Two-Sample Tests

Chapter 5, Lab 1

OpenIntro Biostatistics

Topics

- Distinguishing between paired and independent data
- Two-sample test for paired data
- Two-sample test for independent group data

This lab extends inference for means in the one-sample setting to the two-sample setting. Two-sample data can consist of two samples in which the observations are paired in some way or in which the observations are drawn from two independent groups. In the paired context, inference is made about the population mean of the differences between observations; in the independent context, inference is made about the difference in population means between the two groups.

The material in this lab corresponds to Sections 5.2 - 5.3 of *OpenIntro Biostatistics*.

Distinguishing between paired and independent data

1. The following scenarios describe examples of two-sample data. Identify whether the data are paired (i.e., an observation in one group can be paired to an observation in the other) or independent.

   a) A multimedia program designed to improve dietary behavior among low-income women was evaluated by comparing women who were randomly assigned to intervention and control groups. The intervention was a 30-minute session in a computer kiosk in the Food Stamp office. About 2 months after the program, the participants took a knowledge test about healthy eating practices. The test scores will be used as an outcome in an analysis assessing the efficacy of the multimedia program.

   b) An investigator is studying standardized IQ test scores for third grade students in Massachusetts to see if birth order is associated with test score. He has collected test score data for 25 sets of siblings (first-born versus second-born).

   c) Researchers are interested in analyzing the relationship between oral contraceptive use and blood pressure in women.

      i. A group of women who are not currently oral contraceptive users are identified and their blood pressure is measured. One year later, the women who have become oral contraceptive users are identified; these women are selected as the study population and their blood pressure is measured a second time. The researchers will compare the initial blood pressure and the follow-up blood pressure of the women in the study population.

      ii. A group of oral contraceptive users and a group of non-users are identified, and their blood pressure is measured. The researchers will compare blood pressure between the users and non-users.

Two-sample test for paired data

Did a new wetsuit design allow for increased swim velocities during the 2000 Olympics?

Wetsuits are commonly used in the swimming stage of triathlons; in addition to providing thermal insulation against cold water, wetsuits are also thought to increase swimming speed. In 2008, de Lucas and co-authors conducted a study to assess the effect of wetsuits on swim velocity. In this study, 12 competitive swimmers were asked to swim 1,500 meters at maximal velocity, once wearing a wetsuit and once wearing a standard swimsuit. The order of wetsuit versus swimsuit was randomized for each swimmer.

The mean velocity (m/s) for each 1500m swim is recorded in the `swim` dataset in `oibiostat`.

2. Load and view the data. There are two velocity values for each swimmer: one for the wetsuit trial and one for the swimsuit trial.

   a) For swimmer 1, what is the difference between velocity measured during the wetsuit trial and velocity measured during the swimsuit trial?

   b) How were the values stored in the `velocity.diff` variable calculated? What does a positive value for `velocity.diff` imply, versus a negative value?[1]

   c) Calculate $\overline{d}$, the mean of the observed differences.

3. Conduct a hypothesis test to address the question of interest. Let $\alpha = 0.05$.

   a) Suppose the parameter δ is the population mean of the differences in velocities during a 1500m swim if all competitive swimmers recorded swim velocities with both suit types. State the null and alternative hypotheses.

   b) Calculate the test statistic, where $\overline{d}$ is the mean of the differences, s_d is the standard deviation of the differences, and n is the number of differences (i.e., number of pairs). Note how the formula for the test statistic is identical to the one introduced in the previous chapter; a paired t-test is essentially a one-sample test of difference values.

$$t = \frac{\overline{d} - \delta_0}{s_d / \sqrt{n}}$$

   c) Calculate the p-value from a t-distribution with degrees of freedom $n - 1$.

   d) Draw a conclusion.

4. Calculate a 95% confidence interval. Interpret the interval in the context of the data.

$$\overline{d} \pm t^{\star} \times \frac{s_d}{\sqrt{n}}$$

5. Verify the answers from Questions 3 and 4 using `t.test()`.

[1] Note: there were no negative values in these data.

Two-sample test for independent group data

Does change in non-dominant arm strength after resistance training differ between men and women?

In the Functional polymorphisms Associated with Human Muscle Size and Strength study (FAMuSS), researchers investigated the relationship between muscle strength and genotype at a location on the ACTN3 gene. The `famuss` dataset in `oibiostat` contains a subset of the data collected from the study, which includes additional demographic and phenotypic information. The percent change in non-dominant arm strength, comparing strength after training to before training, is stored as `ndrm.ch`.

6. Load and view the data. Create a plot that shows the association between `ndrm.ch` and `sex`. Describe what you see.

7. Conduct a hypothesis test to address the question of interest. Let $\alpha = 0.05$.

   a) Suppose the parameter μ_F is the population mean change in non-dominant arm strength for women, and μ_M is the population mean change in non-dominant arm strength for men. State the null and alternative hypotheses.

   b) Calculate the test statistic for the difference in means of the two groups, where $\overline{x}_1$ and $\overline{x}_2$ are the means of each sample, s_1 and s_2 are the standard deviations of each sample, and n_1 and n_2 are the number of observations per sample.

   $$t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

   c) Calculate the p-value. The degrees of freedom for the t-statistic in this setting can be approximated as the smaller of $n_1 - 1$ and $n_2 - 1$; i.e., $\min(n_1 - 1, n_2 - 1)$.

   d) Draw a conclusion.

8. Calculate a 95% confidence interval. Interpret the interval in the context of the data.

   $$(\overline{x}_1 - \overline{x}_2) \pm t^{\star} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

9. Verify the answers to Questions 7 and 8 using `t.test()`. Where do the values returned by `t.test()` differ from the hand calculation?

Additional practice

10. Suppose the `swim` data had been incorrectly identified as independent two-group data.

    a) Use `t.test()` to analyze the data accordingly and summarize the results. Do the results suggest a different conclusion than those of the original analysis?

    b) Conjecture as to why a paired study design has more capacity to show evidence of a difference between two groups than an independent group design. (Hint: think about the variance between individuals, versus the variance between two measurements on the same individual).

11. Chapter 1 (Section 7.1, Lab 2) presented an exploratory analysis of the relationship between age, ethnicity, and the amount of expenditures for supporting developmentally disabled residents in the State of California. When age is ignored, the expenditures per consumer is larger on average for White non-Hispanics than Hispanics, but the average differences by ethnicity were much smaller within age cohorts. Hypothesis testing can be used to conduct a more formal analysis of the differences in expenditures by ethnicity, both overall (i.e., ignoring age) and within age cohorts.

    The data are in the `dds.discr` dataset in `oibiostat`. Descriptions of the variables are provided in the documentation file.

    a) Is there evidence of a difference in mean expenditures by ethnic group, comparing Hispanics to White non-Hispanics? There is substantial skew in the distribution of expenditures within each group, so it is advisable to apply a natural log transformation before conducting a t-test. Summarize the results of the test.

    b) One way to account for the effect of age is to compare mean expenditures within age cohorts. When comparing individuals of similar ages, are the differences in mean expenditures by ethnic group larger than would be expected by chance alone?

       Conduct two t-tests to examine the evidence against the null hypothesis of no difference in mean expenditures within the two largest age cohorts: 13-17 years and 22-50 years. Summarize the results.