

Curso de Estadística para Data Scientists con R

Trabajo final – Fecha de entrega 16 de octubre de 2017

Se pretende realizar un análisis descriptivo de la actividad que siguen las ventas de dispositivos móviles y electrónicos de una franquicia en España (toda la información está desglosada por provincias en el archivo Excel correspondiente).

Para el desarrollo de este estudio se sugiere seguir las siguientes instrucciones:

1. Para tener una primera visión global de la estructura que presentan las diferentes variables de las que se dispone, determinar los estadísticos que miden la posición, la dispersión y la forma, así como todos los que se crean convenientes para realizar un análisis descriptivo.
2. Para obtener una idea de las diferentes relaciones que existen entre las variables, construir una matriz de correlaciones entre todas ellas. Confirmar que, efectivamente, existe una fuerte relación positiva entre las variables “Ventas” y “Número de Empleados” (las tiendas más grandes en cantidad de empleados son las que más negocio tienen).
3. Como consecuencia de la conclusión anterior, queremos hacer foco en las variables “Ventas” y “Número de Empleados” y realizar los correspondientes histogramas y estudios de la dispersión.
4. Para ello, será necesario crear intervalos: a modo de ejemplo, para el caso de las ventas se podrían generar rangos homogéneos a partir de la fórmula: $\sqrt[n]{\text{Recorrido}}$ Cantidad de datos, y para el número de empleados se podría tomar una categorización convencional; por ejemplo, la utilizada a nivel Europeo, la cual divide a las empresas en Microempresas (1-9 trabajadores), Pequeña empresa (10-49 trabajadores), Mediana empresa (50-249 trabajadores) y Gran empresa (250 y más trabajadores).
5. Considerando los tramos del apartado anterior, realizar una tabla de contingencia entre ambas variables para observar de forma más precisa la relación entre ambas.
6. Realizar las correspondientes regresiones y análisis multivariantes y explicarlas detalladamente.
7. Con los datos de Madrid y Barcelona, se pide:
 - a. Desarrollar un análisis descriptivo siguiendo los pasos anteriormente descritos.
 - b. Analizar comparativamente Madrid y Barcelona, y también cada una de ellas

respecto al total.

- c. Decidir si existe alguna evidencia para poder afirmar que las ventas se comportan de forma diferente entre las dos provincias.
- d. Considerando las variables “Ventas” y “Productividad”, plantear una regresión lineal y describirla adecuadamente.

Se valorará positivamente el análisis estadístico para el resto de provincias

Obligatorio:

Presentar todo el código de R utilizado en los análisis