

Taller 10

Universidad Externado de Colombia

Departamento de Matemáticas

Estadística 2

Juan Sosa, Ph. D.

June 28, 2018

Brecha salarial

Con los datos dados en `SALHM.txt` correspondientes a los salarios mensuales (en millones de pesos) de una compañía del Estado.

1. Calcular e interpretar un intervalo de confianza para el cociente de varianzas de los salarios entre hombres y mujeres. ¿Existen diferencias significativas?
2. Calcular e interpretar un intervalo de confianza para la diferencia de medias poblacional de los salarios entre hombres y mujeres. ¿Existen diferencias significativas?
3. Graficar un histograma y un qqplot normal para los salarios de los hombres. Probar el sistema de hipótesis correspondiente usando la prueba de Shapiro-Wilk. ¿Los salarios de los hombres parecen provenir de una distribución normal?
4. Repetir el numeral anterior para el salario de las mujeres.
5. ¿El supuesto de normalidad de las poblaciones se satisface? ¿Es absolutamente necesario que se satisfaga dados los tamaños de muestra?

Razonamiento cuantitativo

Considerar la base de datos dada en `saber11 2015 1.csv` correspondiente a la información oficial de las pruebas saber 11 correspondientes al primer semestre del año 2015. Esta base de datos consta de 965 instituciones educativas y 16 variables. Las variables del conjunto de datos son: código del municipio; nombre del municipio; departamento; calendario; naturaleza; jornada; número de estudiantes evaluados; calificación promedio de la institución en: lectura crítica, matemáticas, sociales, ciencias naturales, inglés, razonamiento cuantitativo, competencias ciudadanas. Utilizando un 95% de confianza:

1. Calcular e interpretar un intervalo de confianza para el cociente de varianzas del promedio en matemáticas entre colegios oficiales y no oficiales. ¿Existen diferencias significativas?
2. Calcular e interpretar un intervalo de confianza para la diferencia de medias poblacional del promedio en matemáticas entre colegio oficiales y no oficiales. ¿Existen diferencias significativas?
3. Graficar un histograma y un qqplot normal para el promedio en matemáticas de los colegios oficiales. Probar el sistema de hipótesis correspondiente usando la prueba de Shapiro-Wilk. ¿El promedio en matemáticas de los colegios oficiales parecen provenir de una distribución normal?
4. Repetir el numeral anterior para el promedio en matemáticas de los colegios no oficiales.
5. ¿El supuesto de normalidad de las poblaciones se satisface? ¿Es absolutamente necesario que se satisfaga dados los tamaños de muestra?

Prueba chi cuadrado de bondad de ajuste

1. De Keller (2014), leer y sintetizar la Sección 15.1 (*Chi-squared goodness-of-fit test*, p. 592).
2. (Keller 2012, problem 15.14) In an election held last year that was contested by three parties. Party A captured 31% of the vote, party B garnered 51%, and party C received the remaining votes. A survey of

1,200 voters asked each to identify the party that they would vote for in the next election. These results were recorded where 1 = party A, 2 = party B, and 3 = party C. Can we infer at the 10% significance level that voter support has changed since the election?. Data set available in `Xr15-14.xlsx`.

3. (*Keller 2012, problem 15.17*) According to the Statistical Abstract of the United States 2012, Table 56, the proportions for each category of marital status in 2007 was: Never married (including partnered, not married) 26.9% Married (including separated, but not divorced) 56.6% Widowed 6.2% Divorced 10.3% Can we infer that the American National Election Survey in 2012 overrepresented at least one category of marital status (MARITAL)? Data set available in `ANES2008.xlsx`.
4. (*Anderson 2011, Chap. 12.2, problem 7*) The Wall Street Journal's Shareholder Scoreboard tracks the performance of 1000 major U.S. companies (The Wall Street Journal, March 10, 2003). The performance of each company is rated based on the annual total return, including stock price changes and the reinvestment of dividends. Ratings are assigned by dividing all 1000 companies into five groups from A (top 20%), B (next 20%), to E (bottom 20%). Shown here are the one-year ratings for a sample of 60 of the largest companies. Do the largest companies differ in performance from the performance of the 1000 companies in the Shareholder Scoreboard? Use $\alpha = 0.05$.

