

Olimpiada Estadística 2017

Miembros del Grupo:

Blanca Ancosta Jiménez

Héctor Chueca Oliver

David Sarnago Ojuel

Categoría:

Bachillerato

Tutor:

Íñigo Picaza Cortés

Navarra



IN
e



Equipo: HAW



Comenzamos llevando a cabo una detallada observación global de datos de la "Encuesta Cuatrienal de Estructura Salarial 2014" con el propósito de determinar cuales nos pueden resultar útiles y cuáles por el contrario, pueden ser desechados.

El paso siguiente es establecer una serie de relaciones entre las variables y realizar una observación visual y un estudio general que nos permitan descartar aquellas comparativas no relevantes o cuyos resultados no nos resultan significativos o atractivos.

Una vez aclaradas cuales son las variables a emplear y qué relaciones vamos a definir, debemos organizarlas en función de los cuatro grandes temas estipulados.

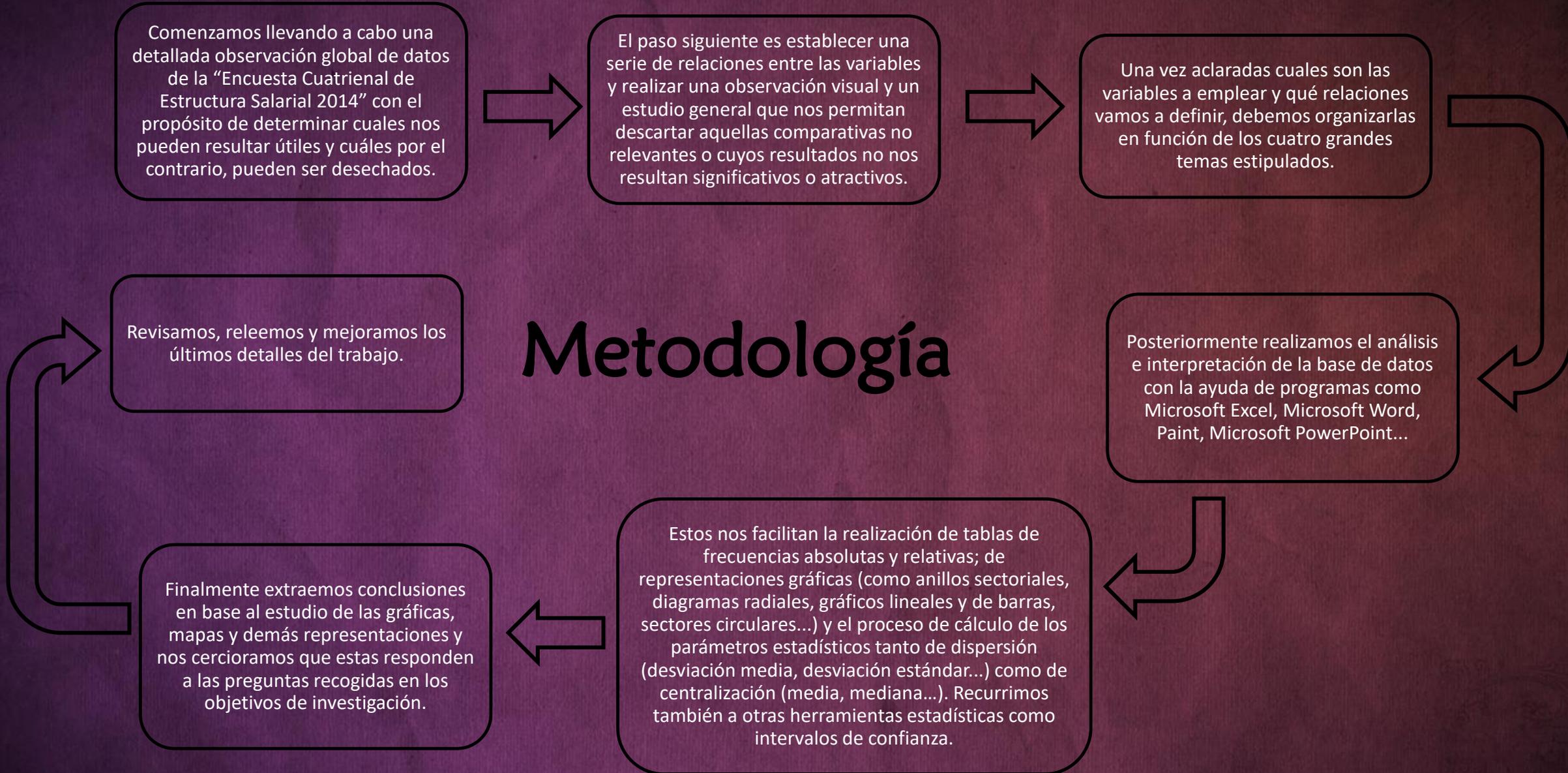
Posteriormente realizamos el análisis e interpretación de la base de datos con la ayuda de programas como Microsoft Excel, Microsoft Word, Paint, Microsoft PowerPoint...

