

6

*

Este fichero excel es de elaboración propia, y se construye en base a los datos entregados por la "Encuesta de estructura salarial año 2006".

Las observaciones tomadas corresponden a las personas cuyo nivel de educación es "licenciados, ingenieros superiores y doctores".

Estos datos se utilizan en los cursos de econometría para analizar cuáles son los principales factores que determinan las diferencias salariales de los trabajadores en España. En la siguiente web esta la información de esta encuesta <https://www.ine.es/dyngs/INEbase/es/op>

En nuestro fichero las variables utilizadas son: Salario, logaritmo del salario, antigüedad, antigüedad al cuadrado, sexo, edad, número de horas trabajadas en el mes de octubre, tipo de jornada laboral, tipo de duración del contrato y responsabilidad.

Logaritmo del salario, y antigüedad al cuadrado son variables que se utilizan siguiendo la teoría de las ecuaciones de Mincer para analizar la estructura de los salarios.;

libname pm"F:\tarea sas\pm";

```
proc import out=pm.encuestaine
datafile ="F:\tarea sas\varpm.xlsx"
dbms=xlsx replace;
getnames=yes;
run;
```

ok 1

```
proc print data=pm.encuestaine (obs=10);
run;
```

```
proc format library = pm;
value sexo 1="hombre" 6="mujer";
value edad 1="menor de 19 años" 2="entre 20 a 29 años" 3="entre 30 y 39 años" 4="entre 40 y 49 años";
value nac 1="españa" 2="no españa";
value control 1="público" 2="privado";
value tipoj 1="completa" 6="parcial";
value tipoc 1="indefinido" 2="determinado";
value respon 1="si" 6="no";
run;
```

ok 1

```
options fmtsearch = (pm);
```

```
data pm.encuestaine2;
set pm.encuestaine;
label nht="horas trabajadas" sal = "salario" lsal = "logaritmo del salario"
      anti= "antigüedad en la empresa" anti2="antigüedad al cuadrado"
      nac= "nacionalidad" control="empresa estatal o privada" tipoj="tipo jornada"
      tipoc="tipo contrato" respon="responsabilidad";
format sexo sexo. ;
format edad edad.;
format nac nac. ;
format control control.;
format tipoj tipoj. ;
format tipoc tipoc.;
format respon respon. ;
run;
```

```
ods rtf file='F:\tarea sas\tareacom.rtf';
proc contents data=pm.encuestaine2;
```

```
run;
3) y 4)
proc means data=pm.encuestaine2 min mean q3 max;
run;
```

① punto por los dos apartados, es mejor hacer proc univariate para variables cuantitativas para analizar valores extremos

*Todo ha salido bien;

5)

```
proc univariate data=pm.encuestaine2;
var sal anti nht;
histogram;
run;
```

ok ①

```
proc tabulate data=pm.encuestaine2 ;
class edad sexo;
var sal nht;
tables edad*sexo,(n sal*mean nht*mean);
run;
```

6) ① No lo hace

7)

```
proc iml;
reset print;
proc corr data=pm.encuestaine2;
var sal anti nht;
run;
```

① No lo hace con proc iml propiamente, que era lo que se pedía.

```
quit;
*aca veo que la diferencia entre hombre y mujeres y españoles o no españoles,
no es significativa, explicado en informe;
```

8)

```
PROC GCHART DATA = pm.encuestaine2;
VBAR nac /DISCRETE TYPE = mean space = 4 mean
sumvar = lsal GROUP = sexo;
RUN;
```

0'5 Solo hace proc gchart, representando media y también se pedía un proc estadístico, no solo gráfico.

```
PROC GCHART DATA = pm.encuestaine2;
VBAR nac /DISCRETE TYPE = mean space = 4 mean
sumvar = nht GROUP = sexo;
RUN;
```

9)

```
* aca planteo el modelo de regresion explicar el lsal. El modelo es muy sencillo,
explicado en informe;
```

0'5 Las variables cualitativas están mal introducidas.

```
proc reg data=pm.encuestaine2;
model lsal=sexo edad anti anti2 nht nac control tipoj tipoc respon;
run;
```

10) *macro sencilla, hace descriptivos e histogramas de 3 variables numericas;

```
%macro uni(bbdd,vnum1,vnum2,vnum3);
proc univariate data=&bbdd ;
var &vnum1 &vnum2 &vnum3;
histogram;
run;
%mend;
```

ok ①

```
%uni(pm.encuestaine2,sal,anti,nht);
```

```
ods rtf close;
```