Taller 09

Universidad Externado de Colombia

Departamento de Matemáticas

Estadística 2

Juan Sosa, Ph. D.

June 27, 2018

Inferencia sobre la varianza poblacional

1. De Keller (2014), leer y sintetizar la Sección 12.2 (*Inference about a population variance*, p. 401).

2. (*Keller 2012, problem 12.77*) The job placement service at a university observed the not unexpected result of the variance in marks and work experience of the university's graduates: Some graduates received numerous offers whereas others received far fewer. To learn more about the problem, a survey of 90 recent graduates was conducted wherein each was asked how many job offers they received. Estimate with 90% confidence the variance in the number of job offers made to the university's graduates. Data set available in `Xr12-77.xlsx`.

3. (*Anderson 2011, Chap. 11, problem 8*) March 4, 2009, was one of the few good days for the stock market in early 2009. The Dow Jones Industrial Average went up 149.82 points (The Wall Street Journal, March 5, 2009). The following table shows the stock price changes for a sample of 12 companies on that day. Data set available in `PriceChange.csv`.

   (a) Compute the sample variance for the daily price change.

(b) Compute the sample standard deviation for the price change.

(c) Provide 95% confidence interval estimates of the population variance and the population standard deviation.

4. (*Anderson 2011, Chap. 11, problem 10*) The average standard deviation for the annual return of large cap stock mutual funds is 18.2% (The Top Mutual Funds, AAII, 2004). The sample standard deviation based on a sample of size 36 for the Vanguard PRIMECAP mutual fund is 22.2%. Construct a hypothesis test that can be used to determine whether the standard deviation for the Vanguard fund is greater than the average standard deviation for large cap mutual funds. With a .05 level of significance, what is your conclusion?

Inferencia sobre la diferencia de proporciones poblacional

1. De Berenson et al. (2012), leer y sintetizar la Sección 10.3 (*Comparing the Proportions of Two Independent Populations*, p. 385).

2. (*Bereson et al. 2012, problem 10.35*) Where people turn for news is different for various age groups. (Data extracted from Cellphone Users Who Access News on Their Phones, USA Today, March 1, 2010, p. 1A.) A study was conducted on the use of cell phones for accessing news. The study reported that 47% of users under age 50 and 15% of users age 50 and over accessed news on their cell phones. Suppose that the survey consisted of 1,000 users under age 50, of whom 470 accessed news on their cell phones, and 891 users age 50 and over, of whom 134 accessed news on their cell phones.

   (a) Is there evidence of a significant difference in the proportion of users under age 50 and users 50 years and older that accessed the news on their cell phones? (Use $\alpha = 0.05$).

   (b) Determine the p-value in (a) and interpret its meaning.

   (c) Construct and interpret a 95% confidence interval estimate for the difference between the population proportion of users under 50 years old and those 50 years or older who access the news on their cell phones.

Inferencia sobre la diferencia de medias poblacional

1. De Berenson et al. (2012), leer y sintetizar la Sección 10.1 (*Comparing the Means of Two Independent Populations*, p. 366).

2. (*Bereson et al. 2012, problem 10.12*) A bank with a branch located in a commercial district of a city has the business objective of developing an improved process for serving customers during the noon-to- 1 P.M. lunch period. Management decides to first study the waiting time in the current process. The waiting time is defined as the time that elapses from when the customer enters the line until he or she reaches the teller window. Data are collected from a random sample of 15 customers, and the results (in minutes) are as follows (and stored in `Bank1.xlsx`): Suppose that another branch, located in a residential area, is also concerned with improving the process of serving customers in the noon-to-1 P.M. lunch period. Data are collected from a random sample of 15 customers, and the results are as follows (and stored in `Bank2.xlsx`):

   (a) Is the population variances from both banks equal or different? (Use $\alpha = 0.05$).

   (b) Is there evidence of a difference in the mean waiting time between the two branches? (Use $\alpha = 0.05$).

   (c) Determine the p-value in (b) and interpret its meaning.

   (d) What assumption is necessary in (b)? Test this assumption using $\alpha = 0.05$.

   (e) Construct and interpret a 95% confidence interval estimate of the difference between the population means in the two branches.

