Practica_final.pdf

Prácticas (3-7) de Estadística RESUELTAS (M.D. Huete)

1º Estadística

Grado en Ingeniería Informática

Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación
UGR - Universidad de Granada

PRÁCTICA 3

Ejercicio 1

b)

	media	desv típica		Cuartiles					total
	mean	sd	IQR	0%	25%	50%	75%	100%	n
	177.322	7.808596	10	159	172	178	182	200	295

c)

	mean	sd	IQR	cv	0%	25%	50%	75%	100%	data:n
Mujer	167.5714	7.156601	10.25	0.04270776	159	162	165	172.25	187	28
Varón	178.3446	7.150976	10.00	0.04009641	164	173	179	183.00	200	267

El CV es menor en los varones, por lo que esa distribución más homogénea. Por tanto, la media es más representativa en los hombres.

d) Para representar el CV según la edad, hay que convertir la edad en una variable cualitativa:

Datos → Modificar variables del conjunto → Convertir variable numérica en factor → Utilizar números

	mean	sd	IQR	cv	0%	25%	50%	75%	100%	data:n
17	184.0000	NA	0.00	NA	184	184.00	184.0	184.00	184	1
18	176.8992	8.1641795	11.00	0.046151585	160	171.00	175.0	182.00	200	129
19	178.1852	8.9739120	14.75	0.050362840	159	170.25	179.5	185.00	197	54
20	179.1818	6.7727491	3.50	0.037798194	167	176.50	180.0	180.00	200	22
21	179.1000	7.2902242	9.75	0.040704769	168	174.50	179.0	184.25	195	20
22	177.1111	7.5450715	7.25	0.042600780	160	173.50	178.0	180.75	190	18
23	176.4000	6.5692888	8.00	0.037240866	165	173.50	177.5	181.50	185	10
24	177.3571	5.0930895	6.00	0.028716574	170	174.00	176.0	180.00	189	14
25	172.5000	4.8476799	4.75	0.028102492	165	170.50	172.5	175.25	179	6

Los CV más pequeños se observan en las edades de 24 y 25 años, con lo que son más homogéneas y por tanto la media de la altura es más representativa.

e) Para ver el porcentaje de alumnos según la edad y según la preferencia sobre cada plataforma de docencia:

Estadísticos → Resúmenes → Distribución de frecuencias → Edad_Factor

counts:
Edad_Factor

17	18	19	20	21	22	23	24	25	26	27	28	29	32	35	38	51	← x_i
1	129	54	22	20	18	10	14	6	8	2	2	2	4	1	1	1	← n_i

percentages:
Edad_Factor

17	18	19	20	21	22	23	24	25	26	27	28	29	32	35	38	51
0.34	43.73	18.31	7.46	6.78	6.10	3.39	4.75	2.03	2.71	0.68	0.68	0.68	1.36	0.34	0.34	0.34

counts:
Plataf

Moodle	PRADO	Swad	Tablón de docencia
9	96	188	2

percentages:
Plataf

Moodle	PRADO	Swad	Tablón de docencia
3.05	32.54	63.73	0.68

¡PREMIO ASEGURADO!

The Power MBA
FUTURE LEADERS.

SUBE UNA STORY COMPLETANDO LA FRASE:

SIEMPRE HE QUERIDO SER

....

Y MENCIONA A **@THEPOWERMBA**

f) Media y CV del peso

	mean	sd	IQR	cv	0%	25%	50%	75%	100%	data:n
Mujer	62.14286	6.731431	10.5	0.1083219	53	56.75	61	67.25	79	28
Varón	78.19101	8.122941	13.5	0.1038859	63	71.50	78	85.00	98	267

La altura (CV = 0.04) es más homogénea que el peso (CV = 0.1), con lo que la altura es más representativa.

g) Peso mínimo del 20% de los alumnos que más pesan → Percentil 80

P_{80} → Modificar los cuantiles añadiendo 0.8

	mean	sd	IQR	cv	0%	80%	100%	data:n
Mujer	62.14286	6.731431	10.5	0.1083219	53	69.2	79	28
Varón	78.19101	8.122941	13.5	0.1038859	63	85.0	98	267

*Nunca pueden faltar ni el 0 ni el 1

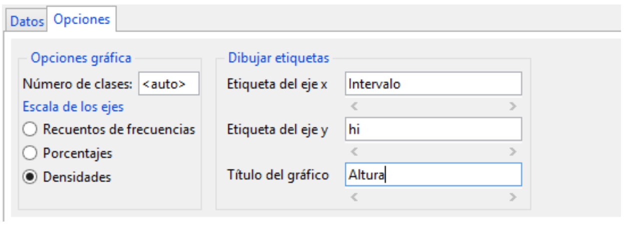
h) Peso máximo que debería tener un alumno para estar dentro del 30% de los alumnos que menos pesan

P_{30}= 57.3 (M) y 72.8 (H) ← Pesos máximos para estar dentro del 30% que menos pesa

	mean	sd	IQR	cv	0%	30%	100%	data:n
Mujer	62.14286	6.731431	10.5	0.1083219	53	57.3	79	28
Varón	78.19101	8.122941	13.5	0.1038859	63	72.8	98	267

Ejercicio 2

a) Histograma de frecuencias (porcentajes) de la Altura según el Sexo
Gráficos → Histograma → Altura



Coeficiente de asimetría γ_1

$\gamma_1 \approx 0$ → distribución asimétrica

$\gamma_1 > 0$ → distribución asimétrica a la derecha (cola a la derecha)

$\gamma_1 < 0$ → distribución asimétrica a la izquierda (cola a la izquierda)

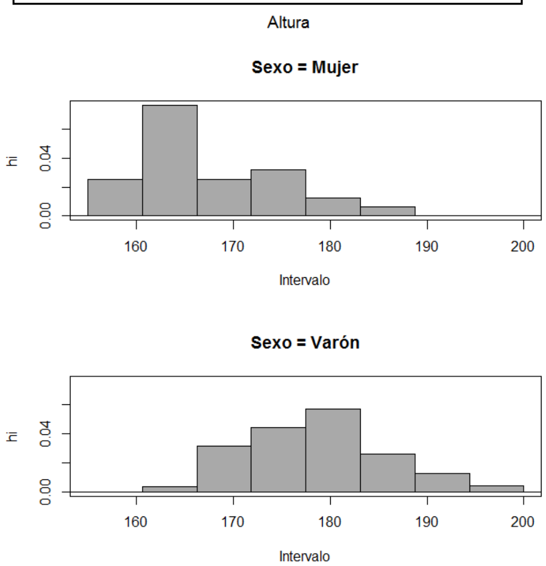
Coeficiente de curtosis o aplastamiento γ_2