Estos nos facilitan la realización de tablas de frecuencias absolutas y relativas; de representaciones gráficas (como anillos sectoriales, diagramas radiales, gráficos lineales y de barras, sectores circulares...) y el proceso de cálculo de los parámetros estadísticos tanto de dispersión (desviación media, desviación estándar...) como de centralización (media, mediana...). Recurrimos también a otras herramientas estadísticas como intervalos de confianza.

Finalmente extraemos conclusiones en base al estudio de las gráficas, mapas y demás representaciones y nos cercioramos que estas responden a las preguntas recogidas en los objetivos de investigación.

Revisamos, releemos y mejoramos los últimos detalles del trabajo.

Metodología



Objetivos

OBJETIVOS DE LA INVESTIGACIÓN:

El objetivo de esta disposición es ofrecer cada tema en su conjunto como un estudio lo más completo, fiel y riguroso posible.

El primer tema: engloba a las variables edad, estudios, salario total y código actividad económica (que hemos reorganizado en sectores) y tiene como principal fundamento:

-Relacionar los intervalos de edad con los distintos niveles de estudios y el salario total anual y por hora.

-Realizar una comparativa entre salario, código de actividad económica y el salario por hora.

-Comprobar si existe una relación entre el código de actividad económica, el nivel de estudios y el salario por hora.

El segundo tema: incluye las variables: región, código de actividad económica (que hemos vuelto a reorganizar), mercado y salario total (que incluye salario bruto, gratificaciones extraordinarias abonadas en el año y valoración en especie). Sus objetivos son:

-Comparativa entre región, código de actividad económica (tipo de trabajo) y tipo de mercado.

-Correlación entre región y salario por hora, para ello, recurrimos a parámetros de dispersión y de centralización.

-Relación entre el código de actividad económica (Sectores), salario medio por hora y mercado predominante.

El tercer tema: consta de las variables: sexo, ocupación económica, responsabilidad, salario bruto anual, gratificación extraordinaria abonadas en el año y valoración en especie y persigue:

-Comparativa entre sexo y salario (por intervalos).

-Correlación entre sexo, cargo ocupado y responsabilidad

-Relación entre salario, ocupación y responsabilidad.

El cuarto tema: incluye las variables antigüedad, edad y salario.

-Comparativa entre edad, intervalos, antigüedad y salario/h

OBJETIVOS DEL TRABAJO

-La finalidad fundamental de trabajo es responder a las preguntas planteadas en los objetivos de investigación.

-Asimismo se pretenden ofrecer los datos de una manera visual y sencilla a través de gráficos significativos y acordes con la información representada.

-Del mismo modo, es importante llevar a cabo una extracción de conclusiones relevantes que completen el mero proceso de cálculo y que dejen patente que se ha llevado a cabo una correcta y exhaustiva interpretación de los datos.

-Obviamente, se persigue también una mayor profundización en la estadística que nos permita ir aprendiendo y disfrutando de los entresijos de esta ciencia.

OBJETIVOS DE LA PRESENTACIÓN

-Configurar un proyecto estadístico que incluya las comparativas, estudios, representaciones y conclusiones obtenidas atendiendo a los criterios de concreción, atractivo y fácil visualización.

Base teórica

-En Estadística, se llama intervalo de confianza a un par de números entre los cuales se estima que estará cierto valor desconocido con una determinada probabilidad de acierto. Formalmente, estos números determinan un intervalo, que se calcula a partir de datos de una muestra, y el valor desconocido es un parámetro poblacional. La probabilidad de éxito en la estimación se representa con $1 - \alpha$ y se denomina nivel de confianza. En estas circunstancias, α es el llamado nivel de significación, esto es, una medida de las posibilidades de fallar en la estimación mediante tal intervalo.

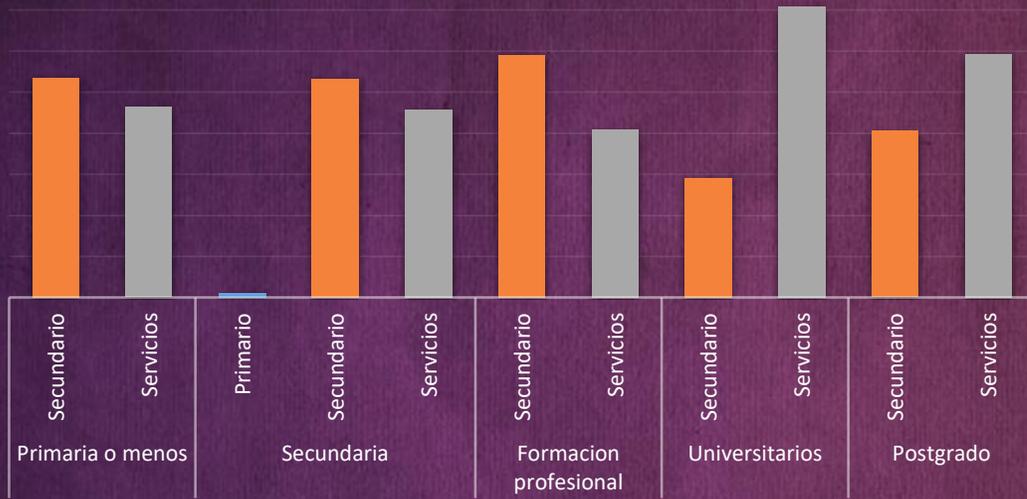
-El nivel de confianza y la amplitud del intervalo varían conjuntamente, de forma que un intervalo más amplio tendrá más probabilidad de acierto (mayor nivel de confianza), mientras que para un intervalo más pequeño, que ofrece una estimación más precisa, aumenta su probabilidad de error.

