

Introducció Bootstrap. Sessio 4 (Bootstrap Confidence Intervals for the mean)

Jordi Ocaña, Sergi Civit,

27 d'abril de 2018

1 Introducció: Situació de laboratori

Construcció d'un interval de confiança per la mitjana d'una distribució normal. Comparació amb el corresponent interval t de Student

Aquesta mostra fa el paper d'unes "dades reals" però en realitat sabem que procedeix d'una $N(15, 3)$

2 Mostra

```
> # Lectura de les dades PROCEDENTS D'UNA NORMAL
> x <- c(15.54, 21.06, 16.52, 13.62, 16.14, 10.98, 13.53, 16.02, 16.79, 15.90)
> n <- length(x)
> sqrt.n <- sqrt(n)
```

2.1 Càlculs BÀSICS

ATENCIÓ: ARA NO "EMPRAREM" AQUESTA INFORMACIÓ EN LA DETERMINACIÓ DEL INTERVAL DE CONFIANÇA BOOTSTRAP-T DONAT QUE EN UN CAS REAL LÒGICAMENT HO DESCONEIXERÍEM

Obtenim:

- Estimació de la mitjana sobre la "mostra real x "
- Estimació de la desviació estàndar sobre la "mostra real x "
- Estimació de l'error estàndard de l'estimador de la mitjana

```
> # Estimació de la mitjana sobre la muestra "real" x:
> media.estim <- mean(x)
> media.estim
```

```
[1] 15.61
```

```

> # Estimació de la desviació estàndar sobre la mostra "real" x:
> sd(x)

[1] 2.62857

> # Estimació de l'error estàndard del estimador de la mitjana:
> sd(x)/sqrt.n

[1] 0.8312267

```

3 PAS 0: Bootstrap NO PARAMÈTRIC de la distribució mostral de l'estadístic t

Aquesta és l'expressió i càlcul de l'estadístic tal i com ho fem de forma usual.

$$t = (media.estim - 15)/se.media$$

Aquesta és l'aproximació bootstrap no paramètric a la distribució muestral de l'estadístic t. Mitjançant simulació de la distribució de t^*

$$t^* = (mean(x.boot) - media.estim)/(sd(x.boot)/sqrt.n)$$

```

> B <- 10000
> t.boot <- replicate(B,
+   {
+       x.boot <- sample(x, replace = TRUE)
+       sqrt.n * (mean(x.boot) - media.estim) / sd(x.boot)
+   }
+ )
> t.boot[1:10]

[1] 0.1074243 0.8671038 -1.3466293 1.1954494 -0.7370950 1.0957835
[7] -1.3201247 -0.2478574 0.4789537 -0.1976826

```

3.1 Bootstrap NO PARAMÈTRIC de la distribució de l'estadístic t: Valors crítics

Obtenim valors crítics que deixen probabilitat $\alpha/2$ en la cua esquerra i dreta de la distribució bootstrap de l'estadístic t:

```

> alfa <- 0.05
> # Valors crítics que deixen probabilidad alfa/2
> # en la cua esquerra i dreta de la distribució
> # bootstrap de l'estadístic t:
>
> quantile(t.boot, probs=c(alfa/2, 1 - alfa/2))

```

```

      2.5%      97.5%
-2.285952  2.096934

> # Comparem-ho amb els corresponents valors t de Student:
> qt(c(alfa/2, 1 - alfa/2), df = n - 1)

[1] -2.262157  2.262157

```

4 Bootstrap NO PARAMÈTRIC: Interval de Confiança bootstrap-t de CUES EQUIPROBABLES

COMPTE: Fixeu-vos el canvi d'ordre dels quantils.

Valor crític inferior: $\text{quantile}(1 - \alpha/2)$

Valor crític superior $\text{quantile}(\alpha/2)$.

