

Unidad 4 - Actividad 4C. Inferencia para datos categóricos.

Alumno: **Luis Fernando Izquierdo Berdugo**

Materia: **Estadística**

Fecha de Entrega: **10 de Octubre de 2024**

Realiza los siguientes ejercicios del capítulo 6 Inferencia de datos numéricos.

Página	Ejercicios
316	6.16
317	6.23
323	6.43
325	6.49
327	6.53

6.16 Is college worth it? Part I. Among a simple random sample of 331 American adults who do not have a four-year college degree and are not currently enrolled in school, 48% said they decided not to go to college because they could not afford school.

(a) A newspaper article states that only a minority of the Americans who decide not to go to college do so because they cannot afford it and uses the point estimate from this survey as evidence. Conduct a hypothesis test to determine if these data provide strong evidence supporting this statement.

La hipótesis nula H_0 será que la proporción de estadounidenses que deciden no ir a la universidad por no poder pagarla es igual o mayor a 0.5

$$H_0: p \geq 0.5$$

La hipótesis alternativa H_1 será que la proporción de estadounidenses que deciden no ir a la universidad por no poder pagarla es menor a 0.5

$$H_1: p < 0.5$$

Se usará la prueba de proporción Z:

$$z = (\hat{p} - p) / \sqrt{(p(1 - p)/n)}$$

donde:

- \hat{p} es la proporción muestral (0.48)
- p es la proporción bajo la hipótesis nula (0.5)
- n es el tamaño de la muestra (331)

Ejecutándose en código de R:

```
> phat = 0.48  
> p1 = .5
```

```
> n = 331
> z <- (phat-p1)/sqrt(p1*(1-p1)/n)
> z
[1] -0.7277362
```

Se calcula el valor de p utilizando la función pnorm:

```
> p_value <- pnorm(z, lower.tail = TRUE)
> p_value
[1] 0.2333875
```

Debido a que el valor de z es menor que el valor crítico de -1.64 y el valor de p es menor que 0.5, se rechaza la hipótesis nula y se acepta la alternativa, entonces la proporción de estadounidenses que deciden no ir a la universidad por no poder pagarla es menor a 0.5.

(b) Would you expect a confidence interval for the proportion of American adults who decide not to go to college because they cannot afford it to include 0.5? Explain.

Debido a que el valor de p calculado no es tan pequeño, no hay una evidencia demasiado contundente contra la hipótesis nula, por lo cual, un intervalo de confianza del 95% probablemente incluya el valor de 0.5.

6.23 Social experiment, Part I. A “social experiment” conducted by a TV program questioned what people do when they see a very obviously bruised woman getting picked on by her boyfriend. On two different occasions at the same restaurant, the same couple was depicted. In one scenario the woman was dressed “provocatively” and in the other scenario the woman was dressed “conservatively”. The table below shows how many restaurant diners were present under each scenario, and whether or not they intervened.

		<i>Scenario</i>		Total
		Provocative	Conservative	
<i>Intervene</i>	Yes	5	15	20
	No	15	10	25
	Total	20	25	45

Explain why the sampling distribution of the difference between the proportions of interventions under provocative and conservative scenarios does not follow an approximately normal distribution.

A primera vista, se puede observar que la ejecución del experimento sugiere que las observaciones no fueron independientes, al tomar a los presentes en un lugar público, las reacciones de algunas personas pudieron influenciar a las demás, con esto la distribución del modelo normal no puede ser aplicada. De igual manera, se nota que las muestras son relativamente pequeñas. La muestra del escenario “provocativo” no cumple con la condición de éxito-fracaso con proporción nula, viendo que su valor np_0 no llega a ser mayor a diez.

