

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	16
7. BIBLIOGRAFÍA	18

## 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 referenciado 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.

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).

STATGRAPHICS Plus - Untitled StatFolio - [datosprojectum1.sf3]						
File Edit Plot Describe Compare Relate Special View Window Help						
	Sexo	Estudios	H_Hogar	Salario	Num_hijos	
1	M	9	8	54000	3	
2	F	5	22	4000	1	
3	M	9	8	11000	0	
4	F	9	22	9000	1	
5	M	9	9	11000	1	
6	F	9	22	21000	2	
7	F	5	26	4700	2	
8	F	8	22	4700	2	
9	M	5	8	14000	0	
10	F	9	28	9400	0	
11	F	9	22	12700	2	
12	M	9	25	21000	2	
13	M	6	28	21000	0	
14	M	9	22	11000	0	
15	M	9	22	7000	1	
16	F	9	42	0	0	
17	F	9	26	9700	0	
18	F	9	28	21000	2	
19	F	9	22	12000	1	
20	F	9	28	21000	0	
21	F	9	22	12700	2	
22	M	9	22	14000	0	
23	F	5	22	14000	2	
24	F	5	7	11700	3	
25	M	6	8	11000	1	
26	F	9	22	0	1	
27	M	9	7	11000	0	
28	M	6	8	12000	0	
29	F	9	26	9700	0	
30	F	5	26	8700	3	
31	F	5	26	0	2	
32	F	9	42	0	1	

### 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).

Frequency Table for Sexo

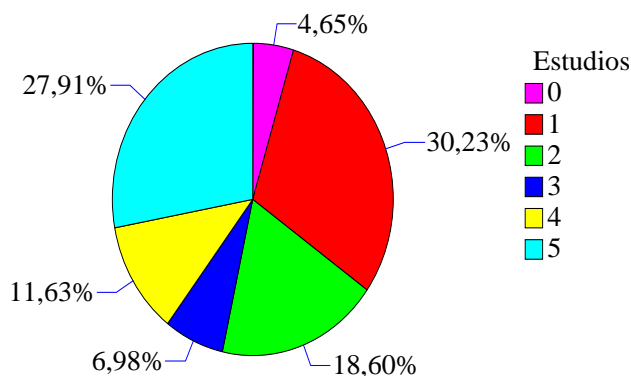
Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	0	18	0,4186	18	0,4186
2	1	25	0,5814	43	1,0000

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 sectores y diagrama de barras), podemos destacar que casi el 35% no tiene estudios más allá de los obligatorios, y el 30% realizó formación profesional (media o superior).

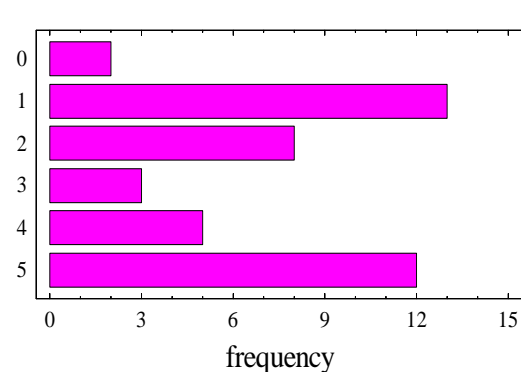
Frequency Table for Estudios

Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	0	2	0,0465	2	0,0465
2	1	13	0,3023	15	0,3488
3	2	8	0,1860	23	0,5349
4	3	3	0,0698	26	0,6047
5	4	5	0,1163	31	0,7209
6	5	12	0,2791	43	1,0000

Piechart for Estudios



Barchart for Estudios



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

Frequency Tabulation for Num\_hijos

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		0,0		14	0,3256	14	0,3256
1	0,0	1,0	0,5	10	0,2326	24	0,5581
2	1,0	2,0	1,5	12	0,2791	36	0,8372
3	2,0	3,0	2,5	5	0,1163	41	0,9535
4	3,0	4,0	3,5	2	0,0465	43	1,0000
above	4,0			0	0,0000	43	1,0000

Mean = 1,32558    Standard deviation = 1,18966

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.

