

Página www

Página de Abertura

Contenido



Página 1 de 22

Regresar

Full Screen

Cerrar

Abandonar

Datos Categóricos: Clase 6

Juan Carlos Correa

23 de marzo de 2022

Punto 2 del primer parcial

La Encuesta de calidad de vida 2007 presenta información sobre si en un hogar tienen vehículo particular (0:No y 1: Sí), si tienen motocicleta y si tienen algún negocio en la casa además del estrato. La información se presenta en la siguiente tabla. Realice un análisis descriptivo.

Negocio en la casa			0	1
Estrato	VehíPar	Motocicleta		
1	0	0	1732	92
		1	101	5
	1	0	11	1
		1	1	0
2	0	0	5109	385
		1	466	47
	1	0	131	14
		1	30	5
3	0	0	3816	254
		1	446	58
	1	0	217	17
		1	35	7
4	0	0	1268	131
		1	148	6
	1	0	155	4
		1	31	2
5	0	0	94	3
		1	6	1
	1	0	919	50
		1	87	4
6	0	0	0	0
		1	0	0
	1	0	416	2
		1	31	3

Prueba LRT para $H_0 : \pi_1 = \pi_2$ vs. $H_1 : \pi_1 \neq \pi_2$ La prueba de la razón de verosimilitud para este problema la construimos así:

$$L(\pi_1, \pi_2) = \pi_1^a (1 - \pi_1)^b \pi_2^c (1 - \pi_2)^d$$

Bajo H_0 tenemos

$$L(\pi_1 = \pi_2 = \pi) = \pi^{a+c} (1 - \pi)^{b+d}$$

- El e.m.v. bajo el modelo restringido es $\hat{\pi} = (a + c)/N$.
- Los e.m.v. bajo el modelo irrestricto son $\hat{\pi}_1 = a/(a + b)$ y $\hat{\pi}_2 = c/(c + d)$.

Por lo tanto

$$G^2 = -2 \log(\lambda) = 2 \log \left(\frac{\hat{\pi}_1^a (1 - \hat{\pi}_1)^b \hat{\pi}_2^c (1 - \hat{\pi}_2)^d}{\hat{\pi}^{a+c} (1 - \hat{\pi})^{b+d}} \right) \sim \chi_{\dim(\Omega) - \dim(\omega)}^2$$

donde $\dim(\Omega) - \dim(\omega) = 2 - 1 = 1$.

La siguiente tabla presenta información sobre el uso de seda dental entre los estudiantes de primer semestre (2008) en UNal-Sede Medellín

	Sí usa	No usa
Hombres	334	318
Mujeres	248	117

Queremos ver si el uso de la seda dental es el mismo entre hombres y mujeres.

Página www

Página de Abertura

Contenido



Página 5 de 22

Regresar

Full Screen

Cerrar

Abandonar

Comparación de Proporciones de Dos Poblaciones

```
dif.prop.TLC<-function(x1,n1,x2,n2,nivel=0.05,bilateral=T){  
  if(x1==0) x1<-0.5  
  p1<-x1/n1  
  if(x2==0) x2<-0.5  
  p2<-x2/n2  
  n<-n1+n2  
  p.p<-(x1+x2)/n  
  z<-(p1-p2)/sqrt(p.p*(1-p.p)/n)  
  proba<-pnorm(abs(z),lower.tail=F)  
  if(bilateral==T) valor.p<-2*proba  
  else valor.p<-proba  
  dif.prop<-p1-p2  
  list(p1=p1,p2=p2,dif.prop=dif.prop,z=z,valor.p=valor.p)  
}
```

