

MATEMÁTICAS II - DI1009 - GRADO EN INGENIERÍA EN
DISEÑO INDUSTRIAL Y DESARROLLO DE PRODUCTOS

PROJECTUM

1



LISBETH SALANDER
LOS HOMBRES QUE
NO PAGABAN BIEN
A LAS MUJERES

ÍNDICE

1. OBJETIVOS DEL TRABAJO	2
2. LOS DATOS	2
3. BLOQUE DE DESCRIPTIVA	4
4. BLOQUE DE PROBABILIDAD	10
5. BLOQUE DE INFERENCIA	13
6. CONCLUSIONES	17
7. BIBLIOGRAFÍA	19

1. OBJETIVOS DEL TRABAJO

En este trabajo se pretende dar respuesta a la siguiente pregunta: ¿Hay diferencias de salario entre hombres y mujeres? Pienso que es un tema importante que debo investigar.

Mi interés surgió tras ver un vídeo que encontré en el aula virtual de la asignatura Matemáticas II (a la que entré sin invitación, pues al fin y al cabo soy una hacker). En el vídeo se narra una historia real sobre la discriminación salarial por sexo. Esta historia acaeció a principios de los 80's: desde 1980 hasta que tras bastantes esfuerzos, 6 años después (en julio de 1986), por fin consiguieron la igualdad salarial en este ayuntamiento gracias al uso de la Estadística. A pesar de tener la misma valoración del puesto de trabajo, la diferencia de salario mensual entre delineantes y secretarias era de varios cientos de euros, si lo actualizamos a hoy en día.

2. LOS DATOS

La población de estudio que voy a considerar son hombres y mujeres residentes en España mayores de edad. Las variables que consideraré, con sus códigos respectivos, son:

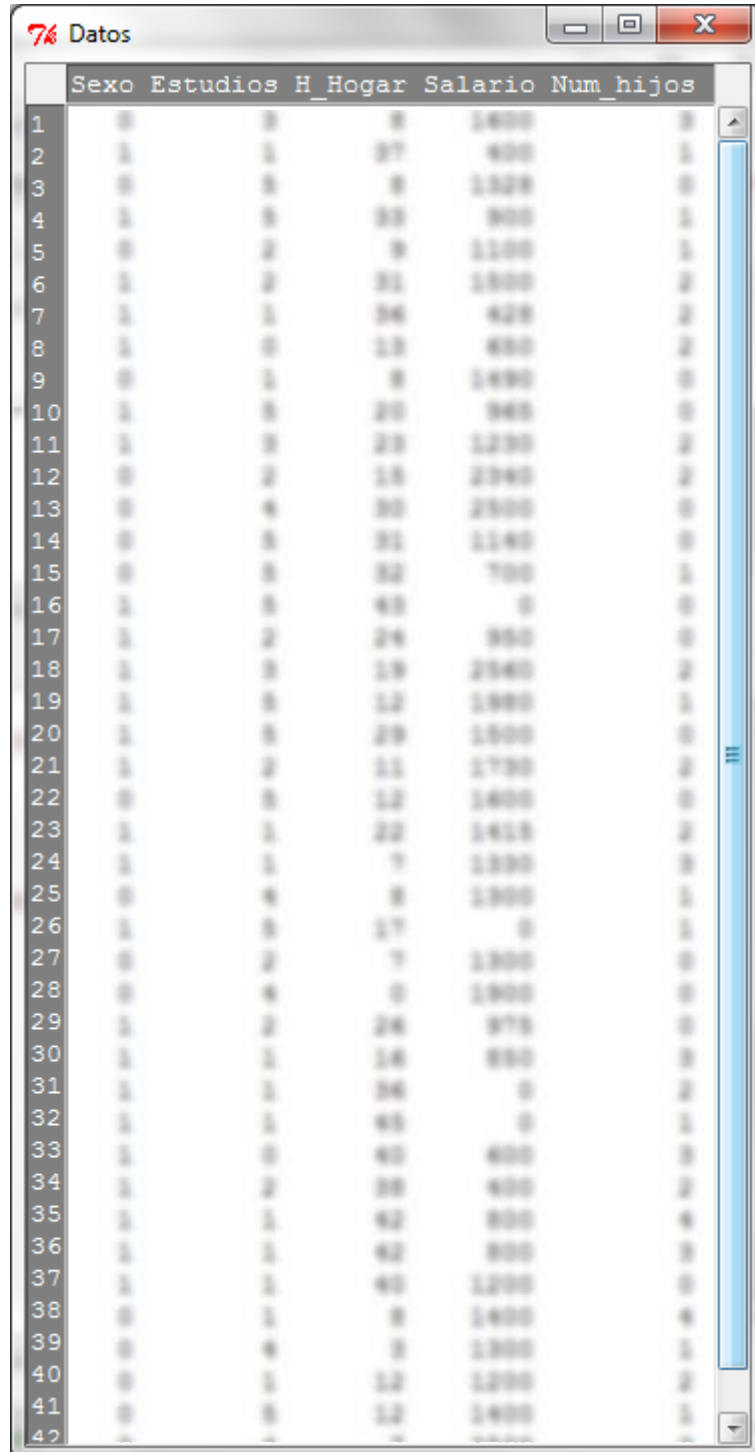
- Sexo: 0 = Hombre, 1 = Mujer
- Estudios: estudios máximos oficiales que posee, 0 = No dispone del graduado escolar, 1 = Graduado escolar (enseñanza obligatoria, EBG, ESO), 2 = Formación profesional de grado medio, 3 = Bachillerato, 4 = Formación profesional superior, 5 = Estudios universitarios
- H_Hogar: horas dedicadas a la semana a las tareas domésticas y al cuidado de familiares (niños y ancianos) (en la última semana).
- Salario: ingresos mensuales propios en euros (en el último mes)
- Número de hijos

Para recolectar los datos debo tomar una muestra aleatoria (punto 5.3. del temario) de dicha población. Alguien convencional, preguntaría a estas personas por las variables expuestas, sin embargo voy a aprovechar que soy una detective y una hacker, para conseguir esta información. He conseguido el censo electoral y como se explica en la primera práctica de Estadística (problema 2), he seleccionado aleatoriamente la muestra, después he investigado a estas personas para conocer sus ingresos (fue fácil hackear los bancos y conocer sus cuentas), las horas dedicadas al hogar y la familia sin que me mintieran (con un teleobjetivo pude vigilarlos y conocerlas exactamente,) etc.

