

# Estudio del salario base

Claudia Negrín León y Katia Álvarez Muñoz

## INTRODUCCIÓN Y OBJETIVOS

Se pretende determinar si factores como los años de antigüedad, las horas de trabajo, las horas extra, los días de vacaciones, la edad y el sexo, influyen de alguna manera sobre el salario base anual de los empleados.

### Descripción de los datos

Número de observaciones muestrales: 100

Fuente: Instituto Nacional de Estadística

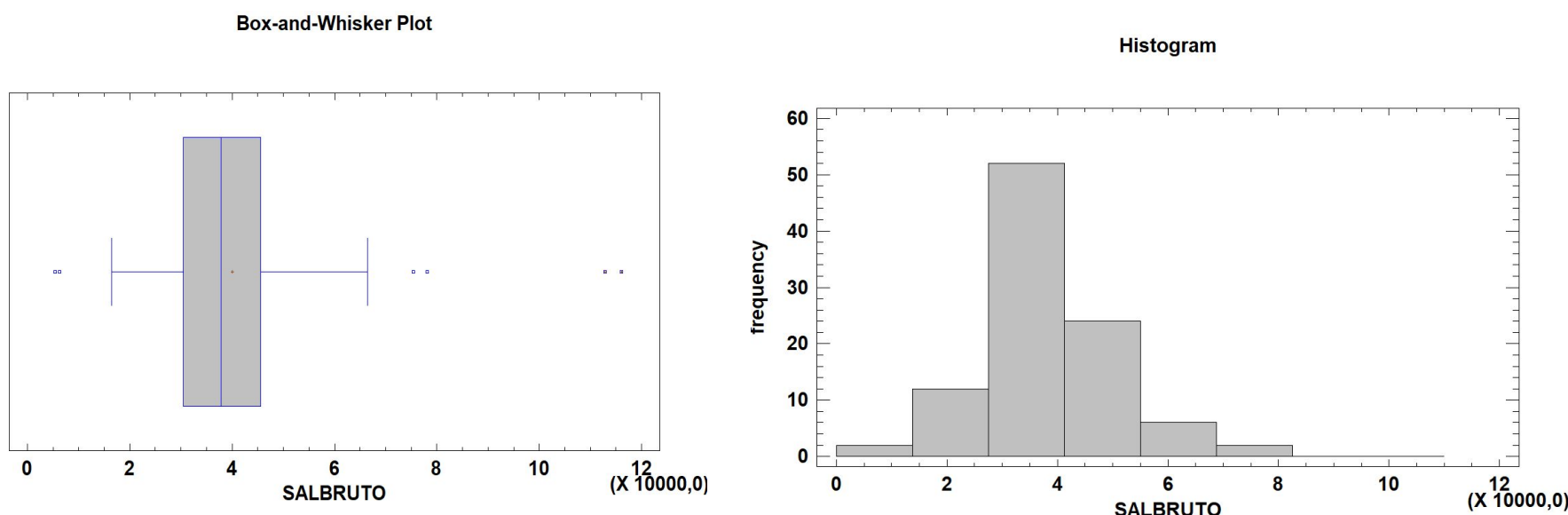
### Variables:

- Variable principal
  - SALBRUTO**: Salario bruto anual
- Variables secundarias
  - ANOANTI** : Años de antigüedad
  - VAN** : Días de vacaciones
  - JSP1** : Horas de jornada semanal
  - HEXTRA** : Horas extra
  - SEXO** : Sexo

