Class Worksheet

Optional, not done in class
Math 125 *Kovitz* Spring 2025

More Tests for Averages

The Standard Error for a Difference

The expected value for the difference of two independent quantities is $b - a$, where

- b is the expected value of the quantity from which the other is being subtracted;
- a is the expected value of the quantity which is being subtracted from the other.

The standard error for the difference of two independent quantities is $\sqrt{a^2 + b^2}$, where

- a is the SE for the first quantity;
- b is the SE for the second quantity.

Comparing Two Sample Averages

To test whether two populations (boxes) have the same average, one sets up the null hypothesis to be that the averages of the two boxes are equal. On this basis, the difference between the sample averages is expected to be 0, and the observed difference just reflects the luck of the draw. The alternative hypothesis says that the average of one box is smaller that the average of the other box. The two-sample z-statistic is computed as

$$z = \frac{\text{observed difference} - \text{expected difference}}{\text{SE for difference}}$$

The expected difference will in this case be equal to 0.

The two sample z-statistic is computed from—

- the sizes of the two samples,
- the averages of the two samples,
- the SDs of the two samples.

The test assumes two independent simple random samples.

Comparing Two Sample Percentages

The same procedure can be used to test whether the percentages of two boxes are equal.

Problems to think about

Box A has an average of 200 and an SD of 15. Box B has an average of 130 and an SD of 24. Now 25 draws are made at random with replacement from box A, and independently 36 draws are made at random with replacement from box B. Find the expected value and standard error for the difference between the average of the draws from box A and the average of the draws from box B.

A coin is tossed 4100 times. Find the expected value and the SE for the difference between the percentage of heads in the first 1600 tosses and the percentage of heads in the last 2500 tosses.

As part of a survey of the attitudes of college students regarding sexual harassment, investigators took independent random samples of 625 males and 1600 females, and presented them with a hypothetical case. 349 of the males felt that the scenario was sexual harassment, compared to 975 of the females; 55.84% versus 60.94%. Is this difference real or can it be explained as a chance variation? Formulate the null and alternative hypotheses in terms of a box model before answering the question.

In the previous problem, to find out whether the difference in percentages is statistically significant, the investigators started by computing $z = (60.94 - 55.84)/.0233$. Is anything wrong?

A researcher claims that SAT mathematics scores are lower among New Hampshire students than among Maine students. In his survey, New Hampshire students only averaged 471 on the test, and their SD was 100; while Maine students averaged 481 on the test with the same SD of 100. Can this difference be explained as a chance variation? You may assume that the researcher took a simple random sample of 400 New Hampshire students who took the SAT, and an independent simple random sample of 400 Maine students who took the SAT. Formulate the null and alternative hypotheses in terms of a box model before answering the question.