

GRADO EN ESTADÍSTICA Y EMPRESA
TÉCNICAS DE INFERENCIA ESTADÍSTICA II
Curso 2015/2016

PRUEBA FINAL

1. En el archivo `Caballos.csv`, se recoge en las variables “time” y “location” el tiempo de 8122 carreras de caballos de 1609.344 metros y el hipódromo de cada carrera.
 - (a) (1 pto.) Contrastar la hipótesis de que el tiempo medio de una carrera en Dundalk es inferior al tiempo medio en Southwell.
 - (b) (0.5 ptos.) Indicar si es necesario asumir normalidad en las distribuciones de los tiempos en el contraste anterior. En caso afirmativo, realizar los contrastes de normalidad correspondientes.
 - (c) (0.5 ptos.) Examinar la hipótesis de que existen diferencias entre la distribución de los tiempos de este tipo de carreras en Dundalk y Southwell.
 - (d) (0.5 ptos.) Examinar si existe evidencia estadística de que la distribución del tiempo de una carrera depende del hipódromo.
2. En el archivo `Pan.csv`, se recoge en la variable “NumeroBarrasPanMiercoles” el número de barras compradas por 100 clientes en una panadería a lo largo de diferentes miércoles escogidos al azar y en la variable “NumeroBarrasPanDomingo” el número de barras compradas por otros 100 clientes diferentes en la misma panadería a lo largo de distintos domingos escogidos al azar.
 - (a) (1 pto.) Contrastar la hipótesis de que la distribución del número de barras que se venden los miércoles en esa panadería es distinta de la del número de barras que se venden los domingos.
 - (b) (0.5 pto.) Examinar si existe evidencia de que el número de barras vendidas depende del día de la semana.
3. Dos analistas se disponen a estudiar si el salario medio de los hombres es superior al de las mujeres de una determinada empresa. Para ello se les proporciona a ambos una misma muestra de salarios de 200 trabajadores de la empresa, 100 de cada sexo. El objetivo es resolver el contraste:

$$H_0 : \mu_X = \mu_Y$$

$$H_1 : \mu_X > \mu_Y$$

donde μ_X y μ_Y representan el salario medio de hombres y mujeres, respectivamente, en toda la empresa. El primer analista decide que concluirá que en la empresa el salario medio de los hombres es superior al de las mujeres si la media muestral observada en los varones supera en 500 euros la media muestral de los salarios en la mujeres, es decir, propone la siguiente región de rechazo:

$$R_1 = \{\bar{X} - \bar{Y} > 500\}$$

Alternativamente, el segundo analista decide concluir esta diferencia en caso de que la media de los varones supere en 1000 euros a la de las mujeres, es decir, propone la siguiente región de rechazo:

$$R_2 = \{\bar{X} - \bar{Y} > 1000\}$$

- (a) (1 pto.) Justificar cuál de los dos analistas cometerá con mayor probabilidad un error de tipo I.
- (b) (1 pto.) Justificar cuál de las dos regiones de rechazo da lugar a un contraste más potente.

4. En el archivo `Internet.csv`, se muestra en las variables “Número” y “Edad” el número de compras realizadas a través de Internet en los últimos tres meses y la edad de 231 individuos encuestados a través de un cuestionario de *on line*.
- (a) (0.5 ptos.) Dibujar esquemáticamente un diagrama de puntos que relacione la edad y el número de compras realizadas.
 - (b) (0.5 ptos.) Calcular e interpretar los coeficientes de correlación de Pearson, Kendall y Spearman.
 - (c) (1 pto.) Examinar si existe evidencia estadística de relación entre las variables e indicar, en caso afirmativo, el tipo de relación: monótona, lineal, positiva, negativa, etc.
5. En el archivo `Metro.csv`, se recoge en la variable “PARADAS” el número de paradas de metro que realiza a la semana cada uno de los individuos encuestados en una muestra de 73 usuarios de metro.
- (a) (0.5 ptos.) Dibujar esquemáticamente un diagrama de caja de la variable “PARADAS” e indicar si aparentemente parece que se trata de una distribución simétrica.
 - (b) (1 ptos.) Contrastar la hipótesis de que la mediana del número de paradas sea inferior a 21 a la semana.
 - (c) (0.5 pto.) Examinar si existe evidencia estadística de que más de la mitad de los usuarios realiza menos de 21 paradas a la semana.