

Universidad del Valle de Guatemala
Departamento de Matemática
Modelos Estadísticos I - CU109

Proyecto: Prueba t de Hipótesis

Integrantes del grupo:
David Coronado 07325
David Hsieh 08225
Carlos López 08107
Héctor Ruano 09063
Guatemala 17 de Mayo del 2010

Índice

1. Marco Teórico	2
2. Descripción del experimento	2
3. Procedimiento	3
4. Resultados	3
5. Análisis	5
6. Conclusiones	8
7. Recomendaciones	8

1. Marco Teórico

En el estudio de la estadística, sabemos como se pueden estimar parámetros tales como la media. Sin embargo, muchos problemas en Ingeniería y Ciencia requieren que se tome una decisión entre aceptar o rechazar una proposición sobre alguno de estos parámetros, a esta proposición le llamamos **hipótesis**[1].

Para todo tipo de investigación, se establece una hipótesis nula H_0 que se presume verdadera hasta que haya una evidencia estadística en la forma de una prueba de hipótesis que indique lo contrario i.e. rechaze H_0 ; para esto, se formula una o hipótesis alternativa H_a que difiera de H_0 .

Una prueba de hipótesis es aquella que permite, siguiendo los pasos adecuados, rechazar o aceptar una hipótesis. La prueba de hipótesis en la que fundamentamos nuestros resultados, es llamada prueba de t de Student [2][4], la cual está basada en una distribución de t, que no es más que una distribución de probabilidad que estima la media de una distribución normal [3], por lo tanto suponemos que nuestras poblaciones siguen una **distribución normal**.

Para rechazar o aceptar la hipótesis nula en una prueba de T, encontramos una región crítica en la distribución, la cual es el conjunto de eventos en las que, si ocurren, nos llevan a decidir o no si hay una diferencia, para encontrar la región crítica se usa el criterio de un intervalo de confianza. Un intervalo de confianza es un tipo de intervalo estimado de un parámetro poblacional. En vez de estimar el parámetro en un sólo valor, un intervalo es probable que contenga al valor, por lo tanto los intervalos de confianza los usamos para indicar la confiabilidad del valor estimado [5].

2. Descripción del experimento

Los estudiantes de la UVG tienen derecho a utilizar las instalaciones de la biblioteca. Existe la idea de que la biblioteca es visitada más por estudiantes mujeres que por estudiantes hombres. El fin de este experimento es comprobar dicha hipótesis midiendo el promedio de estudiantes de cada género que visitan la biblioteca de la UVG por unidad de tiempo y hacer las pruebas correspondientes.

3. Procedimiento

Se midió promedio de estudiantes que utilizan la biblioteca de la UVG por unidad de tiempo (5 minutos). La primera poblacion esta compuesta por los estudiantes hombres de la UVG y la otra poblacion por las estudiantes mujeres de la UVG.

Para realizar el experimento se empezó recolectando los datos. Los datos se tomaron manualmente por los miembros del grupo. Hubo un evento de recolección de datos en el cual dos miembros contaban a los estudiantes que ingresaban mientras que otro contaba los intervalos de tiempo. Este procedimiento duro hora y media (150 minutos) y la entrada de estudiantes era registrada en intervalos de 5 minutos. Por lo tanto, se tuvieron 30 muestras, las cuales se componen de la cantidad de estudiantes de cada población que entran a la biblioteca cada 5 minutos.

Luego estos datos fueron tomados para llevar a cabo el procedimiento de prueba de hipótesis, se realizó antes la Prueba F para ver si los datos tienen varianzas diferentes o iguales y se calculó el coeficiente de correlación para confirmar si los datos eran o no pareados.