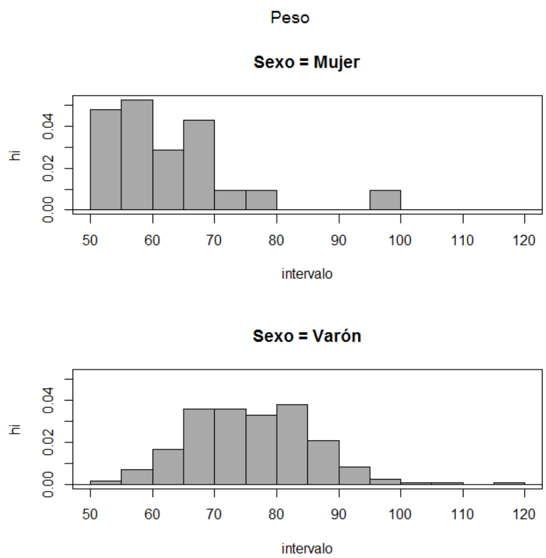
$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$ siendo $\mu_4 = \sum n_i \cdot (x_i - \overline{x})^4$

$\gamma_2 \approx 0$ → distribución igual de apuntada que Gauss

$\gamma_2 > 0$ → distribución más apuntada

$\gamma_2 < 0$ → distribución más aplastada





Se puede observar que la gráfica de las mujeres tiene más cola a la derecha y que la de los hombres es algo más simétrica, pero que ambas tienen cola hacia la derecha.

Asimismo podemos observar que la gráfica de los hombres es como la campana de Gauss, aunque levemente apuntada, mientras que la de las mujeres es algo más apuntada.

Calculemos el coeficiente de asimetría (skewness) y el de curtosis o aplastamiento (kurtosis):

```
          mean       sd  skewness    kurtosis data:n
Mujer 169.1429 7.791714 0.9037401  0.11138115     42
Varón 178.3205 6.908248 0.3826909 -0.01107763    365
```

Como era de esperar en las mujeres γ_1=0.9, que es bastante mayor que 0, y en los hombres γ_1=0.38, es decir, que hay algo de cola. En cuanto al coeficiente de aplastamiento, en las mujeres γ_2=0.11 (algo apuntada) y en los hombres γ_2=-0.011, es decir, que tal y como se visualizaba, es como la campana de Gauss.

LAS MEDICIONES FÍSICAS SON CAMPANAS DE GAUSS

b) Histograma de frecuencias del Peso según el Sexo. Coeficiente de asimetría y aplastamiento.

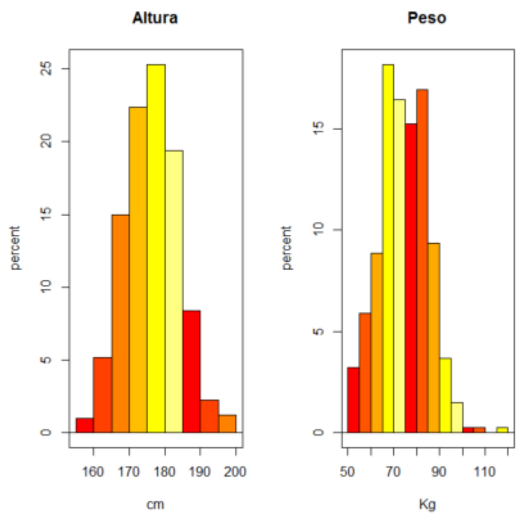
```
          mean        sd  skewness  kurtosis data:n
Mujer 63.80952 10.816063 1.5602788 3.0157662     42
Varón 76.72329  9.626408 0.3459086 0.7999412    365
```

Como era de esperar en las mujeres γ_1=1.56, que es bastante mayor que 0 (muchas cola hacia la derecha), y en los hombres γ_1=0.34, es decir, que hay algo de cola.

Se puedes observar que en el gráfico de la mujeres del Peso hay bastante apuntamiento (γ_2=3.01), mientras que el gráfico de los hombres es levemente apuntado. (γ_2=0.8).

La conclusión al comparar los coeficientes de asimetría y de aplastamiento de la Altura y el Peso es que las alturas y los pesos de los hombres son más simétricas que las de la mujeres, aunque en éstas las alturas tienen a dejar menos cola a la derecha, y que en las alturas hay mayor coincidencia con la campana de Gauss, especialmente en los hombres, mientras que en las mujeres hay más apuntamiento.

c)

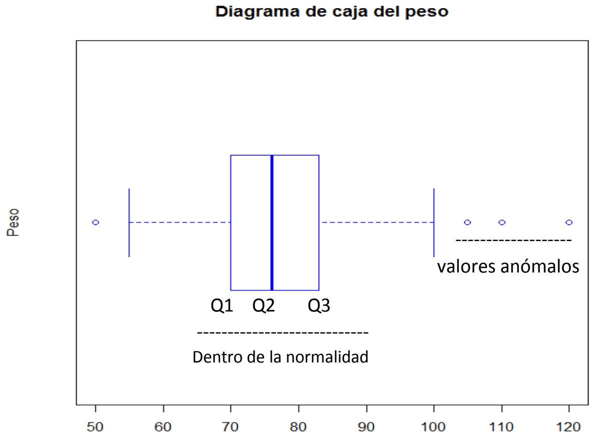


d) Histogramas anteriores sólo con los datos de los varones

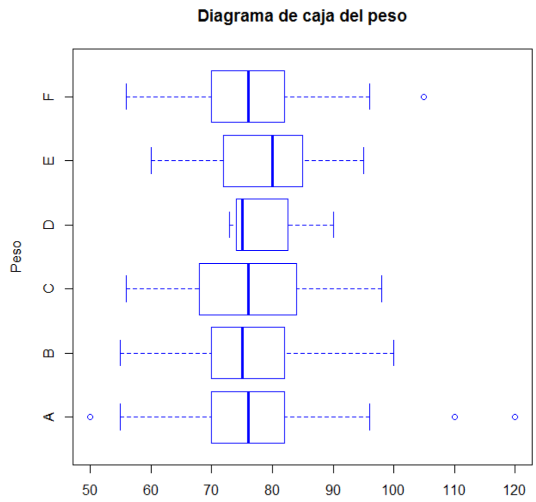
Datos → Conjunto de datos activo → Filtrar el conjunto de datos

```
           mean  skewness    kurtosis   n
Altura 178.32055 0.3826909 -0.01107763 365
Peso    76.72329 0.3459086  0.79994116 365
```

e) Diagrama de caja del Peso de los varones



f) Diagrama de caja del Peso según el grupo



g) Gráfico de tallos y hojas de la variable edad

```
  1    17* | 0
       17. |
177    18* | 000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000
       18. |
188    19* | 000000000000000000000000000000000000000000000000000000000000000000000
       19. |
123    20* | 00000000000000000000000000000000
       20. |
 92    21* | 00000000000000000000000
       21. |
 70    22* | 0000000000000000000
       22. |
 51    23* | 000000000
       23. |
 42    24* | 000000000000000
       24. |
 27    25* | 00000
```

PRÁCTICA 4 - Ejercicio 1

b) Tabla de doble entrada para determinar si influye el sexo en haber estudiado estadística previamente.

