

TRABAJO PRÁCTICO N° 4

Primera parte: revisión de conceptos de pruebas de hipótesis

Ejercicio 1

A- Elija la alternativa correcta

- 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 tratamiento llega a la conclusión de que éste es mejor que el tradicional con $p < 0,05$ ¿Cuál es la interpretación correcta de este resultado?
 - f) Con toda seguridad, el nuevo tratamiento es mejor que el tradicional
 - g) La probabilidad de que el nuevo tratamiento sea mejor que el anterior supera el 95%
 - h) El nuevo tratamiento es un 95% más efectivo que el tradicional
 - i) La probabilidad que el tradicional sea mejor que el nuevo tratamiento es menor de 5%
 - j) Si el tratamiento no fuese efectivo, existe menos del 5% de probabilidad de observar unas muestras tan contrarias a dicha hipótesis como las obtenidas

- 3) En una prueba de hipótesis el p-valor es:
 - k) Un número pequeño
 - l) Fijado antes de realizar la prueba
 - m) La probabilidad de rechazar la hipótesis nula
 - n) La probabilidad de error al rechazar la hipótesis alternativa
 - o) Conocido al extraer la muestra y calcular el estadístico experimental

- 4) Una prueba de hipótesis se considera significativa si:
 - p) Una muestra aleatoria es coherente con la hipótesis nula.
 - q) Una muestra aleatoria no es coherente con la hipótesis nula.
 - r) La hipótesis alternativa es más probable que la nula.
 - s) Todo lo anterior es cierto.
 - t) Son ciertas (b) y (c).

- 5) Se realiza 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?
 - u) Los métodos ofrecen tiempos de duración muy diferentes.
 - v) El nivel de significación es demasiado alto.
 - w) Las muestras son demasiado numerosas.
 - x) Las muestras son demasiado pequeñas.
 - y) Nada de lo anterior.

6) De las siguientes, cuál se corresponde con un error de tipo II:

- z) Aceptar que un tratamiento ineficaz produce efectos útiles.
- aa) Rechazar que un tratamiento ineficaz produce efectos útiles.
- bb) Aceptar que un tratamiento eficaz produce efectos útiles.
- cc) Rechazar que un tratamiento eficaz produce efectos útiles.
- dd) Nada de lo anterior es cierto.

7) Elija la afirmación falsa:

- ee) El nivel de significación es normalmente un valor pequeño.
- ff) La significación de una prueba es conocida tras analizar los datos.
- gg) El nivel de significación de una prueba debe ser fijado antes de analizar los datos.
- hh) Una prueba debe ser declarada significativa antes de recoger los datos.
- ii) Una prueba es declarada significativa si se obtiene una muestra que discrepa mucho de la hipótesis nula.

8) El error de tipo I consiste en:

- jj) rechazar H_0 cuando es falsa.
- kk) rechazar H_0 cuando es cierta.
- ll) No rechazar H_0 cuando es falsa.
- mm) No rechazar H_0 cuando es cierta.
- nn) La probabilidad de rechazar H_0 cuando es falsa

Segunda parte: Pruebas de Independencia y Homogeneidad

Ejercicio 2

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 fueron los siguientes:

	<u>Especialidad</u>				
Zona de la comunidad	A	B	C	D	Total
Norte	20	18	12	17	67
Sur	6	22	15	13	56
Este	4	6	14	11	35
Oeste	10	19	23	40	92
Total	40	65	64	81	250

- a. Obtenga una tabla con las frecuencias relativas observadas. Señale las tres distribuciones que ofrece la tabla y preséntelas cada una de ellas en una nueva tabla.
- b. Obtenga los porcentajes por filas y compare las diferentes zonas.
- c. Obtenga los porcentajes por especialidad y compare las diferentes especialidades.

- d. Calcule las frecuencias esperadas bajo independencia y compárelas con las observadas.
- e. ¿Proporcionan estos datos la evidencia suficiente que indique una falta de independencia entre los dos criterios de clasificación? Considerar un nivel de significación del 1 %.

Ejercicio 3

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

	Sexo		Total
	Masculino	Femenino	
Usa twitter	14	25	39
No usa twitter	159	152	311
Total	173	177	350

¿Sugieren estos datos que existe diferencia de proporciones entre mujeres y hombres que acceden o no a Twitter? Considerar $\alpha = 0,05$.

Ejercicio 4

Para poder aplicar las pruebas de Chi Cuadrado de Pearson de homogeneidad o independencia es necesario que las frecuencias esperadas en las casillas resulten superiores a 5 en al menos el 80% de las casillas.

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:

	Con angioma	Sin angioma
Embarazo normal	37	1334
Embarazo patológico	11	223

Hay evidencia para suponer que Testear la independencia de ambos atributos mediante el estadístico χ^2 .

Tercera parte: 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 siguiente tabla se resumen las respuestas.

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

- 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.
- Realice un análisis de correspondencias para estos datos. ¿Cuántos factores tiene sentido considerar?
- Realizar los gráficos perfiles que considere adecuados.
- Explique la calidad de la representación y las relaciones entre las variables y los ejes. (inercia, calidad, cosenos).
- Hacer una síntesis de sus conclusiones, inspeccione relaciones entre perfiles fila, entre perfiles columna, asociaciones entre filas y columnas de manera adecuada.
- ¿Cual es la inercia total?

