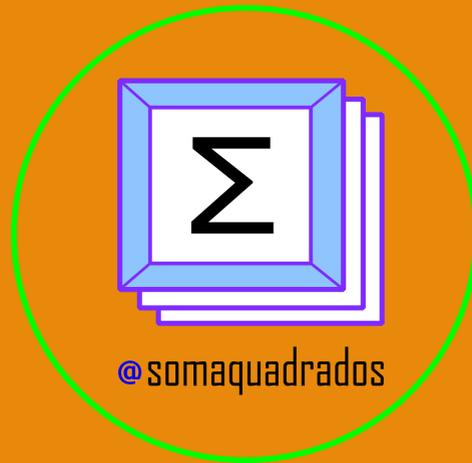


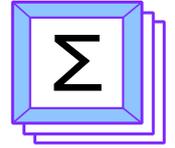
Introducción al análisis de datos biológicos con R

Clase 3



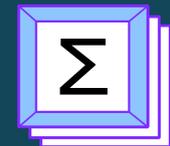
Eliana F. Burgos

Contenidos



@somaquadrados

1. Variables estadísticas
2. Distribución de probabilidades
3. Estadística descriptiva: medidas de posición
4. Estadística descriptiva: medidas de dispersión
5. Ejercicios

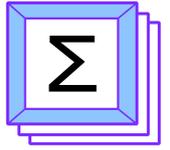


@somaquadrados

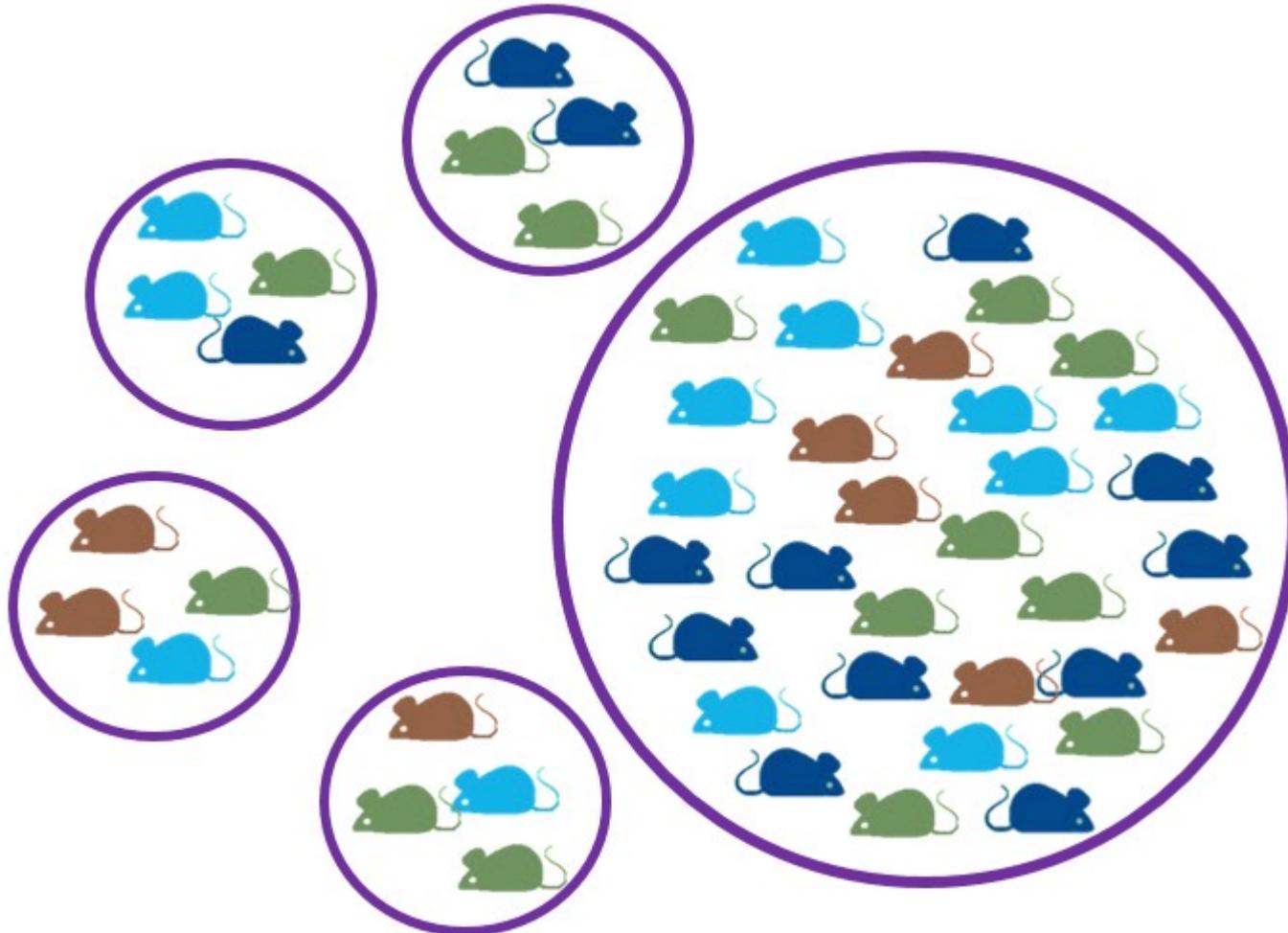
Variables estadísticas

Variables estadísticas

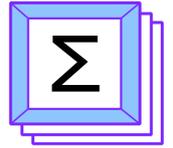
Elementos de la población -> unidades estadísticas



@somaquadrados



Variables estadísticas

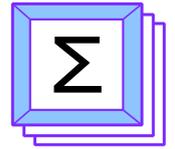


@somaquadrados

- **Variables cualitativas**
 - *ordinales*
 - *nominales*
- **Variables cuantitativas**
 - *discretas*
 - *continuas*



Variables estadísticas



@somaquadrados

- **Variables cualitativas**

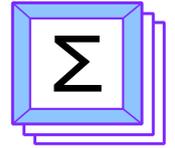
- *ordinales* -> tamaño (grande, mediano, pequeño); distancia (muy lejos, lejos, cerca); clases de edades; año de muestreo; tratamiento (1,2,3;a,b,c)
- *nominales* -> sexo, especie, sitio, color, uso del suelo, cuadrante, presencia/ausencia

- **Variables cuantitativas**

- *discretas* -> abundancia, indiv. positivos, cantidad de huevos/embriones
- *continuas* -> índices, peso, largo, temperatura, humedad, intensidad luminica



Ejemplo



@somaquadrados



Contents lists available at ScienceDirect

Mammalian Biology

journal homepage: www.elsevier.de/mambio



Original Investigation

Bat frugivory in two subtropical rain forests of Northern Argentina:
Testing hypotheses of fruit selection in the Neotropics

Mariano S. Sánchez^{a,b,*}, Norberto P. Giannini^{a,b,c}, Rubén M. Barquez^{a,b}

^a Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina

^b Programa de Investigaciones de Biodiversidad Argentina, Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán, Miguel Lillo 205,
C. P. 4000, Tucumán, Argentina

^c Department of Mammalogy, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, USA

OBJETIVO: evaluar la dieta y el nicho dietario de especies de murciélagos frugívoros.