-Para la construcción de un determinado intervalo de confianza es necesario conocer la distribución teórica que sigue el parámetro a estimar. Es habitual que el parámetro presente una distribución normal.

-En este trabajo vamos a trabajar con un nivel de confianza del 95% para realizar comparaciones de salarios en dos casos, salarios por sexo y salarios por nivel de estudios. *(Usaremos únicamente los datos de salario de trabajadores a jornada completa para evitar desviaciones y añadiremos a los salarios las gratificaciones y las valoraciones en especie)*

1. Salario: Edad, sector y estudios

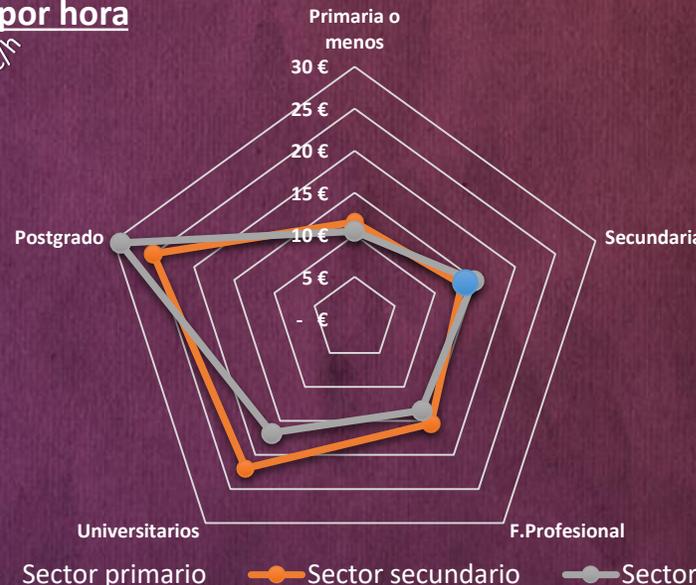
Porcentajes entre nivel de estudios y sector al que pertenecen



Relación entre edad, estudios y salarios

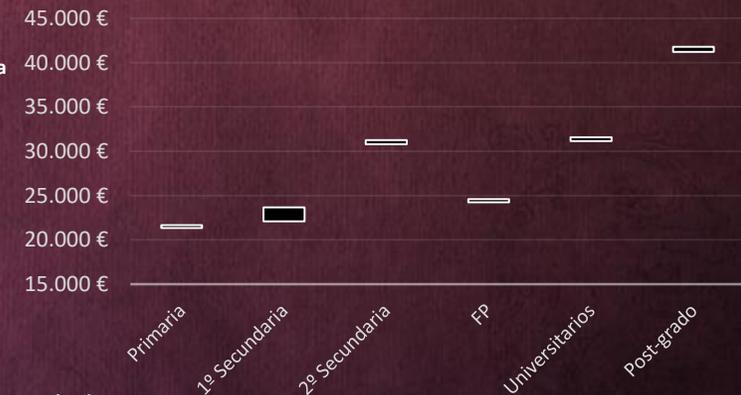
Edad (Años)	%	Estudios	%	Salario anual	Salarios hora
De 20 a 29	11,5	Obligatorio	69%	7.805,86 €	6,58 €/h
		F.Profesional	9%	9.291,44 €	5,29 €/h
		Universitarios	22%	17.641,83 €	9,93 €/h
De 30 a 39	33,5	Obligatorio	51%	20.642,59 €	12,28 €/h
		F.Profesional	18%	24.457,85 €	14,1 €/h
		Universitarios	31%	30.773,59 €	18,7 €/h
De 40 a 49	31	Obligatorio	57%	24.295,10 €	14,63 €/h
		F.Profesional	12%	28.678,10 €	17,35 €/h
		Universitarios	31%	43.275,71 €	26,2 €/h
De 50 a 59	16	Obligatorio	75%	25.043,54 €	15,34 €/h
		Universitarios	25%	43.397,26 €	26,64 €/h
Mayor de 60	8	Obligatorio	81%	23.656,14 €	17,5 €/h
		F.Profesional	6%	8.700,00 €	20 €/h
		Universitarios	13%	27.123,83 €	24,24 €/h

Comparación entre el nivel de estudios, sector al que pertenecen y salario por hora