4. Resultados

Durante la hora y media de observación, obtuvimos los siguientes resultados:

Número de muestra	Población 2: Cantidad de Hombres	Población 2: Cantidad de Mujeres
#1, minuto 0:	5	2
#2, minuto 10:	3	3
#3, minuto 15:	2	5
#4, minuto 20:	4	4
#5, minuto 25:	3	6
#6, minuto 30:	6	1
#7, minuto 35:	1	2
#8, minuto 40:	4	4
#9, minuto 45:	6	2
#10, minuto 50:	2	5
#11, minuto 55:	1	6
#12, minuto 60:	0	2
#13, minuto 65:	2	5
#14, minuto 70:	4	2
#15, minuto 75:	2	5
#16, minuto 80:	5	4
#17, minuto 85:	2	2
#18, minuto 90:	0	1
#19, minuto 95:	1	4
#20, minuto 100:	5	0
#21, minuto 105:	7	2
#22, minuto 110:	4	4
#23, minuto 115:	2	3
#24, minuto 120:	5	6
#25, minuto 125:	4	7
#26, minuto 130:	8	4
#27, minuto 135:	5	2
#28, minuto 140:	3	4
#29, minuto 145:	1	7
#30, minuto 150:	2	4

Durante una 1 hora y 30 minutos, entraron un total de 99 hombres y 108 mujeres, implicando que, las medias muestrales de cada población son:

Población 1 (hombres): 3.6 hombres por minuto

Población 2 (mujeres): 3.3 por minuto

5. Análisis

Para poder realizar la prueba de hipótesis t de Student, necesitamos saber si las poblaciones que estamos estudiando se comportan normalmente, es decir se pueden representar con una distribución normal de Gauss.

Por lo tanto, construimos la siguiente tabla

i	x_i Hombres	x_i Mujeres	F_{acm}	z
1	0	0	0.0167	-2.03
2	0	1	0.05	-1.64
3	1	1	0.0833	-1.38
4	1	2	0.1167	-1.20
5	1	2	0.15	-1.04
6	1	2	0.183	-0.91
7	2	2	0.2167	-0.78
8	2	2	0.25	-0.68
9	2	2	0.2833	-0.57
10	2	2	0.3167	-0.48
11	2	2	0.35	-0.39
12	2	2	0.383	-0.30
13	2	3	0.4167	-0.21
14	3	3	0.45	-0.13
15	3	4	0.483	-0.04
16	3	4	0.5167	0.04
17	4	4	0.55	0.13
18	4	4	0.583	0.21
19	4	4	0.6167	0.30
20	4	4	0.65	0.39
21	4	4	0.683	0.48
22	5	5	0.7167	0.57
23	5	5	0.75	0.68
24	5	5	0.783	0.78
25	5	5	0.8167	0.91
26	5	6	0.85	1.04
27	6	6	0.883	1.20
28	6	6	0.9167	1.38
29	7	7	0.95	1.64
30	8	7	0.9833	2.03

Se formula la hipótesis alternativa, en donde se dice que la media población de las mujeres que visitan la biblioteca por cada cinco minutos es mayor a la media poblacional de los hombres que visitan la biblioteca por cada cinco minutos.

$$H_a : \mu_M > \mu_H$$

Luego se formula la hipótesis nula, en la cual se dice lo contrario, junto a la igualdad.

$$H_0 : \mu_M \leq \mu_H$$

Se obtienen los parámetros, las medias muestrales de los hombres y las mujeres que visitan la biblioteca por cada cinco minutos.

$$\bar{X}_H = \frac{\sum_{i=1}^n X_{i_H}}{n} = \frac{5 + 3 + 2 + \dots + 2}{30} = 3.3$$

$$\bar{X}_M = \frac{\sum_{i=1}^n X_{i_M}}{n} = \frac{2 + 3 + 5 + \dots + 4}{30} = 3.6$$

$$S_H = \frac{\sum_{i=1}^n (X_{i_H} - \bar{X}_H)^2}{n - 1} = \frac{(5 - 3.3)^2}{30 - 1} + \frac{(3 - 3.3)^2}{30 - 1} + \dots + \frac{(2 - 3.3)^2}{30 - 1} = 1.830771$$

$$S_M = \frac{\sum_{i=1}^n (X_{i_M} - \bar{X}_M)^2}{n - 1} = \frac{(2 - 3.6)^2}{30 - 1} + \frac{(3 - 3.6)^2}{30 - 1} + \dots + \frac{(4 - 3.6)^2}{30 - 1} = 2.053592$$

$$f(n; \lambda) = \frac{\lambda^n e^{-\lambda}}{n!} = \frac{3.3^{30} * e^{-3.3}}{30!} =$$

Como $n = 30$, se tienen muestras pequeñas, por lo tanto se realiza la prueba t.