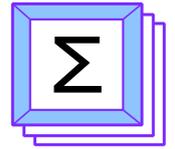


Ejemplo

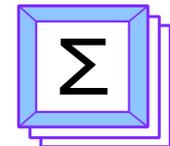
METODOLOGÍA

Datos: recopilaron datos novedosos y utilizaron datos de estudios previos

- **Murcielagos**
 - masa corporal
 - largo del antebrazo
 - sexo (macho, hembra)
 - edad (juvenil, adulto)
- **Vegetación:** colectaron ejemplares en cada sitio
 - tamaño individual de las semillas
 - color de la fruta
 - forma de la fruta
 - n° de semillas por fruto
 - habitat (bosque primario vs bosque secundario ribereño)



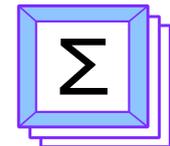
@somaquadrados



@somaquadrados



Parámetro	Tipo de variable
Murcielagos	
masa corporal	cuantitativa continua
largo del antebrazo	cuantitativa continua
sexo	cualitativa nominal de dos niveles: macho/hembra
edad	cualitativa ordinal de dos niveles: juvenil/adulto

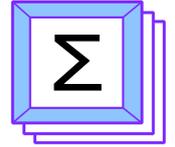


@somaquadrados



Parámetro	Tipo de variable
Vegetación	
tamaño de semilla	cuantitativa continua
color del fruto	cualitativa nominal
forma del fruto	cualitativa nominal
n° de semillas por fruto	cuantitativa discreta
habitat	cualitativa nominal de dos niveles: bosque primario/bosque secundario

Ejemplo



@somaquadrados

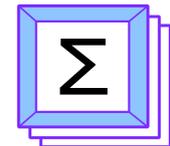
PLOS NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

A hidden deadly venomous insect: First eco-epidemiological assessment and risk mapping of *Lonomias* in Argentina

Milena Gisela Casafús ^{1‡}, Marília Melo Favalesso ^{1,2‡}, Micaela Andrea Gritti¹, Juan Manuel Coronel ³, Ana Tereza Bittencourt Guimarães ², Maria Elisa Peichoto ^{1,4*}

OBJETIVO: analizar el perfil eco-epidemiológico de accidentes y ocurrencias de taturanas (*Lonomias sp*) en Misiones (Argentina) entre enero de 2014 y mayo de 2020



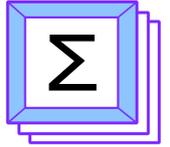
@somaquadrados

Ejemplo

METODOLOGÍA

Datos: epidemiológicos y registros a campo

- **Accidentes**
 - aspectos socio-demográficos
 - localidad del accidente
 - hora del día
 - circunstancias del accidente
 - área anatómica afectada

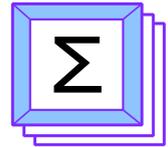


@somaquadrados

Ejemplo

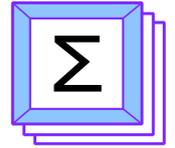
METODOLOGÍA

- **Ocurrencia**
 - planta hospedadora
 - tipo de ambiente
- **Nicho ecologico de las taturanas**
 - Temperaturas máximas y mínimas
 - Precipitaciones
 - Evapotranspiración



@somaquadrados

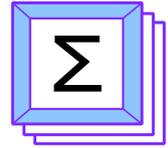
Parámetro	Tipo de variable
Accidentes	
sexo	cualitativa nominal de dos niveles: m/f
grupo etario	cualitativa ordinal de 7 niveles: 0-10; 11-20; 21-30; 31-40; 41-50; 51-60; 61-70
hora del día	cualitativa ordinal de tres niveles: 6-12; 12:01-18:00; desconocido
circunstancias del accidente	cualitativa nominal de tres niveles: durante el trabajo; durante recreación; desconocido
área anatómica afectada	cualitativa nominal de tres niveles: miembros inferiores; miembros superiores; tronco



@somaquadrados

Parámetro	Tipo de variable
Ocurrencia	
Planta hospedadora	cualitativa nominal
Tipo de ambiente	cualitativa nominal de cuatro niveles: urbano, periurbano, rural, bosque
Nicho	
Temperaturas máx y min	cuantitativa continua
Precipitaciones	cuantitativa continua
Evapotrasnpiracion	cuantitativa continua

IMPORTANTE



@somaquadrados

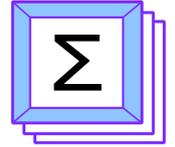
Cuando cargamos la planilla al R, el programa lee cada columna con la categoría asignada en el procesador de bases de datos

No siempre esa categoría es la correcta y hay que configurarla

```
##
## -- Column specification -----
## cols(
##   .default = col_double(),
##   fecha = col_character(),
##   sitio = col_character(),
##   uso_suelo = col_character(),
##   ambiente = col_character(),
##   estacion = col_character(),
##   anio.estacion = col_character()
## )
## i Use `spec()` for the full column specifications.
```

Entonces

variable cualitativa



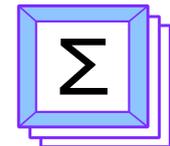
@somaquadrados

```
class(datos$uso_suelo)
```

```
## [1] "character"
```

```
datos$uso_suelo <- as.factor(datos$uso_suelo)  
class(datos$uso_suelo)
```

```
## [1] "factor"
```



@somaquadrados

variable numérica continua

```
datos$IDR_total <- as.numeric(datos$IDR_total)
class(datos$IDR_total)
```

```
## [1] "numeric"
```

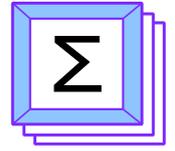
variable numérica discreta

```
class(datos$n_total)
```

```
## [1] "numeric"
```