Se muestra primero cómo se obtuvo la muestra aleatoria del censo y a continuación se muestran algunos de los datos recogidos, a cuyos

dueños mantendré en el anonimato, y que trataré de forma confidencial (por ello la base de datos aparece desenfocada).

```
> dni=12345678
> set.seed(dni)
> sample(34301332,43)
[1] 29996283 9905885 32490624 30340239 297680 17901589 13676128 32929951 27917979 3171124 7749429
[12] 34066649 7712726 18749060 28722345 28278017 29643462 24057586 34117636 19282528 10284342 9806775
[23] 14145117 13890235 13369172 13678715 16433953 25740707 10625509 23820873 32007837 2164091 12745435
[34] 20806205 32874947 8324434 21484588 31006599 23887112 9431505 14498603 20497446 710438
```



	Sexo	Estudios	H_Hogar	Salario	Num_hijos
1	0	0	0	1400	0
2	1	1	37	400	1
3	0	0	0	1328	0
4	1	0	33	900	1
5	0	2	9	1100	1
6	1	2	31	1500	2
7	1	1	34	428	2
8	1	0	13	400	2
9	0	1	0	1490	0
10	1	0	20	945	0
11	1	0	23	1230	2
12	0	2	15	2340	2
13	0	4	30	2500	0
14	0	0	31	1140	0
15	0	0	32	700	1
16	1	0	43	0	0
17	1	2	24	950	0
18	1	0	19	2340	2
19	1	0	12	1980	1
20	1	0	28	1500	0
21	1	2	11	1730	2
22	0	0	12	1400	0
23	1	1	22	1415	2
24	1	1	7	1330	0
25	0	4	0	1300	1
26	1	0	17	0	1
27	0	2	7	1300	0
28	0	4	0	1900	0
29	1	2	24	975	0
30	1	1	14	850	0
31	1	1	34	0	2
32	1	1	45	0	1
33	1	0	40	400	0
34	1	2	38	400	2
35	1	1	42	800	4
36	1	1	42	800	0
37	1	1	40	1200	0
38	0	1	0	1400	4
39	0	4	3	1300	1
40	0	1	12	1200	2
41	0	0	12	1400	1
42	-	-	-	-

3. BLOQUE DE DESCRIPTIVA

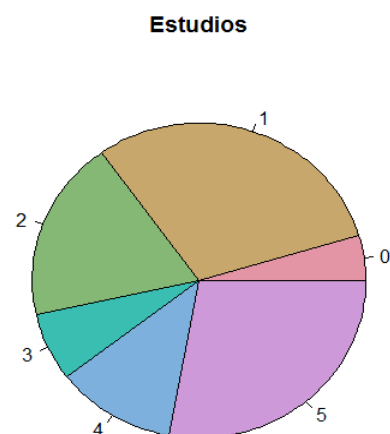
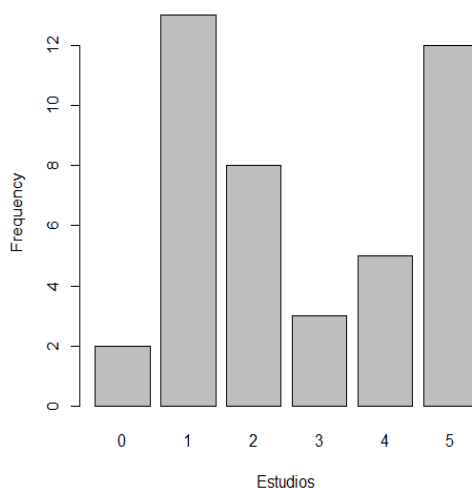
Describiremos en primer lugar cada una de las variables por separado, y después analizaremos algunas combinaciones.

Comencemos con las variables categóricas, Sexo (no ordinal) y Estudios (que podemos considerar ordinal). (Su combinación se estudiará en el bloque de probabilidad).

	Frec. absoluta	Frec. relativa	Frec. acumulada	Frec. relativa acumulada
0	18	41.86047	18	41.86047
1	25	58.13953	43	100.00000

Aproximadamente el 42% de la muestra son hombres, y el 58% mujeres. En lo que respecta a Estudios, según la siguiente tabla de frecuencias y gráficos (diagrama de barras y diagrama de sectores), podemos destacar que casi el 35% no tiene estudios más allá de los obligatorios, y el 30% realizó formación profesional (media o superior).

	Frec. absoluta	Frec. relativa	Frec. acumulada	Frec. relativa acumulada
0	2	4.651163	2	4.651163
1	13	30.232558	15	34.883721
2	8	18.604651	23	53.488372
3	3	6.976744	26	60.465116
4	5	11.627907	31	72.093023
5	12	27.906977	43	100.000000



Analicemos ahora las variables numéricas, empezando por la discreta, Número de hijos, y después analizaremos las dos continuas: Horas en el hogar y Salario.

```
| Frec. absoluta Frec. relativa Frec. acumulada Frec. relativa acumulada
0          14      32.558140          14      32.55814
1          10      23.255814          24      55.81395
2          12      27.906977          36      83.72093
3           5      11.627907          41      95.34884
4           2       4.651163          43     100.00000

      mean      sd IQR      cv skewness  kurtosis 0% 25% 50% 75% 100% n
1.325581 1.189656  2 0.8974601 0.4482993 -0.8237354 0  0  1  2  4 43
```

En cuanto al número de hijos, la moda correspondería a cero hijos, aunque tanto los valores dos hijos y un hijo son bastante frecuentes. Las familias numerosas (tres o más hijos), en cambio, no son muy frecuentes (aproximadamente el 16%). El número medio de hijos es de 1.33 y el número mediano de 1.

Primero analizaremos las variables Horas en el hogar y Salario globalmente, y después sería muy interesante analizarlas según el Sexo.

