

Comparing Two Means

In our first homework assignment we saw data on the humerus lengths of house sparrows presented by Hermon Bumpus to the Anatomical Laboratory of Brown University after an uncommonly severe winter storm. The humerus lengths (in thousands of an inch) of 24 sparrows who died in the storm and 35 sparrows who survived the storm were shown. The sample mean and sample standard deviation for the sparrows who died were 727.9 and 23.5, respectively. The sample mean and standard deviation for the survivors were 738.0 and 19.8, respectively. There is a difference between the sample averages, but can we say there is a difference in the population means?

Exactly what we mean by populations is a bit subtle. Most of the house sparrows alive at the time of the storm were either not exposed to the storm or not sampled by Dr. Bumpus. We can imagine all those house sparrows, along with those observed, falling into two groups: those who would have died and those who would have survived had they been exposed to the storm. We assume that our observed groups are like random samples from those two populations. We want to know if there is a difference between the mean humerus length of those who would have died and those who would have survived had they been exposed to the storm.

Suppose the following conditions hold.

- Population 1 has mean μ_1 and standard deviation σ_1 .
- Population 2 has mean μ_2 and standard deviation σ_2 .
- σ_1 is approximately equal to σ_2 .
- A simple random sample of size n_1 from population 1 yields sample average and standard deviation \bar{Y}_1 and s_1 , respectively.
- A simple random sample of size n_2 from population 2 (independent of the sample from population 1) yields sample average and standard deviation \bar{Y}_2 and s_2 , respectively.
- The sample sizes n_1 and n_2 are large AND/OR populations 1 and 2 are roughly normal.

Then....

- The standard error of $\bar{Y}_2 - \bar{Y}_1$ is given by

$$\text{SE}(\bar{Y}_2 - \bar{Y}_1) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{where} \quad s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}.$$

- The ratio

$$t = \frac{(\bar{Y}_2 - \bar{Y}_1) - (\mu_2 - \mu_1)}{\text{SE}(\bar{Y}_2 - \bar{Y}_1)}$$

has an approximate t -distribution with $n_1 + n_2 - 2$ degrees of freedom.

- A $100(1 - \alpha)\%$ confidence interval for $\mu_2 - \mu_1$ is given by

$$\bar{Y}_2 - \bar{Y}_1 \pm t_{n_1+n_2-2}^{(1-\alpha/2)} \text{SE}(\bar{Y}_2 - \bar{Y}_1),$$

where $t_{n_1+n_2-2}^{(1-\alpha/2)}$ is the t -ratio larger than $100(1 - \alpha/2)\%$ of t -ratios from a t -distribution with $n_1 + n_2 - 2$ degrees of freedom.

1. Test $H_0 : \mu_1 = \mu_2$ against the alternative $H_A : \mu_1 < \mu_2$ using the Bumpus data. Suppose the sparrows that died are like a simple random sample from the population of sparrows that would have died had they been exposed to the storm (population 1). Suppose the sparrows that survived are like a simple random sample from the population of sparrows that would have survived had they been exposed to the storm (population 2).

2. During the first week of class you saw hypothetical data from an experiment designed to compare the effect of two diets (new vs. standard) on lean percentage of hogs. The 50 hogs randomly assigned to receive the new diet had an average lean percentage of 52.3. The standard deviation of their 50 lean percentages was 2.1. The sample average and standard deviation in the control group of 50 hogs were 50.8 and 1.8, respectively. Is there significant evidence that the new diet increases lean percentage?
 - (a) Write down appropriate null and alternative hypotheses.

 - (b) Compute a test statistic.

 - (c) Find the p -value of this test.

 - (d) State a conclusion.

 - (e) How do your results compare with the results of the randomization test conducted by your professor?

 - (f) Find a 95% confidence interval for the difference between the mean lean percentage of hogs fed the new diet and the mean lean percentage of hogs fed the standard diet.