3

Inferencia

*Apellidos, nombre: Garcia Salinas, Daniel Grupo de prácticas: L3*

Notas:

1. Esta práctica se realizará los d as 28, 29 y 30 de noviembre, y 1 de diciembre de 2023.
2. La fecha límite de entrega de esta práctica, que debe presentarse en el Campus Virtual UVa, son las 14 horas del lunes 11 de diciembre de 2023.

En esta práctica se utilizará el fichero de datos DatosInferencia.sgd, que es un archivo en formato *Statgraphics* que contiene el conjunto de datos necesarios para resolver los ejercicios propuestos.

Las variables que aparecen en este chero son las siguientes:

# CALIFICACIONES

*Fundamentos de Programación* Calificaciones en la asignatura *Fundamentos de Programación*

*Programación I* Calificaciones en la asignatura *Programación I*

# TIEMPOS DE RESPUESTA

*Ordenador A* Tiempo de respuesta del *Ordenador A*

*Ordenador B* Tiempo de respuesta del *Ordenador B*

# NOTAS MATERIAS

*Notas i* (0 *i* 10) Columna (*i* + 1)-ésima de la tabla de contingencia de las notas obtenidas por 1102 estudiantes en las *Materias A* y *B*

*Materia A* Distribución de frecuencias de las notas obtenidas por una

m.a.s. de 180 estudiantes de la *Materia A*

*Materia B* Distribución de frecuencias de las notas obtenidas por una

m.a.s.de 240 estudiantes de la *Materia B*

# SERVICIO INFORM`TICO

*Número de Fallos* Número semanal de fallos por problemas de software en un

sistema informático

Este fichero de datos se encuentra disponible en Moodle (https://aulas.inf.uva.es/).

# **EJERCICIO 1**

Un profesor de Informática que ha impartido docencia durante los últimos cuatro cursos académicos en las asignaturas de *Fundamentos de Programación* y *Programación I* desea estudiar las calificaciones obtenidas por sus alumnos. Los resultados obtenidos con una muestra aleatoria simple de sesenta y nueve estudiantes se encuentran en las variables *Fundamentos de Programación* y *Programación I*, respectivamente.

* 1. Texto, Carta

     Descripción generada automáticamenteEl profesor desea saber, en primer lugar, si las calificaciones en la asignatura *Programación I* siguen una distribución normal. ¿Qué contraste debe plantearse si queremos verificar la hipótesis anterior? ¿Cuál es el procedimiento más adecuado para su resolución en este caso? ¿Por qué?

Un conjunto de letras blancas en un fondo blanco

Descripción generada automáticamente con confianza media

Como es un contraste no paramétrico, el procedimiento más adecuado para resolverlo es mediante un Test de Kolmogorov-Smirnov

* 1. Resolver el contraste del apartado anterior con un nivel de significación del 3% utilizando el *p* valor.

Tabla

Descripción generada automáticamente

Como el p-valor es muy grande, no podemos rechazar H0, por lo que se puede concluir que las calificaciones de Programación I si siguen una distribución normal

Supongamos, a partir de ahora, que las calificaciones en las asignaturas *Fundamentos de Programación* y *Programación I* están normalmente distribuidas.

* 1. Calcular un IC del 94% para la calificación media en *Fundamentos de Programación*. ¿Se puede admitir, como asegura el profesor, que la calificación media en la asignatura de *Fundamentos de Programación* es 6*.*2, a un nivel del 6%? Razonar la respuesta.

Texto

Descripción generada automáticamente

Sí se puede admitir que la calificación media de la asignatura de Fundamentos de Programación I es 6,2 ya que este valor si que pertenece al IC de la media

* 1. Un conjunto de letras blancas en un fondo blanco

     Descripción generada automáticamente con confianza mediaA la vista de las calificaciones obtenidas, puede parecer más razonable plantear el contraste unilateral izquierdo *H*0 : = 6*.*2 frente a *H*1 : *f <* 6*.*2 (siendo *f* la calificación media obtenida en *Fundamentos de Programación*). ¿Es ahora cierta la afirmación del profesor? Razonar la respuesta

Como el p-valor es mayor que el α, no podemos rechazar H0, por lo cual se puede que la afirmación del profesor es verdadera, luego, admitimos que la media de la asignatura fundamentos de programación es 6.2

* 1. Con un nivel de confianza del 94%, ¿se puede afirmar que la calificación media en la asignatura *Programación I* es 5.8, o inferior, como dicen los alumnos? Razonar la respuesta.