## ESTUDIO DESCRIPTIVO

### Summary Statistics for SALBRUTO

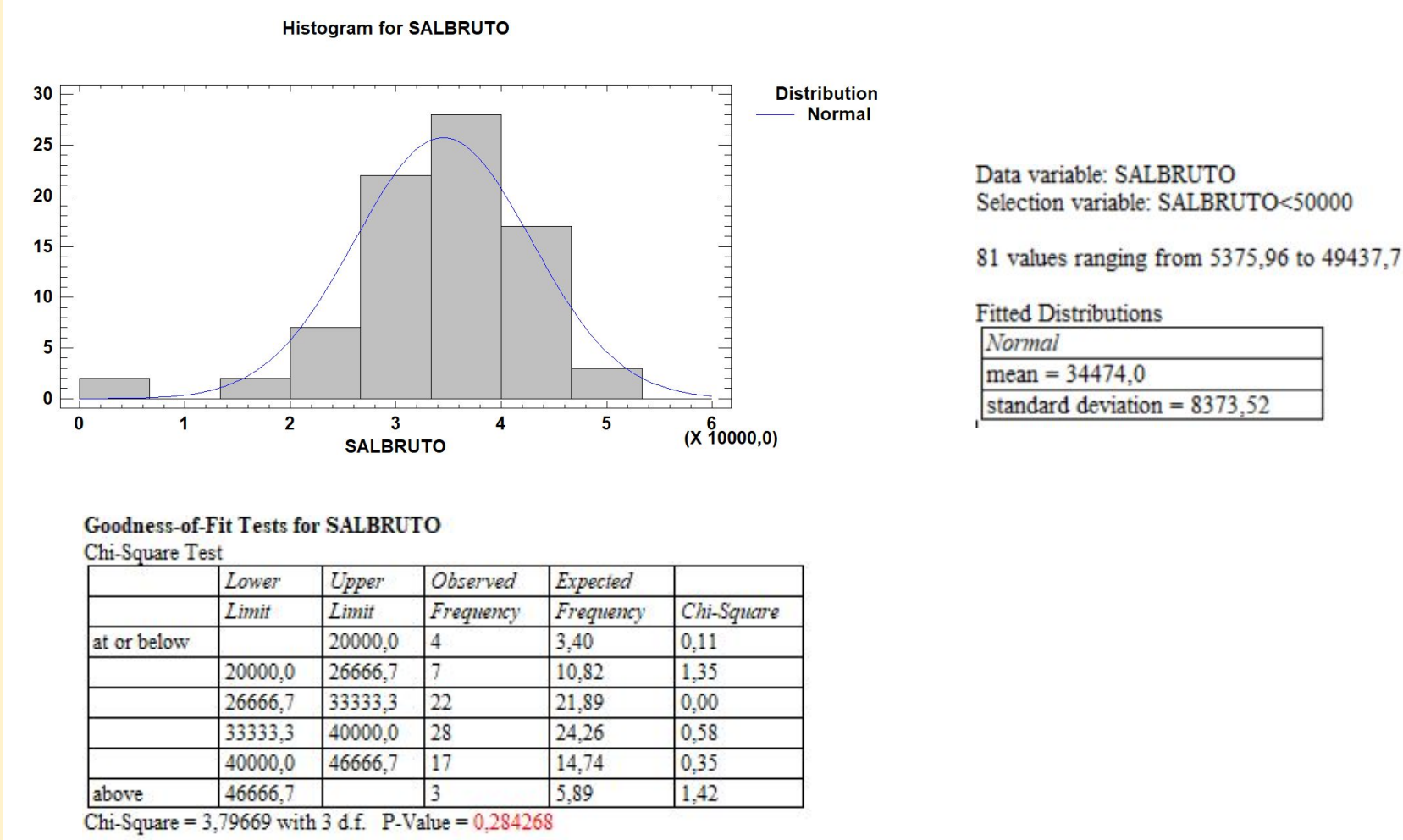
Count	100
Average	40091,4
Median	37842,4
Standard deviation	16189,9
Coeff. of variation	40,3824%
Minimum	5375,96
Maximum	116002,
Range	110626,
Lower quartile	30421,0
Upper quartile	45569,7
Interquartile range	15148,7



Haciendo el análisis descriptivo de la muestra obtenemos los valores más relevantes en la tabla.

También obtenemos el diagrama de caja, en el que se puede observar que hay algunos valores atípicos en la muestra; Y el histograma donde hemos elegido como número de clases 6, que nos parece el valor más adecuado para nuestra muestra de 100 valores.