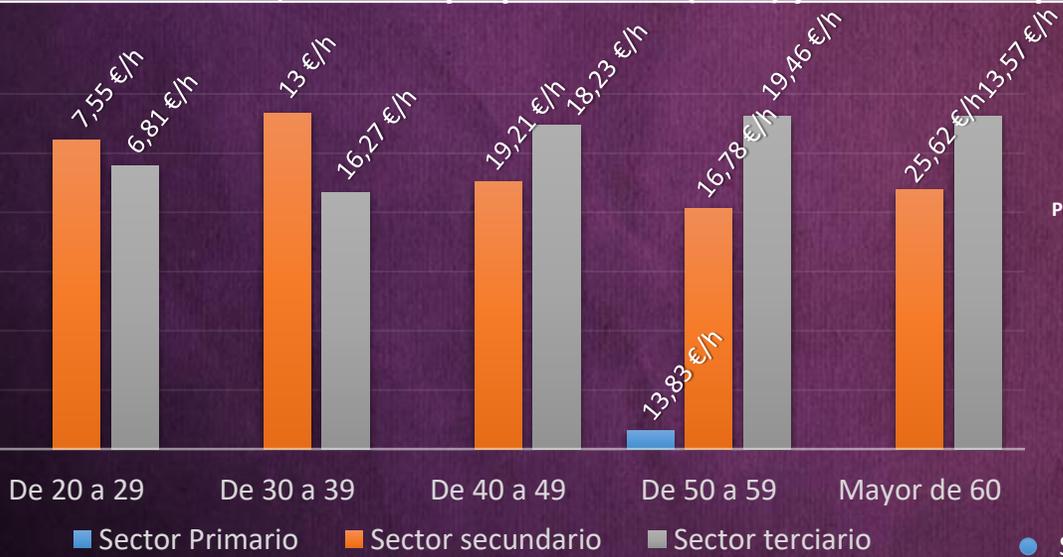


Estudios	Media	Desviación	Intervalo
Primaria	21514,45 €	9891,11	[21354,31 , 21674,6]
1º Secundaria	22222,81 €	12342,29	[22083,82 , 2361,81]
2º Secundaria	31019,63 €	23534,81	[30789,15 , 31250,11]
FP	24420,17 €	12138,42	[24254,07 , 24586,27]
Universitarios	31361,51 €	16570,53	[31149,41 , 31573,61]
Post-grado	41484,2 €	24025,74	[41194,26 , 41774,14]