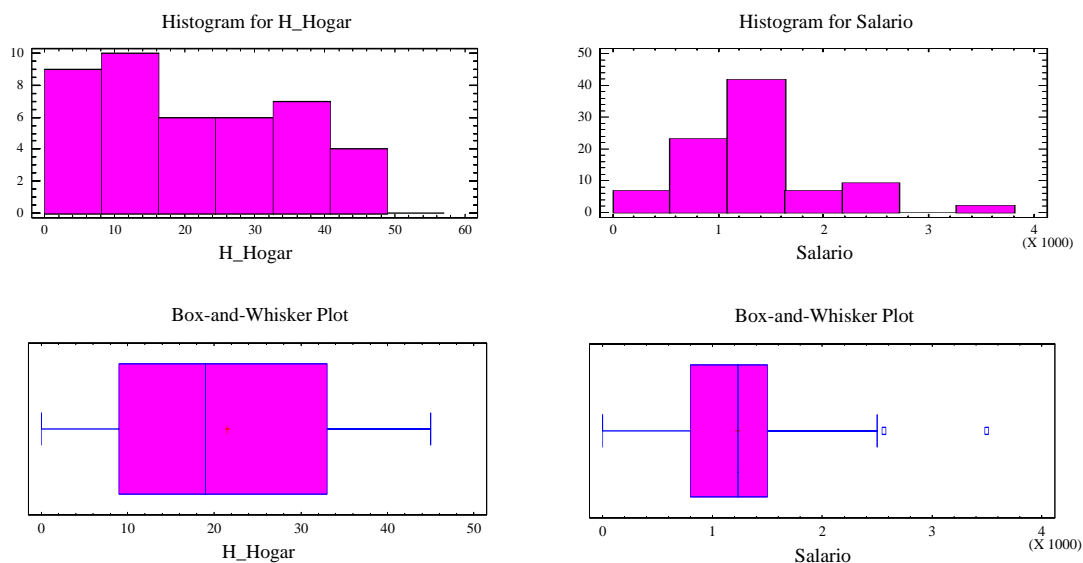
Primero analizaremos las variables Horas en el hogar y Salario globalmente, y después sería muy interesante analizarlas según el Sexo.

Summary Statistics for H_Hogar	Summary Statistics for Salario
Count = 43	Count = 43
Average = 21,4419	Average = 1227,0
Median = 19,0	Median = 1230,0
Mode = 8,0	Mode = 0,0
Variance = 174,586	Variance = 552490,0
Standard deviation = 13,2131	Standard deviation = 743,297
Minimum = 0,0	Minimum = 0,0
Maximum = 45,0	Maximum = 3500,0
Range = 45,0	Range = 3500,0
Lower quartile = 9,0	Lower quartile = 800,0
Upper quartile = 33,0	Upper quartile = 1500,0
Interquartile range = 24,0	Interquartile range = 700,0
Skewness = 0,292242	Skewness = 0,653126
Kurtosis = -1,32371	Kurtosis = 1,10611
Coeff. of variation = 61,6229%	Coeff. of variation = 60,5784%

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%) también corrobora esta gran variación. También podemos apreciar una asimetría positiva (0.65), y una curtosis positiva (en el Statgraphics 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%). 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.

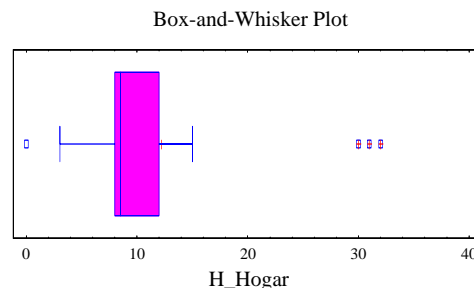
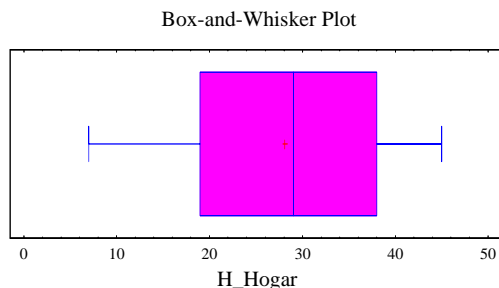
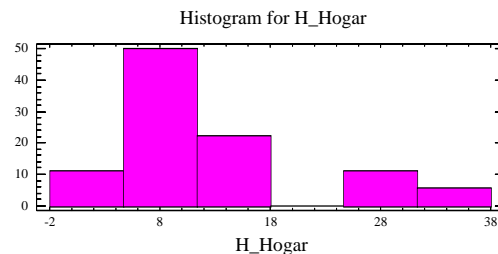
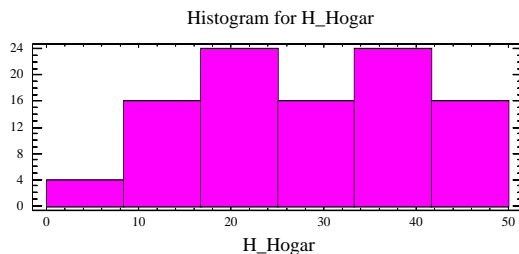
Estas descriptivas pueden apreciarse también en los histogramas de frecuencias relativas y diagramas de cajas para estas dos variables que se muestran seguidamente. La forma del histograma para Salario efectivamente es más picuda con una cola superior más alargada (la asimetría), de hecho en el diagrama 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 plano y la ligera separación entre media (en rojo en el diagrama de cajas) y la mediana (línea interior de la caja).



