

# AID - Trabajo Práctico Nro. III

29 de junio de 2022

## 1. Revisión de conceptos de pruebas de hipótesis

**Ejercicio 1.** Seleccionar la alternativa correcta en cada caso:

1. El nivel de significación de un test de hipótesis:
  - (a) Suele ser pequeño y lo fija el investigador o un convenio generalmente aceptado.
  - (b) Da la probabilidad de declarar significativo el resultado de un test, cuando esto es falso.
  - (c) Al disminuir hace aumentar la probabilidad del error de tipo II.
  - (d) Todo lo anterior es cierto.
  - (e) Todo lo anterior es falso.
2. Un estudio sobre la efectividad de un tipo de campaña llega a la conclusión de que éste significativamente distinto del tradicional con  $p < 0,05$   
¿Cuál es la interpretación correcta de este resultado?
  - (a) Con toda seguridad, el nuevo estilo supera al tradicional
  - (b) La probabilidad de éxito con la nueva campaña supera a la probabilidad del anterior en un 95 %.

- (c) El nuevo estilo es un 95 % mejor que el tradicional.
  - (d) Si la campana no fuese efectiva, existe menos del 5 % de probabilidad de observar muestras tan contrarias a dicha hipótesis como las obtenidas.
  - (e) Ninguna de las anteriores es correcta.
3. En una prueba de hipótesis el p-valor es:
- (a) Un número pequeño.
  - (b) Fijado antes de realizar la prueba.
  - (c) La probabilidad de rechazar la hipótesis nula.
  - (d) La probabilidad de error al rechazar la hipótesis alternativa.
  - (e) Conocido al extraer la muestra y calcular el estadístico experimental.
4. Una prueba de hipótesis se considera significativa si:
- (a) Una muestra aleatoria es coherente con la hipótesis nula.
  - (b) Una muestra aleatoria no es coherente con la hipótesis nula.
  - (c) La hipótesis alternativa es más probable que la nula.
  - (d) Todo lo anterior es cierto.
  - (e) Son ciertas (b) y (c).
5. Se realizó un estudio para comparar la duración de lamparas de bajo consume utilizando dos métodos de fabricación diferentes. No se encontró diferencia estadísticamente significativa. ¿Cuál de las siguientes razones podrían ser causantes del resultado?
- (a) Los métodos ofrecen tiempos de duración muy diferentes.

- (b) El nivel de significación es demasiado alto.
  - (c) Las muestras son demasiado numerosas.
  - (d) Las muestras son demasiado pequeñas.
  - (e) Nada de lo anterior.
6. Elija la afirmación falsa:
- (a) El nivel de significación es normalmente un valor pequeño.
  - (b) La significación de una prueba es conocida después de analizar los datos.
  - (c) El nivel de significación de una prueba debe ser fijado antes de seleccionar la muestra.
  - (d) Una prueba puede resultar significativa antes de recoger los datos.
  - (e) Una prueba se señala como significativa cuando se obtiene una muestra que discrepa mucho de la hipótesis nula.
7. El error de tipo I consiste en:
- (a) Rechazar  $H_0$  cuando es falsa.
  - (b) Rechazar  $H_0$  cuando es cierta.
  - (c) No rechazar  $H_0$  cuando es cierta.
  - (d) No rechazar  $H_0$  cuando es falsa
  - (e) La probabilidad de rechazar  $H_0$  cuando es falsa.

## 2. Pruebas de Independencia y Homogeneidad

**Ejercicio 2.** A un grupo de 350 adultos, quienes participaron en una encuesta, se les preguntó si accedían o no a Internet. Las respuestas clasificadas por sexo fueron las siguientes:

Tabla 1: Twiter por Sexo

| sexo          | F   | M   | Total |
|---------------|-----|-----|-------|
| usa Twiter    | 14  | 25  | 39    |
| no usa Twiter | 159 | 152 | 311   |
| Total         | 173 | 177 | 350   |

- (a) Represente gráficamente esta información. Interprete el gráfico.
- (b) Obtenga los porcentajes por filas y compare las diferentes zonas.
- (c) Obtenga los porcentajes por zona y compare las diferentes zonas.
- (d) Calcule las frecuencias esperadas bajo independencia y compárelas con las observadas, para el grupo general y por zona.
- (e) ¿Sugieren estos datos que existe diferencia de proporciones entre mujeres y hombres que acceden o no a Twiter? Considerar  $\alpha = 0,05$ .