Hint: Desenvolupar $IC(\mu)$ *σ coneguda*

```

> # Interval de confiança bootstrap-t de CUES EQUIPROBABLES, amb nivell de
> # confiança nominal 1 - alfa (= 0.95)
> # Fixeu-vos el canvi d'ordre dels quantils
>
> icBoot.t <- media.estim -
+   quantile(t.boot, probs=c(1 - alfa/2, alfa/2)) * sd(x) / sqrt.n
> names(icBoot.t) <- NULL
> attr(icBoot.t, "conf.level") <- 0.95
> icBoot.t

[1] 13.86697 17.51014
attr(,"conf.level")
[1] 0.95

> # Comparació AMB Intervalo de Confianza per la mitjana a partir
> # de suposició de normalitat --> distribución muestral t de Student
>
> t.test(x, var.equal = TRUE)$conf.int

[1] 13.72963 17.49037
attr(,"conf.level")
[1] 0.95

>

```

Resultats: Els 2 intervals son molt similars

5 Bootstrap NO PARAMÈTRIC: Interval de confiança bootstrap-t SIMETRITZAT

Aquest Interval de Confiança bootstrap-t que anomenarem **simetritzat** (simètric respecte de l'estimació del paràmetre), PERO AMB LES CUES de la distribució de t^* desiguals pel que fa a la seva probabilitats (NO EQUIPROBABLES).

```
> # Aquest Interval de Confianza bootstrap-t que
> # anomenarem {\bf simetritzat} (simétric respecte de l'estimació
> # del paràmetre), PERO AMB LES CUES de la
> # distribució de  $t^*$  desiguals NO EQUIPROBABLES.
> # Determinaremos el valor "t_alfa" tal que
> #  $P\{|t^*| \leq t\_alfa\} = 1 - \alpha$ 
> t_alfa <- quantile(abs(t.boot), probs = 1 - alfa)
> t_alfa

      95%
2.20006

> # Interval de Confianza Bootstrap-t simétric:
> icBoot.t.sim <- media.estim - c(t_alfa, -t_alfa) * sd(x) / sqrt.n
> names(icBoot.t.sim) <- NULL
> attr(icBoot.t.sim, "conf.level") <- 0.95
> icBoot.t.sim

[1] 13.78125 17.43875
attr(,"conf.level")
[1] 0.95

> # Comparació AMB Intervalo de Confianza per la mitjana
> # a partir de suposició de normalitat --> distribució mostral t de Student
>
> t.test(x, var.equal = TRUE)$conf.int

[1] 13.72963 17.49037
attr(,"conf.level")
[1] 0.95

>
```

Resultats: Aquest intervals ENCARA és més similar al interval t de Student

6 Bootstrap PARAMÈTRIC: Interval de confiança bootstrap-t

Comencem amb l'Interval de Confiança bootstrap-t de CUES EQUIPROBABLES amb nivell de confiança nominal $1 - \alpha$ ($= 0.95$)

```

> sd.x = sd(x)
> t.boot <- replicate(B,
+ {
+   x.boot <- rnorm(n, mean = media.estim, sd = sd.x)
+   sqrt.n * (mean(x.boot) - media.estim) / sd(x.boot)
+ }
+ )
> # Interval de Confianza bootstrap-t de CUES EQUIPROBABLES
> # amb nivell de confiança nominal 1 - alfa (= 0.95)
>
> icBootParam.t <- media.estim -
+   quantile(t.boot, probs=c(1 - alfa/2, alfa/2)) * sd(x) / sqrt.n
> names(icBootParam.t) <- NULL
> attr(icBootParam.t, "conf.level") <- 0.95
> icBootParam.t

[1] 13.72716 17.48089
attr("conf.level")
[1] 0.95

> # Comparació AMB Interval de Confiança per la mitjana
> # a partir de suposició de normalitat --> distribución muestral t de Student
>
> t.test(x, var.equal = TRUE)$conf.int

[1] 13.72963 17.49037
attr("conf.level")
[1] 0.95

>

```

7 Bootstrap PARAMÈTRIC: Interval de confiança bootstrap-t SIMETRITZAT

Interval de Confiança Bootstrap-t simetritzat. Recordeu que hem de determinar el valor t_α tal que

$$P^*|t^*| \leq t_\alpha = 1 - \alpha$$

```

> # Interval de confiança bootstrap-t simetritzat.
> # Trobarem el valor "t_alfa" tal que
> #  $P\{|t^*| \leq t_{\alpha}\} = 1 - \alpha$ 
>
> t_alfa <- quantile(abs(t.boot), probs = 1 - alfa)
> icBootParam.t.sim <- media.estim - c(t_alfa, -t_alfa) * sd(x) / sqrt.n
> names(icBootParam.t.sim) <- NULL
> icBootParam.t.sim

```

```
[1] 13.73458 17.48542

> # Recordem, IC paramètric t de Student:
> t.test(x, var.equal = TRUE)$conf.int

[1] 13.72963 17.49037
attr(,"conf.level")
[1] 0.95
```