Calendario

Descripción generada automáticamente

Como el p-valor es menor que el α, rechazamos H0, luego podemos decir que la afirmación de los estudiantes es falsa, luego, admitimos que la media es mayor que 5.8

* 1. ¿Se puede admitir, como afirma el profesor, que la calificación media en *Fundamentos de Programación* es superior en a lo más quince centésimas a la calificación media en *Programación I*, con un nivel de significación del 6%? ¿Contradice esto las respuestas a los apartados anteriores? Razonar las respuestas.

Pizarrón blanco con letras negras sobre fondo blanco

Descripción generada automáticamente con confianza media

Como el p-valor es mayor que el α, no se rechaza H0, luego podemos decir que la afirmación del profesor es cierta, luego, admitimos que la calificación media de Fundamentos de Programación es superior en a lo mas 15 centésimas a la media de Programación I

* 1. El profesor en cuestión también afirma que el porcentaje de alumnos que suspenden *Fundamentos de Programación* (obtienen una calificación inferior a 5) es a lo más del 6%. Plantear el contraste de hip tesis adecuado y calcular el *p* valor correspondiente. ¿Qué decisión debe adoptarse al nivel = 0.02? Razonar la respuesta.

Texto, Calendario

Descripción generada automáticamente

Como el p-valor es mayor que el α, no se rechaza H0, luego podemos decir que la afirmación del profesor es cierta, admitiendo que el porcentaje de alumnos que suspenden Fundamentos de Programación no suspera el 6%

* 1. Si el profesor desea realizar el contraste del apartado anterior con un nivel de confianza del 98% y una probabilidad de aceptar la hipótesis nula cuando la verdadera proporción de suspensos es 0*,*13 del 5%, ¿cuál deber a ser el tamaño muestral mínimo? Razonar la respuesta.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

# **EJERCICIO 2**

El tiempo de respuesta de un ordenador se define como el tiempo que un usuario debe esperar mientras la computadora accede a la información en el disco. Se desea averiguar si los tiempos de respuesta medios de dos ordenadores *A* y *B* son distintos. Para ello se han seleccionado muestras aleatorias simples independientes de 85 tiempos de respuesta en cada computadora (registrados en milisegundos). Los resultados obtenidos se encuentran en las variables *Ordenador A* y *Ordenador B*.

2.1 Suponiendo que el tiempo de respuesta de un ordenador tiene una distribución simétrica, plantear el contraste de hipótesis que permita resolver el problema.

Texto

Descripción generada automáticamenteTexto, Carta

Descripción generada automáticamente

Dibujo de un pizarrón blanco

Descripción generada automáticamente con confianza baja2.3 Para saber si el tiempo de respuesta del Ordenador *A* sigue una distribución normal, un ingeniero de sistemas aplica el test *chi-cuadrado* de bondad de ajuste y el test de Kolmogorov-Smirnov, ambos a nivel = 0*.*05, y decide rechazar la hipótesis de normalidad cuando alguno de los dos detecte falta de ajuste. ¿Qué decisión adoptará? ¿Cómo es el error de tipo I de esta técnica de contraste: α = 0*,*05, α *>* 0*,*05 o α *<* 0*,*05? Justificar la respuesta.

Tabla

Descripción generada automáticamente

En cualquiera de las 2 pruebas, el p-valor obtenido es grande, por lo cual aceptamos H0, luego el tiempo de respuesta del ordenador A sigue una distribución normal

Para saber cómo es el error de tipo 1, definimos las siguientes regiones críticas:

α = Pr {Error de tipo I} = Pr {(C1 ∪ C2) / H0} = 0.05

C1 = región crítica con chi-cuadrado

C2 = región crítica con Kolmogorov-Smirnov

2.4 Calcular un IC de nivel 0*:*95 para la desviación típica del tiempo de respuesta del Ordenador *A* y un IC de nivel 0*:*99 para la varianza del tiempo de respuesta del Ordenador *B*.

