Práctica 6

Partimos de la base de datos “bankoko” [www.uv.es/mlejarza/azar/bankoko.RData](http://www.uv.es/mlejarza/azar/bankoko.RData).

que consta de 29543 observaciones de los supuestos clientes de un banco sobre estas 13 variables cuyas distribuciones que podemos suponer POBLACIONALES resumimos en este cuadro (summary)



Sin embargo, cada alumno trabajará, en cada ejercicio, con muestras aleatorias (MAS) de estos datos que cada alumno generará fijando su correspondiente semilla aleatoria según su número ya asignado. En cada caso se llevarán a cabo distintos contrastes sobre distintas hipótesis .

En los contrastes sobre medias y diferencias de medias se aconseja emplear los procedimientos del paquete {estadística}. En el caso de contrastes de diferencias con la opción introducir. (También puede usarse el procedimiento t.test de R)

En los contraste de proporciones se aconseja usar la función correspondiente del paquete {estadística} en los de diferencias de proporciones puede usarse la función de este paquete pero quizá sea preferible usar la función p.test2 (<http://www.uv.es/mlejarza/azar/zp.R> ). No usar en ningún caso la función prop.test del paquete {stat} que se basa en un estadístico diferente.

En los contrastes sobre varianza puede usarse la función del paquete {estadística} o no, pero no se recomienda la del paquete {stat}

* **Una vez realizados los contrastes se comentarán los resultados y se extraerán las conclusiones oportunas en todos los casos. Tras la realización de ya cinco prácticas, debe exigirse algo más que un código para hacer los cálculos.**
* **Si siempre se va   a exigir que se comente y se extraigan conclusiones en todos los análisis, cuando nos encontremos ante contrastes que “comparan” esto se hace imprescindible.**

**Cuestiones/objetivos/claves/preguntas de investigación a resolver**

1.- A partir de una muestra aleatoria simple de los ingresos de 250 clientes ( debe fijarse la semilla correspondiente) ¿Puede admitirse que los ingresos medios de los clientes son de 24000 u.m a un nivel de significación del 5%? ¿ y a uno del 1%? Caso de no ser admisible ¿Puede asegurarse que los ingresos medios son significativamente mayores/menores? ¿ a qué nivel de significación? ( sólo debe realizarse el caso que **aparentemente** pudiera ser )

2.- A partir de una MAS de los gastos de 100 clientes proceder de la misma forma que en 1 pero con la media de gastos y el valor 20000 u.m.

3.- Estudiar si hombres y mujeres tienen o no, por término medio, el mismo saldo-medio-ponderado, a partir de 2 MAS independientes de 225 hombres y 250 mujeres ¿Hay diferencias significativas?¿en qué sentido? ¿a qué nivel? Comentar adecuadamente los resultados

4.- Contestar ,según un esquema similar al de tres, y dos muestras de los mismos tamaños , (eventualmente pueden ser las mismas que antes dependiendo de cómo se hayan diseñado) a la cuestión de si los clientes femeninos tienen más niños que los masculinos.

5.- ¿Es similar la proporción de impagos entre hombres y mujeres? . ¿Hay diferencias significativas?¿en qué sentido? ¿a qué nivel? Comentar adecuadamente los resultados. Usar dos muestras de los mismo tamaños que en 3,4 ( eventualmente las mismas muestras)

6.- Se sospecha que los clientes con tarjeta de crédito tienen gastos superiores que los clientes que no la tienen. Estudiar esta cuestión con metodología de CH aplicada a datos muestrales de dos muestras de 500 clientes con tarjeta y otros 100 sin tarjeta.

7.- Utilizando las informaciones muestrales de los ejercicios 1 y 2 ¿Es la variabilidad entre ingresos y gastos similar en nuestros clientes o hay diferencias significativas en algún sentido?

**Para los siguientes ejercicios tomamos una muestra aleatoria de tamaño 100 ( fijar semilla) considerando todas las características ( variables) . En cada ejercicio se usaran los datos muestrales de las variables pertinentes**.

8.- Utilizando esa muestra como una única muestra ( emparajeda) de 100 clientes ,analizar si existe correlación entre ingresos y gastos (sentido, significación, en su caso) Comentar resultados.

9.- A partir de esa muestra ( n=100) puede admitirse que ingresos, gastos y saldos siguen una distribución normal (usar como estimaciones de las medias y las desviaciones típicas, las muestrales)

10.- ¿Son edad y región de residencia factores independientes? ¿Y género y disponer de coche? ¿y tener una hipoteca y tener tarjeta de crédito?

11.- ¿Influye el factor edad en los ingresos? En caso afirmativo ¿cómo? Los gastos son diferentes entre los clientes según sus lugares de residencia ¿cómo? (ANOVA)