8 Construcció d'Interval de confiança Bootstrap per la mitjana d'una distribució EXPONENCIAL.

Construcció d'un interval de confiança per la mitjana d'una distribució EXPONENCIAL. Comparació amb el corresponent interval t de Student

Aquesta mostra fa el paper d'unes "dades reals" però en realitat sabem que procedeix d'una EXPONENCIAL(15), $Exp(\mu = 15)$

8.1 Mostra

```
> # Lectura de les dades PROCEDENTS D'UNA EXPONENCIAL
> x <- c(8.51, 8.71, 69.19, 10.05, 23.64, 8.67, 1.51, 20.36, 1.23, 5.27)
> n = length(x)
> sqrt.n <- sqrt(n)
```

8.2 Càlcul BÀSIC

ATENCIÓ: ARA NO "EMPRAREM" AQUESTA INFORMACIÓ EN LA DETERMINACIÓ DELS INTERVALS DE CONFIANÇA BOOTSTRAP DONAT QUE EN UN CAS REAL LÒGICAMENT HO DESCONEIXERÍEM

Obtenim:

- Estimació de la mitjana sobre la "mostra real x"

```
> # Estimació de la mitjana sobre la muestra "real" x:
> media.estim <- mean(x)
> media.estim

[1] 15.714
```

9 Bootstrap NO PARAMÈTRIC: Interval de Confiança bootstrap-t de CUES EQUIPROBABLES

```
> t.boot <- replicate(B,
+ {
```

```

+             x.boot <- sample(x, replace = TRUE)
+             sqrt.n * (mean(x.boot) - media.estim) / sd(x.boot)
+         }
+     )
> # Valors crítics en la "tabulació" de la distribució
> # mostrat bootstrap de l'estadístic:
>
> quantile(t.boot, probs = c(alfa/2, 1 - alfa/2))

      2.5%      97.5%
-7.973394  1.481645

> # NO S'ASSEMBLEN alss corresponents valors de la distribució t de Student:
> qt(c(alfa/2, 1 - alfa/2), df = n - 1)

[1] -2.262157  2.262157

> # Interval de confiança bootstrap-t
> icBoot.t <- media.estim - quantile(t.boot, probs=c(1 - alfa/2, alfa/2)) * sd(x) / sqrt.n
> names(icBoot.t) <- NULL
> attr(icBoot.t, "conf.level") <- 0.95
> icBoot.t

[1]  6.281819 66.472784
attr("conf.level")
[1] 0.95

> # Molt diferent de l'interval de confiança t de Student
> t.test(x, var.equal = TRUE)$conf.int

[1]  1.313062 30.114938
attr("conf.level")
[1] 0.95

```

10 Bootstrap NO PARAMÈTRIC: Interval de Confiança bootstrap-t SIMETRITZAT

En realitat fixeuvos que la probabilitat es concentra en la cua esquerra

```

> # Interval de confiança bootstrap-t simetritzat
> t_alfa <- quantile(abs(t.boot), probs = 1 - alfa)
> # A cada cua deixa la següent probabilitat:
> sum(t.boot < -t_alfa) / B

[1] 0.05

> sum(t.boot > t_alfa) / B

```

```

[1] 0

> # En realitat fixeuvos que la probabilitat es concentra en la cua esquerra
>
> icBoot.t.sim <- media.estim - c(t_alfa, -t_alfa) * sd(x) / sqrt.n
> names(icBoot.t.sim) <- NULL
> attr(icBoot.t.sim, "conf.level") <- 0.95
> icBoot.t.sim

[1] -13.77555 45.20355
attr(,"conf.level")
[1] 0.95

> # Molt diferent de l'interval de confiança t de Student
> t.test(x, var.equal = TRUE)$conf.int

[1] 1.313062 30.114938
attr(,"conf.level")
[1] 0.95

>

```

11 Bootstrap PARAMÈTRIC: Interval de Confiança bootstrap-t de CUES EQUIPROBABLES

Interval de confiança bootstrap-t

```

> t.boot <- replicate(B,
+ {
+   x.boot <- rexp(n, rate = 1 / media.estim)
+   sqrt.n * (mean(x.boot) - media.estim) / sd(x.boot)
+ }
+ )
> # Valors crítics en la "tabulació" de la distribució
> # mostrat bootstrap de l'estadístic:
>
> quantile(t.boot, probs = c(alfa/2, 1 - alfa/2))

      2.5%      97.5%
-4.126604  1.621250

> # NO S'ASSEMBLEN alss corresponents valors de la distribució t de Student:
> qt(c(alfa/2, 1 - alfa/2), df = n - 1)

[1] -2.262157  2.262157