A	B	C	D	E
5%	8%	15%	20%	12%

Prueba chi cuadrado de independencia

1. De Keller (2014), leer y sintetizar la Sección 15.2 (*Chi-squared test of a contingency table*, p. 599).
2. (*Keller 2012, problem 15.37*) The relationship between drug companies and medical researchers is under scrutiny because of possible conflict of interest. The issue that started the controversy was a 1995 case control study that suggested that the use of calcium-channel blockers to treat hypertension led to an increase risk of heart disease. This led to an

intense debate both in technical journals and in the press. Researchers writing in the New England Journal of Medicine (Conflict of Interest in the Debate over Calcium Channel Antagonists, January 8, 1998, p. 101) looked at the 70 reports that appeared during 1996–1997, classifying them as favorable, neutral, or critical toward the drugs. The researchers then contacted the authors of the reports and questioned them about financial ties to drug companies. The results were recorded in the following way: Column 1: Results of the scientific study; 1 = favorable, 2 = neutral, 3 = critical Column 2: 1 = financial ties to drug companies, 2 = no ties to drug companies Do these data allow us to infer that the research findings for calcium-channel blockers are affected by whether the research is funded by drug companies? Data set available in `Xr15-37.xlsx`.

3. (*Anderson 2011, Chap. 12, problem 34*) A Pew Research Center survey asked respondents if they would rather live in a place with a slower pace of life or a place with a faster pace of life (USA Today, February 13, 2009). Consider the following data showing a sample of preferences expressed by 150 men and 150 women.

Respondent	Preferred Pace of Life		
	Slower	No Preference	Faster
Men	102	9	39
Women	111	12	27

- (a) Combine the samples of men and women. What is the overall percentage of respondents who prefer to live in a place with a slower pace of life? What is the overall percentage of respondents who prefer to live in a place with a faster pace of life? What is your conclusion?
- (b) Is the preferred pace of life independent of the respondent? Use $\alpha = 0.05$. What is your conclusion? What is your recommendation?

A Bipartisan Agenda for Change

(*Anderson 2011, Chap. 12, Case Problem*) In a study conducted by Zogby International for the Democrat and Chronicle, more than 700 New Yorkers were

polled to determine whether the New York state government works. Respondents surveyed were asked questions involving pay cuts for state legislators, restrictions on lobbyists, term limits for legislators, and whether state citizens should be able to put matters directly on the state ballot for a vote (Democrat and Chronicle, December 7, 1997). The results regarding several proposed reforms had broad support, crossing all demographic and political lines. Suppose that a follow-up survey of 100 individuals who live in the western region of New York was conducted. The party affiliation (Democrat, Independent, Republican) of each individual surveyed was recorded, as well as their responses to the following three questions.

- Should legislative pay be cut for every day the state budget is late?
- Should there be more restrictions on lobbyists?
- Should there be term limits requiring that legislators serve a fixed number of years?

The responses were coded using 1 for a Yes response and 2 for a No response. The complete data set is available in the file named NYReform.

1. Use descriptive statistics to summarize the data from this study. What are your preliminary conclusions about the independence of the response (Yes or No) and party affiliation for each of the three questions in the survey?
2. With regard to question 1, test for the independence of the response (Yes and No) and party affiliation. Use $\alpha = 0.05$.
3. With regard to question 2, test for the independence of the response (Yes and No) and party affiliation. Use $\alpha = 0.05$.
4. With regard to question 3, test for the independence of the response (Yes and No) and party affiliation. Use $\alpha = 0.05$.
5. Does it appear that there is broad support for change across all political lines? Explain.