Estadísticos → Tablas de contingencia → Tabla de doble entrada → Fila: Est_est / Col: Sexo

Sin porcentajes

```
        Sexo
Est_Est Mujer Varón
     No    22   232
     Sí    20   132
```

Con porcentajes totales $\left(\frac{n_{ij}}{N}\right)$

```
Total percentages:
      Mujer Varón Total
No      5.4  57.1  62.6
Sí      4.9  32.5  37.4
Total  10.3  89.7 100.0
```

Con porcentajes por columnas

```
Est_Est Mujer Varón
  No     52.4  63.7
  Sí     47.6  36.3
  Total 100.0 100.0
```

*Si las distribuciones condicionadas fueran iguales → serían independientes

c) Grado de relación/asociación entre las variables. Estadístico chi-cuadrado

$$\boxed{0 \leq \chi^2 \leq N \cdot mín\{p-1, q-1\}}$$

La cercanía al 0 significa independencia
La cercanía al máximo es máxima relación

$$p \equiv nº\ filas$$
$$q \equiv nº\ columnas$$

```
Column percentages:
            Sexo
Est_Est Mujer Varón
  No     52.4  63.7
  Sí     47.6  36.3
  Total 100.0 100.0
  Count  42.0 364.0

        Pearson's Chi-squared test

data:  .Table
X-squared = 2.073, df = 1, p-value = 0.1499
```

$$0 \leq \chi^2 \leq N \cdot mín\{p-1, q-1\}$$
$$0 \leq \chi^2 \leq 407 \cdot mín\{2-1\,, 2-1\}$$
$$0 \leq \chi^2 \leq 407 \cdot mín\{1\,, 1\}$$
$$0 \leq \chi^2 \leq 407$$

χ^2está muy cerca del 0, luego hay muy poca asociación → son prácticamente independientes

d) Tabla de contingencia con el número de vehículos y el número de miembros en la unidad familiar.

En primer lugar hay que convertir ambas variables en factores numéricos

```
Frequency table:
            n_miem_factor
n_veh_factor  1  2  3   4  5  6  7  8
           0  0  3  4   4  0  0  0  0
           1  2 10 27  55 12  4  0  0
           2  1  2 37 103 31  3  2  1
           3  0  1  9  41 15  1  1  0
           4  0  0  5  13  3  0  0  0
           5  1  0  1   8  4  1  0  0
           6  0  0  1   0  1  0  0  0

        Pearson's Chi-squared test

data:  .Table
X-squared = 59.704, df = 42, p-value = 0.03732
```

x: vehículos
y: miembros

A simple vista no se ve ninguna relación.

Veamos un ejemplo con una distribución condicionada del número de miembros cuando las familias solamente tienen un vehículo en su hogar

```
Row percentages:
            n_miem_factor
n_veh_factor   1   2    3  4    5   6 7 8 Total Count

           1 1.8 9.1 24.5 50 10.9 3.6 0 0  99.9   110
```

Porcentajes por filas:

χ^2 = 59.7 filas=7, columnas = 8

$$0 \leq \chi^2 \leq N \cdot mín\{p-1, q-1\}$$

$$0 \leq \chi^2 \leq 407 \cdot mín\{7-1\,, 8-1\}$$

$0 \leq \chi^2 \leq 407 \cdot 6$ → $0 \leq \chi^2 \leq 2442$ → Apenas hay asociación

Ejercicio 2

Se quiere estudiar la posible asociación entre el nivel de estudios y el hábito de fumar.

	Primarios	Medios	Superiores
Sí	20	10	4
No	16	12	2

b) Distribuciones marginales (porcentajes totales)

	Primarios	Medios	Superiores	Total
Sí	31.2	15.6	6.2	53.1
No	25.0	18.8	3.1	46.9
Total	56.2	34.4	9.4	100.0

c) ¿Son las variables independientes? (Distribución por filas)

	Primarios	Medios	Superiores	Total	Count
Sí	58.8	29.4	11.8	100	34
No	53.3	40.0	6.7	100	30

χ^2 = 1.047

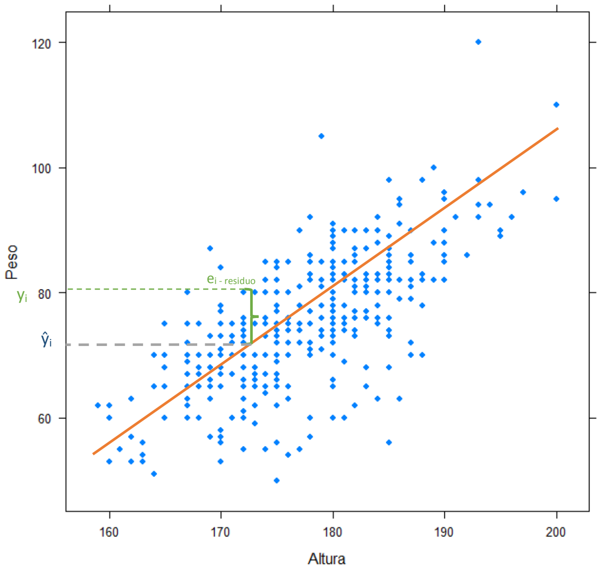
$0 \leq \chi^2 \leq 64$ → Apenas hay asociación

```
        Pearson's Chi-squared test

data:  .Table
X-squared = 1.047, df = 2, p-value = 0.5924
```

Ejercicio 3

Modelo de regresión lineal para predecir el peso en función de la altura.
Introducir una Gráfica XY (x: altura , y: peso). **Y = a + b·X**



b) Realizar ajuste: Estadísticos → Ajuste de modelos → Regresión lineal

```
Coefficients:
              Estimate
(Intercept) -99.81416   ← a
Altura        0.98777   ← b
```