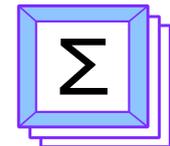
```
datos$n_total <- as.integer(datos$n_total)
```

Distribución de probabilidades



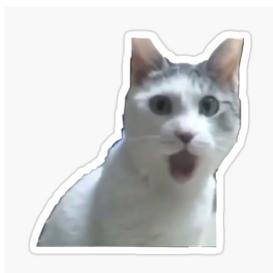
@somaquadrados

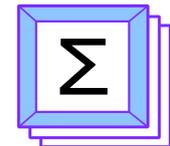
- **función**
- asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra.
- está definida sobre el conjunto de todos los sucesos y cada uno de los sucesos es el rango de valores de la **variable aleatoria**
- tiene una relación estrecha con las distribuciones de frecuencia, ya que describe **cómo se espera** que **varíen** los resultados



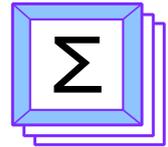
@somaquadrados

Distribución	Alias
Distribución normal	norm
Distribución Binomial	binom
Distribución de Poisson	pois
Distribución exponencial	exp
Distribución t de Student	t
Distribución Chi cuadrado	chisq
Distribución F	f



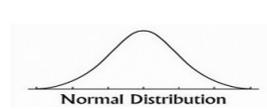


Funciones	Sintaxis	Explicacion
densidad	dnorm	devuelve el valor de la función de densidad de probabilidad
	dbinom	probabilidad de obtener un determinado número de aciertos en un determinado número de ensayos
	dpois	probabilidad puntual para un determinado valor
probabilidad	pnorm	
	pbinom	probabilidad de que la variable tome valores menores o iguales que x
	ppois	
cuantiles	qnorm	
	qbinom	permite encontrar el cuantil (percentil) para cualquier probabilidad p
	qpois	



@somaquadrados

Distribucion normal



- es una de las distribuciones de probabilidad de variable continua más frecuente
- permite modelar numerosos fenómenos naturales, sociales y psicológicos
- está identificada por una combinación de valores de dos parámetros: μ y σ
- definida por una **Función de densidad**

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

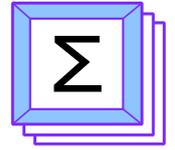
μ = media, mediana y la moda de la distribución

σ = la desviación estándar ($\sigma > 0$)

σ^2 = la variación

Estandarizar $\rightarrow N(\mu, \sigma) \rightarrow N(0,1)$

Distribucion normal



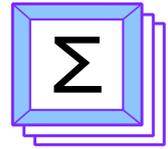
@somaquadrados

```
library(stats)  
help("Distributions")
```

Vamos a modelar una distribución normal

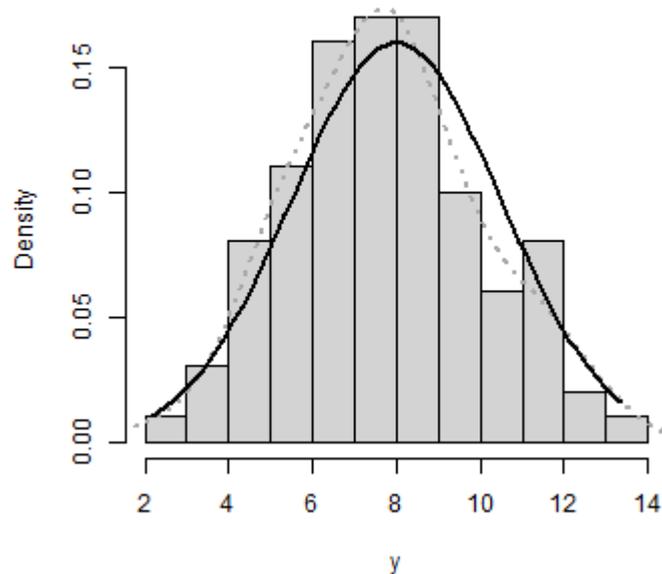
```
hist(y, breaks = 40, freq = FALSE, main="")  
lines(density(y), lwd = 2, lty=3, col="darkgrey")  
xvec <- seq(min(y)-0.5,max(y)+0.5, by=0.1)  
lines(xvec, dnorm(xvec, mean = mu, sd = sigma), lwd = 2)
```

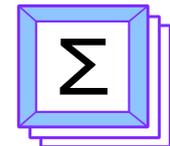
Ejemplo



@somaquadrados

Estamos estudiando el proceso de invasión de árboles en la sabana del Parque Nacional El Palmar, y para eso desarrollamos un experimento en el que colocamos plántulas de *Allophylus edulis* cerca de las palmeras y medimos su altura al cabo de un año. Sabemos que la media de la altura de esta especie es de 8 cm y la varianza es de 6.25 cm.





@somaquadrados

Sabemos que nuestra variable altura tiene una distribución normal. Si la plantula logra pasar la altura media en cercania de una palmera, tiene mayor probabilidad de asentarse y queremos evaluar qué probabilidad hay de que la plántula crezca por debajo de la media+su desvio estandar (es decir, por debajo de $8-2.5= 5.5$) al cabo de un año de comenzado el experimento.

```
pnorm(5.5, mean=8, sd=2.5)
```

```
## [1] 0.1586553
```

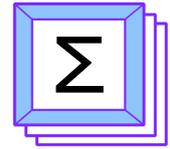
```
pnorm(12, mean=8, sd=2.5)
```

```
## [1] 0.9452007
```

```
qnorm(0.9452007, mean=8, sd=2.5, lower.tail = FALSE)
```

```
## [1] 4
```

Distribucion Binomial

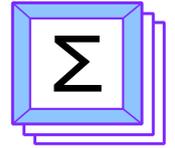


@somaquadrados

- una distribución de probabilidad de una variable discreta.
- Describe el resultado de ensayos independientes de n
- puede generar dos resultados:
 - positivo vs. negativo; presente vs ausente -> **Bernoulli**
 - cantidad de positivos vs. cantidad de examinados -> **Binomial**
- la variable toma valores discretos cuya probabilidad depende de n , definido como el número de ensayos, y p , definido como la probabilidad del resultado a contar en cada ensayo.

$$f(k, n, p) = \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \dots, n$$

Ejemplo



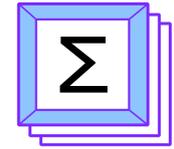
@somaquadrados

Además de ver que tanto crecieron las plántulas de *A. edulis* queremos saber también cuántas sobrevivieron en un año.