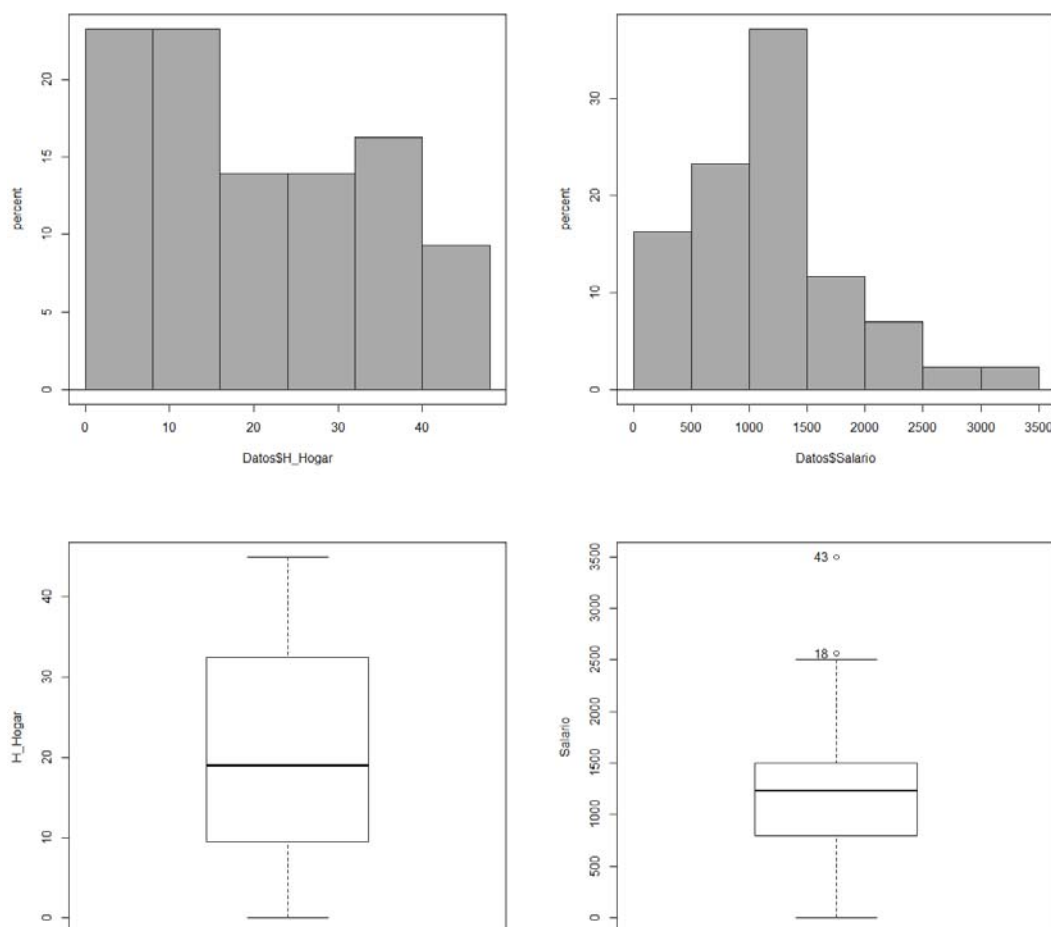
```
      mean      sd IQR      cv skewness  kurtosis 0% 25% 50% 75% 100% n
H_Hogar  21.44186 13.21309 23 0.6162288 0.2721691 -1.3887270 0  9.5 19 32.5 45 43
Salario 1227.00000 743.29697 700 0.6057840 0.6082655  0.6684817 0 800.0 1230 1500.0 3500 43
```

El salario medio mensual y el mediano son muy cercanos, en torno a 1230 euros, aunque la variación es grande (desviación típica de 743 euros), de hecho el rango intercuartílico es de 700, es decir, la diferencia entre el percentil 25 (el 25% tiene un salario inferior a 800) y el 75 (el 25% tiene un salario superior a 1500) es de 700 euros. En términos relativos, el coeficiente de variación (60.6%) también corrobora esta gran variación. También podemos apreciar una asimetría positiva (0.61), y una curtosis positiva (en el R le resta 3 respecto de la vista en clase), así que es más picuda que la campana de Gauss, con colas más alargadas.

En cuanto al número de horas en el hogar, hay una ligera discrepancia entre la media y mediana (algún valor o valores han hecho aumentar algo la media respecto de la mediana), aunque no muy destacable, ambos valores están en torno a las 20 horas semanales. Al igual que con el salario mensual, la variación (desviación típica de 13 horas) del número de horas dedicadas al cuidado de familiares (hijos y ancianos) y a las tareas domésticas es muy grande, de hecho varía desde 0 (el mínimo) a 45 (el máximo), y también en términos relativos, mediante la medida adimensional del coeficiente de variación que compara media y desviación típica podemos apreciarlo (61.6%). Hay una ligera asimetría, aunque lo que sí es destacable a nivel de forma de la

distribución es la curtosis, teniendo un apuntamiento negativo, que corresponde a una distribución más plana que la campana de Gauss.

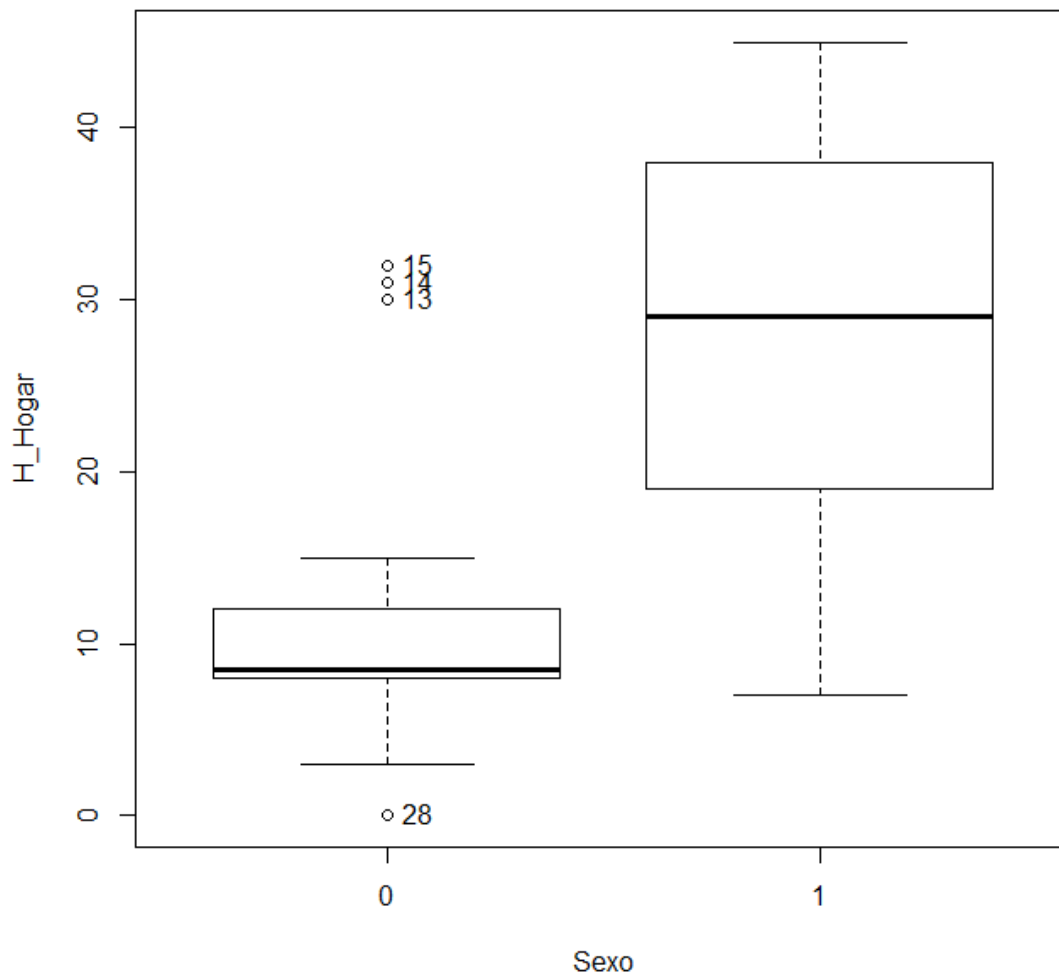
Estas descriptivas pueden apreciarse también en los histogramas de frecuencias relativas y diagramas de cajas para estas dos variables que se muestran seguidamente. La forma del histograma para Salario efectivamente es más picuda con una cola superior más alargada (la asimetría), de hecho en el diagrama de cajas podemos ver dos outliers (correspondientes a los salarios 2560 y 3500 euros). Para las horas en el hogar también podemos observar un histograma bastante más plano y una mayor simetría mediante el diagrama de cajas.