Y = -99.81 + 0.98·X

```
Multiple R-squared:  0.5015
```

R^2=0.5 → El ajuste es aceptable / regular

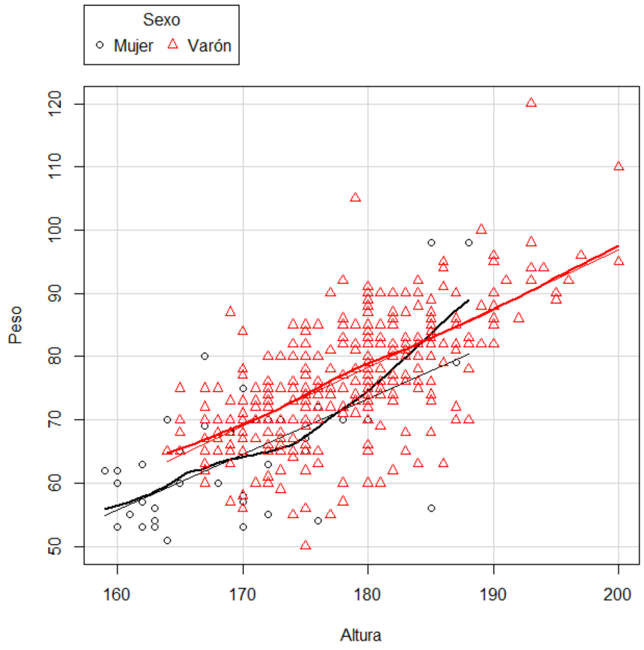
c) Ajuste por sexos

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -88.49773    9.74858  -9.078   <2e-16 ***
Altura        0.92654    0.05463  16.961   <2e-16 ***
```

Y = -88.498 + 0.9265·X

```
Adjusted R-squared:  0.4406
```

R^2=0.4406 → El ajuste no es muy bueno



d) Regresión múltiple

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \ldots$$

Intercept Altura Sexo

Estadísticos → Ajuste de modelos → Modelo lineal

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -91.93587    8.86628 -10.369  < 2e-16 ***
Altura         0.92079    0.05199  17.711  < 2e-16 ***
Sexo[T.Varón]  4.46302    1.28568   3.471 0.000574 ***
```

Varón: 1 – Mujer: 0

$$Y = -91.93 + 0.92 \cdot X_1 + 4.463 \cdot X_2$$

```
Multiple R-squared:  0.516, Adjusted R-squared:  0.5136
```

$R^2 = 0.516$ → El ajuste es regular

e) Incluyamos también el número de hermanos (n_herm). ¿Cree que aporta algo dicha variable al modelo construido?

Contrastes para coeficientes

Un contraste en Estadística es una afirmación que se hace. Según los datos obtenidos, se acepta o se rechaza dicha afirmación.

H_0 (hipótesis nula): $\beta_0 = 0$

H_1 (hipótesis alternativa): $\beta_0 \neq 0$

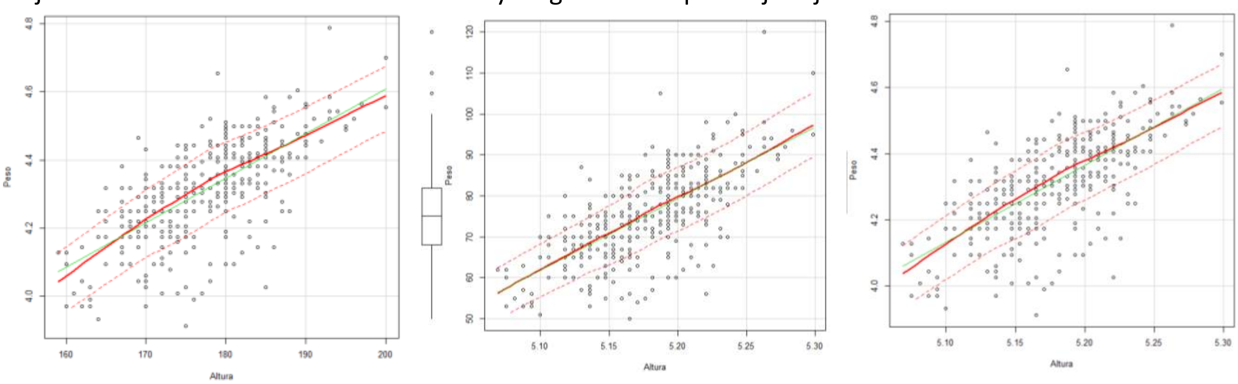
Si p-valor < 0.05 se rechaza la hipótesis nula [Pr (>|t|)]

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -92.28764    8.90074 -10.369  < 2e-16 ***
Altura         0.92035    0.05205  17.684  < 2e-16 ***
Sexo[T.Varón]  4.44361    1.28740   3.452 0.000616 ***
n_herm         0.20138    0.39197   0.514 0.607694
```

En intercept p-valor <2e-16 √ se rechaza
En la altura p-valor <2e-16 √ se rechaza
En el sexo p-valor = 0.0006 √ se rechaza
En el n_herm p-valor = 0.6>0.05 **X** se acepta

<u>**Los asteriscos indican que están bien metidas las variables</u>

f) También se pueden realizar modelos "linealizables", como por ejemplo, el modelo exponencial **Y** = $e^{a+b \cdot X}$, el modelo logarítmico **Y = a + b·ln(X)** o modelo multiplicativo **Y = a·X^b**, entre otros. Vamos a ajustarlos al Peso en función de la Altura y elegiremos el que mejor ajuste nos ofrezca.

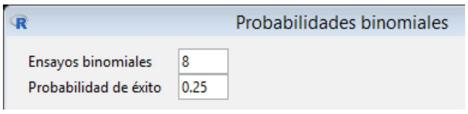


PRÁCTICA 5

Ejercicio 1 X="Nº de familias reacias de 8" ~> B(8 , 0.25)

a) P[X=1] = 0.2669678 ≈ 0.2700

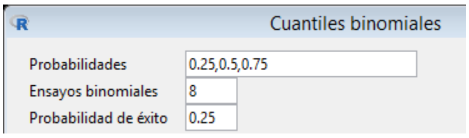
```
  Probability
0 1.001129e-01
1 2.669678e-01
2 3.114624e-01
```



b) P[X=2] = 0.3114654 ≈ 0.3115

c) P[X≤2] = 0.6785