- Entonces, si instalamos 5 plantas por parcela, en un total de 5 parcelas, nuestro $n = 25$.
- Sabemos por estudios previos que la probabilidad de que una planta al azar colocada cerca de una palmera sobreviva es de $p = 0.3$
- Queremos saber cual es la probabilidad de que ninguna sobreviva

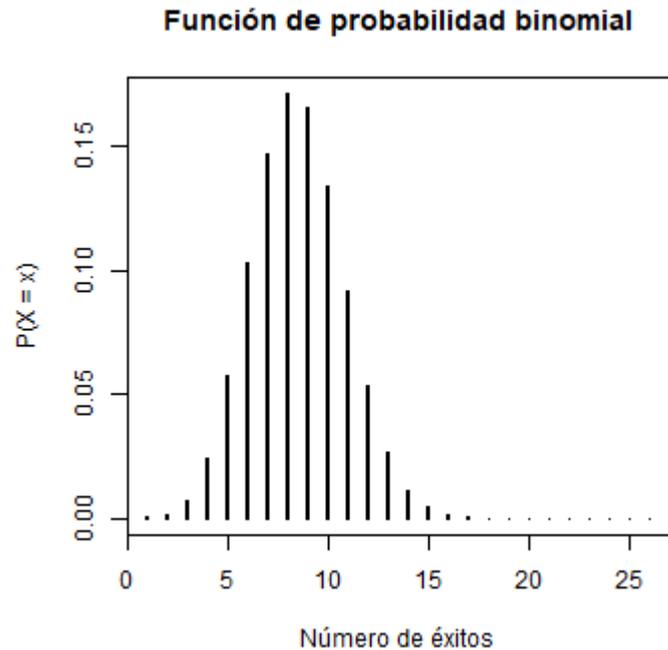
```
p <- 0.3
n <- 25
k <- 0
dbinom(k,size=n,prob=p)
```

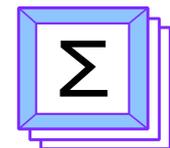
```
## [1] 0.0001341069
```



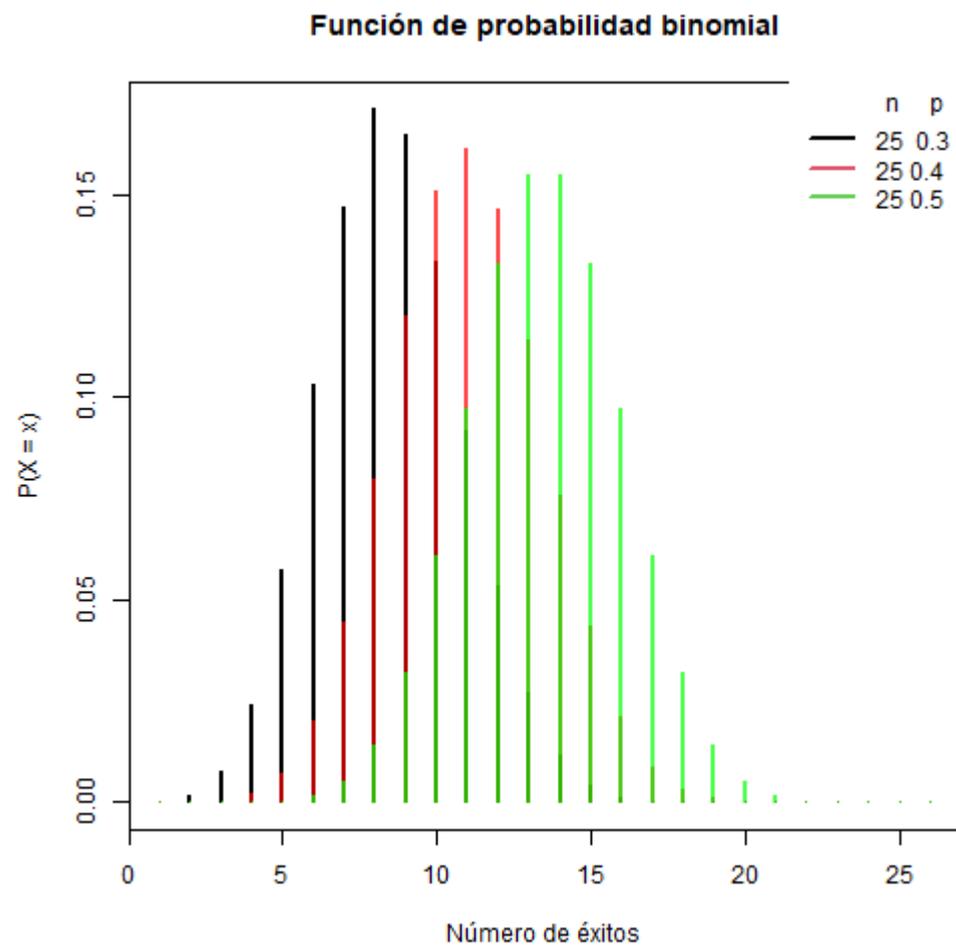
@somaquadrados

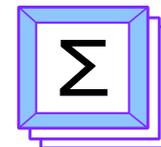
```
n <- 0:25  
  
plot(dbinom(n, size = 25, prob = 0.3), type = "h", lwd = 2,  
     main = "Función de probabilidad binomial",  
     ylab = "P(X = x)", xlab = "Número de éxitos")
```





@somaquadrados





@somaquadrados

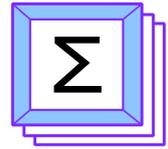
- Tambien podemos calcular la probabilidad con la siguiente funcion

```
pbinom(n,25,0.3)
```

```
## [1] 0.0001341069 0.0015709661 0.0089605279 0.0332405166 0.0904719186 0.1934884421
## [7] 0.3406549043 0.5118485440 0.6769281251 0.8105639765 0.9021999889 0.9557535026
## [13] 0.9825302595 0.9940060124 0.9982215951 0.9995464926 0.9999013758 0.9999818955
## [19] 0.9999972326 0.9999996543 0.9999999656 0.9999999974 0.9999999999 1.0000000000
## [25] 1.0000000000 1.0000000000
```

```
plot(pbinom(n, size=25, p=0.3))
```

Comparamos



@somaquadrados

dbinom	pbinom
0.000134	0.99986

¿Cuál es la diferencia?