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 la primera columna aparecen los resultados para mujeres y en la segunda para hombres.

Summary Statistics for H_Hogar	Summary Statistics for H_Hogar
Count = 25	Count = 18
Average = 28,08	Average = 12,2222
Median = 29,0	Median = 8,5
Mode =	Mode = 8,0
Variance = 134,993	Variance = 85,9477
Standard deviation = 11,6187	Standard deviation = 9,2708
Minimum = 7,0	Minimum = 0,0
Maximum = 45,0	Maximum = 32,0
Range = 38,0	Range = 32,0
Lower quartile = 19,0	Lower quartile = 8,0
Upper quartile = 38,0	Upper quartile = 12,0
Interquartile range = 19,0	Interquartile range = 4,0
Skewness = -0,20206	Skewness = 1,38546
Kurtosis = -1,29318	Kurtosis = 1,04041
Coeff. of variation = 41,377%	Coeff. of variation = 75,852%

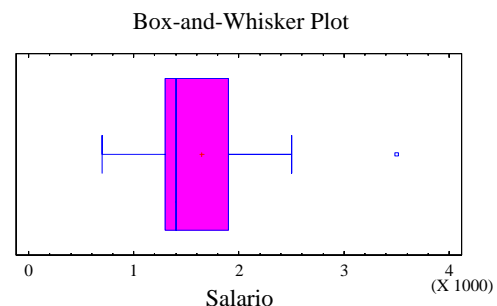
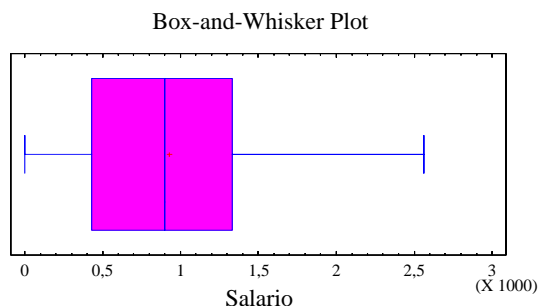
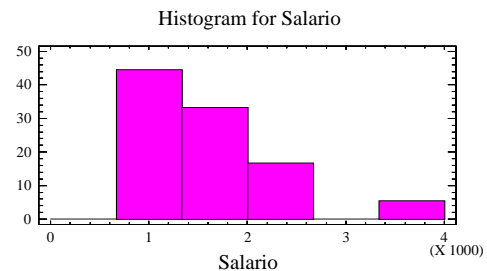
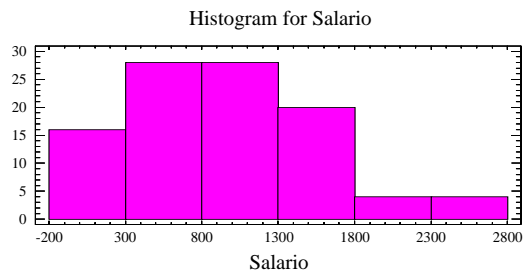


Es claramente apreciable la distinta distribución de las horas en hombres y mujeres (basta en fijarse en la distinta escala de las X's, que para las mujeres llega hasta el 50). De hecho, en las mujeres, el mínimo número de horas corresponde a 7, mientras que en los hombres a 0, y en el diagrama de cajas 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).



A continuación aparecen en la primera columna los resultados para mujeres y en la segunda para hombres en lo concerniente a Salario.