```
> pbinom(c(2), size=8, prob=0.25, lower.tail=TRUE)
[1] 0.6785431
```

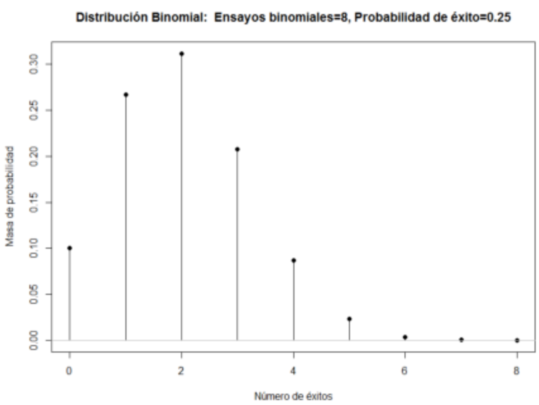


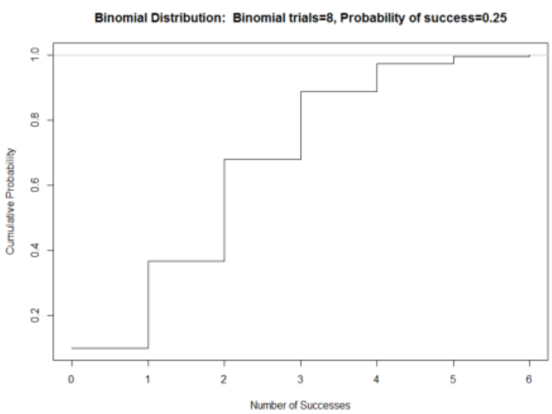
d) Cuartiles

$Q_1 = 1$

$Q_2 = 2$

$Q_3 = 3$





e) Graficar f.m.p. y función de distribución

local({ .x <- 0:**8** (…)

f) Muestra de una distribución binomial

Ejercicio 2 X="Nº de averías nuevas diarias" ~> P(1.2)

a) P[X=0] = 0.3012

b) P[X≥2] = 1 – P[X≤1] = 0.3373727

1-ppois(c(1), lambda=1.2, lower.tail=TRUE) **- o -** ppois(c(1), lambda=1.2, lower.tail=TRUE) && 1- 0.6626273

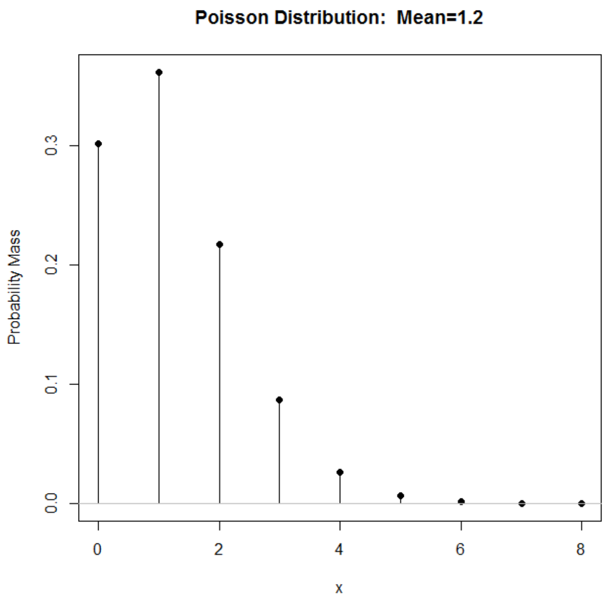
c) Mediana y percentil 75

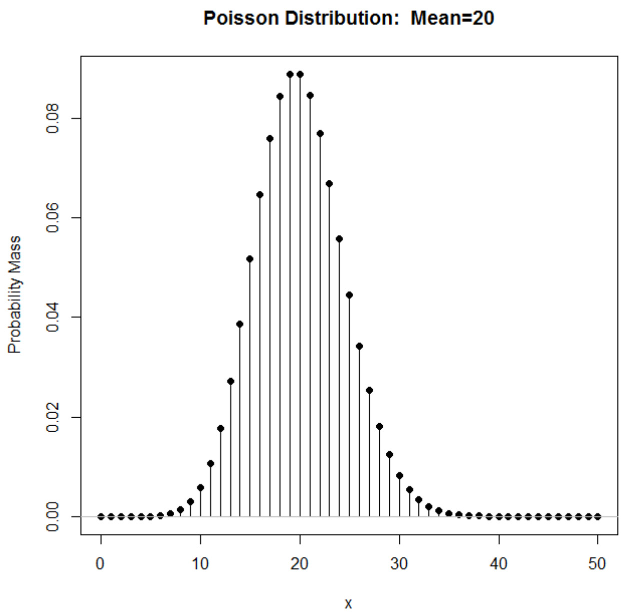
Me = Q_2 = 1 P_{75} = Q_3 = 2

d) Y ~> P(20) --- **Mediana de la distribución**

Me = 20

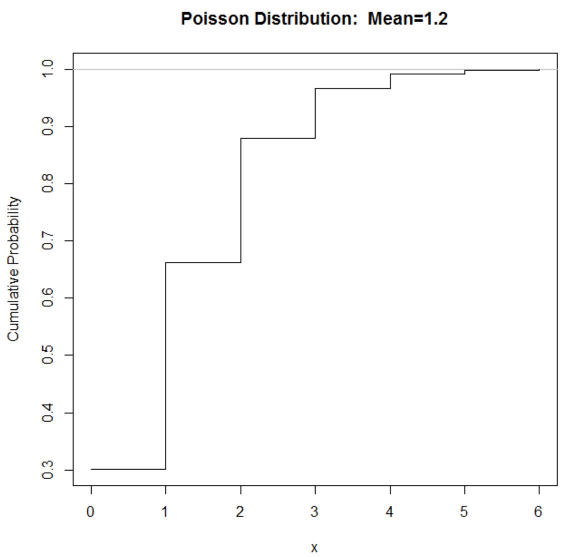
e) f.m.p. de Granada y Almería, con medias de 1.2 y 20, respectivamente





Recuerda a una campana de Gauss la distribución de Almería

f) Función de distribución en el caso de Granada



Ejercicio 3 X="Longitud de las alas de las águilas" ~> N(2.1 , 0.2)

a) **P[X > 2.5] =** 1 – P[X < 2.5] = 0.02275
b) **P[X < 2] =** 0.3085375
c) **P[1.8 < X < 2.2]** = P[X < 2.2] < P[X < 1.8] = 0.6914625 - 0.0668072 = 0.6246553

pnorm(c(2.2), mean=2.1, sd=0.2, lower.tail=TRUE)- pnorm(c(1.8), mean=2.1, sd=0.2, lower.tail=TRUE)

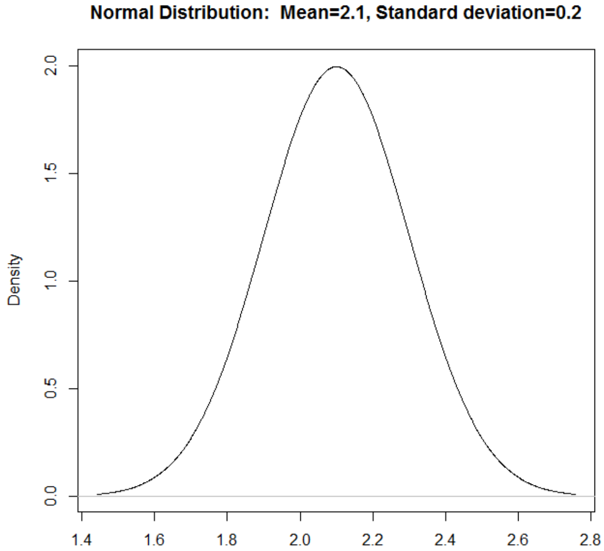