```



```

> # Interval de confiança bootstrap-t
> icBoot.t <- media.estim - quantile(t.boot, probs=c(1 - alfa/2, alfa/2)) * sd(x) / sqrt.n
> names(icBoot.t) <- NULL
> attr(icBoot.t, "conf.level") <- 0.95
> icBoot.t

[1] 5.393088 41.984041
attr("conf.level")
[1] 0.95

> # Molt diferent de l'interval de confiança t de Student
> t.test(x, var.equal = TRUE)$conf.int

[1] 1.313062 30.114938
attr("conf.level")
[1] 0.95

```

12 Bootstrap PARAMÈTRIC: Interval de Confiança bootstrap-t SIMETRITZAT

Interval de confiança bootstrap-t simetritzat

En realitat fixeu-vos que la probabilitat es concentra en la cua esquerra

```

> # Interval de confiança bootstrap-t simetritzat
> t_alfa <- quantile(abs(t.boot), probs = 1 - alfa)
> # A cada cua deixa la següent probabilitat:
> sum(t.boot < -t_alfa) / B

[1] 0.0497

> sum(t.boot > t_alfa) / B

[1] 3e-04

> # En realitat fixeu-vos que la probabilitat es concentra en la cua esquerra
>
> icBoot.t.sim <- media.estim - c(t_alfa, -t_alfa) * sd(x) / sqrt.n
> names(icBoot.t.sim) <- NULL
> attr(icBoot.t.sim, "conf.level") <- 0.95
> icBoot.t.sim

[1] -3.87603 35.30403
attr("conf.level")
[1] 0.95

> # Molt diferent de l'interval de confiança t de Student
> t.test(x, var.equal = TRUE)$conf.int

```

```
[1] 1.313062 30.114938
attr(,"conf.level")
[1] 0.95
>
```

13 Conclusions

- Quin d'ells és més correcte?

Quan la veritable distribució de les dades no és normal, possiblement els intervals **bootstrap-t** són preferibles ja que l'enfocament **bootstrap** aproxima millor la veritable distribució mostral de

$$t = (media.estim - 15)/se.media$$

és a dir, són més correctes els valors crítics obtinguts en la "tabulació bootstrap de l'estadístic" que els obtinguts en la taula de la "t de Student amb n - 1" g.d.l.

- Simulació per tal d'estudiar recobriment dels IC
 - Simulació de 10000 valors t procedents de $rexp(n, rate = 1/15)$
 - COMPTE: PARÀMETRE FIXAT PER NOSALTRES, NO ESTIMAT A PARTIR DE LES DADES.
 - Com ho farem?
1. Per cada mostra exponencial simulada
 2. Calcularem els 4 intervals bootstrap (no paramètric i paramètric)(no simetritzat i simetritzat)
 3. Calcularem l'intervalo Student t,
 4. Determinarem si **recobreixen el veritable valor de la mitjana, que és 15**