6.43 Rock-paper-scissors. Rock-paper-scissors is a hand game played by two or more people where players choose to sign either rock, paper, or scissors with their hands. For your statistics class project, you want to evaluate whether players choose between these three options randomly, or if certain options are favored above others. You ask two friends to play rock-paper-scissors and count the times each option is played. The following table summarizes the data:

Rock	Paper	Scissors
43	21	35

Use these data to evaluate whether players choose between these three options randomly, or if certain options are favored above others. Make sure to clearly outline each step of your analysis, and interpret your results in context of the data and the research question.

Hipótesis nula H_0 : Los jugadores eligen sus opciones de manera aleatoria.

Hipótesis alternativa H_1 : Los jugadores tienen una preferencia por alguna de las opciones.

Se utilizará una prueba de chi-cuadrado de bondad de ajuste, la cual compara la distribución observada de las frecuencias con una distribución esperada bajo la hipótesis nula de aleatoriedad.

```
> total_observaciones <- 43 + 21 + 35
> frecuencias_esperadas <- rep(total_observaciones / 3, 3)
> frecuencias_observadas <- c(43, 21, 35)
> estadistico_chi2 <- sum((frecuencias_observadas -
frecuencias_esperadas)^2 / frecuencias_esperadas)
> grados_libertad <- 2
> valor_p <- 1 - pchisq(estadistico_chi2, grados_libertad)
> valor_p
[1] 0.02334025
```

Debido a que el valor de p es menor que el nivel de significancia 0.05 se rechaza la hipótesis nula y se concluye que existe evidencia suficiente para afirmar que los jugadores tienen una preferencia por alguna de las opciones.

6.49 Shipping holiday gifts. A December 2010 survey asked 500 randomly sampled Los Angeles residents which shipping carrier they prefer to use for shipping holiday gifts. The table below shows the distribution of responses by age group as well as the expected counts for each cell (shown in parentheses).

		Age						Total
		18-34		35-54		55+		
Shipping Method	USPS	72	(81)	97	(102)	76	(62)	245
	UPS	52	(53)	76	(68)	34	(41)	162
	FedEx	31	(21)	24	(27)	9	(16)	64
	Something else	7	(5)	6	(7)	3	(4)	16
	Not sure	3	(5)	6	(5)	4	(3)	13
Total		165		209		126		500

(a) State the null and alternative hypotheses for testing for independence of age and preferred shipping method for holiday gifts among Los Angeles residents.

Hipótesis nula H0: Edad y método de envío son independientes.

Hipótesis alternativa H1: Edad y método de envío no son independientes.

(b) Are the conditions for inference using a chi-square test satisfied?

Las condiciones para la prueba de chi-cuadrado son dos:

1. Las observaciones deben ser independientes entre sí.
2. Se espera que cada escenario tenga al menos 5 casos esperados

Debido a que el tamaño de la muestra es grande y se especifica que se seleccionaron de manera aleatoria, se puede decir que la primera condición se cumple, sin embargo, se puede observar que en la columna de Edad mayor a 55 años, nos encontramos con frecuencias menores que 5, por lo cual la segunda condición no se cumple, por lo cual, las condiciones para inferencia de la prueba de chi-cuadrado no se satisfacen.

6.53 The Egyptian Revolution. A popular uprising that started on January 25, 2011 in Egypt led to the 2011 Egyptian Revolution. Polls show that about 69% of American adults followed the news about the political crisis and demonstrations in Egypt closely during the first couple weeks following the start of the uprising. Among a random sample of 30 high school students, it was found that only 17 of them followed the news about Egypt closely during this time.

(a) Write the hypotheses for testing if the proportion of high school students who followed the news about Egypt is different than the proportion of American adults who did.

Hipótesis nula H0: La proporción de estudiantes que siguieron las noticias de Egipto es igual a la proporción de adultos estadounidenses que siguieron las noticias

Hipótesis alternativa H1: La proporción de estudiantes que siguieron las noticias de Egipto no es igual a la proporción de adultos estadounidenses que siguieron las noticias

(b) Calculate the proportion of high schoolers in this sample who followed the news about Egypt closely during this time.

$$\text{Proporción} = 17/30 = 0.567$$

(c) Based on large sample theory, we modeled p using the normal distribution. Why should we be cautious about this approach for these data?