Como nos piden el IC de la varianza, simplemente elevamos al cuadrado los extremos superiores e inferiores del intervalo: IC = [0.914464^2, 1.80572^2] = [0.836244,3.260625]

Texto

Descripción generada automáticamente

2.5 Calcular un IC de nivel 0*.*95 para la diferencia de tiempos medios de ejecución. Justificar la respuesta.

Texto, Pizarra

Descripción generada automáticamente

Como calculamos en el apartado anterior: Texto

Descripción generada automáticamente

El p-valor es muy pequeño, luego no podemos asumir que las desviaciones sean iguales.

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

2.6 Resolver a un nivel = 0*:*05 el contraste planteado en el punto primero, utilizando la respuesta del apartado anterior. Justificar la respuesta.

Como vemos en el apartado anterior, el IC de la diferencia de medias es

[-0.051949,0.745114]. Podemos comprobar que el 0 pertenece al intervalo, no rechazando H0, es decir, podemos dar por hecho que las medias de los tiempos de respuesta de los ordenadores A y B son iguales.

# **EJERCICIO 3**

Se quiere averiguar si existe alguna relación entre las notas obtenidas en dos materias *A* y *B*, pertenecientes a la titulación de *Informática Financiera*, que se cursan el mismo semestre. Por este motivo, se seleccionaron al azar 1102 estudiantes que habían cursado durante varios años ambas materias y se anotaron los resultados obtenidos. Los datos se muestran en la siguiente tabla:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Notas Materia *A* |  |  |  | Notas Materia *B* | | | | |  |  |  |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 0 | 2 | 9 | 2 | 4 | 4 | 2 | 1 | 1 | 3 | 1 | 0 |
| 1 | 6 | 13 | 11 | 20 | 8 | 9 | 7 | 2 | 4 | 4 | 2 |
| 2 | 6 | 10 | 14 | 25 | 9 | 14 | 16 | 4 | 4 | 3 | 3 |
| 3 | 6 | 10 | 17 | 18 | 13 | 18 | 11 | 10 | 10 | 5 | 5 |
| 4 | 4 | 6 | 12 | 12 | 12 | 19 | 10 | 11 | 16 | 5 | 10 |
| 5 | 2 | 8 | 11 | 18 | 17 | 22 | 11 | 14 | 18 | 9 | 13 |
| 6 | 3 | 4 | 8 | 13 | 13 | 13 | 11 | 7 | 23 | 12 | 16 |
| 7 | 2 | 1 | 6 | 10 | 13 | 11 | 10 | 17 | 17 | 8 | 9 |
| 8 | 5 | 4 | 3 | 9 | 13 | 11 | 10 | 13 | 9 | 10 | 11 |
| 9 | 2 | 3 | 3 | 7 | 9 | 10 | 16 | 11 | 10 | 10 | 14 |
| 10 | 0 | 3 | 1 | 7 | 10 | 10 | 8 | 2 | 12 | 9 | 14 |

Pizarrón blanco con texto en letras negras sobre fondo blanco

Descripción generada automáticamente con confianza media3.1 Plantear un contraste de hipótesis que permita decidir si las variables están o no relacionadas.

Texto

Descripción generada automáticamente con confianza media3.2 Resolver el contraste anterior y tomar una decisión con un nivel de significación del 1%, utilizando para ello el *p* valor. ¿Cuál ha sido el valor observado en el estadístico de contraste y los grados de libertad?

NOTA: Se dijo en clase que sin agrupar, (se puede ver que Statgraphics se “queja” al haber celdas con menos de 5 casos) luego dejaremos este resultado así, se resolverá el contraste más adelante por otros métodos

Como el p-valor es 0, rechazamos H0

3.3 Con el fin de estudiar si la distribución de las notas es la misma en ambas asignaturas, se seleccionaron muestras aleatorias de 180 estudiantes de la Materia *A* y 240 estudiantes de la Materia *B*. Las notas obtenidas se muestran en la siguiente tabla:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | Notas | |  |  |  |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Materia *A* | 3 | 5 | 14 | 37 | 48 | 44 | 26 | 3 | 0 |
| Materia *B* | 0 | 6 | 15 | 47 | 58 | 58 | 35 | 15 | 6 |

Texto

Descripción generada automáticamentePlantear el contraste de hip tesis que permita resolver la cuestión planteada. Resolver el contraste y tomar una decisión con un nivel de significación del 8% utilizando para ello el *p* valor. Indicar el valor observado del estadístico de contraste y los grados de libertad.