Página www

Página de Abertura

Contenido



Página 6 de 22

Regresar

Full Screen

Cerrar

Abandonar

```
dif.prop.LRT <-function(x1,n1,x2,n2,nivel=0.05){
  if(x1==0) x1<-0.5
  p1<-x1/n1
  if(x2==0) x2<-0.5
  p2<-x2/n2
  n<-n1+n2
  p.p<-(x1+x2)/n
  G2<-2*log(p1^x1*(1-p1)^(n1-x1)*p2^x2*(1-p2)^(n2-
x2)/(p.p^(x1+x2)*(1-p.p)^(n-x1-x2)))
  valor.p<-pchisq(G2,1,lower.tail=F)

  dif.prop<-p1-p2
  list(p1=p1,p2=p2,dif.prop=dif.prop,G2=G2,valor.p=valor.p)
}
```

```
> # Datos de odontología de estudiantes UNal (2008-I)
> # población 1: Hombres 334+318
> # población 2: Mujeres 248+117
> # Respuesta: Uso de seda dental Hombres 334, Mujeres 248
>
> dif.prop.TLC(334,652,248,365)
$p1
[1] 0.5122699

$p2
[1] 0.679452

$dif.prop
[1] -0.1671821

$z
[1] -10.77619

$valor.p
[1] 4.459911e-27
```

Página www

Página de Abertura

Contenido



Página 8 de 22

Regresar

Full Screen

Cerrar

Abandonar

```
> dif.prop.LRT(334,652,248,365)
```

```
$p1
```

```
[1] 0.5122699
```

```
$p2
```

```
[1] 0.679452
```

```
$dif.prop
```

```
[1] -0.1671821
```

```
$G2
```

```
[1] 27.15201
```

```
$valor.p
```

```
[1] 1.880696e-07
```


Página www

Página de Abertura

Contenido



Página 9 de 22

Regresar

Full Screen

Cerrar

Abandonar

La siguiente tabla presenta el uso de cepillo dental por parte de los primíparos I-2008 UNal-Sede Medellín.

	Sí usa cepillo	No usa
Hombres	649	4
Mujeres	364	1

Página www

Página de Abertura

Contenido

◀

▶

◀

▶

Página 10 de 22

Regresar

Full Screen

Cerrar

Abandonar

```
> p1<-4/653
> p2<-1/365
>
> muestra.hombres<-rbinom(10000,653,p1)
> muestra.mujeres<-rbinom(10000,365,p2)
> dife<-muestra.hombres/653-muestra.mujeres/365
> summary(dife)
      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
-0.0149100  0.0006461  0.0033860  0.0034210  0.0061260  0.0199100
> length(dife[dife>0])/length(dife)
[1] 0.8149
> hist(dife,main='p1-p2',ylab='Densidad',xlab='Diferencia')

> quantile(dife,probs=c(0.025,0.975))
      2.5%      97.5%
-0.005156391  0.011042816
```

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



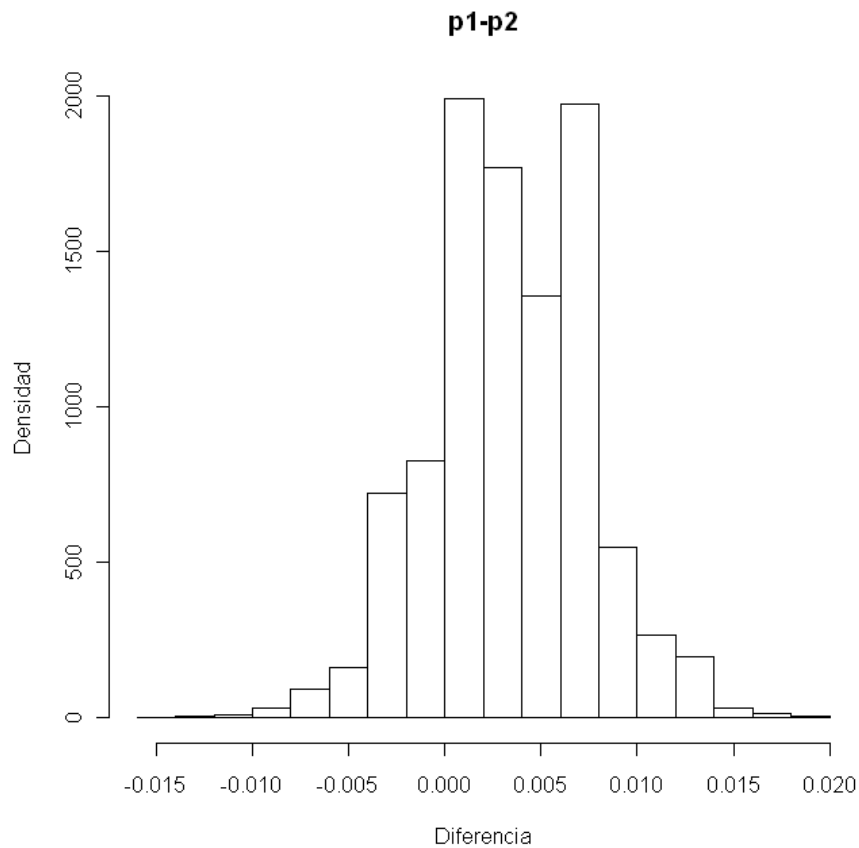
[Página 11 de 22](#)

