

ESTADISTICA II

Evaluación final

Ejercicio ANOVA de dos factores

El Departamento de nutrición de la Ciudad desea analizar los efectos del tipo de harina y del porcentaje de endulzamiento sobre los atributos físicos de un determinado tipo de torta, a la venta actualmente. Para ello, se diseñó un experimento utilizando dos tipos de harina (multipropósito y harina para tortas), y distintos porcentajes de endulzamiento, en cuatro niveles diferentes. Los siguientes datos corresponden a la información de la densidad específica de muestras de tortas preparadas con las distintas combinaciones posibles. Se prepararon 3 tortas con cada una de las combinaciones posibles

		Tipo de harina	
		General	Para tortas
Concentración de azucar	0	0.90	0.91
		0.87	0.90
		0.90	0.80
	50	0.86	0.88
		0.89	0.82
		0.91	0.83
	75	0.93	0.86
		0.88	0.85
		0.87	0.80
	100	0.79	0.86
		0.82	0.85
		0.80	0.85

Los resultados obtenidos son los que se muestran en el archivo *terapia.sav*

- 1) Existen diferencias estadísticamente significativas debidas al tipo de harina utilizada?
- 2) Existen diferencias estadísticamente significativas debidas al nivel de endulzamiento utilizado?
- 3) Existe interacción estadísticamente significativa entre los factores considerados?
- 4) Realizar los gráficos de medias e interpretar.

EJERCICIO DE REGRESION. Consumo de cigarrillos

El siguiente conjunto de datos refieren al consumo de cigarrillos en los Estados Unidos. La base de datos que se presenta (archivo p081.sav) corresponde a los datos de los 50 estados del país, y las variables consignadas son las siguientes:

Variable	Definición
Age	Edad mediana de las personas que viven en el Estado
HS	Porcentaje de personas de más de 25 años con secundario completo
Income	Ingreso personal per capital
Black	Porcentaje de individuos de raza negra
Female	Porcentaje de mujeres
Price	Precio promedio de un paquete de cigarrillos
Sales	Número de paquetes vendidos per cápita en el estado

Se requiere construir una ecuación de regresión que relacione el consumo de cigarrillos per-cápita en todo el estado (variable Sales), con diversas variables socioeconómicas y demográficas, y determinar si estas variables son útiles para predecir el consumo de los cigarrillos.

- 1.Cuál es el nivel de asociación lineal de las variables predictoras con la variable Sales? Comentar.
2. Realizar una regresión lineal múltiple, seleccionando los mejores predictores entre las variables independientes disponibles, utilizando un método de selección automática. Describir el proceso de selección automática utilizado. (Sug. Considerar como probabilidad de entrada 0.10 y de salida del modelo 0.15.)
3. Qué información da el coeficiente de determinación?
4. Cuáles son los supuestos necesarios para definir la prueba inferencial de los estimadores de los parámetros?
5. Analizar la bondad del ajuste del modelo obtenido, comentando los indicadores y/o test que considera.
6. Realizar un análisis de los residuos del modelo para evaluar el cumplimiento de los supuestos. Para esto, realizar gráficos de los residuos con el valor predicho.
7. Analizar la colinealidad de las variables predictoras presentes en la ecuación.
8. Analizar la presencia de observaciones atípicas y/o influyentes. Comentar y resolver según el caso.

Ejercicio de regresión logística

Factores de riesgo de bajo peso al nacer

El bajo peso al nacer, definido como por un peso al nacer inferior a 2500 gr., ha sido una preocupación de los médicos durante años debido a que tanto las tasas de mortalidad como la de nacimientos defectuosos son muy altas para los niños con bajo peso al nacer. El comportamiento de la mujer durante el embarazo (incluyendo la dieta, los hábitos tabáquicos y los cuidados prenatales) pueden alterar las chances de un parto de un niño con bajo peso.

Los datos que se presentan en este ejercicio corresponden a 189 nacimientos de los cuales 59 han resultado en niños con bajo peso. El objetivo de este ejercicio es determinar cuáles de las variables presentes en la base de datos que se adjunta son factores de riesgo de bajo peso al nacer.

La base de datos que se presenta (archivo LOWBWT.sav) contiene las siguientes variables:

ID: Código de identificación

LOW: Bajo peso al nacer. (0 = ≥ 2500 g; 1 = < 2500 g) (variable dependiente)

AGE: Edad de la madre

LWT: Peso de la madre el momento de la última menstruación (en libras)

RACE: Raza (1 = White; 2 = Black; 3 = Other)

SMOKE: Fumó durante el embarazo (0 = No 1 = Yes)

PTL: Antecedentes de embarazos prematuros (0 = None; 1 = One; 2 = Two, etc).

HT: Antecedentes de hipertensión arterial (0 = No; 1 = Yes)

UI: Irritabilidad uterine (0 = No; 1 = Yes)

FTV: Cantidad de consultas obstétricas durante el primer trimestre (0 = None; 1 = One; 2 = Two, etc.)

BWT: Peso al nacer del bebé en gramos

Se requiere construir una ecuación de regresión logística que relacione la variable dicotómica que indica si se trata de un nacimiento con bajo peso al nacer con el resto de las variables que corresponda, y determinar si estas variables son útiles para predecir la variable dependiente.

1. Calcular el riesgo relativo y los odds ratio de la variable dependiente con cada una de las variables dicotómicas. Analizar los resultados.
2. Cuál es la definición de odds ratio? Qué información suministra y de qué manera puede calcularse utilizando la regresión logística?
3. Calcule los odds ratio de cada una de las variables predictoras disponibles con la variable dependiente? Comentar.
4. Realizar una regresión logística múltiple, seleccionando los mejores predictores entre las variables independientes disponibles, utilizando un método de selección automática.
5. Según el modelo obtenido, cuáles son los principales factores de riesgo del bajo peso y cuál es la magnitud de su efecto?
6. Cuáles son los supuestos necesarios para definir la prueba inferencial de los estimadores de los parámetros?
7. Indicar porcentaje de casos bien predichos por el modelo.