Estudios, salario e intervalos de confianza

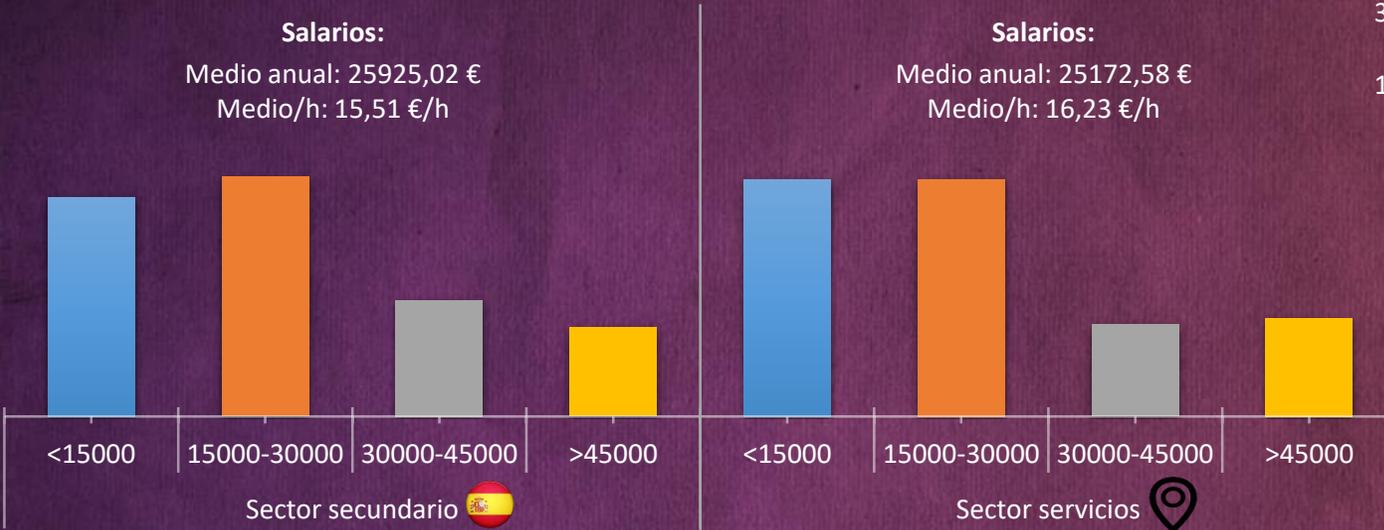


Relación entre edad, sector al que pertenecen (en %) y salario medio por hora



2. Región: Mercado, sector y salario

Relación entre el sector y el salario (en intervalos), Mercado predominante

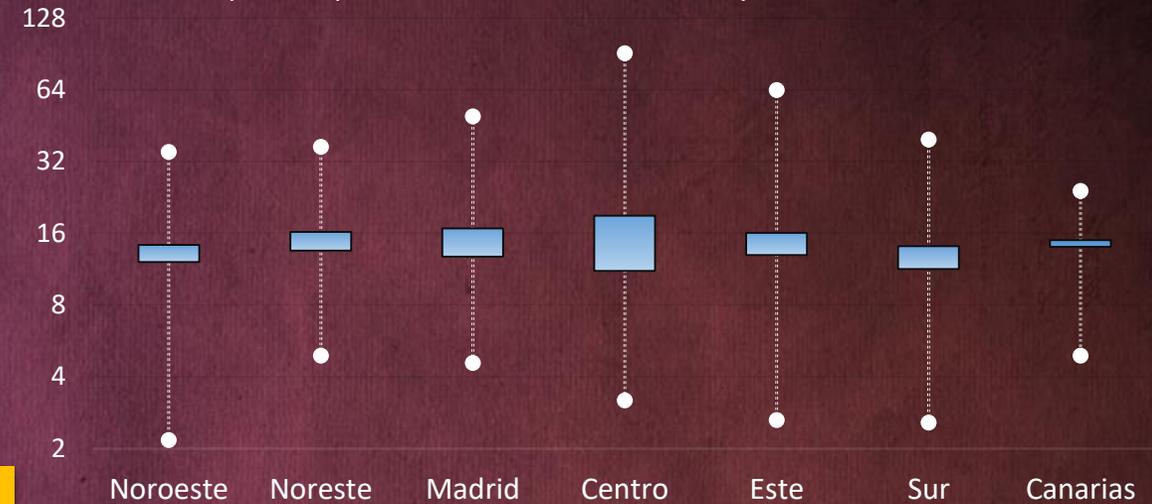


Tipo de trabajo en cada región



Región y salario por hora:

Los extremos de cada línea representan los mínimos y máximos absolutos, la parte superior del "cuadrado" la media y la inferior la mediana



Mercado predominante por región



3. Sexo: Salario, cargo y responsabilidad

Top 10 Salarios

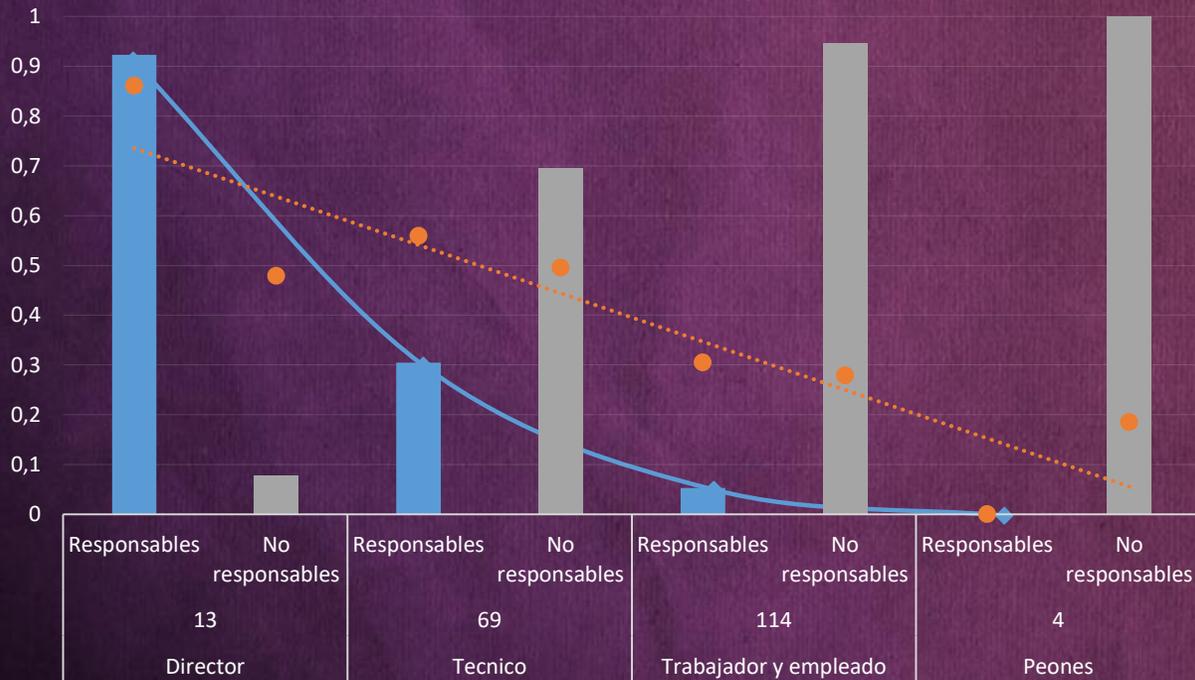
Sexo	Salario Hora
H	91,03 €/h
H	63,9 €/h
M	49,48 €/h
H	41,40 €/h
H	39,53 €/h
H	38,43 €/h
H	38,31 €/h
H	37,23 €/h
H	36,82 €/h
H	36,15 €/h