[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)



Muestreando sólo con N fijo

El modelo de probabilidad observado es

$$P(A = a, B = b, C = c) = \frac{N!}{a!b!c!d!} \pi_{11}^a \pi_{12}^b \pi_{21}^c \pi_{22}^d$$

donde $d = N - a - b - c$ y π_{ij} es la probabilidad de la (i, j) -ésima celda, $i, j = 1, 2$.

La hipótesis de interés corriente es

H_0 : Independencia de las dos respuestas ó

H_0 : $\pi_{ij} = \pi_{i+}\pi_{+j}$, $i, j = 1, 2$.

Página www

Página de Abertura

Contenido



Página 13 de 22

Regresar

Full Screen

Cerrar

Abandonar

La siguiente tabla presenta información sobre el problema de la onicofagia (comerse las uñas) y el de morder cosas entre primiparos I-2008 UNal-Sde Medellín

	Morder Cosas	
	SI	NO
Onicofagia SI	155	91
NO	293	434

```
> chisq.test(matrix(c(155,91,293,434),ncol=2))
```

Pearson's Chi-squared test with Yates' continuity correction

```
data:  matrix(c(155, 91, 293, 434), ncol = 2)
X-squared = 37.2339, df = 1, p-value = 1.048e-09
```

```
> chisq.test(matrix(c(155,91,293,434),ncol=2),simulate.p.value=T)
```

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

```
data:  matrix(c(155, 91, 293, 434), ncol = 2)
X-squared = 38.1424, df = NA, p-value = 0.0004998
```

Prueba de hipótesis Monte Carlo

Las pruebas de hipótesis vía Monte Carlo no dependen de resultados asintóticos y algunos las consideran pruebas exactas. Vamos a ilustrar el procedimiento en un esquema de muestreo multinomial.

Clasificación II				
		1	2	Total
Clasificación I	1	n_{11}	n_{12}	n_{1+}
	2	n_{21}	n_{22}	n_{2+}
Total		n_{+1}	n_{+2}	n_{++}

la hipótesis a verificar es

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j} \quad \forall i, j$$

Algoritmo

1. Genere una muestra $(n_{11}^*, n_{12}^*, n_{21}^*, n_{22}^*)$ de una distribución multinomial

$$MN \left(n_{++}, \frac{n_{1+}n_{+1}}{n_{++}}, \frac{n_{1+}n_{+2}}{n_{++}}, \frac{n_{2+}n_{+1}}{n_{++}}, \frac{n_{2+}n_{+2}}{n_{++}}, \right)$$

2. Calcule el estadístico χ^2 para la muestra anterior

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij}^* - e_{ij}^*)^2}{e_{ij}^*}$$

