

Práctica 4

Estadística bidimensional.

- Fichero de datos: **datospractica2.txt**

1. Calcular una nueva variable donde se tengan los valores

“Pequeña” para las empresas con menos de 35 trabajadores

“Mediana” para las empresas con un número de trabajadores entre 35 y 75

“Grande” para las empresas con más de 75 trabajadores

Datos - Recodificar variable

```
Datos$tipoempresa <- recode(Datos$n_trbs,
  '0:34 = "Pequeña"; 35:74 = "Mediana"; 75:122 = "Grande"; ', as.factor.result=TRUE)
```

2. Obtener una tabla de doble entrada con los datos del tipo de empresa junto con la compañía telefónica que incluya el porcentaje sobre el total. Hacer otras dos tablas que incluyan los porcentajes de filas y los porcentajes de columnas

Estadísticos - Tabla de contingencia - Tabla de doble entrada

```
> .Table
      tipoempresa
compa Grande Mediana Pequeña
Movistar 17 15 2
Orange 15 11 2
Vodafone 17 15 6

> rowPercents(.Table) # Row Percentages
      tipoempresa
compa Grande Mediana Pequeña Total Count
Movistar 50.0 44.1 5.9 100 34
Orange 53.6 39.3 7.1 100 28
Vodafone 44.7 39.5 15.8 100 38

> totPercents(.Table) # Percentage of
Total
      Grande Mediana Pequeña Total
Movistar 17 15 2 34
Orange 15 11 2 28
Vodafone 17 15 6 38
Total 49 41 10 100

> colPercents(.Table) # Column Percentages
      tipoempresa
compa Grande Mediana Pequeña
Movistar 34.7 36.6 20
Orange 30.6 26.8 20
Vodafone 34.7 36.6 60
Total 100.0 100.0 100
Count 49.0 41.0 10
```

a) ¿Qué porcentaje de empresas operan con MoviStar y son medianas?

b) ¿Qué porcentaje de las empresas que operan con MoviStar son medianas?

c) ¿Qué porcentaje de las empresas medianas operan con MoviStar?

3. Hacer una tabla donde se presente la facturación anual media para cada tipo de empresa

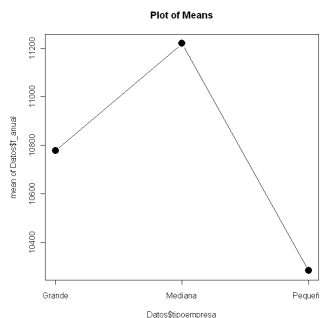
Estadísticos - Resumen - Tabla de estadísticas

```
      tipoempresa
      Grande Mediana Pequeña
10777.17 11219.09 10283.59
```

4. Hacer un gráfico de la facturación anual media en cada tipo de empresa

Gráficas - Gráfica de las medias

```
plotMeans(Datos$f_anual, Datos$tipoempresa, error.bars="none")
```



5. Estudiar de forma conjunta las variables f_marzo, f_abril, f_mayo, f_junio, f_julio, f_agosto y f_septiembre, calculando la matriz de correlaciones y la matriz de varianza-covarianzas.
Estadísticos – Resúmenes – Matriz de correlaciones

```
cor(Datos[,c("f_marzo", "f_abril", "f_mayo", "f_junio", "f_julio", "f_agosto",
"f_septiembre")], use="complete.obs")
```

f_marzo	f_abril	f_mayo	f_junio	f_julio	f_agosto	f_septiembre
f_marzo	1.00000000	-0.7449234	0.19068425	0.3224841	0.05425256	0.06754282
f_abril	-0.74492337	1.00000000	-0.11853354	-0.1562924	0.39232697	0.35984007
f_mayo	0.19068425	-0.1185335	1.00000000	0.8023663	0.09682221	0.10160490
f_junio	0.32248412	-0.1562924	0.80236635	1.00000000	0.23098348	0.23786045
f_julio	0.05425256	0.3923270	0.09682221	0.2309835	1.00000000	0.99818777
f_agosto	0.06754282	0.3598401	0.10160490	0.2378605	0.99818777	1.00000000
f_septiembre	0.06035944	0.3776355	0.09905465	0.2342348	0.99962069	0.99946643

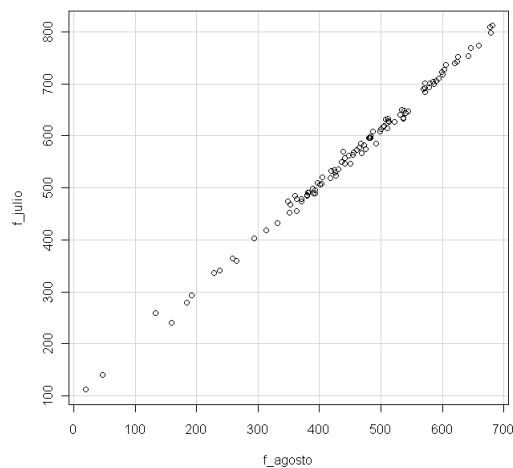
```
cov(Datos[,c("f_marzo", "f_abril", "f_mayo", "f_junio", "f_julio", "f_agosto",
"f_septiembre")], use="complete.obs")
```

f_marzo	f_abril	f_mayo	f_junio	f_julio	f_agosto	f_septiembre
f_marzo	30290.285	-20170.060	18672.822	68825.89	1302.696	1553.944
f_abril	-20170.060	24204.037	-10375.968	-29817.67	8420.993	7400.448
f_mayo	18672.822	-10375.968	316582.485	553615.18	7516.051	7557.230
f_junio	68825.894	-29817.666	553615.181	1503777.86	39079.053	38558.368
f_julio	1302.696	8420.993	7516.051	39079.05	19034.546	18204.890
f_agosto	1553.944	7400.448	7557.230	38558.37	18204.890	17474.676
f_septiembre	2670.167	14933.387	14166.413	73010.42	35054.849	33582.605

6. Hacer un diagrama de dispersión donde se observen las dos variables con mayor correlación lineal de las estudiadas.

Gráficas – Diagrama de dispersión

```
> scatterplot(f_julio~f_agosto, reg.line=FALSE, smooth=FALSE, spread=FALSE,
boxplots=FALSE,
+ span=0.5, data=Datos)
```



Ejercicios propuestos

Utilizando el archivo de datos de la encuesta a los alumnos:

1. Hacer una tabla cruzando las variables “red social preferida” y “sexo”. Estudiar qué porcentaje de estudiantes prefieren cada una de las redes sociales entre los dos sexos.
2. Obtener el peso medio para hombres y para mujeres y representarlo gráficamente
3. Hacer una matriz de correlaciones para el peso y la altura de los estudiantes y representar la nube de puntos asociada.