Sin embargo, en las anteriores descriptivas aparecen mezcladas las dos muestras, las correspondientes a hombres y mujeres, y es posible que no se distribuyan igual, por ello repetiremos los análisis pero esta vez desagregando los resultados y gráficas por Sexo. Comencemos por las horas dedicadas al hogar y familia. En las descriptivas numéricas, aparecen primero los resultados

para hombres (codificados con un 0), y luego para mujeres (codificadas con un 1).

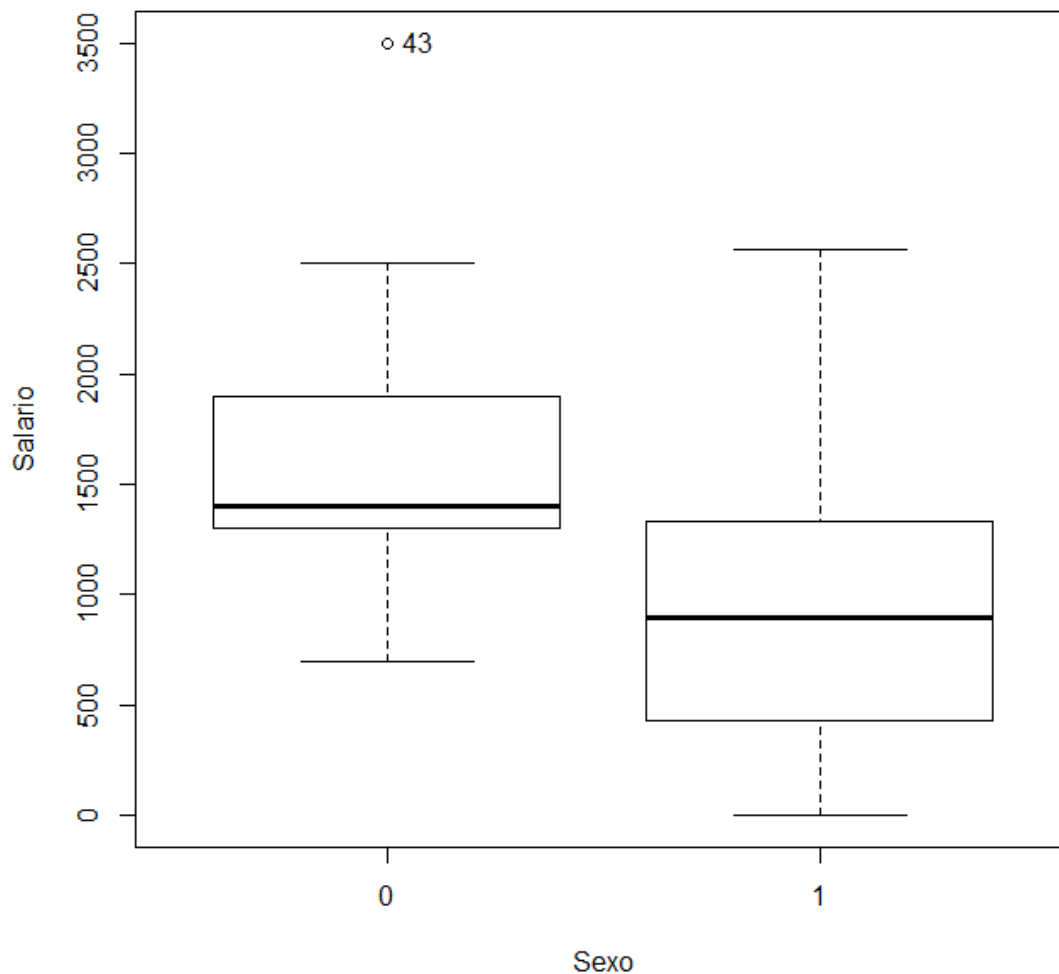
	mean	sd	IQR	cv	skewness	kurtosis	0%	25%	50%	75%	100%	data:n
0	12.22222	9.270799	4	0.7585199	1.1631000	0.08379824	0	8	8.5	12	32	18
1	28.08000	11.618663	19	0.4137701	-0.1784592	-1.41429648	7	19	29.0	38	45	25



Es claramente apreciable la distinta distribución de las horas en hombres y mujeres (basta en fijarse en la distinta localización y anchura de las cajas). De hecho, en las mujeres, el mínimo número de horas corresponde a 7, mientras que en los hombres a 0, y en el diagrama de cajas para hombres vemos que el valor máximo 32 es un outlier (hay otros tres outliers, uno en el 30 y 31 y otro inferior en el 0), en cambio para mujeres el valor máximo es de 45 horas. No sólo son diferentes en el rango, sino también en la localización (en la mujeres la media es de 28 horas, mientras que en los hombres de 12) y en la forma (en las mujeres hay simetría y la curtosis es negativa, mientras que en los hombres hay asimetría positiva y la curtosis es positiva, aunque casi nula).

A continuación aparecen los resultados en lo concerniente a Salario desagregados por Sexo.

	mean	sd	IQR	cv	skewness	kurtosis	0%	25%	50%	75%	100%	data:n
0	1644.333	673.0049	525	0.4092874	1.2010013	0.8494469	700	1300	1400	1825	3500	18
1	926.520	649.1172	902	0.7005971	0.4382668	-0.2371475	0	428	900	1330	2560	25



De nuevo, se observa una clarísima diferencia entre ambas distribuciones, especialmente en lo que respecta a la localización: el primer cuartil de los hombres es 1300, o sea el 75% de los hombres cobra más de 1300, mientras que justamente esa cantidad (1330 euros) es el tercer cuartil de las mujeres, o sea el 25% de las mujeres cobra más de 1330. En cuanto a centralidad, el salario mediano (que es similar al medio para las mujeres) es de 900 euros en el caso de las mujeres, en cambio sube a 1400 euros para los hombres (en este caso el salario medio es bastante superior, 1644, a la mediana, puesto que hay valores que hacen subir la media, pues la distribución es claramente asimétrica: muchos cobran poco, y unos pocos cobran mucho), es decir, el salario femenino mediano es un 35.7% inferior al masculino ($1400 \cdot 0.643 = 900$).