**Ejercicio 3.** Se clasificó en forma cruzada una muestra de 250 técnicos en telecomunicaciones en base a su especialidad y a la zona de la comunidad en que estaban trabajando. Los resultados están tabulados a continuación:

Tabla 2: Especialidad por Zona

| Zona  | A  | B  | C  | D  | E   | Totales |
|-------|----|----|----|----|-----|---------|
| Norte | 20 | 18 | 12 | 17 | 67  | 134     |
| Sur   | 6  | 22 | 15 | 13 | 56  | 112     |
| Este  | 4  | 6  | 14 | 11 | 35  | 70      |
| Oeste | 10 | 19 | 23 | 40 | 92  | 184     |
| Total | 40 | 65 | 64 | 81 | 250 | 500     |

- (a) Considera adecuado un test de homogeneidad o de independencia? Fundamente su respuesta considerando el tipo de muestreo realizado.
- (b) Establezca las hipótesis de interés, realice el contraste y concluya considerando un nivel de significación del 1 %.

**Ejercicio 4.** Entre 1605 recién nacidos registrados en una maternidad, se han presentado 48 con un angioma cuya presencia, se sospecha puede estar relacionada con el carácter (normal o patológico) del embarazo de la madre. Los resultados son los siguientes:

Tabla 3: Presencia de Angioma por Tipo de Embarazo

|                     | con angioma | sin angioma |
|---------------------|-------------|-------------|
| Embarazo normal     | 37          | 1334        |
| Embarazo patológico | 11          | 223         |

Plantear y testear las hipótesis correspondientes considerando un nivel de significación del 5 %.

### 3. Análisis de correspondencias

**Ejercicio 5.** Se ha realizado una encuesta entre el personal de una empresa. Se le preguntó a cada uno el cargo que desempeña y la cantidad de cigarrillos diarios que fuma. La frecuencia de fumador fue categorizada con las categorías: No fuma- Fuma Poco – Fumador Medio y Fuma Mucho

En la Tabla4 se resumen las respuestas.

**Tabla 4: Hábito por Puesto**

|                  | Categoría de fumador |      |       |       |               |
|------------------|----------------------|------|-------|-------|---------------|
| Grupo            | No fuma              | Poco | Medio | Mucho | Total de Fila |
| Gerente Senior   | 4                    | 2    | 3     | 2     | 11            |
| Gerente Junior   | 4                    | 3    | 7     | 4     | 18            |
| Empleados Senior | 25                   | 10   | 12    | 4     | 51            |
| Empleados Junior | 18                   | 24   | 33    | 13    | 88            |
| Secretarias      | 10                   | 6    | 7     | 2     | 25            |
| Total columna    | 61                   | 45   | 62    | 25    | 193           |

Estamos interesados en estudiar la relación, si existiera entre las variables: “puesto de trabajo” y “nivel de fumador” en el contexto de esta empresa.

1. La primera pregunta que nos hacemos es si la distribución de la variable fumador es similar en todos los niveles de la variable puesto de desempeño. Construya para eso las distribuciones condicionales de fumador a cada grupo de trabajo.
2. Realice un análisis de correspondencias para estos datos. ¿Cuántos factores tiene sentido considerar?

3. Realizar los gráficos perfiles que considere adecuados.
4. Explique la calidad de la representación y las relaciones entre las variables y los ejes (inercia, calidad, cosenos).
5. Hacer una síntesis de sus conclusiones, inspeccione relaciones entre perfiles fila, entre perfiles columna, asociaciones entre filas y columnas de manera adecuada.
6. ¿Cual es la inercia total?

**Ejercicio 6.** En el archivo de datos **Autos.xls** (Infostat) están los datos de 339 usuarios de auto Las variables que se han preguntado refieren al origen del auto (americano, japonés o europeo), estado civil (soltero, casado-hijo) , relación con la casa(dueño, alquila), tipo de auto(familiar, sport), sexo(Hombre-Mujer), tamaño del auto(chico, mediano y grande) e ingreso familiar(en dos niveles 1 y 2).

- (a) Elegir tres variables y construir la matriz disyuntiva y la matriz de Burt. Explicar el significado de los valores diagonales y verificar las propiedades de la matriz.
- (b) Realizar un análisis de correspondencias múltiples con estas variables y explicar los resultados.

**Ejercicio 7.** La opinión de los ingleses de los europeos se encuentra en el archivo opinión **ingleses.xls**. Conducir un análisis de correspondencias para caracterizar a los europeos desde la mirada de los ingleses.

1. Qué características son las más usuales?
2. Qué características son las más raras?
3. En función de estos datos, le parece justo decir que Paris es la ciudad del glamour?

**Ejercicio 8.** En el archivo **proteinas.xls** se tabuló el consumo de distintos tipos de proteínas per cápita de los habitantes de distintos países de Europa.

Conduzca un análisis de correspondencias múltiples para describir el tipo de consumo de proteínas de los países. Vincule este análisis con la posición geográfica de los mismos.