Distribución salarial entre sexos



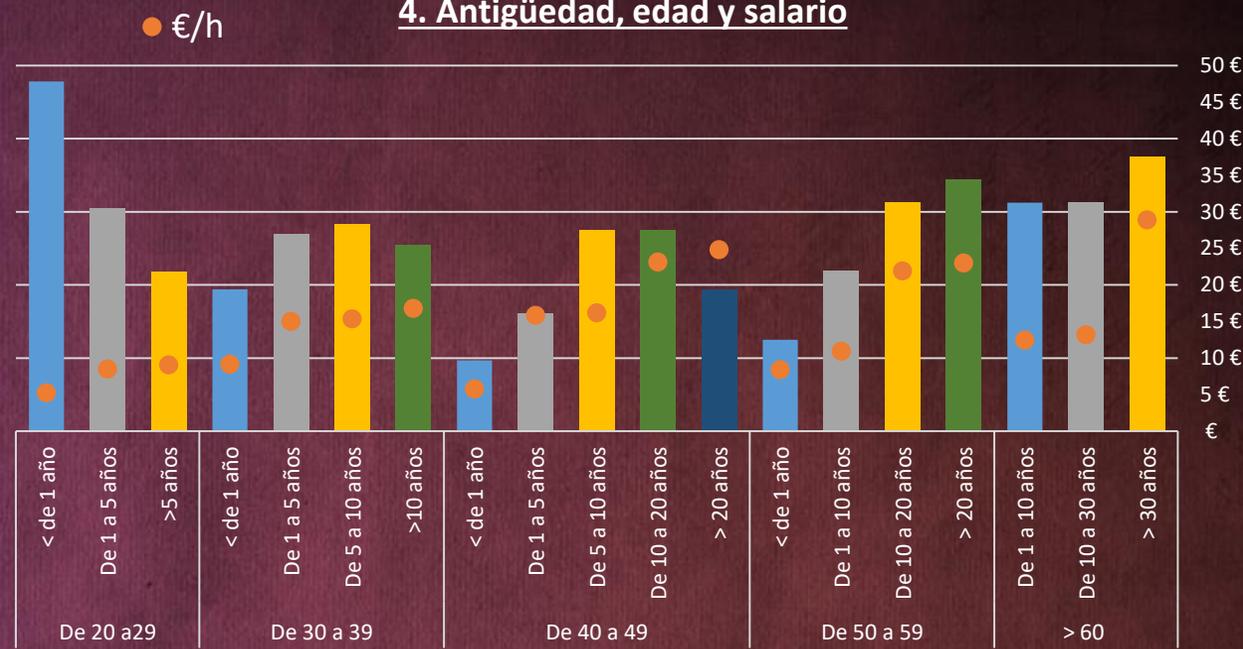
Cargo, salario y responsabilidad

$y = -3,8844x + 33,299$ $y = 0,1415x^2 - 1,0097x + 1,7829$
 $R^2 = 0,82$ $R^2 = 0,997$