3. Repita los dos pasos anteriores M veces, digamos $M = 1000$, y calcule la distribución empírica del χ^2 . Calcule el percentil α superior, digamos χ_α^2 .
4. Para los datos originales calcule χ_c^2 . Si Este es valor es mayor que el valor crítico hallado en el punto anterior, rechace H_0 .


```
# Datos de primiparos
# Morder Objetos
# Si No
# 155 91 Si Onicofagia
# 283 434 No

tabla<-matrix(c(155,91,283,434),ncol=2,byrow=T)

# Prueba bootstrap

prueba.boot.ind<-function(tabla,nboot=10000){
  n<-sum(tabla)
  proba<-tabla/n
  marg1<-colSums(proba)
  marg2<-rowSums(proba)
  prob.ho<-marg1%*%t(marg2)
  res<-rmultinom(nboot,n,c(prob.ho[1,],prob.ho[2,]))
  calc.chi<-function(x)chisq.test(matrix(x,ncol=2,byrow=T))$statistic
  res<-apply(res,2,calc.chi)
  chi.obs<-chisq.test(tabla)$statistic
  valor.p<-sum(ifelse(res>=chi.obs,1,0))/length(res)
  plot(density(res,from=0),main='Distribución Nula Bootstrap',ylab='Densidad')
  abline(v=0)
  list(chi.obs=chi.obs,valor.p=valor.p)
}
```

Página www

Página de Abertura

Contenido



Página 18 de 22

Regresar

Full Screen

Cerrar

Abandonar

```
prueba.boot.ind(tabla)
```

```
$chi.obs
```

```
X-squared
```

```
39.98141
```

```
$valor.p
```

```
[1] 0
```

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



[Página 19 de 22](#)

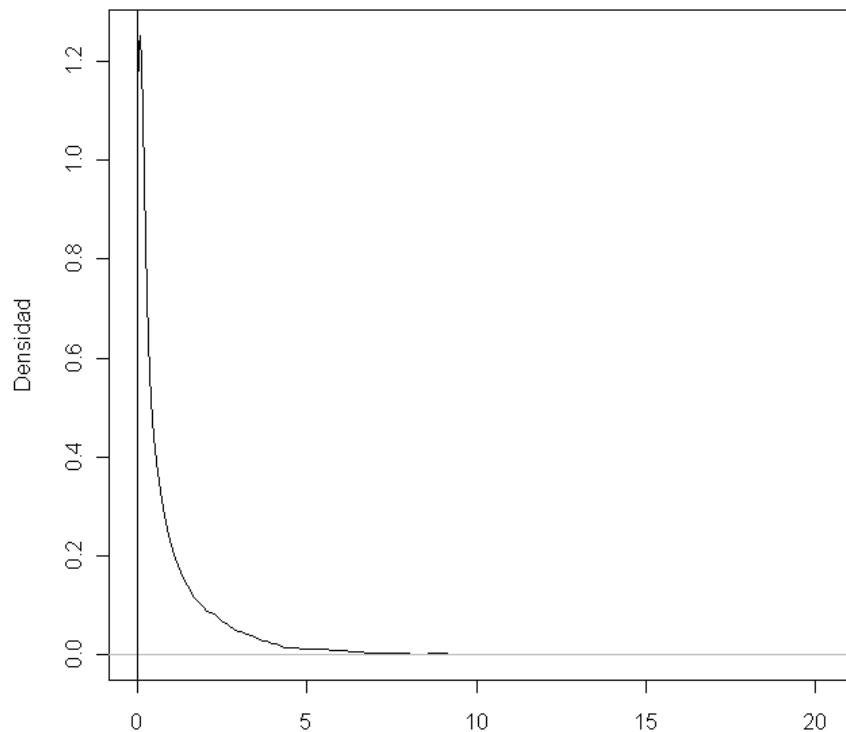
[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)

