UDS

Mi Universidad

Nombre del Alumno: juan Antonio Espinosa Hernández

Nombre del tema: super nota

Parcial: 3

Nombre de la Materia: estadística inferencial

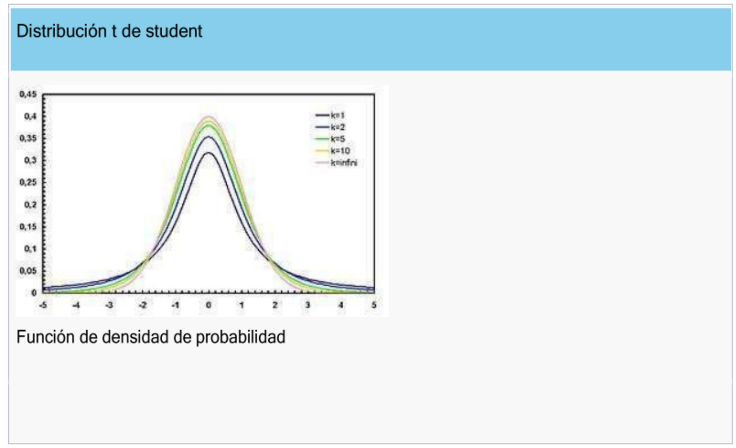
*Nombre del profesor: **Jorge***

Nombre de la Licenciatura: administración y estrategias de negocios

*Cuatrimestre: **3***

Pruebas de hipótesis con dos muestras y varias muestras de datos numéricos

Distribuciones normal y t de Student: Aparece de manera natural al realizar la prueba t de Student para la determinación de las diferencias entre dos varianzas muestrales y para la construcción del intervalo de confianza para la diferencia entre las partes de dos poblaciones cuando se desconocela desviación típica de una población y esta debe ser estimada a partir de los datos de una muestra



La distribución t de Student es la distribución de probabilidad del cocientdonde

 Z es una variable aleatoria distribuida según una normal típica (de media nula

y varianza 1).

 V es una variable continua que sigue una distribución χ² con grados de libertad.

 Z y V son independientes

Distribución t de Student no estandarizada La distribución t puede generalizarse a 3 parámetros, introduciendo un parámetro vocacional y otro de escala . El resultado es una distribución t de Student Las pruebas de significación estadística sirven para comparar variables entre distintas muestras. Si la distribución de la muestra es normal se aplican los llamados tests paramétricos. Si la distribución no puede asumirse normal se aplican las pruebas no paramétricas.

Pruebas de significación: Pruebas paramétricas

Prueba t de Student: Con esta prueba se pretende averiguar si dos muestras que tienen medias iguales, provienen de la misma

Y esta prueba permite comparar la media con su valor verdadero o bien las medias de dos poblaciones

Tabla 2. Valores críticos de t (p= 0.05, entre paréntesis p=0.1).

gl	1	2	3	4	5	6	7	8	9
t	12.71	4.3	3.18	2.78	2.57	2.45	2.36	2.31	2.26
	(6.31)	(2.92)	(2.35)	(2.13)	(2.02)	(1.94)	(1.89)	(1.86)	(1.83)
gl	**10**	**12**	**14**	**16**	**18**	**20**	**30**	**50**	**α**
T	2.23	2.18	2.14	2.12	2.1	2.09	2.04	2.01	1.96
	(1.81)	(1.78)	(1.76)	(1.75)	(1.73)	(1.72)	(1.70)	(1.68)	(1.64)

Nota: *los valores críticos de t son estimados para una prueba de 2 colas. Para una prueba de una cola se toma el valor que corresponde a p=0.1, es decir, el doble del valor de p deseado (0.05).*

Figura 1. Esquema del cálculo de t para datos apareados y no apareados

A) t student con datos apareados

X_{11}	X_{21}	D_1
X_{12}	X_{22}	D_2
$\bar{X}_{1i}$	$\bar{X}_{2i}$	$\bar{D}_i$

$\bar{D}_i = \sum D_i/n$

$S=\sqrt{\sum (D_i-\bar{D}_i)^2/(n-1)}$

$t=\bar{D}_i/EE \Rightarrow EE=s/\sqrt{n}$

Si $p \leq 0.05$ rechazar H_0

B) t student con datos no apareados

X_{11}	X_{21}
X_{12}	X_{22}
$\bar{X}_{1i}$	$\bar{X}_{2i}$
$\bar{X}1$	$\bar{X}2$
S_1	S_2

$EE_{(x_1-x_2)}=\sqrt{(\bar{s})^2(1/n_1+1/n_2)}$

siendo $(\bar{s})^2=((n_1-1)s_1^2+(n_2-1)s_2^2)/(n_1+n_2-2)$

$t=(x_1-x_2)/EE$ [t tiene (n_1+n_2-2)° de libertad]

Si $p \leq 0.05$ rechazar H_0

Datos apareados: tienen la ventaja de permitir trabajar simplificando a una sola muestra (cuyos valores corresponden a la diferencia "Di" entre cada par de datos apareados). Sustituimos x - µ (Ec. 2.2) por Di - 0 porque el valor real de las diferencias, suponiendo quelas dos muestras tienen la misma media, es 0. La DE se calcula con la muestra de diferencias.

Datos no apareados: como no se puede simplificar a una sola muestra, se ha de introducir el concepto de desviación estándar ponderada "sp" (Ec. 2.3). En la ecuación 2.2 se sustituyes por sp y x - µ por x1 - x2 y el tamaño de muestra "n" se sustituye por N ponderado "(N1 + N2)/ N1N2"

Figura 3. Tabla 2x2. Prueba $\aleph^2$

Se tabula, como ejemplo, los niños nacidos en un hospital durante un año en función de su peso y de las semanas de gestación

		Prematuros 0	Prematuros 1	
Bajo peso	0	**352** 342,7(E)	**35** 44.3 (E)	$\sum_i$=387 **O**
	0	**12** 21.3 (E)	**12** 2.7 (E)	$\sum_i$=24 **O**
		$\sum_j$=364 **O**	$\sum_j$=47 **O**	$\sum_t$=411

Prematuros: 1→semanas de gestación < 36;0→≥36.
Bajo peso: 1→< 2500 g; 0→ ≥2500 g.
O: observados; E: esperados

Prueba de chi-cuadrado de Karl-Pearson: Consiste en un contraste de frecuencias o contajes observados (O) con los teóricos o esperados

(E) representados en la tabla de contingencia de dimensiones 2x2 de la figura 3. Así, se define el

parámetro chi cuadrado de la siguiente forma:

χ

2= Σ(O-E)2

/E (Ec. 2.23.a)

Las frecuencias se ajustan a la distribución de Poisson, que tiene una propiedad poco habitual, la variancia coincide con la media. χ 2 , por lo tanto, es una razón de tipo señal-ruido entre el cuadrado de la diferencia entre las frecuencias observada y esperada respecto a la media esperada.

Tabla 9

Valores críticos de $\aleph^2$ (p = 0.05)

número g l	Valor crítico	número g l	Valor crítico
1	3.84	6	12.59
2	5.99	7	14.07
3	7.81	8	15.51
4	9.49	9	16.92
5	11.07	10	18.31

Tabla tomada de HR Neave. Elementary Statistics Tables. George Allen & Unwin Ltd

la prueba sólo se puede aplicar cuando el N.º total de observaciones es mayor de 50 y las frecuencias individuales esperadas no son menores de 5