Summary Statistics for Salario	Summary Statistics for Salario
Count = 25	Count = 18
Average = 926,52	Average = 1644,33
Median = 900,0	Median = 1400,0
Mode = 0,0	Mode = 1300,0
Variance = 421353,0	Variance = 452936,0
Standard deviation = 649,117	Standard deviation = 673,005
Minimum = 0,0	Minimum = 700,0
Maximum = 2560,0	Maximum = 3500,0
Range = 2560,0	Range = 2800,0
Lower quartile = 428,0	Lower quartile = 1300,0
Upper quartile = 1330,0	Upper quartile = 1900,0
Interquartile range = 902,0	Interquartile range = 600,0
Skewness = 0,496226	Skewness = 1,4306
Kurtosis = 0,281979	Kurtosis = 2,19564
Coeff. of variation = 70,0597%	Coeff. of variation = 40,9287%



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.

Regression Analysis - Linear model:  $Y = a + b \cdot X$

Dependent variable: Salario  
Independent variable: H\_Hogar

Parameter	Estimate	Standard Error	T Statistic	P-Value
Intercept	1890,03	184,252	10,2579	0,0000
Slope	-30,9224	7,33914	-4,21336	0,0001

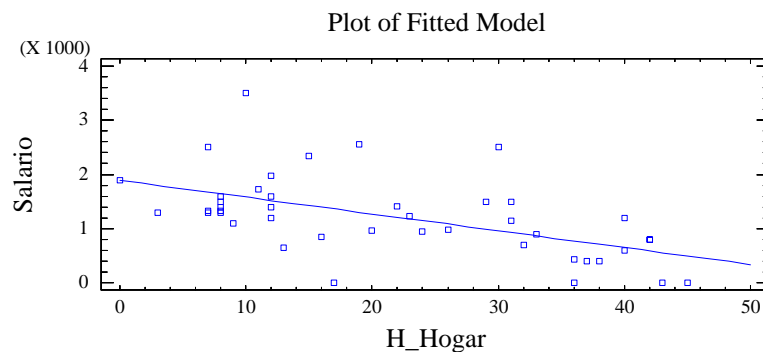
#### Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	7,01142E6	1	7,01142E6	17,75	0,0001
Residual	1,61932E7	41	394956,0		
Total (Corr.)	2,32046E7	42			

Correlation Coefficient = -0,549687

R-squared = 30,2156 percent

Standard Error of Est. = 628,455



La recta de regresión sería:  $\text{Salario} = 1890.03 - 30.9224 \cdot \text{H\_Hogar}$ . De la gráfica, a través de la pendiente de la recta (- 30.9224), 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 ( $R^2 = 0.302$ ) 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 marginales de ambas variables (aunque las marginales ya las habíamos considerado en el bloque de descriptiva).

Frequency Table for Sexo by Estudios

	0	1	2	3	4	5	Row Total
0	0 0,00%	3 6,98%	3 6,98%	1 2,33%	5 11,63%	6 13,95%	18 41,86%
1	2 4,65%	10 23,26%	5 11,63%	2 4,65%	0 0,00%	6 13,95%	25 58,14%
Column Total	2 4,65%	13 30,23%	8 18,60%	3 6,98%	5 11,63%	12 27,91%	43 100,00%

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

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

Statistic	Value	P-Value	Df
Contingency Coeff.	0,4472		

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:

$$P(X=0) = 2 \cdot 10^{-38}$$

#### Cumulative Distribution

Distribution: Binomial

Variable	Lower Tail Area (<)				
	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,0				
49	0,0277642				
50	0,0432146				

Variable	Probability Mass (=)				
	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	2,11314E-38				
49	0,0154504				
50	0,021763				

Variable	Upper Tail Area (>)				
	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	1,0				
49	0,956785				
50	0,935022				

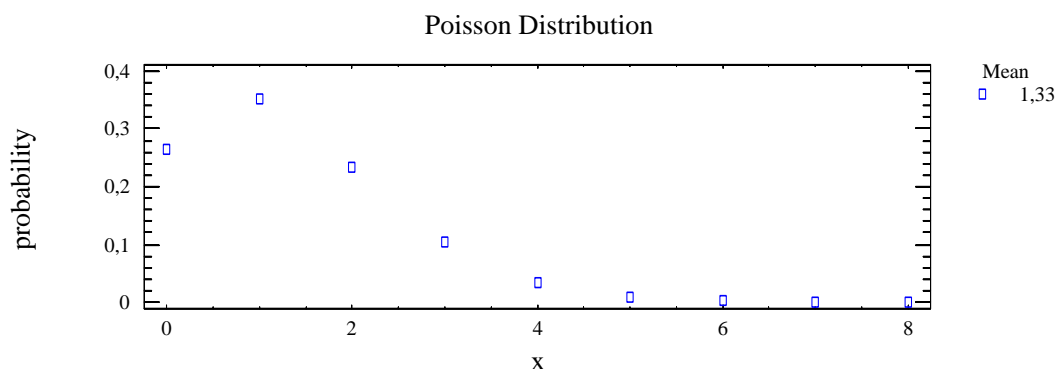
La probabilidad de que haya más de 50 mujeres y al menos 50 son respectivamente:

$P(X > 50) = 0.935022$  y  $P(X \geq 50) = P(X > 49) = 0.956785$  que también podía haberse obtenido como  $P(X \geq 50) = P(X > 50) + P(X = 50) = 0.935 + 0.022$ .

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  $P(Y=0) = 0.264$ , de tener al menos uno es  $P(Y \geq 1) = P(Y > 0) = 0.735$ , de tener más de cuatro hijos  $P(Y > 4) = 0.01$ , y de no ser familia numerosa  $P(Y < 3) = 0.85$ .

# Cumulative Distribution

Distribution: Poisson

Variable	Lower Tail Area (<)				
	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,0				
1	0,264477				
2	0,616232				
3	0,850149				
4	0,953852				

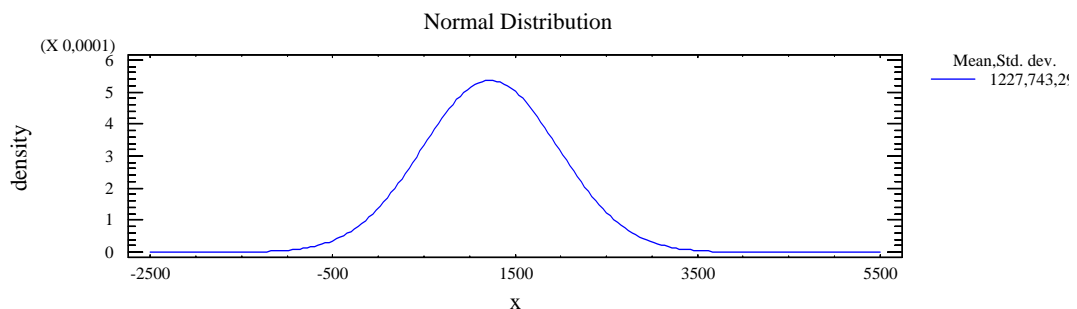
  

Variable	Probability Mass (=)				
	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,264477				
1	0,351755				
2	0,233917				
3	0,103703				
4	0,0344813				

Variable	Upper Tail Area (>)				
	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,735523				
1	0,383768				
2	0,149851				
3	0,0461479				
4	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:



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 0.15.

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.

#### Cumulative Distribution

-----

Distribution: Normal

Variable	Lower Tail Area (<)				
	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
1000	0,38003				
2000	0,850822				

Variable	Probability Density				
	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
1000	0,000512265				
2000	0,000312535				

Variable	Upper Tail Area (>)				
	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
1000	0,61997				
2000	0,149178				

#### Inverse CDF

-----

Distribution: Normal

CDF	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0,1	274,424				
0,25	725,652				
0,5	1227				
0,9	2179,58				
0,99	2956,17				

## 5. BLOQUE DE INFERENCIA

A lo largo de todo este bloque consideraremos un  $\alpha$  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