d) **Longitud máxima del 90% de las águilas con menor longitud de sus alas**

P_{90} = 2.3563

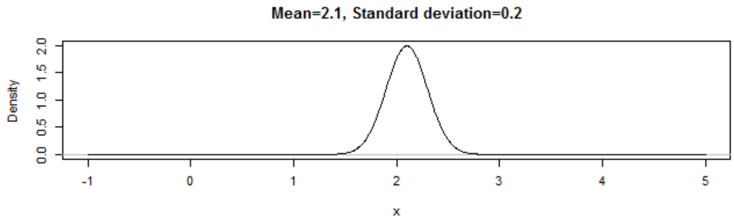
e) **Longitud mínima del 5% de las águilas con mayor longitud**

P_{95} = 2.4290

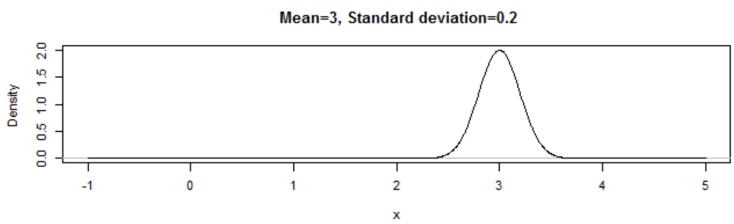
f) **Función de densidad**



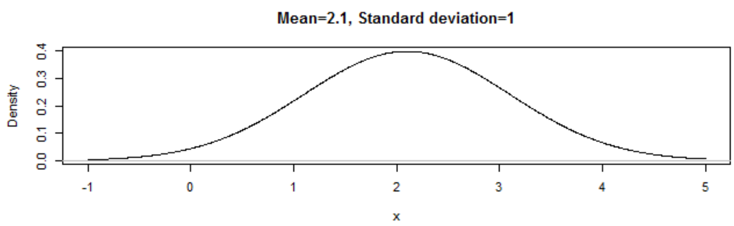
g) **Comparación de densidades de varias normales**



La media está centrada en 2.1



En las dos primeras, la desviación típica es la misma, por lo que la amplitud de la campana de Gauss es igual. La diferencia es que en la segunda, la media está centrada en 3.



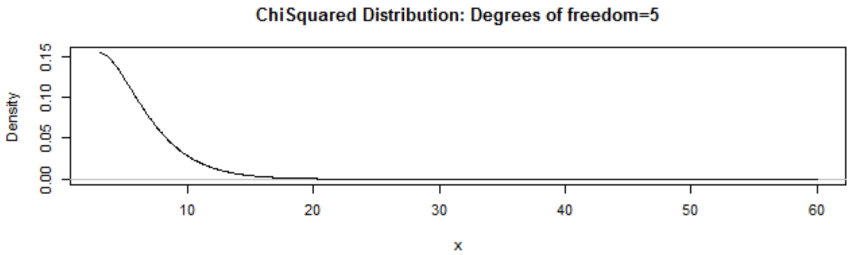
En la tercera la media vuelve a estar centrada en 2.1, pero al ser 1 la desviación típica, es mucho más amplia la campana.

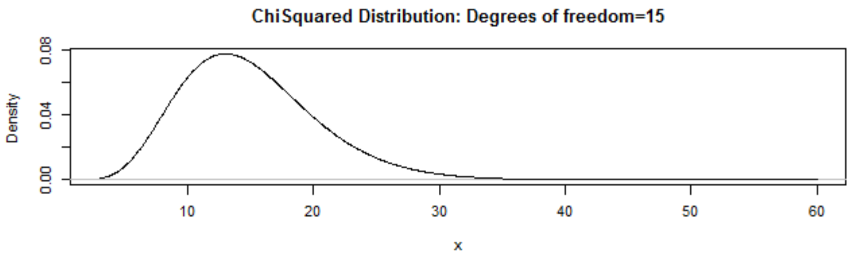
Ejercicio 4 $X \sim> \chi^2_{15}$

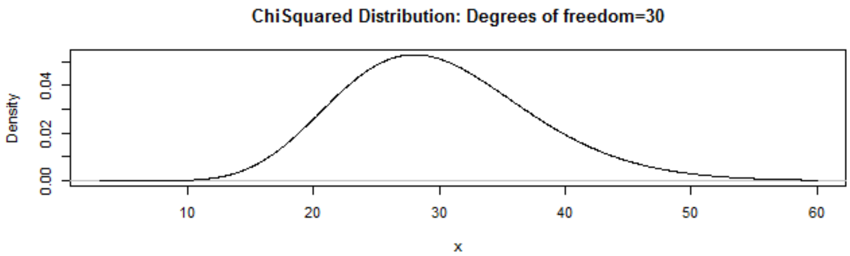
a) **P[X ≤ 8.55]** = 0.10015
b) **P[X ≤ 27.6]** = 0.9757887
c) **Cuartiles**

Q_1 = 11.03654 **Q_2 =** 14.33886 **Q_3 =** 18.24509

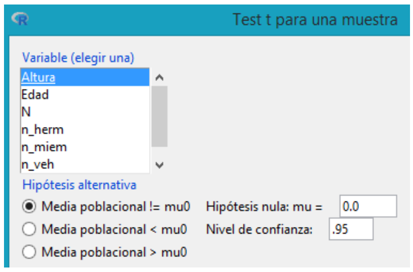
d) **Represente la densidad de esta distribución, así como las de otras distribuciones Chi-cuadrado con 5 y con 30 grados de libertad**







Ejercicio 1



b) i) X="Alturas" ~> N(μ,σ) -- $IC(\mu)_{95\%}$

<u>Estadísticos → Medias → Test t para una muestra</u>