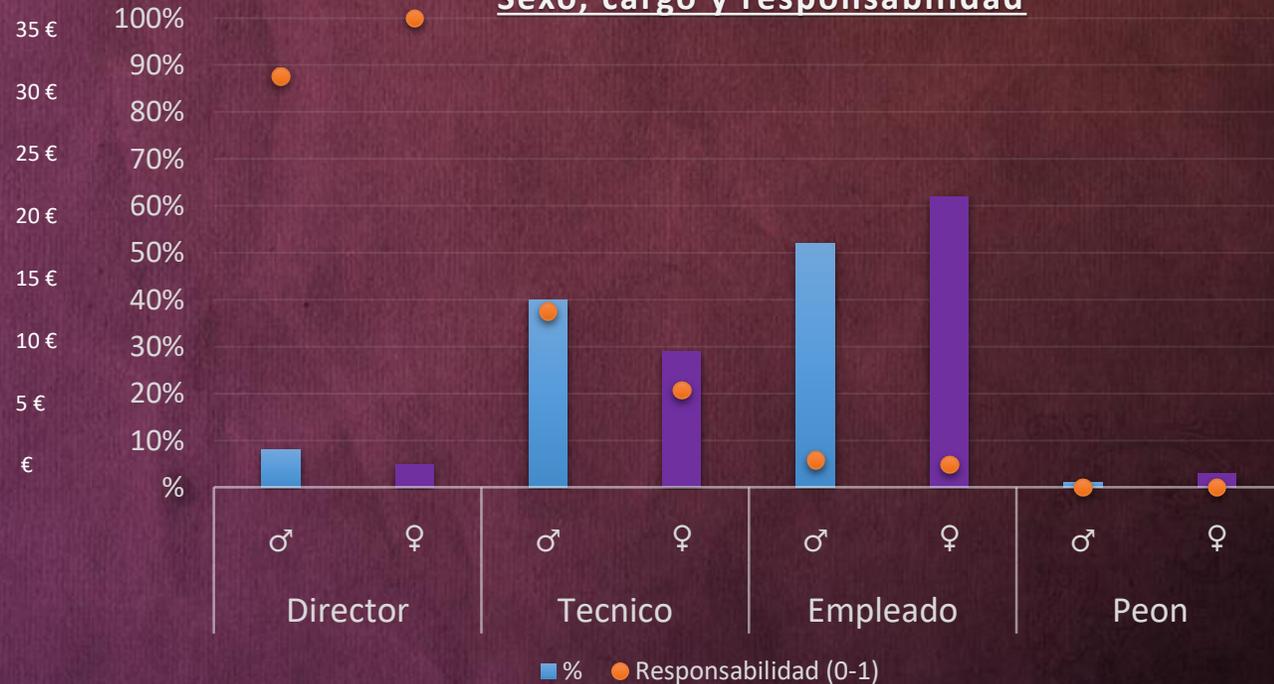


■ % ● €/h ⋯ Lineal (€/h)

4. Antigüedad, edad y salario



Sexo, cargo y responsabilidad



■ % ● Responsabilidad (0-1)

Conclusiones diapositiva 1 y 2

1. Salario: Edad, sector y estudios

-Con respecto al primer intervalo de edad (20-29) existe una relación de proporcionalidad directa entre el nivel de estudios y el salario anual. A medida que aumenta el nivel de estudios también aumenta el salario por hora resultando una excepción a esta regla aquellos que tienen estudios de FP puesto que cobran exactamente 1,29 € menos la hora que los que tienen simplemente estudios obligatorios. Esta conclusión no podemos afirmar que sea totalmente congruente porque la frecuencia absoluta de este grupo es muy pequeña (2 personas).

-En lo que se refiere al resto de intervalos, se mantiene la tendencia ascendente de salario anual y por hora, cuanto mayor es el nivel de estudios. Además conforme avanzamos en los intervalos de edad, aumentan ambos salarios.

-De nuevo, observamos una excepción, al comparar el intervalo 50-59 con el de mayores de 60. A pesar de mantener prácticamente igual el salario por hora, el anual se reduce enormemente, hecho que nos induce a pensar que los mayores de 60 trabajan menos horas.

-Asimismo, resulta destacable mencionar que la proporción de universitarios es mínima, debido a que la mayoría aún están estudiando, el porcentaje aumenta considerablemente en los dos intervalos siguientes (30-39 y 40-49) y cae en los dos últimos a favor de los estudios obligatorios.

-Del mismo modo, cabe destacar que el sector secundario supone el porcentaje mayoritario en los dos primeros intervalos de edad, mientras que, a partir de los 40 años, el sector terciario se convierte en el predominante.

-Por otra parte, podemos deducir que no existe relación directa entre sector y salario. Esta gráfica de barras nos permite reafirmar nuestra conclusión de que conforme avanzamos entre los intervalos de edad, mayor es el salario.

-Es importante citar que cuando hablamos de sectores, nos referimos al secundario y servicios, puesto que contamos con un único encuestado del sector primario, de un total de 200.

-Tal y como cabía esperar, la representación radial confirma que a mayor nivel de estudios, mayor salario.

-El sueldo se mantiene prácticamente igual entre sectores en aquellas personas con estudios de primaria, secundaria y formación profesional. En los universitarios existe un amplio rango salarial entre los dos sectores, siendo mayor en el secundario. Por el contrario, en los posgraduados, el salario es mayor si se trabaja en el sector terciario.

-Observando la última gráfica, podemos indicar que el sector mayoritario en aquellos que tienen estudios de primaria, secundaria y FP es el secundario, mientras que los que cuentan con mayor número de estudios trabajan fundamentalmente en el sector terciario.

2. Región: Mercado, sector y salario

-En todas las regiones el tipo de trabajo predominante es la industria de manufacturas, salvo en la región de Madrid donde predomina el comercio y en la zona de Canarias donde el mayor porcentaje corresponde a las actividades profesionales, científicas y técnicas.

-Asimismo, se puede apreciar que el mercado predominante es el local o regional en el Norte (Noroeste y Noreste) y en el Centro, mientras que en Madrid, el Este y el Sur, es el nacional.

-La región con un salario medio mayor es el Centro (18,93 €/h), pero al mismo tiempo, también es la de mayor desviación típica (19,79), debido a la presencia de un individuo con un salario de 91,03 €/h que aumenta la media de manera considerable. Si este fuese excluido, la media descendería hasta 15,13 €/h. Esto también es demostrable con la mediana de la región, que resulta ser la menor de toda la nación (11,11 €/h).

-Por otra parte, en todas las regiones, el salario medio se encuentra dentro del intervalo predominante, exceptuando en las regiones del Centro y del Sur. En el primer caso, el motivo ya ha quedado explicado anteriormente, mientras que en el Sur se debe a que la media del primer intervalo (menos de 15000) se encuentra próxima a este valor y por tanto, hace que el salario medio no se encuentre comprendido en ese intervalo.