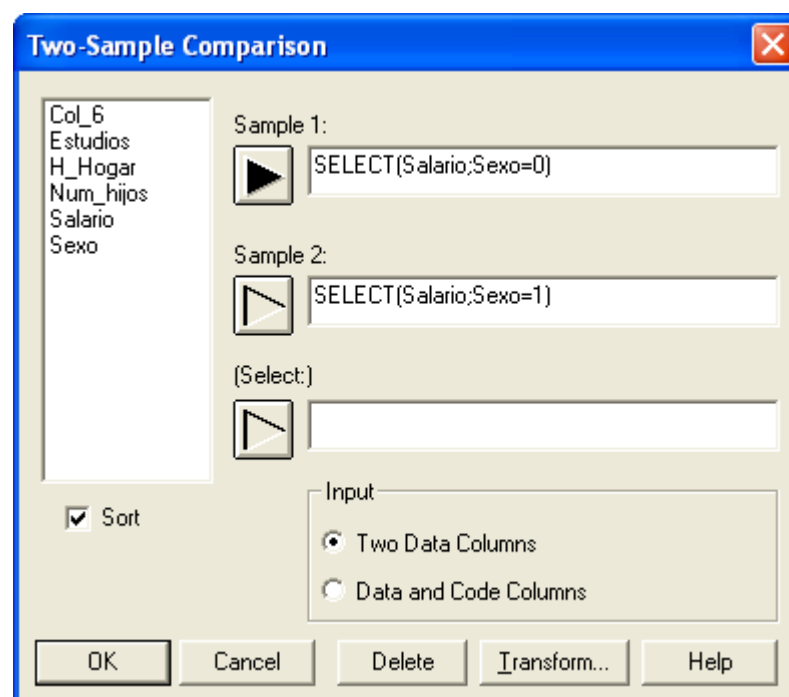
#### Chi-Square Test

Chi-Square	Df	P-Value
10,75	5	0,0566

En base al p-valor, 0.0566, superior al  $\alpha$ , 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. 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).

Como son muestras independientes, usaremos la opción *Two simple comparison* del Statgraphics, y hemos empleado Select para no tener que introducir de nuevo los datos en dos variables (Salario para hombres y Salario para mujeres):



En primer lugar, veremos si las varianzas poblacionales difieren:

```
95,0% Confidence Intervals
Standard deviation of SELECT(Salario;Sexo=0): [505,015;1008,93]
Standard deviation of SELECT(Salario;Sexo=1): [506,849;903,021]
Ratio of Variances: [0,450435;2,7517]
```

El intervalo de confianza al 95% para el cociente de varianzas ( $\sigma_1^2 / \sigma_2^2$ ) 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:

#### Comparison of Means

```
-----
95,0% confidence interval for mean of SELECT(Salario;Sexo=0): 1644,33 +/- 334,678 [1309,66,1979,01]
95,0% confidence interval for mean of SELECT(Salario;Sexo=1): 926,52 +/- 267,943 [658,577,1194,46]
95,0% confidence interval for the difference between the means
    assuming equal variances: 717,813 +/- 411,482 [306,332,1129,3]
```

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,  $\mu_1$ , va de 1309.7 a 1979, mientras que el de las mujeres va de 658.6 a 1194.5.

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 e (0.27, 1.63), como se ve en la imagen:

```
95,0% Confidence Intervals
    Standard deviation of SELECT(H_Hogar;Sexo=0): [6,95669;13,8983]
    Standard deviation of SELECT(H_Hogar;Sexo=1): [9,07218;16,1633]
    Ratio of Variances: [0,266787;1,62979]
```

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,  $\mu_1$ , va de 7.6 a 16.8, mientras que en las mujeres va de 23.3 a 32.9.

#### Comparison of Means

```
-----
95,0% confidence interval for mean of SELECT(H_Hogar;Sexo=0): 12,2222 +/- 4,61027 [7,61195,16,8325]
95,0% confidence interval for mean of SELECT(H_Hogar;Sexo=1): 28,08 +/- 4,79596 [23,284,32,876]
95,0% confidence interval for the difference between the means
    assuming equal variances: -15,8578 +/- 6,68471 [-22,5425,-9,17307]
```

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 última 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-Wallis Test for Salario by Sexo

Sexo	Sample Size	Average Rank
0	18	29,0278
1	25	16,94

```
-----
Test statistic = 9,71374    P-Value = 0,0018282
```



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  $\alpha=0.05$ ).

Kruskal-Wallis Test for H\_Hogar by Sexo

Sexo	Sample Size	Average Rank
0	18	12,8333
1	25	28,6

Test statistic = 16,5475    P-Value = 0,0000474478

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-Wallis Test for Salario by Estudios

Estudios	Sample Size	Average Rank
0	2	8,5
1	13	16,2308
2	8	23,375
3	3	32,8333
4	5	33,2
5	12	22,2083

Test statistic = 11,3846    P-Value = 0,0442662

Kruskal-Wallis Test for H\_Hogar by Estudios

Estudios	Sample Size	Average Rank
0	2	28,25
1	13	26,6923
2	8	20,9375
3	3	18,3333
4	5	8,8
5	12	23,0

Test statistic = 8,24938    P-Value = 0,143017

## 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*.
- 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/>)