## AJUSTE DE DISTRIBUCIONES



Al estudiar los datos de nuestra muestra, no se ajustaba a ninguna distribución conocida, probablemente esto se deba a que en nuestra muestra hay dos poblaciones. Por lo tanto dividimos la muestra en dos partes, para encontrar una muestra que se ajustara a una distribución normal. Utilizamos la variable SALBRUTO para hacer esta división, cogiendo los valores menores de 50000. Esta muestra sí se ajusta a una distribución normal (p valor > 0,05) y será por tanto, con la que continuemos el estudio.

## INTERVALOS DE CONFIANZA Y CONTRASTE DE HIPÓTESIS

### Confidence Intervals for SALBRUTO

95,0% confidence interval for mean: 34474,0 +/- 1851,54 [32622,5; 36325,6]  
95,0% confidence interval for standard deviation: [7252,98; 9906,79]

### Hypothesis Tests for SALBRUTO

Sample mean = 34474,0  
Sample median = 35840,9  
Sample standard deviation = 8373,52

### t-test

Null hypothesis: mean = 34000,0  
Alternative: not equal

Computed t statistic = 0,5095

P-Value = 0,611804

Do not reject the null hypothesis for alpha = 0,05.

### Hypothesis Tests for SALBRUTO

Sample mean = 34474,0  
Sample median = 35840,9  
Sample standard deviation = 8373,52

### t-test

Null hypothesis: mean = 30000,0  
Alternative: not equal

Computed t statistic = 4,80877

P-Value = 0,00000701646

Reject the null hypothesis for alpha = 0,05.

Intentamos hacer un contraste con el dato real de la media del salario bruto en España en el año 2014, pero la diferencia de los valores era muy grande. Creemos que esto se debe a que nuestra muestra es atípica y los valores de los salarios brutos son más altos en nuestra muestra, de lo que son habitualmente; Y por lo tanto no tendría mucho sentido hacer este contraste.

También hemos intentado hacer un contraste, haciendo algún tipo de división con las otras variables, pero no hemos encontrado ningún caso en el que los datos de cada parte de la división siguieran distribuciones normales, por lo tanto tampoco hemos podido utilizar este método.

Por ello hemos hecho dos contrastes guiándonos con el intervalo de confianza obtenido en nuestro estudio. El primer contraste con un valor cercano a la media de la muestra, que acepta la hipótesis nula planteada. Y en el segundo contraste un valor que no pertenece al intervalo de confianza y que rechaza la hipótesis nula planteada.

## REGRESIÓN MÚLTIPLE

### Regresión múltiple inicial

Dependent variable: SALBRUTO

Independent variables:

ANOANTI  
VAN  
JSP1  
HEXTRA

Selection variable: SALBRUTO<50000

Parameter	Estimate	Error	Statistic	P-Value
CONSTANT	-154,663	8281,41	-0,0186759	0,9851
ANOANTI	188,225	69,5483	2,70638	0,0084
VAN	-205,142	58,1975	-3,52494	0,0007
JSP1	864,881	206,613	4,18599	0,0001
HEXTRA	0,189172	58,9086	0,00321128	0,9974