En las mujeres la distribución también es asimétrica, aunque no tan pronunciada como en los hombres, donde la curtosis también es bastante elevada.

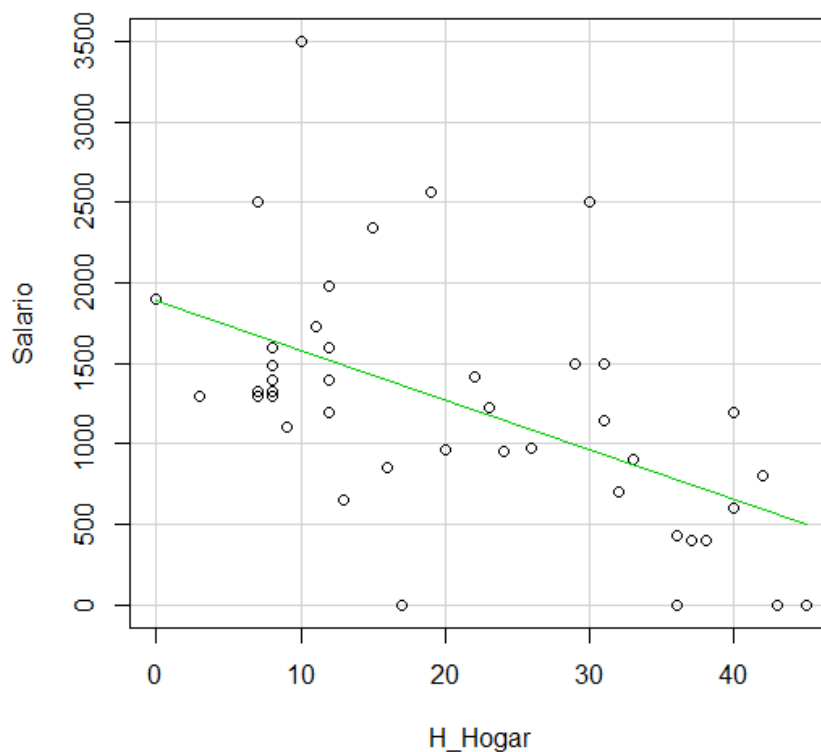
Como tenemos dos variables continuas, podemos estudiar si existe relación entre ambas, mediante un análisis de regresión. Vamos a calcular la recta de regresión del Salario sobre el Número de horas en el hogar.

```
Call:
lm(formula = Salario ~ H_Hogar, data = Datos)

Residuals:
    Min       1Q   Median       3Q      Max
-1364.3  -344.7  -119.0   208.6  1919.2

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1890.035    184.252   10.258 6.91e-13 ***
H_Hogar      -30.922      7.339   -4.213 0.000135 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 628.5 on 41 degrees of freedom
Multiple R-squared:  0.3022, Adjusted R-squared:  0.2851
F-statistic: 17.75 on 1 and 41 DF, p-value: 0.0001346
```



```

      H_Hogar  Salario
H_Hogar 1.0000000 -0.5496875
Salario -0.5496875  1.0000000

```

La recta de regresión sería: $\text{Salario} = 1890.035 - 30.922 \cdot H_{\text{Hogar}}$. De la gráfica, a través de la pendiente de la recta (- 30.922), o bien mediante el coeficiente de correlación (-0.55), podría desprenderse una relación lineal negativa entre ambas variables (a más horas en el hogar, menos salario). Aunque si miramos la gráfica y el coeficiente de determinación ($R^2 = 0.3022$) bastante alejado del uno y cercano a cero, veremos que la relación entre ambas variables es muy débil, con lo cual el Número de horas en el hogar no es una buena variable para predecir el Salario.

4. BLOQUE DE PROBABILIDAD

Consideremos las variables categóricas Sexo y Estudios, y la siguiente tabla de contingencia, que nos proporciona las frecuencias cruzadas y (relativas) marginales de ambas variables (aunque las marginales ya las habíamos considerado en el bloque de descriptiva).

```

      Estudios
Sexo  0  1  2  3  4  5
0     0  0  3  3  1  5  6
1     2 10  5  2  0  6

> totPercents(.Table) # Percentage of Total
      0      1      2      3      4      5 Total
0     0.0  7.0  7.0  2.3 11.6 14.0  41.9
1     4.7 23.3 11.6  4.7  0.0 14.0  58.1
Total  4.7 30.2 18.6  7.0 11.6 27.9 100.0

```

En base a dicha tabla podemos estimar algunas probabilidades, por ejemplo:

$P(\text{Estudios superiores} | \text{Mujeres}) = 6/25 = 14/58.1 = 0.24$, es decir, entre las mujeres, el 24% tiene estudios superiores, mientras que en los hombres, $P(\text{Estudios superiores} | \text{Hombres}) = 6/18 = 14/41.9 = 0.33$, es el 33%.

$P(\text{Estudios superiores} \cap \text{Mujer}) = 6/43 = 13.95\%$, casi el 14% es mujer con estudios universitarios. De igual forma, $P(\text{Estudios superiores} \cap \text{Hombre}) = 6/43$, el 14% es hombre con estudios universitarios.

Por último, la probabilidad de que tenga estudios superiores o sea mujer es: $P(\text{Estudios superiores} \cup \text{Mujer}) = P(\text{Estudios superiores}) + P(\text{Mujer}) - P(\text{Estudios superiores} \cap \text{Mujer}) = 0.2791 + 0.5814 - 0.1395 = 0.721$.

El coeficiente de contingencia es de 0.45. Su interpretación se deja para el bloque de inferencia:

```

Pearson's Chi-squared test

data: .Table
X-squared = 10.7479, df = 5, p-value = 0.05661

> chi2=10.7479
> N=43
> sqrt(chi2/(N+chi2))
[1] 0.4471786

```

Como se señaló en el bloque de descriptiva, el 58% son mujeres, podemos tomar como 0.58 la probabilidad de que se sea mujer en una muestra de adultos. Si consideramos una muestra de 100 adultos, y consideramos la variable X = Número de mujeres de la muestra, tendríamos que esta variable sería una Binomial($n=100$, $p=0.58$), y podemos calcular algunas probabilidades. Por ejemplo, la probabilidad de que no haya ninguna mujer en la muestra es:

$P(X=0) = 2 \cdot 10^{-38}$. La probabilidad de que hayan exactamente 10 mujeres es $P(X=10) = 9 \cdot 10^{-24}$.

```

      Pr
0  2.113144e-38
1  2.918151e-36
2  1.994765e-34
3  8.998605e-33
4  3.013461e-31
5  7.989977e-30
6  1.747015e-28
7  3.239702e-27
8  5.200879e-26
9  7.341771e-25
10 9.226158e-24
11 1.042436e-22

> pbinom(c(0,49,50), size=100, prob=0.58, lower.tail=TRUE)
[1] 2.113144e-38 4.321458e-02 6.497755e-02
> pbinom(c(49,50), size=100, prob=0.58, lower.tail=FALSE)
[1] 0.9567854 0.9350225