El p-valor es menor que el α, por lo cual rechazamos H0, es decir, las notas de las materias no siguen una distribución común

NOTA: Al igual que en el apartado anterior, se dijo en clase que sin agrupar, (se puede ver que Statgraphics se “queja” al haber celdas con menos de 5 casos) luego dejaremos este resultado así, se resolverá el contraste más adelante por otros métodos

H0: Las notas de la Materia A y la Materia B siguen una distribución común

H1: Las notas de la Materia A y la Materia B no siguen una distribución común

3.4 Calcular un IC de nivel aproximado 0*.*95 para la proporción de estudiantes que aprueban cada una de las materias.

MATERIA A:

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

MATERIA B:

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Un conjunto de letras blancas en un fondo blanco

Descripción generada automáticamente con confianza mediaInterfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente3.5 ¿Se puede aceptar que las proporciones de alumnos que aprueban son las mismas en ambas materias? Razonar la respuesta.

Debido a que el p-valor es alto, no se puede rechazar la hipótesis nula, por lo que se puede aceptar que las proporciones son las mismas en ambas materias

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente3.6 Calcular un IC de nivel aproximado 0*:*97 para la diferencia de proporciones de estudiantes de las Materias *A* y *B* que obtienen un 10. ¿Se puede afirmar con una confianza (aproximada) del 97% que la proporción de alumnos de las Materias *A* y *B* que tienen un 10 es la misma? ¿Y que es mayor en la Materia *B*? Justificar la respuesta.

pA = proporción de estudiantes con un 10 en la materia A

pB = proporción de estudiantes con un 10 en la materia B

Podemos ver que, si observamos el p valor, lo que este nos dice es que no debemos rechazar la hipótesis nula, pero observando el IC, el 0 no esta en el intervalo, por lo cual deberíamos rechazarla. Para resolver esto, planteemos un contraste de desigualdades

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamenteImagen en blanco y negro

Descripción generada automáticamente con confianza baja

En este caso no podemos rechazar la hipótesis nula, porque el p-valor es muy grande, por lo que podemos concluir que la proporción de alumnos que obtienen un 10 es mayor en la materia B

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente3.7 ¿Cuántos estudiantes de cada materia deben seleccionarse si se desea estimar la diferencia de proporciones utilizadas en el apartado anterior con una precisión del 5% y un nivel de confianza del 95%? (suponer que se toman muestras de igual tamaño en ambas poblaciones de estudiantes y utilizar las estimaciones del enunciado).

Interfaz de usuario gráfica, Texto

Descripción generada automáticamente3.8 Repetir el apartado anterior si no se tiene una estimación inicial de las proporciones. Justificar la respuesta.

# **EJERCICIO** **4**

El número semanal de fallos por problemas de *software* que se han producido en un sistema informático son los siguientes:

Número de fallos 0 1 2 3 4 5 6 8

Frecuencia 41 62 63 38 12 7 1 1

Un dibujo de un pizarrón blanco

Descripción generada automáticamente con confianza baja4.1 ¿Puede aceptarse que el número semanal de fallos por problemas de *software* que se producen en ese sistema informático es una variable aleatoria con distribución de Poisson? Justificar la respuesta.

El p-valor es alto, por lo cual no podemos rechazar H0 asi que si podemos suponer que el numero semanal de fallos por problemas de software que se producen en un sistema informático sigue una distribución de Poisson

Suponiendo que el número semanal de fallos por problemas de *software* se distribuye según una ley de Poisson:

Texto

Descripción generada automáticamente con confianza media4.2 Calcular una estimación insesgada de la tasa semanal de fallos del sistema por problemas de *software*. Razonar la respuesta.

Como suponemos que la variable tasa de fallos sigue una distribución de Poisson, estimamos la media

Texto, Carta

Descripción generada automáticamente4.3 Calcular un intervalo de confianza de nivel al menos 0*:*94 para el parámetro de la distribución, justificando la respuesta.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente4.4 Calcular ahora un intervalo de confianza de nivel aproximado 0*:*94 para el parámetro de la distribución utilizando el TCL.

4.5 ¿Puede aceptarse, a un nivel del 2% que, como dice el ingeniero informático responsable del sistema, la tasa semanal de fallos es de a lo más 1*:*6? ¿Y a un nivel del 5%? Justificar la respuesta.

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamenteInterfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Se rechaza H0 porque p-valor<α

No se puede rechazar H0 para α, porque p-valor>α