```
95 percent confidence interval:
 176.6395 178.1074
```

ii) $IC(\mu)_{99\%}$

```
99 percent confidence interval:
 176.4072 178.3397
```

Para reducir la amplitud del intervalo habría que:

- Reducir el nivel de confianza
- Aumentar el tamaño de la muestra

c) IC para la altura media poblacional de los hombres:

Datos → Conjunto de datos activo → Filtrar conjunto de datos activo → Expresión: Sexo=="Varón"

```
95 percent confidence interval:
 177.6095 179.0316
```

Ejercicio 2

Construir un IC para la varianza poblacional de una N(0,1), tomando una muestra aleatoria de tamaño 5000

Distribuciones → Distribuciones continuas → Distribución normal → Muestra de una distribución normal → Filas = 5000 // Columnas = 1

```
N01 <- rnorm(5000,mean=0,sd=1)
ci <- 4999*var(Muestra)/qchisq(0.975,4999)
cs <- 4999*var(Muestra)/qchisq(0.025,4999)
c(ci,cs)
```

```
> c(ci,cs)
[1] 0.9736719 1.0530947
```

Como 1∈ IC → Puedo afirmar que la varianza de la población $\sigma^2=1$

Ejercicio 3 - Se desea conocer qué proporción de individuos de esta zona están a favor de instalar la central mediante un intervalo de confianza al 95%

Datos → Nuevo conjunto de datos → Introducir nuevo conjunto de datos: Add row (20 veces) → Conjunto de datos activo → Guardar el conjunto de datos activo

Datos → Mod. variables del conjunto de datos activo → Convertir variable numérica en factor → 0:No // 1:Sí

Estadísticos → Proporciones → Test de proporciones para una muestra

```
95 percent confidence interval:
 0.3420853 0.7418021
sample estimates:
   p
0.55
```

Ejercicio 4

X="Altura mujeres" ~> $N(\mu_x,\sigma_x)$ // Y="Altura hombres" ~> $N(\mu_y,\sigma_y)$

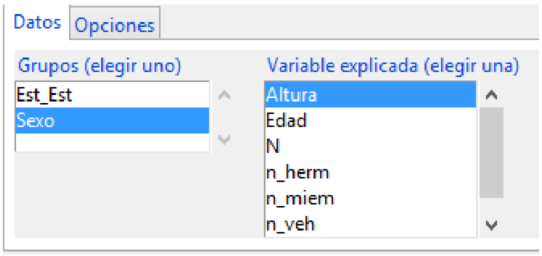
$IC(\mu_x-\mu_y)_{95\%}$

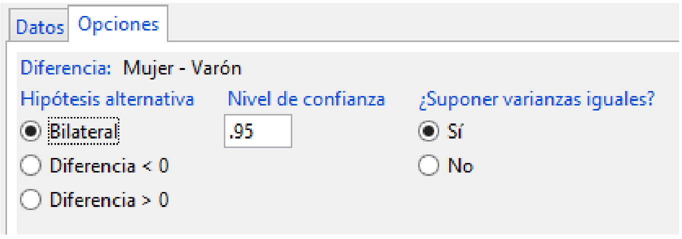
Estadísticos → Varianzas → Test F para dos varianzas

```
95 percent confidence interval:
 0.836581 2.106849
```

Puedo afirmar que $\sigma_y^2=\sigma_y^2$ porque 1∈ IC

Estadísticos → Medias → Test t para muestras independientes






```
95 percent confidence interval:
 -11.420762  -6.934619
```

Las alturas medias de los varones y las mujeres no son iguales. **$\mu_x \neq \mu_y$** El resultado indica que la altura media de los hombres es mayor

Ejercicio 5

$IC(\mu_x-\mu_y)_{95\%}$ para la diferencia media de los pesos entre hombres y mujeres

Comprobamos si las varianzas son iguales:

```
95 percent confidence interval:
 0.8302097 2.0908035
```

→ son iguales →

```
95 percent confidence interval:
 -16.037915  -9.789612
```

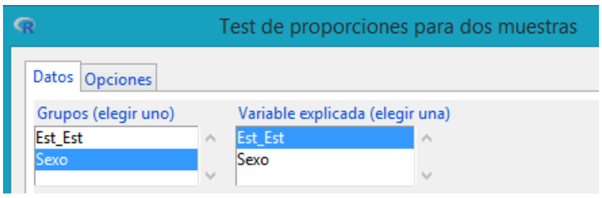
La diferencia de medias de los pesos no es igual para hombres y mujeres

Ejercicio 6

$IC(p)_{95\%}$ para la proporción de alumnos que ha estudiado estadística

```
95 percent confidence interval:
 0.5775689 0.6713078
```

$IC(p_M-p_H)_{95\%}$ para la diferencia de proporciones entre mujeres y hombres que ha estudiado estadística




```
95 percent confidence interval:
 -0.27246571  0.04535948
```

0 ∈ IC → $p_M=p_H$ → No depende del sexo

PRÁCTICA 7

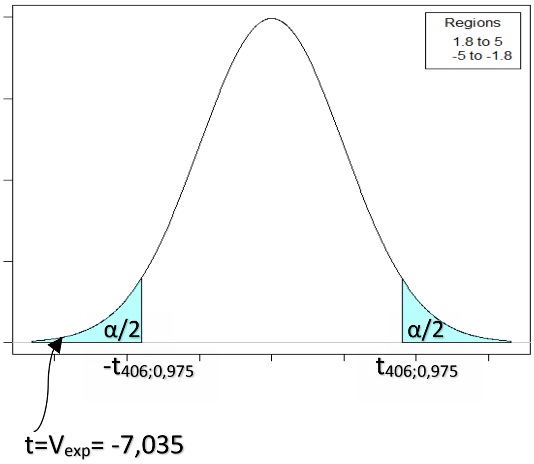
Ejercicio 1 – Contraste la hipótesis de que la altura media es de 180cm – α=0.05

X = "Altura" ~> N(μ,σ)

$$\begin{cases} H_0: \mu = 180 \\ H_1: \mu \neq 180 \end{cases}$$

Estadísticos → Medias → Test t para una muestra → Altura

