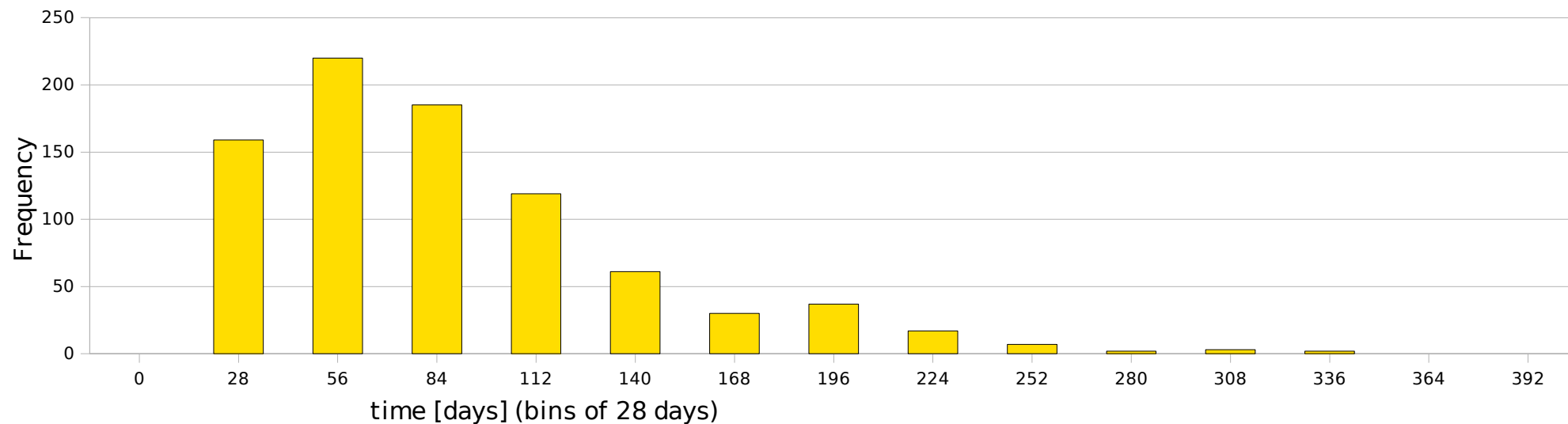
X = 'temps entre actualitzacions d'un component'

\bar{x} = 73,5 dies

stdev(x) = 53,19 dies

75% : [21,4 dies , 4,6 mesos]

Time between updates
i586 - openSUSE 11.0



*Estadístiques calculades amb aquells components que tenen més d'una actualització (41%).

Estudi preliminar: openSUSE 11.0 - i586 - updates rep

X = 'temps entre actualitzacions d'un component'