```

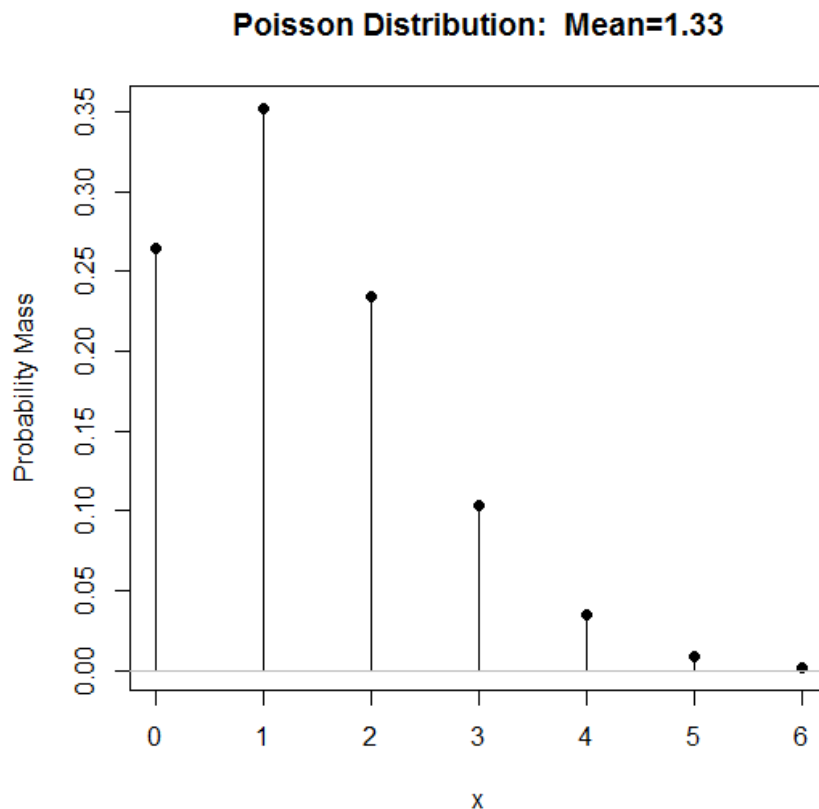
La probabilidad de que haya más de 50 mujeres y al menos 50 son respectivamente:

$P(X>50) = 0.9350225$ y $P(X \geq 50) = P(X>49) = 0.9567854$ que también podía haberse obtenido como $P(X \geq 50) = P(X>50) + P(X=50) = 0.935 + 0.022 = 1 - P(X \leq 49) = 1 - 0.04321458$.

El número medio de mujeres que esperaríamos tener en la muestra es: $100 \cdot 0.58 = 58$.

Cambiamos de variable y consideremos la variable Y = Número de hijos, que podría considerarse Poisson de $\lambda=1.33$ (la media).

Aquí vemos representada, la probabilidad para distintos valores (la función de probabilidad):



La probabilidad de no tener hijos sería $P(Y=0) = 0.2645$, de tener al menos uno es $P(Y \geq 1) = P(Y > 0) = 0.7355 = 1 - 0.2645$, de tener más de cuatro hijos $P(Y > 4) = 0.0117 = 1 - P(Y \leq 4) = 1 - 0.9883$, y de no ser familia numerosa $P(Y < 3) = 0.8501 = P(Y \leq 2)$.

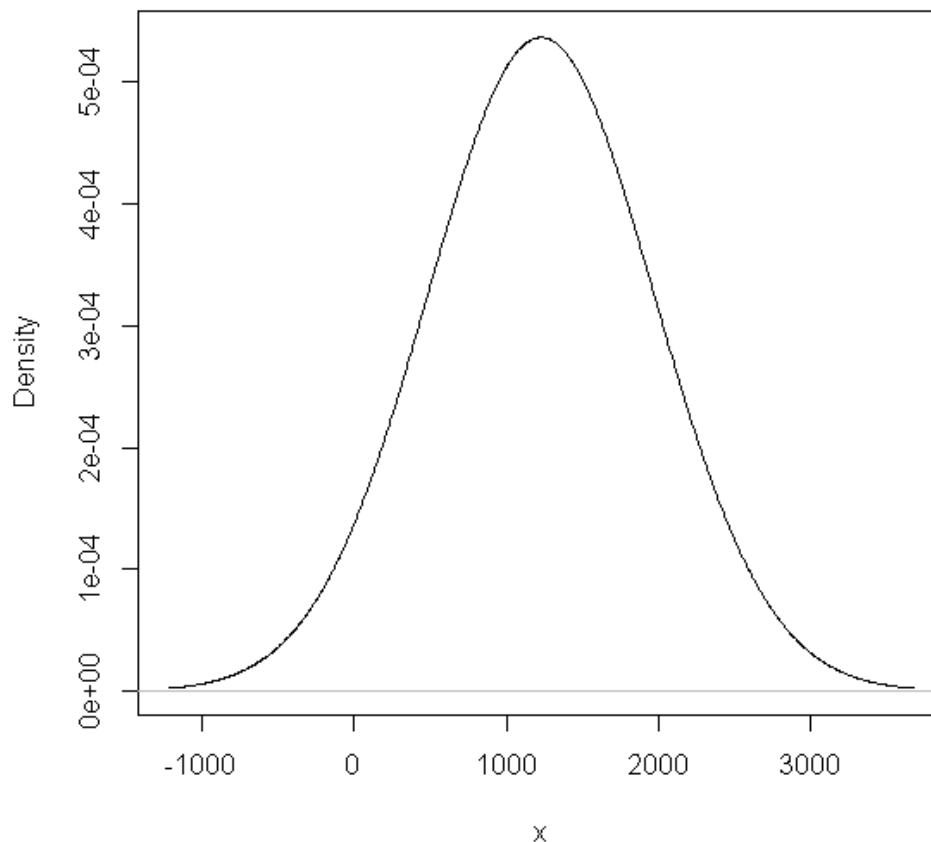
```
> ppois(c(0,1,2,3,4,5), lambda=1.33, lower.tail=TRUE)
[1] 0.2644773 0.6162320 0.8501489 0.9538521 0.9883334 0.9975054

> 1-0.2644773
[1] 0.7355227

> ppois(c(4), lambda=1.33, lower.tail=FALSE)
[1] 0.0116666
```