3. *(Anderson et al. 2011, Chap. 10, problem 42)* In early 2009, the economy was experiencing a recession. But how was the recession affecting the stock market? Shown are data from a sample of 15 companies. Shown for each company is the price per share of stock on January 1 and April 30 (The Wall Street Journal, May 1, 2009). Data set available in `PriceChange2.csv`.

   (a) What is the change in the mean price per share of stock over the four-month period?

(b) Provide a 90% confident interval estimate of the change in the mean price per share of stock. Interpret the results.

(c) What was the percentage change in the mean price per share of stock over the fourmonth period?

(d) If this same percentage change were to occur for the next four months and again for the four months after that, what would be the mean price per share of stock at the end of the year 2009?

Muestras pareadas

1. De Berenson et al. (2012), leer y sintetizar la Sección 10.2 (*Comparing the Means of Two Related Populations*, p. 377).

2. (*Bereson et al. 2012, problem 10.22*) Is there a difference in the prices at a warehouse club such as Costco and store brands? To investigate this, a random sample of 10 purchases was selected, and the prices were compared. (Data extracted from Shop Smart and Save Big, Consumer Reports, May 2009, p. 17.) The prices for the products are stored in `Shopping1.xlsx`.

   (a) At the 0.05 level of significance, is there evidence of a difference between the mean price of Costco purchases and store-brand purchases?

   (b) What assumption is necessary about the population distribution in order to perform this test?

   (c) Perform a simulation using normal qqplots to validate this assumption. Also, test this assumption using the Shapiro-wilk normality test.

   (d) Construct a 95% confidence interval estimate of the mean difference in price between Costco and store brands. Interpret the interval.

   (e) Compare the results of the test and the confidence interval.

3. *(Anderson et al. 2011, Sec. 10.3, problem 23)* Bank of America's Consumer Spending Survey collected data on annual credit card charges in seven different categories of expenditures: transportation, groceries,

dining out, household expenses, home furnishings, apparel, and entertainment (US Airways Attaché, December 2003). Using data from a sample of 42 credit card accounts, assume that each account was used to identify the annual credit card charges for groceries (population 1) and the annual credit card charges for dining out (population 2). Using the difference data, the sample mean difference was \$850, and the sample standard deviation was \$1123.

(a) Formulate the null and alternative hypotheses to test for no difference between the population mean credit card charges for groceries and the population mean credit card charges for dining out.

(b) Use a .05 level of significance. Can you conclude that the population means differ? What is the p-value?

(c) Which category, groceries or dining out, has a higher population mean annual credit card charge? What is the point estimate of the difference between the population means? What is the 95% confidence interval estimate of the difference between the population means?

Inferencia sobre el cociente de varianzas poblacional

1. De Anderson et al. (2011), leer y sintetizar la Sección 11.2 (*Inferences About Two Population Variances*, p. 460).

2. *(Anderson et al. 2011, Chap. 11, problem 18)* Data were collected on the top 1000 financial advisers by Barron's (Barron's, February 9, 2009). Merrill Lynch had 239 people on the list and Morgan Stanley had 121 people on the list. A sample of 16 of the Merrill Lynch advisers and 10 of the Morgan Stanley advisers showed that the advisers managed many very large accounts with a large variance in the total amount of funds managed. The standard deviation of the amount managed by the Merrill Lynch advisers was \$587 million. The standard deviation of the amount managed by the Morgan Stanley advisers was \$489 million. Conduct a hypothesis test at $\alpha = .10$ to determine if there is a significant difference in the population variances for the amounts managed by the

two companies. What is your conclusion about the variability in the amount of funds managed by advisers from the two firms?

3. *(Anderson et al. 2011, Chap. 11, problem 21)* Fidelity Magellan is a large cap growth mutual fund and Fidelity Small Cap Stock is a small cap growth mutual fund (Morningstar Funds 500, 2006). The standard deviation for both funds was computed based on a sample of size 26. For Fidelity Magellan, the sample standard deviation is 8.89%; for Fidelity Small Cap Stock, the sample standard deviation is 13.03%. Financial analysts often use the standard deviation as a measure of risk. Conduct a hypothesis test to determine whether the small cap growth fund is riskier than the large cap growth fund. Use $\alpha = 0.05$ as the level of significance.