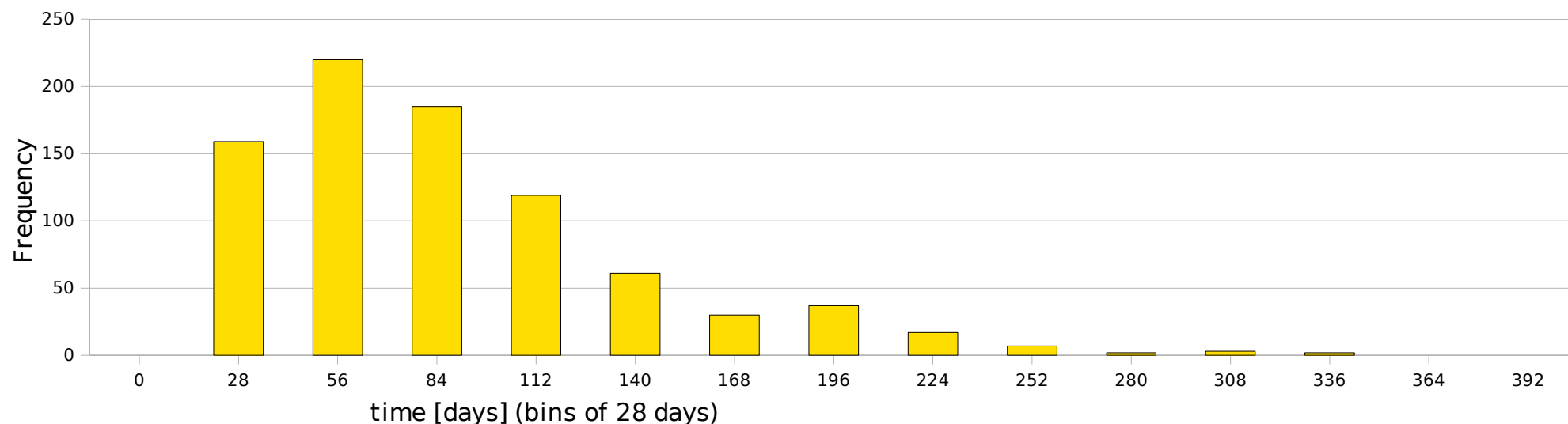
\bar{x} = 73,5 dies

stdev(x) = 53,19 dies

75% : [21,4 dies , 4,6 mesos] → *massa variabilitat!*

→ *considerar que tots els components es comporten igual (x) no es viable.* Time between updates

i586 - openSUSE 11.0



*Estadístiques calculades amb aquells components que tenen més d'una actualització (41%).

Conclusions

Aquest projecte d'R+D es centra en:

- Crear un **model predictiu** de quan seran les pròximes **actualitzacions** dels components que formen una *software appliance* o distribució concreta.
- Implementar una aplicació *on-line* amb el model predictiu.

Lo qual permetrà

- **Planejar** el manteniment de *software appliances* i distribucions Linux.
- Construir **models de negoci** apropiats.

Extra slides

Estudi preliminar: openSUSE 11.0 - i586 - updates rep

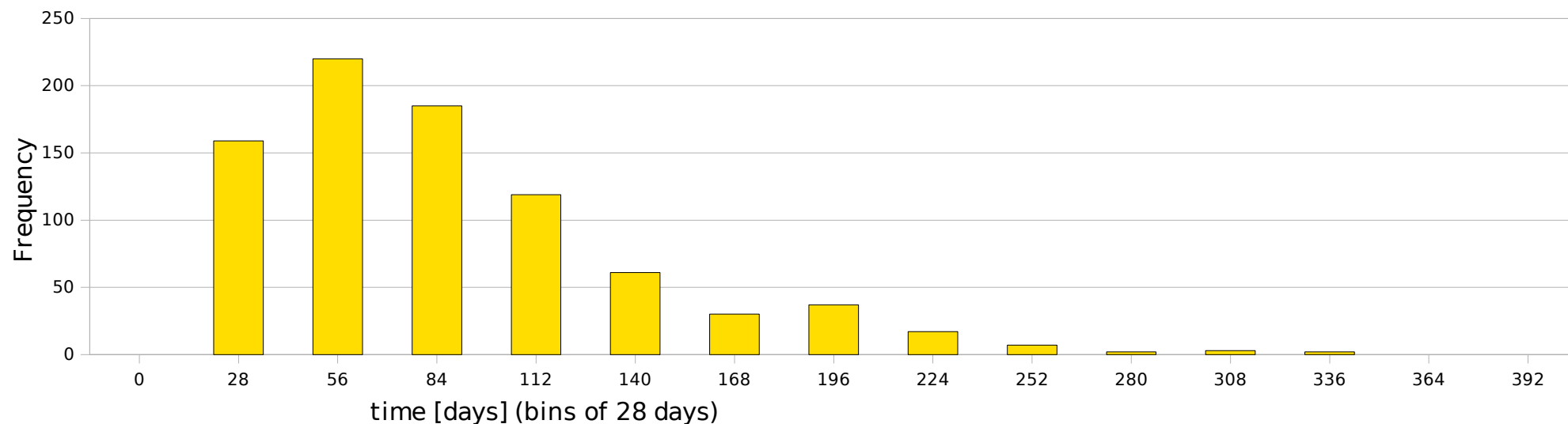
$X =$ 'temps entre actualitzacions d'un component'*

$\bar{x} = 73.5$ dies

stdev(x) = 53.19 dies

Time between updates

i586 - openSUSE 11.0



~~$x \sim \exp(\lambda=1/\bar{x})^{**}$~~ → No és un procés Poisson

*assumim que el "temps entre updates" de tots els components es comporten de forma similar. Només tenim en compte aquells components amb més d'una actualització.

**segons els tests d'Anderson-Darling, Kolmogorov-Smirnov, i Chi-squared