R-squared = 34,2616 percent

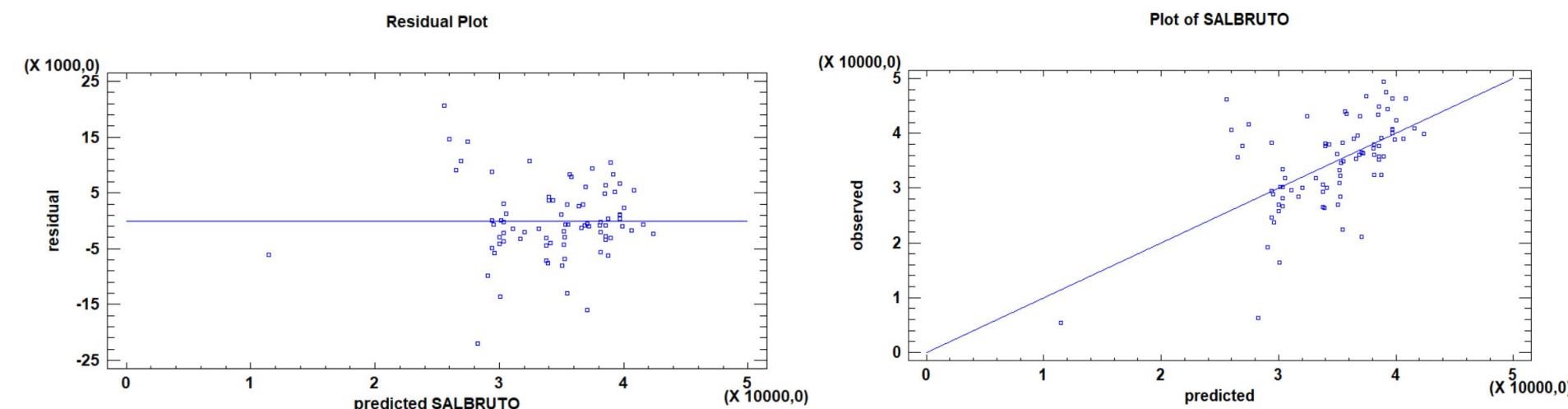
R-squared (adjusted for d.f.) = 30,8017 percent

Standard Error of Est. = 6965,56

Mean absolute error = 4908,78

Durbin-Watson statistic = 1,19824 (P=0,0000)

Lag 1 residual autocorrelation = 0,387818



En la regresión inicial obtuvimos que la única variable no significativa es HEXTRA (p valor > 0,05), además, los gráficos de residuos frente a valores predichos y frente a valores ajustados nos permiten concluir que hay linealidad y homocedasticidad, y por lo tanto podemos continuar mejorando con la regresión.

### Eliminación de variables no significativas y Modelo final

Dependent variable: SALBRUTO

Independent variables:

ANOANTI  
VAN  
JSP1  
SEXO=6

Selection variable: SALBRUTO<50000

Parameter	Estimate	Error	Statistic	P-Value
CONSTANT	-2149,86	7706,72	-0,27896	0,7810
ANOANTI	196,367	64,4131	3,04855	0,0032
VAN	-200,512	53,9108	-3,71932	0,0004
JSP1	880,457	191,74	4,59194	0,0000
SEXO=6	6463,63	1851,37	3,49126	0,0008

R-squared = 43,3475 percent

R-squared (adjusted for d.f.) = 40,3658 percent

Standard Error of Est. = 6466,3

Mean absolute error = 4567,1

Durbin-Watson statistic = 1,4506 (P=0,0029)

Lag 1 residual autocorrelation = 0,253963

SALBRUTO = -2149,86 + 196,367\*ANOANTI - 200,512\*VAN + 880,457\*JSP1 + 6463,63\*SEXO

Eliminamos HEXTRA y añadimos la variable binaria SEXO. El modelo ahora explica el 43,35% de los datos. La variable sexo nos permite concluir que en igualdad de condiciones, una mujer gana una media de 6463,63 € más que un hombre.

Por ejemplo, una mujer que tenga 7 años de antigüedad en el trabajo, 0 días de vacaciones adicionales y 35 horas de jornada, tendrá un salario bruto esperado de 36504,334 €

## CONCLUSIONES

Antes de concluir debemos tomar en cuenta la sospecha que tenemos de que la muestra que tomamos sea atípica y realmente no represente a la población, como podemos observar comparando la media muestral de 34.474,0 €, con la media poblacional de 22.806,9 € (dato publicado por el Instituto Nacional de Estadística).

Con respecto a los objetivos planteados inicialmente y basándonos en los resultados de la recta de regresión múltiple, concluimos que de las variables estudiadas la única que no influye sobre el salario bruto anual de los españoles es la cantidad de horas extra. Las horas de jornada laboral y los años de antigüedad influyen de manera positiva sobre este, mientras que los días de vacaciones tienen un efecto negativo.