Distribución Nula Bootstrap



N = 10000 Bandwidth = 0.1191

```

# Prueba bootstrap mejorada
# Se tiene en cuenta la aleatoriedad del estimador
prueba.boot.ind.mej<-function(tabla,nboot=10000){
  n<-sum(tabla)
  proba<-tabla/n
  marg1<-colSums(proba)
  # Recuerde que p.gorro~N(pi,pi(1-p1)/n)
  pi.1m<-rnorm(nboot,mean=marg1[1],sd=sqrt(marg1[1]*(1-marg1[1])/n))
  marg2<-rowSums(proba)
  pi.m1<-rnorm(nboot,mean=marg2[1],sd=sqrt(marg2[1]*(1-marg2[1])/n))
  prob.multi<-cbind(pi.1m*pi.m1,pi.1m*(1-pi.m1),(1-pi.1m)*pi.m1,(1-pi.1m)*(1-pi.m1))

  genera.multi<-function(pis,n) rmultinom(1,n,pis)
  res<-matrix(unlist(apply(prob.multi,1,genera.multi,n)),ncol=4,byrow=T)

  calc.chi<-function(x)chisq.test(matrix(x,ncol=2,byrow=T))$statistic
  res<-apply(res,1,calc.chi)
  chi.obs<-chisq.test(tabla)$statistic
  valor.p<-sum(ifelse(res>=chi.obs,1,0))/length(res)
  plot(density(res,from=0),main='Distribución Nula Bootstrap',ylab='Densidad')
  abline(v=0)
  list(chi.obs=chi.obs,valor.p=valor.p)
}

```

Página www

Página de Abertura

Contenido



Página 21 de 22

Regresar

Full Screen

Cerrar

Abandonar

```
> prueba.boot.ind.mej(tabla)
```

```
$chi.obs
```

```
X-squared
```

```
39.98141
```

```
$valor.p
```

```
[1] 0
```

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



[Página 22 de 22](#)

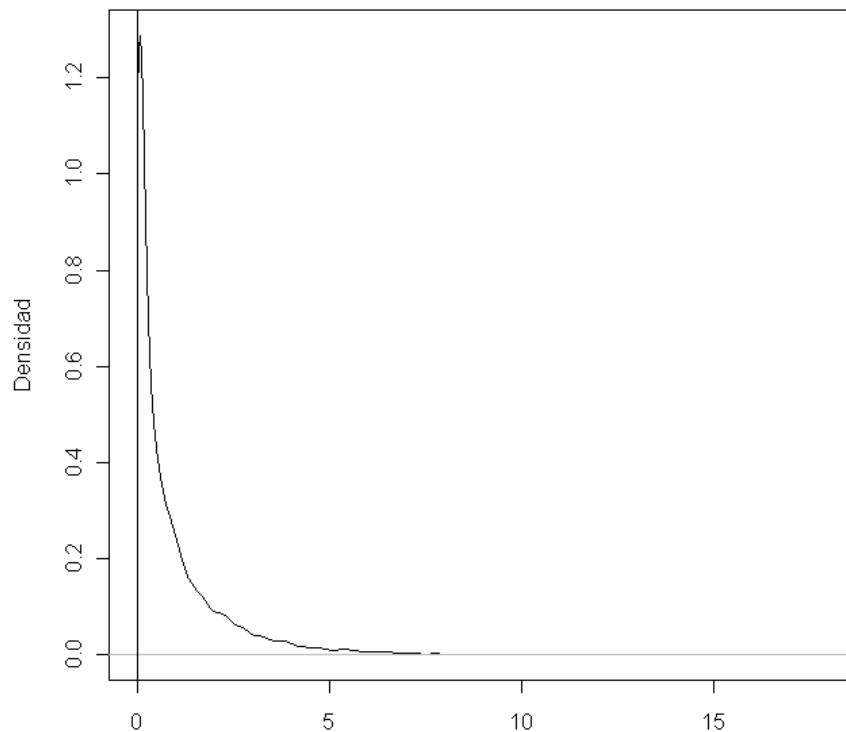
[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)

Distribución Nula Bootstrap



N = 10000 Bandwidth = 0.1142