

Código: Nombre y apellidos del estudiante

Prueba de evaluación continua 5: Intervalos de confianza y contrastes de hipótesis de dos muestras, ANOVA y el contraste de las tablas de contingencia

Criterios de evaluación

Las preguntas cortas valen un punto cada una y los problemas tres puntos cada uno. Se deben desarrollar y razonar las respuestas y se debe demostrar agilidad utilizando el lenguaje de la asignatura. Esto incluye tanto los problemas como las preguntas.

Formato y fecha de entrega

El nombre del fichero que contenga la PEC2 resuelta del estudiante José Gómez Ortega y de la asignatura de Fundamentos de Estadística tendrá el siguiente formato: 71501_PEC2_gomez_ortega.doc (o docx) o bien 71501_PEC2_gomez_ortega.pdf. Recordad que los apellidos se escribirán sin acentos y que este mismo nombre tiene que figurar en el espacio correspondiente de la primera página y en el pie de página de todo el documento de respuestas.

Las respuestas a las pruebas de evaluación continua deben entregarse en el Registro de Evaluación Continua (REC) que se encuentra en la parte superior del aula, en un **único archivo** en formato WORD o PDF. El último día para entregar esta actividad es el **25 de mayo de 2018** (incluido).

Enunciado

En esta PEC utilizaremos dos ficheros:

- Un archivo de datos para analizar, que tenemos en dos formatos (*trabajadores.xlsx* y *trabajadores.RData*).
- Una tabla con los parámetros personalizados para resolver vuestra PEC, diferentes para cada uno (*71501_PEC5_datos.pdf*). Aquí encontraréis los datos que le corresponden utilizando el código que se encuentra colgado en el tablón de vuestra aula.

En los ficheros *trabajadores.xlsx* y *trabajadores.RData* tenemos recogidos datos sobre una muestra representativa de trabajadores de grandes empresas de la República de Zambombia. De cada trabajador hemos recogido las siguientes variables:

- Empresa: Nombre de la compañía donde trabaja.
- Categoría: Categoría laboral
- Antigüedad: Antigüedad en la empresa al inicio del año 2017, en años.
- Sueldo: Ingresos totales percibidos durante el año 2017, en coronas zambombesas.

NOTA: Cada uno de los problemas de contrastes tendremos que solucionarlos siguiendo estos pasos:

- i. Planteamiento del contraste de hipótesis: definición de las hipótesis nula y alternativa
- ii. Expresión del estadístico de contraste así como de la distribución que sigue
- iii. Cuantificación del estadístico y conclusión calculando el p-valor o para el nivel de significación α dado

PREGUNTAS

1,2. Para las preguntas 1 y 2 responder el "Cuestionario PEC5" que encontraréis en el espacio "Actividades de Moodle" del aula. Para hacerlo tenéis el mismo tiempo que para entregar la PEC y hasta entonces podéis cambiar vuestras respuestas (podéis hacer tantos intentos como queráis y se evalúa sólo el último). No hace falta que incluyáis nada del cuestionario en este documento de respuesta de la PEC.

3.- ¿Podemos afirmar que los trabajadores de la compañía **empresa1** tienen un sueldo de media más elevada que los de la compañía **empresa2**? Decidirlo con un nivel de significación del 5%.

Nota 1: Ved el Módulo 8 "[Contraste de dos muestras](#)", y la Ficha F003 "[Contraste de dos muestras](#)".

Nota 2: Para trabajar con sólo dos compañías es conveniente seleccionarlas con el menú **Datos > Tabla de datos activa > Hacer un subconjunto de la tabla de datos activa**. En el campo «Expresión de selección» podéis poner `empresa == "empresa1" | empresa == "empresa2"` (no olvidéis cambiar `empresa1` y `empresa2` por el nombre de vuestras empresas y ponerlos entre comillas). Es aconsejable que le deis un nombre diferente a la tabla resultante (campo "Nombre de la nueva tabla de datos", por ejemplo, `DosEmpresas`) para conservar también la tabla completa, ya que más adelante la volveremos a necesitar.