dbinom expresa la función de masa, es decir la función de probabilidad; muestra la probabilidad de tomar al azar una plantula que no sobrevivió

pbinom expresa la funcion de distribución acumulada; muestra la probabilidad de supervivencia en las plantas

```
## [1] 0.1934884
```

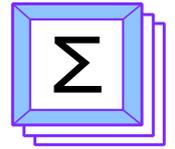
```
## [1] 0.5118485
```

Ademas tenemos la funcion **qbinom** para conocer los cuantiles (devuelve la función de densidad acumulativa inversa):

```
qbinom(p = 0.511848, size = 25, prob = 0.3)
```

```
## [1] 7
```

Distribucion Bernoulli

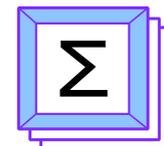


@somaquadrados

- toma valores de 0 y 1

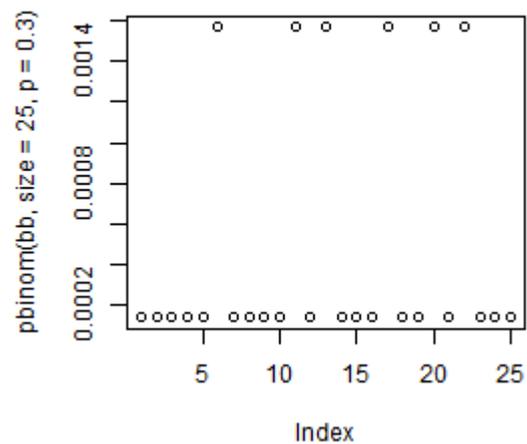
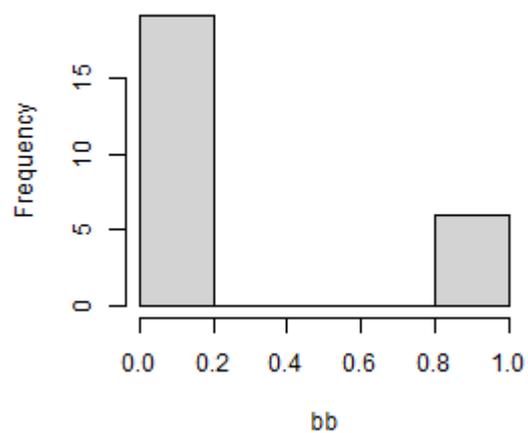
```
bb <- rbinom(25,size=1,prob=0.3)
bb
```

```
## [1] 0 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0 1 0 0 1 0 1 0 0 0
```

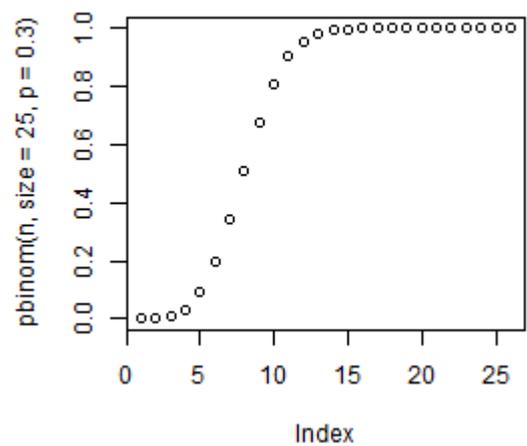
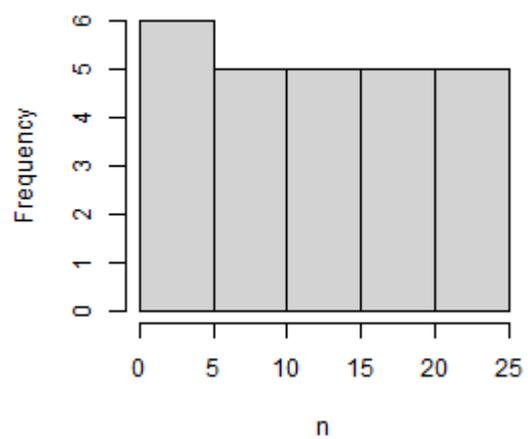


@somaquadrados

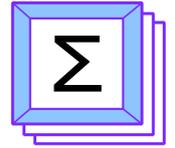
Histogram of bb



Histogram of n



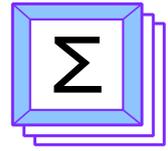
Distribución de Poisson



@somaquadrados

- es una distribución de probabilidad discreta
- se usa en conteos
- La ocurrencia media por intervalo suele representarse con la letra griega lambda λ y, en base a dicho valor, la distribución de Poisson estima la probabilidad de que se produzcan x sucesos dentro de un intervalo dado

$$P(X \leq x | \lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{for } x = 0, 1, 2, \dots$$



@somaquadrados

Ejemplo

- Siguiendo con el ejemplo anterior, cuantas plantulas crecieron al lado de las palmeras despues de un año

```
dpois(0, lambda = 5) #probabilidad de que ninguna crezca
```

```
## [1] 0.006737947
```

```
dpois(12, lambda = 5) #probabilidad de que crezcan 12
```

```
## [1] 0.00343424
```

```
ppois(1, lambda = 5) #probabilidad de que mas de una sobreviva
```

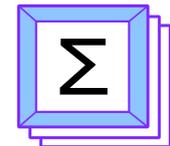
```
## [1] 0.04042768
```

```
ppois(25, lambda = 5) #probabilidad de que sobrevivan todas
```

```
## [1] 1
```

```
p <- c(0.25, 0.50, 0.75)  
qpois(p, lambda = 5)
```

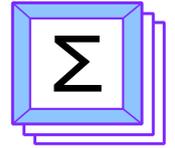
```
## [1] 3 5 6
```



@somaquadrados



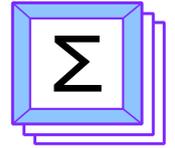
Estadística descriptiva



@somaquadrados

- *descripción de una población*
- descripción de las diferentes variables
- teniendo en cuenta:
 - valor medio
 - dispersión/variación
 - forma

Medidas de posición



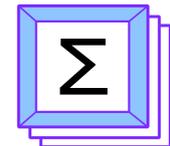
@somaquadrados

- son medidas de tendencia central
- marcan la acumulación de los datos en torno a un valor
- **media, mediana y moda**

Media

- muestra el valor promedio de nuestra variable de interés

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

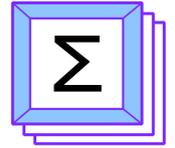


@somaquadrados

Podemos calcular la media para nuestra muestra completa o definir qué valores queremos excluir del cálculo