14 Estudi de simulació

```
> nsim = 1000 # COMPTE HAURIA DE SER 10.000
> # Generació de 'nsim' mostres procedents dels següents paràmetres
> # coneguts, fixats per nosaltres:
> media <- 15
> expRate <- 1 / media
> # Sobre cadascuna d'aquestes mostres es calculassin diversos intervals de
> # confiança bootstrap. Aquests intervals requeriran la generació de
> # 'B' remuostres bootstrap:
```

```

> B <- 100 # COMPTE HAURIA DE SER 1000
> # És a dir, estem fent una simulació de grandària total
> # 10000 x 1000. Trigarà bastant.....
>
> # Estudiarem les propietats d'interval de nivell de confiança
> # "nominal" (és a dir, que almenys en teoria és) 0.95.
> alfa <- 0.05
> alfa1 <- alfa / 2
> alfa2 <- 1 - alfa1
> probColas <- c(alfa2, alfa1)
> conf.level <- 1 - alfa
> set.seed(4317)
> coverages <- replicate(nsim,
+   {
+       x.sim <- rexp(n, rate = expRate)
+       media.estim <- mean(x.sim)
+       se.media <- sd(x.sim) / sqrt.n
+       # Bootstrap no paramètric:
+       t.boot <- replicate(B,
+   {
+       x.boot <- sample(x.sim, replace = TRUE)
+       sqrt.n * (mean(x.boot) - media.estim) / sd(x.boot)
+   }
+   )
+   # IC boot-t
+   icBoot.t <- media.estim - quantile(t.boot, probs = probColas) * se.media
+   # IC boot-t simetritzat:
+   t_alfa <- quantile(abs(t.boot), probs = conf.level)
+   icBoot.t.sim <- media.estim - c(t_alfa, -t_alfa) * se.media
+
+   # Bootstrap paramètric:
+   estim.expRate = 1 / media.estim
+   t.boot.param <- replicate(B,
+   {
+       x.boot <- rexp(n, rate = estim.expRate)
+       sqrt.n * (mean(x.boot) - media.estim) / sd(x.boot)
+   }
+   )
+   # IC boot-t paramètric
+   icBootParam.t.sim <- media.estim - quantile(t.boot.param, probs = probColas) * se.media
+   # IC boot-t simetritzat paramètric:
+   t_alfa <- quantile(abs(t.boot.param), probs = conf.level)
+   icBootParam.t.sim <- media.estim - c(t_alfa, -t_alfa) * se.media
+
+   # IC t de Student:
+   ic.t <- t.test(x.sim, var.equal = TRUE)$conf.int

```

```

+
+      # Recubriment?:
+      c(media >= c(icBoot.t[1], icBoot.t.sim[1], icBootParam.t[1],
+                  icBootParam.t.sim[1], ic.t[1])) &
+      c(media <= c(icBoot.t[2], icBoot.t.sim[2], icBootParam.t[2],
+                  icBootParam.t.sim[2], ic.t[2]))
+    }
+
+ )
> rownames(coverages) <- c("boot-t NP", "boot-t NP SIM",
+   "boot-t P", "boot-t P SIM", "t Student")
> coverages[,1:10]

```

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
boot-t NP	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
boot-t NP SIM	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
boot-t P	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
boot-t P SIM	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE
t Student	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE

```

> # Estimació del veritable nivell de confiança:
> apply(coverages, 1, mean)

```

boot-t NP	boot-t NP SIM	boot-t P	boot-t P SIM	t Student
0.922	0.917	1.000	0.935	0.892

15 Cas Estudi

Nombre d'espècies de dinosaures per família observades durant tres períodes del Cretaci a Dakota del Nord i Montana (Sheehan, P.M., Fastovsky, D.E., Hoffmann, R.G., Berghaus, C.B., Gabriel, D.L. (1991). Sudden extinction of the dinosaurs: Latest Cretaceous, Upper Great Plains, U.S.A. Science, 254, 835-839)

```

> family <- factor(c("Ceratops", "Hadro", "Pachycephalo", "Tyranno", "Ornithomo",
+                   "Sauronith"),
+                 levels=c("Ceratops", "Hadro", "Pachycephalo", "Tyranno",
+                           "Ornithomo", "Sauronith"))
> dinos <- data.frame(
+   upper = c(50, 29, 3, 3, 4, 1),
+   middle = c(53, 51, 2, 3, 8, 6),
+   lower = c(19, 7, 1, 2, 1, 3)
+ )
> rownames(dinos) <- family
> dinos

```

	upper	middle	lower
Ceratops	50	53	19
Hadro	29	51	7
Pachycephalo	3	2	1
Tyranno	3	3	2
Ornithomo	4	8	1
Sauronith	1	6	3

Si la biodiversitat (en termes de nombre d'espècies per família) havia anat decreixent al llarg dels tres períodes, possiblement es podria considerar un índex que els dinosaures estaven immersos en una crisi”que ja venia de lluny. La confrontació d'aquesta possibilitat, davant **la hipòtesi que la biodiversitat s'havia mantingut estable**, requeriria mesurar la diversitat mitjançant un índex objectiu, com ara **l'índex de Shannon - Margalef**.

Proposat per Shannon en el context de la teoria de la informació, Margalef va proposar d'utilitzar-lo en Ecologia com a mesura de biodiversitat.