Nota 3: Si no tenéis activada la opción de `RCommander` correspondiente al contraste que queréis hacer, probar en primer lugar a ejecutar la orden `DosEmpresas$empresa=droplevels(DosEmpresas$empresa)` y después refrescar la tabla mediante el menú **Datos > Tabla de datos activa > Refresca la tabla de datos activa**

Nota 4: Fijaos bien en la ordenación que hace R de las dos empresas, ya que es muy importante a la hora de definir las hipótesis unilaterales. Podéis hacer una descriptiva para ver qué empresa pone primera y cuál segunda.

4.- Sin fijarnos en las diferencias entre los sueldos medios, ¿podemos afirmar que las variabilidades poblacionales de los sueldos de las compañías **empresa1** y **empresa2** son diferentes? Decidir con un nivel de significación del 10%.

Nota 1: Ved el Módulo 9 "[Contraste de variancias](#)" y la Ficha F005 "[Contraste de la variancia](#)".

EJERCICIOS

1.- Queremos estudiar la estructura de las categorías de los trabajadores de las empresas **empresa1** y **empresa2**. Más concretamente, queremos contrastar si la variable empresa y la categoría laboral son independientes y si la proporción de trabajadores de la categoría **tipo1** es igual en la **empresa1** que en la **empresa2**.

Nota 1: Ver la Ficha 023 [Tablas de contingencia](#), el [caso práctico](#) para y la bibliografía complementaria "[Análisis bivariable entre dos variables cualitativas: tablas de contingencia](#)" para la primera parte y el Módulo 8 "[Contraste de dos muestras](#)" y la Ficha F003 "[Contraste de dos muestras](#)" para la segunda parte.

- a) Plantear el contraste de independencia de las dos variables indicando cuáles son las hipótesis del contraste, cuál es el estadístico de contraste y la distribución que sigue.
- b) Indicar el p-valor del contraste y la decisión que se tomaría con un nivel de significación del 1%
- c) Si quisiéramos realizar un contraste sobre la igualdad de proporciones poblacionales de la categoría **tipo1** de las dos empresas, ¿qué valor tomaría el estadístico de contraste? (No hace falta realizar el contraste).

PEC 5: Intervalos de confianza y contrastes de hipótesis de dos muestras, ANOVA y el contraste de las tablas de contingencia

2. Fijémonos ahora sólo en los trabajadores de la compañía **empresa1**. Estamos interesados en estudiar si existen diferencias estadísticamente significativas de sueldo entre las diferentes categorías de la empresa.

Nota 1: Ver el Módulo 12 "[Análisis de la Variancia](#)", y la Ficha F025 "[Análisis de la variancia con un factor controlado](#)".

Nota 2: Para trabajar con los datos de sólo la empresa1, es conveniente eliminar de la tabla los datos de la empresa2. Podéis hacerlo con **Datos > Tabla de datos activa > Haz un subconjunto de la tabla de datos activa**. En el campo «Expresión de selección» podéis poner **empresa == "empresa1"** (no olvidéis ponerlo entre comillas). Es aconsejable que le deis un nombre diferente a la tabla (campo "Nombre de la nueva tabla de datos") para conservar también la tabla con dos empresas.

- a) A nivel descriptivo, elaborar un resumen numérico y uno gráfico para describir la variable sueldo de los trabajadores de la compañía **empresa1** según las diferentes categorías laborales. ¿Qué semejanzas y diferencias apreciáis?
- b) Vamos a realizar un contraste para saber si la población de donde hemos extraído la muestra anterior, presenta o no diferencias en relación al sueldo. ¿Cuáles son las hipótesis del contraste, cuál es el estadístico de contraste y qué distribución sigue? Justificar los grados de libertad.
- c) ¿Qué valor tiene el estadístico de contraste y qué decisión tomáis para los niveles de significación 0.01 y 0.10?

