22S:105
Statistical Methods and Computing

Two independent-sample t-tests

Lecture 17
Mar. 20, 2006

Kate Cowles
374 SH, 335-0727
kcowles@stat.uiowa.edu

Two independent sample problems

- Goal of inference:
  - to compare the characteristics of two different populations
  - to compare responses to two different "treatments"

- Examples of two-independent-sample problems:
  - A medical researcher is interested in the effect on blood pressure of added dietary calcium. She conducts a randomized comparative experiment in which one group of subjects receives a calcium supplement and a control group gets a placebo.
  - A climatologist wishes to test whether seeding with silver nitrate affects the amount of rainfall produced from clouds. He randomly selects 26 clouds to seed, and measures the rain output by each of them as well as the rain output by 26 other randomly selected unseeded clouds.

Which study design?

The following situations require inference about a population mean or means. Identify which type of problem each one is:

- one-sample
- paired-sample
- two independent samples

1. To check a new method of chemical analysis, a chemist gets a reference specimen of known concentration from the National Institute of Standards and Technology. She then makes 20 measurements of the concentration of this specimen using the new method and checks for bias by comparing the mean of her 20 measurements with the known concentration.
2. Another chemist is checking the same new method. He has no reference specimen, but a familiar analytic method is available. He wants to know if the new and old methods agree. He takes a specimen of unknown concentration and measures the concentration 10 times with the new method and 10 times with the old method.

Comparing means from two different populations

Assumptions:

- We have two simple random samples, from two distinct populations.
  - The samples are independent.
    - The selection of one sample has no influence on the selection of the other. In particular, there is no matching.
  - The sizes of the two samples need not be the same.
  - We measure the same variable for both samples.
- The populations are normally distributed.

The set-up for two-independent-sample t-tests

		Group 1	Group 2
Population	Mean	μ_1	μ_2
	Standard deviation	σ_1	σ_2
Sample	Mean	$\bar{x}_1$	$\bar{x}_2$
	Standard deviation	s_1	s_2
	Sample size	n_1	n_2

Example: Cloud seeding

We wish to use our sample data to test whether the population mean of rainfall produced per cloud is the same for unseed clouds as for seeded clouds. We will use a two-sided test assuming that we don't know in advance in what direction a difference is likely to go.

$$H_0 : \mu_u = \mu_s \ \textit{or} \ \mu_u - \mu_s = 0$$
$$H_a : \mu_u \neq \mu_s \ \textit{or} \ \mu_u - \mu_s \neq 0$$

We will conduct our test at $\alpha = .05$.

Thus the quantity we really want to estimate is the difference between the two population means

$$\mu_u - \mu_s$$

As usual, we will use the corresponding sample statistics

$$\bar{x}_u - \bar{x}_s$$

as our best guess of the unknown population value of interest.

Now we need to standardize $\bar{x}_u - \bar{x}_s$ in order to find out whether it is different enough from 0 to provide strong evidence against H_0.

That is, we need to compute

$$\frac{(\bar{x}_u - \bar{x}_s) - (\mu_{u0} - \mu_{s0})}{standard\ error\ of\ (\bar{x}_u - \bar{x}_s)}$$

Suppose:

- we knew the standard deviations σ_u and σ_s in the populations of rainfall amounts from unseeded and seeded clouds
- $$\sigma_u = \sigma_s = \sigma\ (some\ known\ value)$$

Then the standard error of $(\bar{x}_s - \bar{x}_u)$ would be

$$\sqrt{\frac{\sigma^2}{n_u} + \frac{\sigma^2}{n_s}}$$

And the z statistic is

$$z = \frac{(\bar{x}_u - \bar{x}_s) - (\mu_{u0} - \mu_{s0})}{\sqrt{\frac{\sigma^2}{n_u} + \frac{\sigma^2}{n_s}}}$$