Ejemplo

La especie *Steindachneridion melanodermatum*, conocido como "surubí del Iguazú" es una especie endémica de la cuenca del río Iguazú. Queremos saber cual es el peso promedio de la población que estamos estudiando



@somaquadrados

Media de toda la muestra

```
mean(surubi$peso)
```

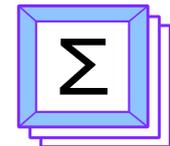
```
## [1] 3.668524
```

Media de un conjunto de datos recortado

```
mean(surubi$peso, na.rm = TRUE, trim = 0.025)
```

```
## [1] 3.636535
```

El comando *trim()* nos permite indicar los datos que queremos excluir de cada extremo de la distribución



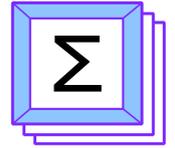
@somaquadrados

- Si además de los datos que tomamos en campo, contamos con estudios previos y queremos conocer la media de esa población utilizando todos los datos, se puede calcular la media ponderada
- Esto se realiza con una adaptación de la formula donde incluimos la media y en n de las bases que vamos a utilizar. Por ejemplo: nuestros datos tienen una media de 3.668 kg ($n=50$), y los estudios previos muestran medias de 2.15 kg ($n=25$), 3.12 kg ($n=47$)

```
mpond <- ((3.668*50)+(2.15*25)+(3.21*47))/(50+25+47)
mpond
```

```
## [1] 3.180492
```

Mediana



@somaquadrados

- es el valor que se encuentra en la mitad en la lista ordenada de nuestros datos

```
median(surubi$peso)
```

```
## [1] 3.583357
```

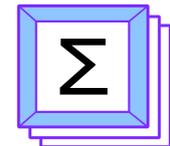
Moda

- la moda es el valor más frecuente en nuestros datos
- se puede programar siguiendo

```
frecuencias <- data.frame(table(surubi$peso))  
moda <- frecuencias[which.max(frecuencias$Freq),1]  
moda
```

```
## [1] 3.583357449
```

```
## 33 Levels: 1.53907058 1.820126645 1.884465124 2.069217919 ... 7.33345747
```



@somaquadrados

```
library(modeest)  
mfv(surubi$peso)
```

```
## [1] 3.583357
```

Estadística descriptiva

Medidas de dispersión

- nos permiten visualizar que tan variables o dispersos son nuestros datos
- valores mínimos y máximos, varianza, el desvío estandar, asimetría y curtosis, cuantiles

Mínimo y máximo

```
min(surubi$peso)
```

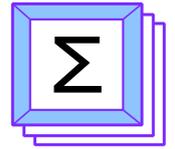
```
## [1] 1.539071
```

```
max(surubi$peso)
```

```
## [1] 7.333457
```

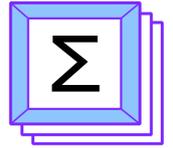
```
range(surubi$peso)
```

```
## [1] 1.539071 7.333457
```



@somaquadrados

Varianza



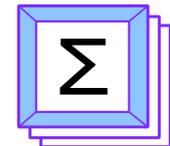
@somaquadrados

- es el promedio de los cuadrados de los desvíos
- Es la esperanza del cuadrado de la desviación típica de dicha variable respecto a su media
- se expresa en la unidad de la variable al cuadrado

$$s^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

```
var(surubi$peso)
```

```
## [1] 1.837311
```



@somaquadrados

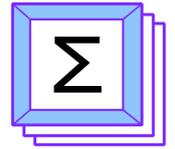
Desvio estandar

- es la raíz cuadrada de la varianza
- se expresa en la misma unidad en la que estan nuestros datos

$$s = \sqrt{s^2}$$

```
sd(surubi$peso)
```

```
## [1] 1.355475
```



@somaquadrados

Error estandar

- error estándar es la desviación estándar de la distribución muestral
- una estimación de la desviación estándar, derivada de una muestra particular usada para computar la estimación.
- es la desviación estándar dividida por la raíz cuadrada del número de observaciones.

```
library(plotrix)  
std.error(surubi$peso)
```

```
## [1] 0.1916931
```

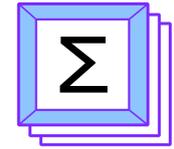
Cálculo manual

- EE es el desvío estandar dividido la raíz cuadrada del n° de observaciones

```
sd(surubi$peso)/sqrt(length(surubi$peso))
```

```
## [1] 0.1916931
```

Coeficiente de variación



@somaquadrados

- Comparar dos grupos de datos de forma estandarizada
- permite comparar datos en diferentes escalas
- a $>CV$, menor representatividad de la media
- es una medida relativa

Datos que tomamos a campo Masa corporal expresada en **kg**

Teniamos una media de 3.67 kg con un desvío de 1.35 kg

```
sd(surubi$peso)/mean(surubi$peso) # (1.35/3.668)
```

```
## [1] 0.3694877
```

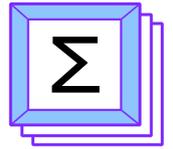
Datos de la bibliografía Masa corporal expresada en **gr**

Teniamos una media de 3120 gr (3.12 kg) con un desvio de 300 gr (0.30 kg)

```
(300/3120)
```

```
## [1] 0.09615385
```

Asimetría y curtosis



@somaquadrados

- dan cuenta de la forma general de los datos
- nos permite identificar ciertas tendencias y comportamiento de los datos
- se utilizan comandos de la librería **psych**

Asimetría

- da cuenta de cómo se organizan los datos alrededor de la media
- pueden ser simétricas o asimétricas positivas o negativas
- los valores deben encontrarse entre -2 y 2.

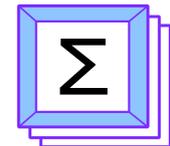
```
skew(surubi$peso)
```

```
## [1] 0.5488607
```

nos devuelve un valor en la escala de la variable que no nos permite comparar entre diferentes sets de datos, entonces lo podemos estandarizar

```
skew(surubi$peso)/sqrt(6/50)
```

```
## [1] 1.584424
```



@somaquadrados

Curtosis o apuntamiento

- mide que tan apuntada o achatada es la distribución de los datos al cercanos a la media

```
kurtosi(surubi$peso)
```

```
## [1] -0.5993963
```

```
kurtosi(surubi$peso)/sqrt(6/50)
```

```
## [1] -1.730308
```

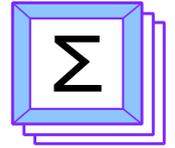
negativa la distribución es **platicúrtica**

igual a cero la distribución es **mesocúrtica**

positiva la distribución es **leptocúrtica**

En nuestro ejemplo, el peso de los surubí estudiados se distribuyen con una forma simétrica platicúrtica

¿Qué pasa si queremos comparar dos variables?