Debido a que no se conocen las varianzas, se realiza la prueba F para determinar si se puede asumir varianzas iguales.

$$F_0 = \frac{S_M^2}{S_H^2} = \frac{2.053592^2}{1.830771^2} = 1.25$$

$$F_{critica} = F_{n_M-1, n_H-1} = F_{29,29} = 1.90$$

$$1.25 < 1.90$$

F_0 se encuentra en la región de aceptación, por lo tanto se asumen las varianzas iguales.

$$\sigma_H^2 = \sigma_M^2$$

Entonces se prosigue a hacer la prueba con el estadístico de prueba:

$$t_0 = \frac{\bar{X}_M - \bar{X}_H - \Delta_0}{S_p \sqrt{\frac{1}{n_M} + \frac{1}{n_H}}}$$

Para ello, es necesario obtener el error de estimación S_p :

$$S_p = \sqrt{\frac{(n_M - 1)S_M^2 + (n_H - 1)S_H^2}{n_M + n_H - 2}} = \sqrt{\frac{(30 - 1)2.053592^2 + (30 - 1)1.830771^2}{30 + 30 - 2}} = 1.945374$$

Al en disposición el error de estimación, se prosigue a obtener el estadístico de prueba:

$$t_0 = \frac{3.6 - 3.3}{1.945374 \sqrt{\frac{1}{30} + \frac{1}{30}}} = 0.597260$$

$$t_{critico} = t_{\alpha, n_M + n_H - 2} = t_{0.05, 30 + 30 - 2} = t_{0.05, 58} = 1.684$$

$$t_0 < t_{critico}$$

Se puede observar, que el valor t_0 obtenido es menor que el valor $t_{critico}$, por lo que se encuentra en el área de aceptación. De ser así, se acepta la hipótesis nula.

Entonces se concluye que, **probablemente con $\alpha = 0.05$, la media poblacional de mujeres que visitan la biblioteca no es mayor a la media poblacional de hombres que visitan la biblioteca.**

6. Conclusiones

- No hay suficiente evidencia para concluir que la media de mujeres que visitan la biblioteca es mayor a la media de hombres que visitan la biblioteca.
- Se utilizó la Prueba t ya que la muestra es pequeña y no se conocen las varianzas de las muestras.
- A pesar de que la media muestral de las mujeres es mayor que la media muestral de los hombres, al realizar la prueba de hipótesis, se observó que no hay diferencia entre las dos muestras.

7. Recomendaciones

- Para obtener mejores resultados se recomienda que se haga el experimento con una muestra más grande, ya que la probabilidad de cometer un error de tipo I disminuye.
- Se recomienda variar las horas de recolección de datos para tener una información más general.

Referencias

- [1] Leticia De La Torre, Capítulo 2, Curso de Estadística I, Ingeniería Industrial, Instituto Tecnológico de Chihuahua, 2002, 1991.
- [2] *Student's t -test*, Wikipedia en Inglés, 16 de Mayo del 2010 http://en.wikipedia.org/wiki/Student's_t-test
- [3] *Student's t -distribution*, Wikipedia en Inglés, 16 de Mayo del 2010, http://en.wikipedia.org/wiki/Student's_t-distribution
- [4] Prueba t de Student, Wikipedia en Español, 16 de Mayo del 2010, http://es.wikipedia.org/wiki/Prueba_t_de_Student

- [5] *Confidence Interval*, Wikipedia en Inglés, 16 de Mayo del 2010,
http://en.wikipedia.org/wiki/Confidence_interval