16 Revisiting the Dinosaur extinction. Authors MNPR 17-18

A progressive decrease of biodiversity across the three periods would agree with an hypothesis of dinosaur progressive decline, while biodiversity stability would be more consistent with the possibility of a catastrophic event as the cause of dinosaur extinction.

Contrasting these possibilities requires the use of an objective biodiversity index, like the Shannon information measure, whose use in a biodiversity context was firstly proposed by Margalef.

Function evaluating the Shannon index from an absolute frequencies vector. These absolute frequencies are internally converted into relative frequencies. In fact, any vector of non-negative values (e.g. surface covered by distinct species) is admissible to compute the index.

```
> #
> shannon <- function(freq.abs)
+ {
+     freq.rels = freq.abs / sum(freq.abs)
+     - sum(ifelse(freq.rels > 0, freq.rels * log2(freq.rels), 0))
+ }
```

Examples of its use:

```
> # Diversity over a given period:
> shannon(dinos[, "middle"])
```

```
[1] 1.746269
```

```

> shannon(dinos[,2])

[1] 1.746269

> # Diversity for all periods:
> apply(dinos, 2, shannon)

      upper   middle   lower
1.596475 1.746269 1.798425

> vapply(dinos, shannon, FUN.VALUE = 0.0)

      upper   middle   lower
1.596475 1.746269 1.798425

```

Biodiversity seems to be stable, even to increase...

In the before mentioned paper, the authors take the approach of testing for differences between the biodiversities in all the periods (that is to say, H0: all diversities are equal vs H1: almost one pair of diversities differ

17 Inference on a single diversity measure

Let us make some inference over these data. First, we will compute confidence intervals for the biodiversity.

We include, as an option, the computation of the standard error of the Shannon index in function 'shannon':

```

> shannon <- function(freq.abs, compute.se = TRUE)
+ {
+   n <- sum(freq.abs)
+   freq.rels = freq.abs / n
+   log2freq.rels <- ifelse(freq.rels > 0, log2(freq.rels), 0)
+   result <- - sum(freq.rels * log2freq.rels)
+   if (compute.se) {
+     attr(result, "se") <- sqrt((sum(freq.rels * log2freq.rels^2) - result^2) / n)
+   }
+   return(result)
+ }
> biodivs <- lapply(dinos, shannon)
> biodivs

$upper
[1] 1.596475
attr(,"se")
[1] 0.1366722

$middle

```

```

[1] 1.746269
attr(,"se")
[1] 0.1091939

$lower
[1] 1.798425
attr(,"se")
[1] 0.234451

> unlist(biodivs)

      upper      middle      lower
1.596475 1.746269 1.798425

> sapply(biodivs, attr, "se")

      upper      middle      lower
0.1366722 0.1091939 0.2344510

> biodivs <- sapply(dinos, shannon, compute.se = FALSE)
> biodivs

      upper      middle      lower
1.596475 1.746269 1.798425

> sapply(biodivs, attr, "se")

$upper
NULL

$middle
NULL

$lower
NULL

```

We assume that each column in 'dinos' is a multinomial sample from unknown "true" probabilities and 'size' parameter = sum(frequencies). Thus, the relative frequencies are an estimate of these probabilities and each (parametric?) bootstrap resample may be generated from a multinomial

Mn(size, relative frequencies)

Caution: in the R function 'rmultinom' the first argument, 'n', stands for the number of multinomial samples to be generated (here, in a bootstrap context, this will correspond to the number of bootstrap resamples, say B) while the second argument, 'size' corresponds to the "n" multinomial parameter

E.g. for the upper period:

```

> freq.abs <- dinos[, "upper"]
> freq.abs

```

```

[1] 50 29 3 3 4 1

> shannon.sample <- shannon(freq.abs)
> shannon.sample

[1] 1.596475
attr(,"se")
[1] 0.1366722

> n <- sum(freq.abs)
> n

[1] 90

> freq.rels <- freq.abs / n
> freq.rels

[1] 0.55555556 0.32222222 0.03333333 0.03333333 0.04444444 0.01111111

> # One bootstrap multinomial resample:
>
> set.seed(12345)
> freq.boot <- rmultinom(1, size = n, prob = freq.rels)
> freq.boot

      [,1]
[1,]    56
[2,]    26
[3,]     2
[4,]     3
[5,]     1
[6,]     2

> # Caution: it is a 'k x one-column matrix' with k = length(freq.rels)

```

Next session

18 BOOTSTRAP-t CONFIDENCE INTERVALS