	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

Ejercicio 6

En un estudio epidemiológico de los decesos por problemas del corazón se registraron los datos de las siguientes variables para 113 individuos en un Instituto de Los Angeles.

Columna	Variables (tabla Shock)	Tipo	Unidades	Modalidades
A	ID	Nominal	Ninguna	Pacientes numerados secuencialmente
B	Sujeto	Nominal	Ninguna	Pacientes según clasificación
C	Edad	Continua	Años	Edad del último cumpleaños
D	Altura	Continua	Cm	
E	Sexo	Nominal		1: masculino; 2: femenino
F	Sobrevida	Nominal		1: sobrevivió; 3: murió
G	Shock	Nominal		2: sin shock 3: shock hipovolémico 4: shock cardíogenico 5: shock bacterial 6: shock neurogénico 7: otro
H	Presión sistólica	Continua	Mm Hg	Aproximado al entero más próximo
I	Presión arterial	Continua	Mm Hg	Aproximado al entero más próximo
J	Latidos	Continua	Latidos por minuto	
K	Presión diastólica	Continua	Mm Hg	Aproximado al entero más próximo
L	Presión venosa	Continua	Mm Hg	Decimal
M	Superficie corporal	Continua	m ²	Decimal
N	Índice cardíaco	Continua	Litro/min/m ²	Decimal
O	Tiempo de aparición	Continua	Seg	Decimal

P	Tiempo promedio de circulación	Continua	Seg	Decimal
Q	Excreción urinaria	Continua	MI/ hr	Aproximado al entero más cercano
R	Índice de volumen de plasma	Continua	MI/ kg	Decimal
S	Índice de glóbulos rojos	Continua	MI/ kg	Decimal
T	Hemoglobina	Continua	Gr	Decimal
U	Hematocrito	Continua	%	Decimal

- Construir tablas de contingencia utilizando las variables Sexo, Sobrevida y Shock, tomadas de a dos.
- Utilizar el método del análisis de correspondencias simples para un par de variables. Graficar.
- Utilizar el método del análisis de correspondencias múltiples para cuatro variables que se elijan. Graficar. Interpretar.
- Estudiar la distribución de la edad y recodificarla en categorías. Repetir b. tomando las variables C, H, I, J, O, P, Q, R, S, T, U ilustrativas y E, F, G y edad como variables activas. Graficar.

Ejercicio 7

En el archivo de datos de prueba de InfoStat Autos 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).

- 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.
- Realizar un análisis de correspondencias múltiples con estas variables y explicar los resultados.

Ejercicio 8

En la siguiente tabla se exhiben países europeos y el consumo de distintos tipos de proteínas per cápita de los habitantes.

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.

País	Carne Vacuna	Carne Cerdo	Huevos	Leche	Pescado	Cereal	Embutidos	Frutos Secos	Frutas y Vegetales
Albania	10,1	1,4	0,5	8,9	0,2	42,3	0,6	5,5	1,7
Austria	8,9	14	4,3	19,9	2,1	28	3,6	1,3	4,3
Bélgica	13,5	9,3	4,1	17,5	4,5	26,6	5,7	2,1	4
Bulgaria	7,8	6	1,6	8,3	1,2	56,7	1,1	3,7	4,2
Checosl	9,7	11,4	2,8	12,5	2	34,3	5	1,1	4
Dinamarca	10,6	10,8	3,7	25	9,9	21,9	4,8	0,7	2,4
AlemaniaE	8,4	11,6	3,7	11,1	5,4	24,6	6,5	0,8	3,6
Finlandia	9,5	4,9	2,7	33,7	5,8	26,3	5,1	1	1,4
Francia	18	9,9	3,3	19,5	5,7	28,1	4,8	2,4	6,5

Grecia	10,2	3	2,8	17,6	5,9	41,7	2,2	7,8	6,5
Hungría	5,3	12,4	2,9	9,7	0,3	40,1	4	5,4	4,2
Irlanda	13,9	10	4,7	25,8	2,2	24	6,2	1,6	2,9
Italia	9	5,1	2,9	13,7	3,4	36,8	2,1	4,3	6,7
P.Bajos	9,5	13,6	3,6	23,4	2,5	22,4	4,2	1,8	3,7
Noruega	9,4	4,7	2,7	23,3	9,7	23	4,6	1,6	2,7
Polonia	6,9	10,2	2,7	19,3	3	36,1	5,9	2	6,6
Portugal	6,2	3,7	1,1	4,9	14,2	27	5,9	4,7	7,9
Rumania	6,2	6,3	1,5	11,1	1	49,6	3,1	5,3	2,8
España	7,1	3,4	3,1	8,6	7	29,2	5,7	5,9	7,2
Suecia	9,9	7,8	3,5	24,7	7,5	19,5	3,7	1,4	2
Suiza	13,1	10,1	3,1	23,8	2,3	25,6	2,8	2,4	4,9
Inglaterra	17,4	5,7	4,7	20,6	4,3	24,3	4,7	3,4	3,3
Rusia	9,3	4,6	2,1	16,6	3	43,6	6,4	3,4	2,9
AlemaniaO	11,4	12,5	4,1	18,8	3,4	18,6	5,2	1,5	3,8