```
data:  Altura
t = -7.035, df = 406, p-value = 8.517e-12
alternative hypothesis: true mean is not equal to 180
95 percent confidence interval:
 176.6395 178.1074
sample estimates:
mean of x
 177.3735
```



$$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$$

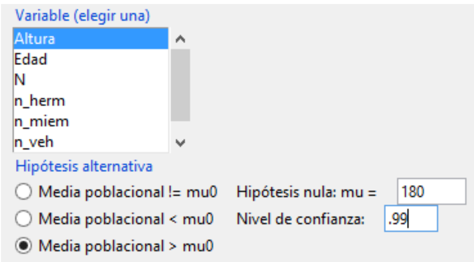
El p-valor = 8,517e-12

Si p-valor < α → V_{exp} cae en <u>Región de rechazo</u>

8,517e-12 < 0,05 → Se rechaza la hipótesis nula.

c) Contraste la hipótesis de que la altura media es inferior a 180 cm a un nivel de significación del 1%

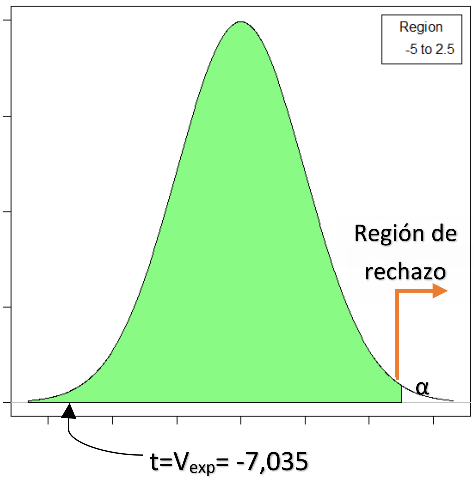
$$\begin{cases} H_0: \mu \leq 180 \\ H_1: \mu > 180 \end{cases}$$




```
data:  Altura
t = -7.035, df = 406, p-value = 1
alternative hypothesis: true mean is greater than 180
99 percent confidence interval:
 176.5015      Inf
sample estimates:
mean of x
 177.3735
```

t = -7.035

p-valor = 1 > 0.01 → **Se acepta la hipótesis nula**



d) Contraste la hipótesis de que la altura media de las alumnas es inferior a la altura media de los alumnos a un nivel de significación del 5%.

Previamente comprobamos si las varianzas poblacionales son iguales en ambos grupos:

Estadísticos → Varianzas → Test F para 2 varianzas

```
data:  Altura by Sexo
F = 1.2721, num df = 41, denom df = 364, p-value = 0.26
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.836581 2.106849
```

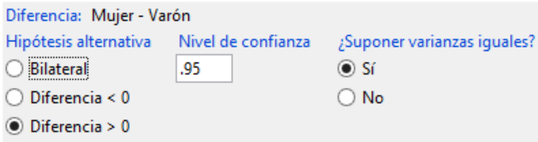
$\sigma_x^2=\sigma_y^2$

$$\begin{cases} H_0: \mu_y - \mu_x \leq 0 \\ H_1: \mu_x - \mu_y > 0 \end{cases}$$

p-valor = 0,26 > 0,05

→ Aceptamos H_0

Estadísticos → Medias → Test t para muestras independientes




```
data:  Altura by Sexo
t = -8.0434, df = 405, p-value = 1
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -11.05881       Inf
sample estimates:
mean in group Mujer mean in group Varón
           169.1429            178.3205
```

p-valor > α → 1 > 0.05

➔ **Se acepta la hipótesis nula**

➔ $\boldsymbol{\mu_y = \mu_x}$

Ejercicio 2 – Contrate del número medio de hermanos en mujeres y hombres

a) Contraste la hipótesis de que las varianzas son iguales a un nivel de significación del 5%

```
data:  n_herm by Sexo
F = 0.90544, num df = 41, denom df = 364, p-value = 0.7204
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5954378 1.4995532
```

p-valor = 0,72 < 0,05

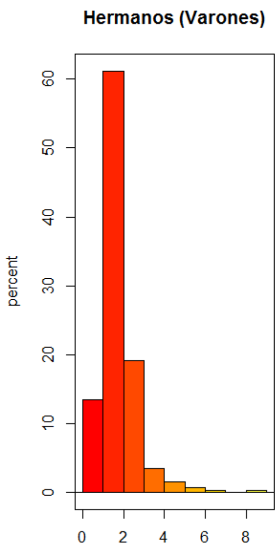
Se acepta $\sigma_x^2=\sigma_y^2$

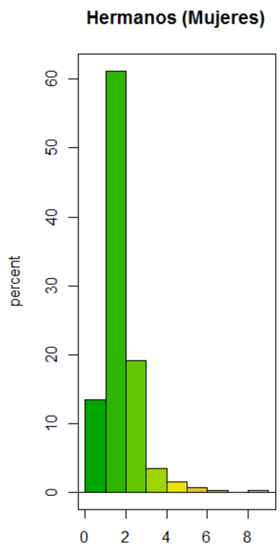
b) $\begin{cases} H_0: \mu_y - \mu_x = 0 \\ H_1: \mu_x - \mu_y \neq 0 \end{cases}$

```
t = -0.76942, df = 405, p-value = 0.4421
```

p-valor = 0,4421 > 0.05 → Se acepta H_0 → $\boldsymbol{\mu_y = \mu_x}$

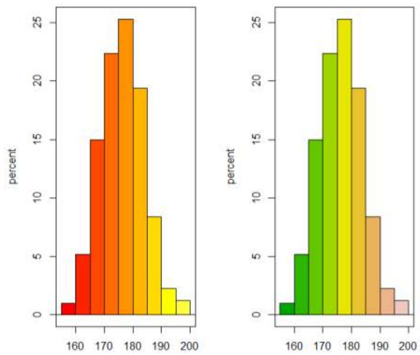
c) Compruebe gráficamente que la hipótesis previa de normalidad de las variables se verifica





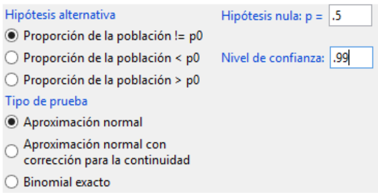
Gráficamente se ve que la distribución no es normal.

Con las alturas sí serviría:



Ejercicio 3

a) Contraste la hipótesis de que la mitad de los alumnos de Informática ha estudiado Estadística en alguna ocasión a un nivel de significación del 1%.



Estadísticos → Proporciones → Test de proporciones para una muestra

$H_0: p = 0,5$

$H_1: p \neq 0,5$

```
data:  rbind(.Table), null probability 0.5
X-squared = 25.626, df = 1, p-value = 4.145e-07
alternative hypothesis: true p is not equal to 0.5
99 percent confidence interval:
 0.5621940 0.6849978
sample estimates:
        p
0.6256158
```

p-valor = 4,14e-07 < 0,01

Rechazamos H_0

b) Contrate la hipótesis de que la proporción de alumnas que han estudiado estadística es superior a la proporción de alumnos a un nivel de significación del 5%.

$$\begin{cases} H_0: \mu_x \geq \mu_y \\ H_1: \mu_x < \mu_y \end{cases}$$

```
data:  .Table
X-squared = 2.073, df = 1, p-value = 0.07496
```

p-valor > α -- 0,07 > 0.05 → Se acepta H_0