@somaquadrados

Covarianza

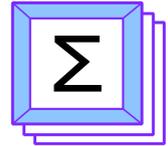
- mide la asociación lineal entre dos variables
- puede ser mayor, igual o menor que cero.
- será positiva cuando la variable respuesta aumente con el aumento de la explicativa
- nos denota el tipo de relacion: positiva, negativa, neutra

```
cov(surubi$peso, surubi$largo)
```

```
## [1] 1.225165
```

```
cov(surubi$peso, surubi$temp)
```

```
## [1] 0.4468726
```



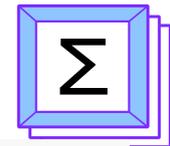
@somaquadrados

Correlacion

- nos permite estimar el grado en que estan correlacionadas dos variables
- la relacion lineal será débil si el valor es cercano a cero
- lineal directa fuerte si el valor es cercano a 1
- lineal inversa fuerte si es cercano a -1

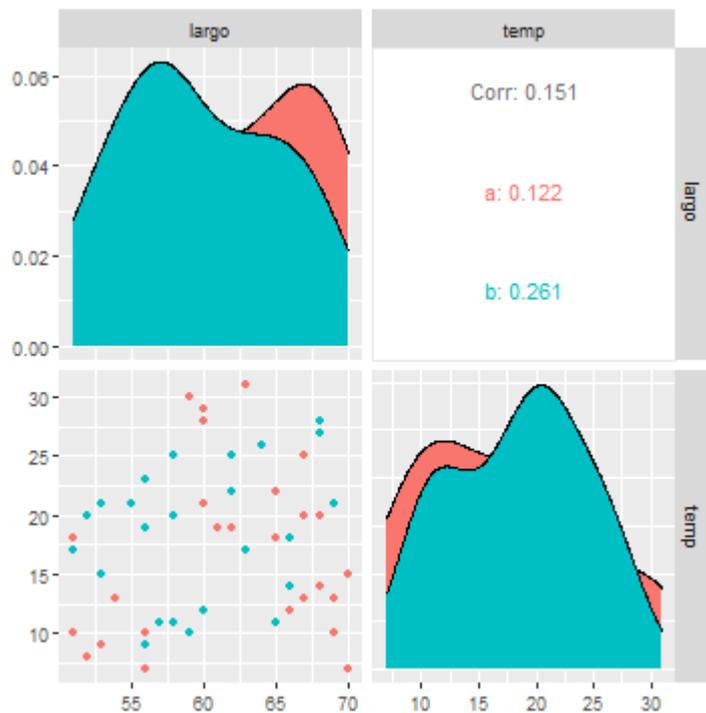
```
cor.test(surubi$largo, surubi$temp)
```

```
##  
##      Pearson's product-moment correlation  
##  
## data:  surubi$largo and surubi$temp  
## t = 1.0602, df = 48, p-value = 0.2943  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
##  -0.1326641  0.4122588  
## sample estimates:  
##          cor  
## 0.1512695
```

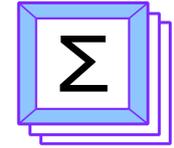


Con la librería **GGally**

```
ggpairs((surubi[,c(4:5)]), mapping=aes(color=surubi$sitio))
```



Cuartiles, deciles y percentiles



@somaquadrados

- se calcula con `quantile()` del paquete *stats*

cuartiles

```
quantile(surubi$peso, prob=seq(0, 1, 1/4))
```

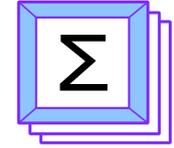
```
##           0%           25%           50%           75%          100%  
## 1.539071 2.343200 3.583357 4.689977 7.333457
```

Deciles

```
quantile(surubi$peso, prob=seq(0, 1, length = 11))
```

```
##           0%           10%           20%           30%           40%           50%           60%           70%           80%  
## 1.539071 2.284228 2.334192 2.675262 3.414819 3.583357 3.583357 4.040215 5.226326  
##           90%          100%  
## 5.527421 7.333457
```

Percentiles

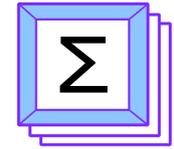


@somaquadrados

```
quantile(surubi$peso,seq(0, 1, length = 101))
```

##	0%	1%	2%	3%	4%	5%	6%	7%	8%
##	1.539071	1.676788	1.814506	1.850366	1.881892	1.967604	2.058133	2.132954	2.205584
##	9%	10%	11%	12%	13%	14%	15%	16%	17%
##	2.247867	2.284228	2.291649	2.291649	2.302450	2.316754	2.325514	2.332056	2.334192
##	18%	19%	20%	21%	22%	23%	24%	25%	26%
##	2.334192	2.334192	2.334192	2.334192	2.334192	2.334192	2.334192	2.343200	2.360856
##	27%	28%	29%	30%	31%	32%	33%	34%	35%
##	2.433705	2.568946	2.654938	2.675262	2.693952	2.710060	2.753246	2.847400	2.913087
##	36%	37%	38%	39%	40%	41%	42%	43%	44%
##	2.914249	2.947201	3.068187	3.208360	3.414819	3.583357	3.583357	3.583357	3.583357
##	45%	46%	47%	48%	49%	50%	51%	52%	53%
##	3.583357	3.583357	3.583357	3.583357	3.583357	3.583357	3.583357	3.583357	3.583357
##	54%	55%	56%	57%	58%	59%	60%	61%	62%
##	3.583357	3.583357	3.583357	3.583357	3.583357	3.583357	3.583357	3.583357	3.583357
##	63%	64%	65%	66%	67%	68%	69%	70%	71%
##	3.583357	3.592402	3.604712	3.681899	3.787709	3.828474	3.834684	4.040215	4.371983
##	72%	73%	74%	75%	76%	77%	78%	79%	80%
##	4.522504	4.537089	4.594564	4.689977	4.814564	4.969541	5.087930	5.161418	5.226326
##	81%	82%	83%	84%	85%	86%	87%	88%	89%
##	5.278791	5.343644	5.429832	5.491607	5.503029	5.511189	5.511189	5.512575	5.518234