-El salario más alto se da en el sector terciario, aunque no es demasiado significativa la diferencia con respecto al secundario. Además, el salario medio en los dos sectores, queda contenido en el intervalo predominante en cada uno de ellos (15-30000)

-En cada sector, merece especial alusión el mercado predominante. En el caso del secundario, es el nacional (40,20%), si bien es cierto, que el resto de mercados tienen también porcentajes elevados. Esto puede ser explicado con el hecho de que este sector comprende los diferentes tipos de industria y esta como podíamos presuponer exporta a distintos niveles.

-Por su parte, el sector terciario tiene al mercado local como predominante (51,48%). El mercado nacional también supone un porcentaje importante en los servicios, con un 40,59%. El resto de mercados son prácticamente despreciables. Esto se puede justificar con que las principales actividades del sector (turismo, hostelería...) abarcan solo regiones reducidas.

Conclusiones diapositiva 3

3. Sexo: Salario, cargo y responsabilidad

-El intervalo de salarios predominante en mujeres es el de menos de 15000€ (Siendo un 46,46%), mientras que el mayoritario en hombres es de 15 a 30000€ (con 35,64%). Por otra parte, el 30% de los encuestados cobran 30000€ o más, siendo de estos, casi un 73% hombres.

Es decir, que por cada mujer que cobra 30000€ o más, hay casi 3 hombres.

-Como información añadida, merece especial mención que, 9 de los 10 mayores salarios encuestados son de hombres. Además estos cobran por término medio 6,47€/h más que las mujeres.

-Todos los cálculos estadísticos muestran que los salarios de las mujeres son inferiores al de los hombres, siendo de especial importancia la enorme diferencia de casi 10000€ en la mediana entre ambos sexos.

-Como cabía esperar tras el análisis anterior, el porcentaje de directores es mayor en hombres que en mujeres. La proporción de empleados es semejante en ambos sexos. - Sorprendentemente, el porcentaje de peones es superior en mujeres que en hombres, mientras que en técnicos sucede lo contrario.

-En lo que respecta a responsabilidad, resulta remarcable mencionar que a medida que disminuye el cargo de ocupación, desciende la responsabilidad. Asimismo, encontramos significativo el hecho de que un director carezca de esta. Además, en lo referido a los sexos, se aprecia una notable diferencia en el porcentaje de responsabilidad en el cargo de técnico, siendo superior en hombres.

-Tal y como se observa, a mayor cargo, el salario es mayor. Además, dentro de un mismo cargo presentan un salario mayor aquellos que tienen responsabilidad en el mismo. Esto puede ser fácilmente observado con el coeficiente de correlación, que, al ser superior a 0,75 (en este caso, 0,82), nos indica que existe una relación de dependencia muy fuerte entre ambos. Esta correlación viene determinada por la recta: $y = -3,8844x + 33,299$

-De igual modo, el porcentaje de responsabilidad decrece cuanto menor es el cargo. En este caso, el coeficiente de determinación es prácticamente 1 (0,997), siendo su representación gráfica una parábola casi perfecta (determinada por la ecuación: $y = 0,1415x^2 - 1,0097x + 1,7829$).

4. Antigüedad, edad y salario

-Para entender esta relación, debemos tener en cuenta que no todos los intervalos de antigüedad son iguales, puesto que cuanto menor es la edad, menos años de antigüedad es posible acumular.

-Tampoco resulta difícil apreciar que conforme avanzamos en los intervalos de edad, el salario percibido es mayor (única excepción el intervalo 50-59). Además, a mayor antigüedad, también se da un salario superior.

Conclusiones Finales

-Como aspecto remarcable, hay que citar la escasez de datos con los que hemos contado para realizar el trabajo. A pesar de que a simple vista estos parecían inabarcables, una vez organizados y agrupados en intervalos, hemos apreciado que la muestra es bastante pequeña así que solo hemos podido establecer conclusiones muy concretas (fieles únicamente a estos datos) que posiblemente no pueden ser aplicables en niveles mucho más generales como poblaciones.

-A esta circunstancia, hay que sumar además la irregularidad de los datos, la gran variedad de frecuencias absolutas en cada categoría y la gran cantidad de datos irrelevantes que hemos tenido que desechar por no aportar información destacable.

-En definitiva, a pesar de haber extraído numerosas conclusiones interesantes debemos tener siempre presente que hacen referencia a un conjunto de datos determinado y reducido.

-Con el objetivo de revisar el trabajo y a modo de comparación con las gráficas realizadas, hemos recurrido a los intervalos de confianza que nos han permitido corroborar las conclusiones obtenidas.

-Estos, apreciables en la diapositiva 4, han sido una herramienta útil y alternativa a las habituales y tan recurridas representaciones gráficas, que nos han permitido ampliar nuestros conocimientos de estadística y trabajar con los datos desde otra perspectiva.

-En conclusión, estamos satisfechos con las relaciones que hemos sido capaces de establecer puesto que nos han permitido obtener resultados interesantes y nos han dotado de nuevos conocimientos sobre la estadística.