Por último, vamos a considerar las variables continuas. Vamos a centrarnos en la variable Salario, cuyo histograma se asemejaba más al de una Normal que no el de Horas en el hogar. Si consideramos la variable Salario como una Normal de media 1227 euros y desviación típica 743.297 euros, la representación de su función de densidad es:

Normal Distribution: Mean=1227, Standard deviation=743.29



La probabilidad de que gane menos de 1000 euros es de 0.38, de que gane entre 2000 y 1000 euros es: $0.85 - 0.38 = 0.47$. La probabilidad de que gane más de 2000 es de $1 - 0.85 = 0.15$.

```
> pnorm(c(1000,2000), mean=1227, sd=743.29, lower.tail=TRUE)
[1] 0.3800312 0.8508233
```

Podemos también calcular diversos percentiles: el 25% cobra menos de 725 euros (pensemos que se está considerando a todos los adultos, no sólo los que tienen alguna remuneración, es decir, por ejemplo, se incluyen estudiantes sin trabajo) y el 90% menos de 2180.

```
> qnorm(c(0.25,0.9), mean=1227, sd=743.29, lower.tail=TRUE)
[1] 725.6585 2179.5645
```

5. BLOQUE DE INFERENCIA

A lo largo de todo este bloque consideraremos un α de 0.05.

Vamos a empezar con el contraste de independencia de las variables Sexo y Estudios:

H_0 : las variables Sexo y Estudios son independientes

H_1 : las variables Sexo y Estudios no son independientes

```
Pearson's Chi-squared test

data: .Table
X-squared = 10.7479, df = 5, p-value = 0.05661
```

En base al p-valor, 0.0566, superior al α , no rechazamos la hipótesis nula, así que no tenemos pruebas suficientes para afirmar que las variables Sexo y Estudios sean dependientes, por tanto, nos quedamos con la hipótesis de partida, la independencia de ambas variables. Eso sí debemos notar que el p-valor es bastante pequeño, y hemos estado al borde de rechazar H_0 .

Centrémonos ahora en la variable Salario y estudiemos la diferencia por sexo (Salario para hombres y Salario para mujeres). Para poder emplear los métodos vistos en clase, asumiremos que cada una de ellas es Normal (aunque por ejemplo en el Salario para hombres vimos que su histograma no se asemejaba mucho a la campana de Gauss, y era bastante asimétrico). Notemos que son muestras independientes.

En primer lugar, veremos si las varianzas poblacionales difieren:

```
> var.test(Salario ~ Sexo, alternative='two.sided', conf.level=.95, data=Datos)

F test to compare two variances

data: Salario by Sexo
F = 1.075, num df = 17, denom df = 24, p-value = 0.853
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4504353 2.7516956
sample estimates:
ratio of variances
 1.074955
```

El intervalo de confianza al 95% para el cociente de varianzas (σ_1^2 / σ_2^2) es: (0.45, 2.75). Como 1 \in (0.45, 2.75), podemos asumir que las varianzas poblacionales (las varianzas de los salarios de las mujeres y hombres adultos) son iguales, $\sigma_1^2 = \sigma_2^2$.

Ahora comparemos las medias:

```
> t.test(Salario~Sexo, alternative='two.sided', conf.level=.95, var.equal=TRUE,
+ data=Datos)

Two Sample t-test

data: Salario by Sexo
t = 3.523, df = 41, p-value = 0.001063
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 306.3324 1129.2943
sample estimates:
mean in group 0 mean in group 1
 1644.333          926.520
```

El intervalo de confianza al 95% para la diferencia de medias ($\mu_1 - \mu_2$) es: (306.3, 1129.3). Como $0 \notin (306.3, 1129.3)$, podemos afirmar que las medias poblacionales (los salarios medios de las mujeres y hombres adultos) son distintas, $\mu_1 \neq \mu_2$.

Con un 95% de confianza el salario medio de los hombres, μ_1 , va de 1309.7 a 1979, mientras que el de las mujeres va de 658.6 a 1194.5. Para poder obtener estos intervalos sin tener que introducir de nuevo los datos en dos variables (Salario para hombres y Salario para mujeres), hemos empleado la ventana de comandos seleccionando los datos de Salario para cada Sexo:

```
> t.test(Datos$Salario[Datos$Sexo==0], alternative='two.sided', mu=0.0, conf.level=.95)

One Sample t-test

data: Datos$Salario[Datos$Sexo == 0]
t = 10.3659, df = 17, p-value = 9.119e-09
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1309.656 1979.011
sample estimates:
mean of x
 1644.333

> t.test(Datos$Salario[Datos$Sexo==1], alternative='two.sided', mu=0.0, conf.level=.95)

One Sample t-test

data: Datos$Salario[Datos$Sexo == 1]
t = 7.1368, df = 24, p-value = 2.239e-07
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 658.5776 1194.4624
sample estimates:
mean of x
 926.52
```

Repitamos estos análisis para la variable, Número de horas dedicadas al hogar y cuidado de familiares, que asumiremos Normal para cada Sexo. De nuevo, las varianzas podemos suponerlas iguales, porque $1 \in (0.27, 1.63)$, como se ve en la imagen:


```

F test to compare two variances

data:  H_Hogar by Sexo
F = 0.6367, num df = 17, denom df = 24, p-value = 0.3408
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2667867 1.6297919
sample estimates:
ratio of variances
 0.6366812

```

Las medias también son distintas, puesto que $0 \notin (-22.54, -9.17)$. Nótese que con un 95% de confianza el número de horas en el hogar medio de los hombres, μ_1 , va de 7.6 a 16.8, mientras que en las mujeres va de 23.3 a 32.9.

```

> t.test(H_Hogar~Sexo, alternative='two.sided', conf.level=.95, var.equal=TRUE, data=Datos)

Two Sample t-test

data:  H_Hogar by Sexo
t = -4.7909, df = 41, p-value = 2.192e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-22.542472 -9.173083
sample estimates:
mean in group 0 mean in group 1
 12.22222      28.08000

> t.test(Datos$H_Hogar[Datos$Sexo==0], alternative='two.sided', mu=0.0, conf.level=.95)

One Sample t-test

data:  Datos$H_Hogar[Datos$Sexo == 0]
t = 5.5933, df = 17, p-value = 3.23e-05
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 7.611962 16.832482
sample estimates:
mean of x
 12.22222

> t.test(Datos$H_Hogar[Datos$Sexo==1], alternative='two.sided', mu=0.0, conf.level=.95)

One Sample t-test

data:  Datos$H_Hogar[Datos$Sexo == 1]
t = 12.084, df = 24, p-value = 1.082e-11
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 23.28405 32.87595
sample estimates:
mean of x
 28.08

```

Nosotros hemos asumido que las variables Salario y Horas en el hogar para ambos Sexos eran Normales, a pesar de que las gráficas y las descriptivas no indicaban un buen ajuste a una Normal en algunos de los casos.