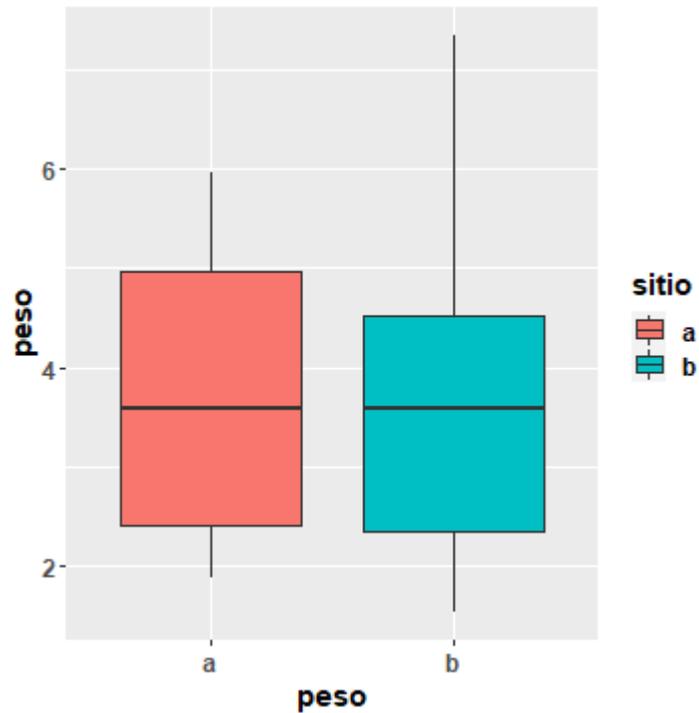
Algunos graficos descriptivos

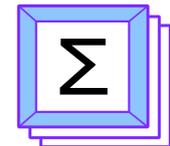


@somaquadrados

Grafico de cajas

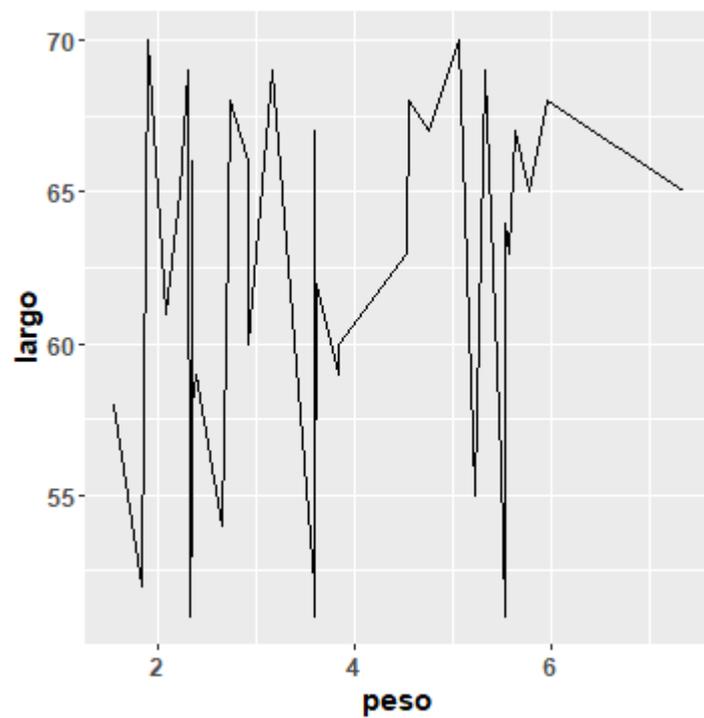
```
f1<- ggplot(surubi, aes(x= sitio, y= peso, fill=sitio))+ geom_boxplot() + xlab("peso")  
f1
```





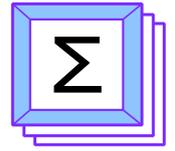
@somaquadrados

```
f2<- ggplot(surubi, aes(x=peso, y=largo))+ geom_line()+ xlab("peso") + ylab ("largo'  
f2
```

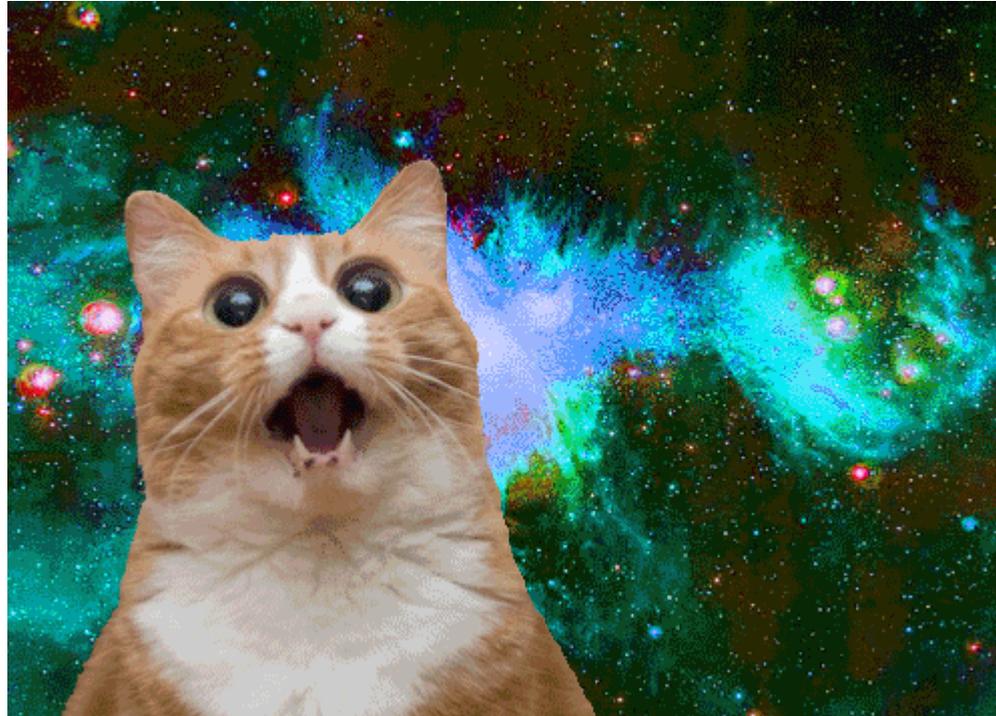


Dudas y/o consultas

efburgos@conicet.gov.ar ; mariliabioufpr@gmail.com



@somaquadrados



Ejercicios -> [Ejercicios.pdf](#)

Formulario-> <https://forms.gle/tmLsFUB5VfVqEcTKA>

FIN