Debido al tamaño de la muestra, como esta es pequeña, podría no ser representativa para la población y esto haría que la distribución normal no sea confiable.

(d) The normal approximation will not be as reliable as a simulation, especially for a sample of this size. Describe how to perform such a simulation and, once you had results, how to estimate the p-value.

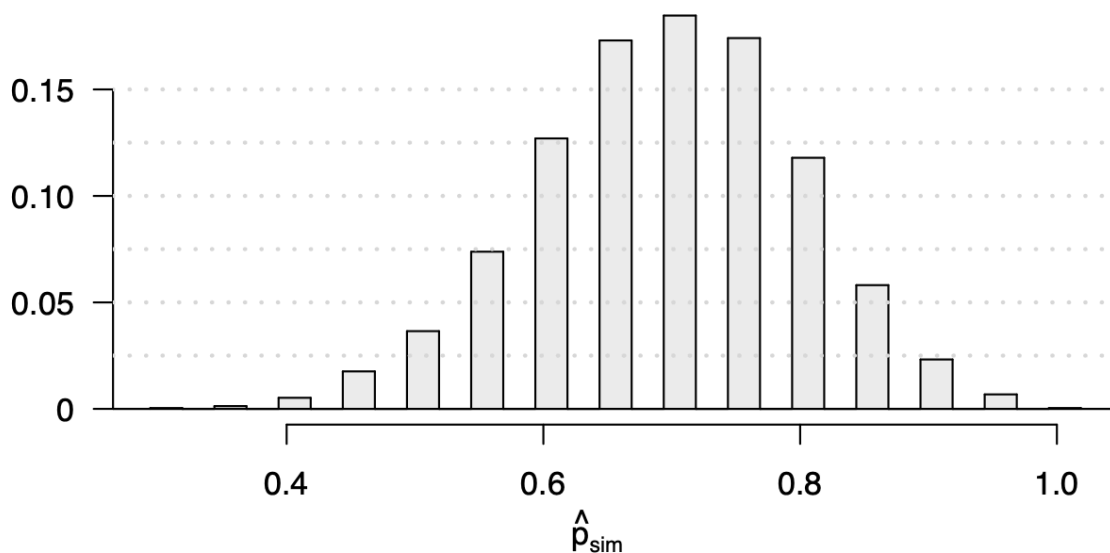
Para la simulación se tendrían que seguir los siguientes pasos:

1. Generación de muestras aleatorias: se simulan muchas cantidades de muestras de tamaño de 30 o más de una población igual a la proporción de adultos (0.69)
2. Cálculo de proporciones: Se calcula la proporción de estudiantes para cada muestra generada previamente.

Para la estimación del valor p :

1. Identificación de valores en la población simulada que sean más extremos que la proporción observada originalmente.
2. Determinar la proporción de simulaciones que resultaron en proporciones más extremas que la observada originalmente.

(e) Below is a histogram showing the distribution of \hat{p}_{sim} in 10,000 simulations under the null hypothesis. Estimate the p-value using the plot and determine the conclusion of the hypothesis test.



El valor p es la probabilidad de observar una proporción muestra tan o más extrema que la observada (0.567). Para estimar el valor p con el histograma, se deben identificar la proporciones de simulaciones que dieron como resultado \hat{p}_{sim} igual o más extremo que 0.567

Del histograma, se observa que hay aproximadamente 1000 simulaciones que cayeron en 0.567. Entonces, el valor p será:

$$p_{value} = 1000/10000 = 0.1$$

Como el valor p es mayor que 0.05, se acepta la hipótesis nula, lo cual significa que no hay evidencia suficiente para concluir que la proporción de adultos estadounidenses que siguieron las noticias de Egipto es diferente de la proporción de estudiantes que siguieron las noticias.