En la penúltima práctica de Estadística (y el último punto del temario de teoría, 6.4), vimos que en caso de no cumplirse la Normalidad podríamos usar el test de Kruskal-Wallis. Vamos a usarlo en cada caso, para asegurarnos de que nuestras conclusiones son válidas.

```

> kruskal.test(Salario ~ Sexo, data=Datos)

Kruskal-Wallis rank sum test

data: Salario by Sexo
Kruskal-Wallis chi-squared = 9.7137, df = 1, p-value = 0.001829

> tapply(Datos$H_Hogar, Datos$Sexo, median, na.rm=TRUE)
 0      1
8.5 29.0

> kruskal.test(H_Hogar ~ Sexo, data=Datos)

Kruskal-Wallis rank sum test

data: H_Hogar by Sexo
Kruskal-Wallis chi-squared = 16.5475, df = 1, p-value = 4.745e-05

```

El p-valor=0.0018 es inferior al $\alpha=0.05$, por tanto, efectivamente en cuanto a Salario hay diferencia entre mujeres y hombres. También, la hay en cuanto a Horas en el hogar (p-valor= 0.000047 inferior al $\alpha=0.05$).

Por último, podemos también fijarnos en las posibles diferencias según el nivel de Estudios. Sí que hay diferencia según el nivel de estudios en el Salario (p-valor= 0.044 inferior al $\alpha=0.05$), como se podría intuir, pero no en el Número de horas en el hogar (p-valor=0.14 > $\alpha=0.05$).

```

> kruskal.test(Salario ~ Estudios, data=Datos)

Kruskal-Wallis rank sum test

data: Salario by Estudios
Kruskal-Wallis chi-squared = 11.3846, df = 5, p-value = 0.04427

> tapply(Datos$H_Hogar, Datos$Estudios, median, na.rm=TRUE)
 0      1      2      3      4      5
26.5 36.0 19.5 19.0  7.0 18.5

> kruskal.test(H_Hogar ~ Estudios, data=Datos)

Kruskal-Wallis rank sum test

data: H_Hogar by Estudios
Kruskal-Wallis chi-squared = 8.2494, df = 5, p-value = 0.143

```

6. CONCLUSIONES

A través de este estudio hemos comprobado la diferencia de Salario entre mujeres y hombres, así como en el distinto número de horas dedicadas a las tareas domésticas y cuidados de familiares.

La muestra seleccionada fue aleatoria (caso de no haberlo sido, y haber tomado una muestra de conveniencia por falta de medios, éste sería un punto a mejorar). El tamaño muestral no fue muy grande (43 personas), pues tuve que estar investigando a cada persona y esto me llevaba bastante tiempo. Si por ejemplo, hubiera querido restringir el error de la estimación de la proporción de mujeres en la población, y que no se fuera en más del 1% del real con un 95% de confianza, habría debido tomar al menos $9359 (0.58 (1-0.58)(1.96/0.01)^2 = 9358.1376)$ muestras.

Aunque el tamaño muestral de nuestro estudio es pequeño, en estudios con un mayor tamaño muestral y realizado con más medios, se llegaron a las mismas conclusiones (véase la Encuesta de Empleo del Tiempo del INE con 24.000 hogares muestreados y la Distribución de los Salarios del INE, con un tamaño muestral de 215.000).

Este trabajo surgió, como se explica en el primer punto, a raíz de un vídeo sobre discriminación salarial en los años 80 en Estados Unidos. Pese a que ha llovido bastante desde entonces, vemos que en Estados Unidos no fue hasta el año 2009 que se aprobó la Ley de Salario Justo Lilly Ledbetter, que firmó Obama el 29 de enero de 2009. La ley recibió su nombre de Lilly Ledbetter, que sufrió discriminación salarial durante casi dos décadas. Trabajó cerca de 20 años para la compañía de neumáticos Goodyear en una fábrica de Alabama. Al cabo de ese tiempo, se enteró de que ganaba el sueldo más bajo de los 16 supervisores de la planta, a pesar de contar con más experiencia que sus compañeros varones.

Cuando llevó finalmente el caso a juicio, fue desestimado por el Tribunal Supremo por 5 votos contra 4. El argumento fue que había tardado demasiado tiempo en presentar la demanda y no tenía derecho por tanto a la compensación. La ley no acaba automáticamente con la discriminación salarial. Sólo permite que las personas damnificadas puedan recurrir ante los tribunales. El plazo para hacerlo es de 180 días después de recibir la nómina en cuestión. En Estados Unidos, las mujeres ganan de media un 22% menos que los varones por realizar el mismo trabajo.

Si ahora miramos a la Unión Europea, desgraciadamente, también nos encontramos con diferencias. De hecho, en el Parlamento Europeo (<http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A6-2008-0325&language=ES>) se reivindicó el 22 de febrero como Día Internacional de la Igualdad Salarial. En concreto, el 22 de febrero ya que ésta es la fecha hasta la que tienen que trabajar las mujeres europeas para equiparar sus salarios al de los varones en 12 meses.

Este trabajo podría ampliarse y mejorarse, además de tomando más muestras (con muestreo aleatorio), considerando otras variables como por ejemplo, recopilando el tipo de trabajo realizado, el número de horas dedicado al trabajo

remunerado o el número de horas dedicado al ocio. Además, en lugar de restringirnos a España (donde también se podrían haber desagregado los datos por comunidad autónoma), podríamos haber ampliado nuestras miras a la Unión Europea, o a otros lugares del globo, donde la situación de la mujer sería bastante peor.

7. BIBLIOGRAFÍA

- Apuntes de la asignatura.
- Programa 3 de la serie *AGAINST ALL ODDS: INSIDE STATISTICS*, titulado *Describing Distributions*, traducido y doblado al castellano por I. Epifanio en el aulavirtual de la asignatura.
- Marcia Weskott, Jody Fitzpatrick, Lynda Dickson, Gay Francis. *Comparable Worth versus the Free Market: A Case of Study*. Frontiers, 1988.
- Encuesta de Empleo del Tiempo del INE (2007) y la Distribución de los Salarios del INE (2005) (<http://www.ine.es>), en el aulavirtual de la asignatura.
- Obama firma la ley contra la discriminación salarial de las mujeres (<http://www.rtve.es/noticias/20090129/obama-firma-ley-contra-discriminacion-salarial-las-mujeres/226331.shtml>).
- Parlamento Europeo. INFORME 28 de julio de 2008 sobre la igualdad entre mujeres y hombres – 2008 (2008/2047(INI)) Comisión de Derechos de la Mujer e Igualdad de Género. Ponente: Iratxe García Pérez. <http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A6-2008-0325&language=ES>
- Wikipedia (<http://es.wikipedia.org/>)