Two Sample Confidence Interval for $\mu_1 - \mu_2$

February 7, 2022

1. Is there a difference in the number of chocolate chips in Tim Hortons cookies versus the number of chocolate chips in Subway Cookies? Suppose that two random samples of Tim Hortons and Subway cookies of size $n = 28$ are collected. The sample of Tim Horton's cookies have a mean of 21.56 chocolate chips, and a standard deviation of 2.49 chocolate chips. The Subway cookies have a mean of 18.22 chocolate chips and a standard deviation of 4.51 chips. It is reasonable to assume that the distribution for chocolate chips is normal amongst the population of Subway and Tim Horton's cookies.

   (a) State all statistics, parameters for this problem.

   > **Solution:** Statistics: $\overline{x}_1$, $\overline{x}_2$, $\overline{x}_1 - \overline{x}_2$.
   > Parameters: μ_1, μ_2, $\mu_1 - \mu_2$.

   (b) Have the conditions for constructing a confidence interval been met? Explain.

   > **Solution:** Random Sampling: It is stated in the problem that the samples of cookies are random and thus representative of the population of all cookies.
   > Independence: Here the two samples of 28 cookies are clearly less than 10% of their respective populations of all cookies.
   > Normality: Require samples n_1, n_2, greater than or equal to 30 for the central limit theorem to apply. Here this is not the case, but it stated in the intro that populations are normal. Therefore we may assume the sampling distribution of $\overline{x}_1 - \overline{x}_2$ is normal with a mean of $\mu_1 - \mu_2$ and a standard deviation of $\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}$.

   (c) Describe the distribution of $\overline{x}_1 - \overline{x}_2$.

   > **Solution:** the sampling distribution of $\overline{x}_1 - \overline{x}_2$ is normal with a mean of $\mu_1 - \mu_2$ and a standard deviation of $\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}$.

   (d) Construct a 95% confidence interval for $\mu_1 - \mu_2$.

   > **Solution:**
   >
   > $$\begin{array}{rcl} \text{Point Estimate} & \pm & \text{Margin of Error} \\ (\overline{x}_1 - \overline{x}_2) & \pm & t^\star\sqrt{\dfrac{s_1}{n_1} + \dfrac{s_2}{n_2}} \\ 3.34 & \pm & 2.018\sqrt{\dfrac{2.49^2}{28} + \dfrac{4.51^2}{28}} \end{array}$$
   >
   > Confidence interval is $(1.38, 5.30)$.

   (e) Is there convincing evidence of a difference between the average number of chocolate chips between Tim Hortons and Subway cookies?

   > **Solution:** Yes, all of the plausible values for $\mu_1 - \mu_2$ are positive.

2. The most recent American Time survey, conducted by the Bureau of Labor Statistics, found that many Americans barely spend any time reading for fun. People ages 15 to 19 average only 7.8 minutes of leisurely reading per day with a standard deviation of 5.4 minutes. However, people ages 75 and over read for an average of 43.8 minutes per day with a standard deviation of 35.5 minutes. These results were based on random samples of 975 people ages 15 to 19 and 1050 people ages 75 and over.

   Construct and interpret a 95% confidence interval for the difference in mean amount of time (minutes) that people age 15 to 19 and people ages 75 and over read per day.

Solution: State: Estimating $\mu_1 - \mu_2$ the true difference between the average time people spend reading between people ages 15 to 19 and people ages 75 and over.

Plan: Two sample t interval for $\mu_1 - \mu_2$.

Check Conditions:
Random: Both samples are independent and random, and thus representative of their respective populations (all people in each age range).
Independence: $n_1 < 10\% N_1$ and $n_2 < 10\% N_2$ so sampling without replacement is fine. (Need to show numbers for compelte response)
Normality: Here both samples are large enough (> 30) so that the central limit theorem applies. Therefore the sampling distribution for $\overline{x}_1 - \overline{x}_2$ is approximately normal.

Do:

$$\begin{array}{rcl} \text{Point Estimate} & \pm & \text{Margin of Error} \\ (\overline{x}_1 - \overline{x}_2) & \pm & t^\star \sqrt{\dfrac{s_1}{n_1} + \dfrac{s_2}{n_2}} \\ (7.78 - 43.8) & \pm & t^\star \sqrt{\dfrac{5.4^2}{975} + \dfrac{35.5^2}{1050}} \end{array}$$

Our confidence interval is $(-38.2, -33.8)$.

Conclude: We are 95% confidence that the interval from -38.2 to -33.8 minutes captures the true difference in mean amount of time spent reading between people age 15 to 19 and people who are 75 and over.