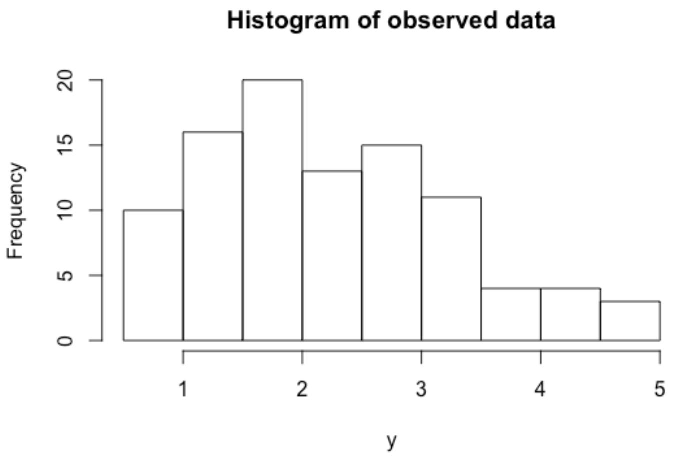
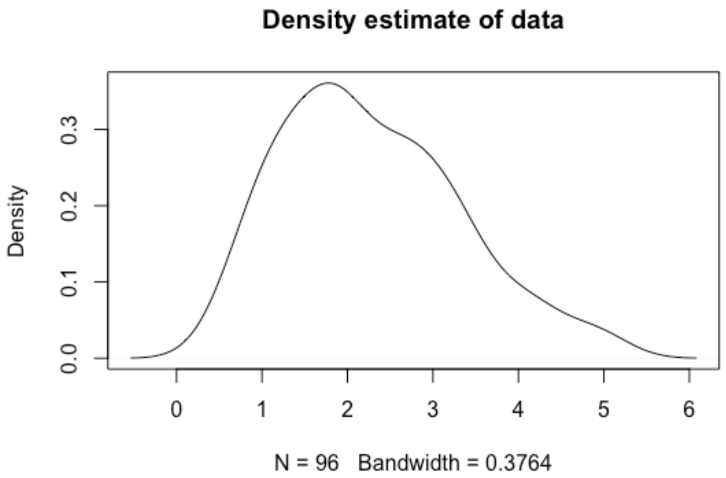
Test loi de poisson

y : le nombre d'ALDs déclarées pour 100000 personnes dans le département.

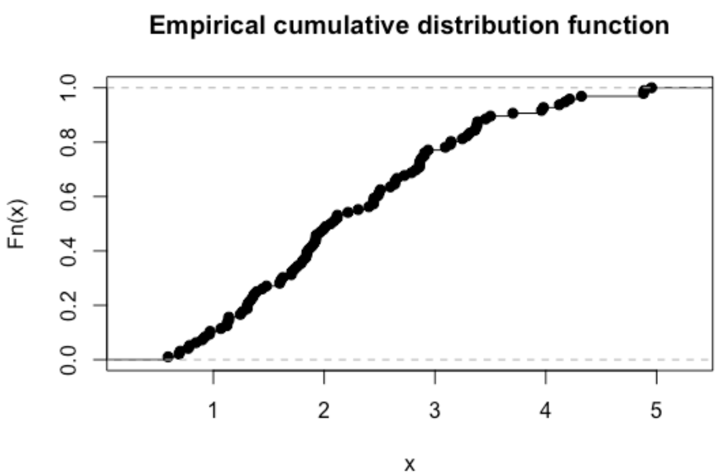
```
hist(y,main="Histogram of observed data")
```




```
plot(density(y),main="Density estimate of data")
```




```
plot(ecdf(y), main="Empirical cumulative distribution function")
```



Le test du *chi-square* est défini pour l'hypothèse:

H0: les données suivent une distribution spécifiée

HA: les données ne suivent pas la distribution spécifiée

Pour le calcul de l'ajustement du *chi-square*, les données sont divisées en *k bins* et la statistique de test est définie de cette manière:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

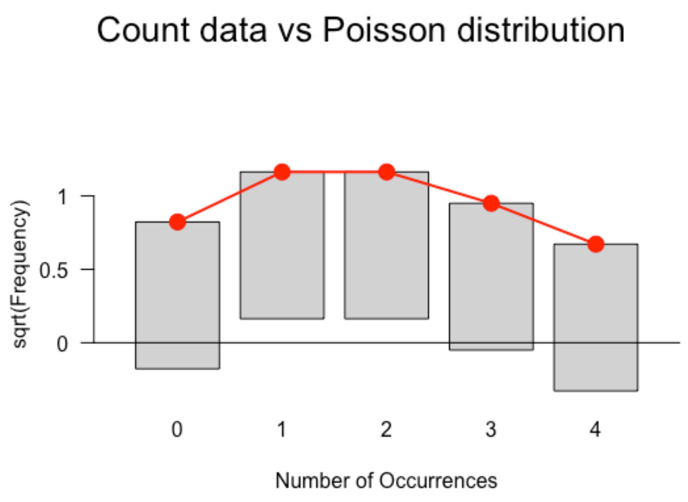
où O_i est la fréquence observée pour le *bin i* et E_i est la fréquence attendue pour le *bin i*.
La fréquence attendue est calculée par la fonction de distribution cumulative. Cette statistique est distribuée sous la forme d'une variable aléatoire χ^2 avec *k-p-1* degrés de liberté (*p* est le nombre de paramètres estimés par les données de l'échantillon). L'hypothèse selon laquelle les données proviennent d'une population avec la distribution spécifiée est acceptée si χ^2 est inférieur à la fonction du point de pourcentage du *chi-square* avec *k-p-1* degrés de liberté et un niveau de signification de α. H0 est accepté car la valeur de pvalue est supérieure à un niveau de signification fixé au moins à 5%.
Dans l'environnement R, en cas de données de comptage, nous pouvons utiliserun fonction `goodfit()`

```
> install.packages("vcd")
> library(vcd) ## loading vcd package
> gf = goodfit(y,type= "poisson",method= "ML")
> summary(gf)

      Goodness-of-fit test for poisson distribution

                      X^2 df  P(> X^2)
Likelihood Ratio 1.368598  3 0.7129117

> plot(gf,main="Count data vs Poisson distribution”)
```




```
> # to automatically get the pvalue
> gf.summary = capture.output(summary(gf))[[5]]
> pvalue = unlist(strsplit(gf.summary, split = " "))
```

```
> pvalue = as.numeric(pvalue[length(pvalue)]); pvalue
[1] 0.7129117

> # to mannualy compute the pvalue
> chisq = sum(  (gf$observed-gf$fitted)^2/gf$fitted )
> df = length(gf$observed)-1-1
> pvalue = pchisq(chisq,df); pvalue
[1] 0.2029661
```

Nous ne pouvons pas rejeter l'hypothèse nulle car la valeur de *p-value* est plutôt élevée, donc les données d'échantillon appartiennent probablement à une distribution de poisson avec paramètre $\lambda = \text{mean}(y)$

```
> mean(y)
[1] 2.280301

> qqPlot(y, distribution = "pois", lambda = mean(yy), xlab = 'Theoretical Quantiles', ylab = 'Empirical Quantiles', main = 'Q-Q plot Poisson')
```