Bioinformatics 525: Module 2

Introduction to Statistics

Lab #2

Read TROPHY.csv data in RStudio using "Import Dataset" on the Environment Window.

IMPORTATNT: type attach(TROPHY) to have the variables accessible for analysis.

1. **One-sample test:** Compare the mean of HDL cholesterol at baseline to μ_0=50. Test both the two-sided hypothesis H_0: Mean of HDL = 50 vs. H_A: Mean of HDL ≠ 50 and the one sided hypothesis H_0: Mean HDL < 50 vs. H_A: Mean HDL ≥ 50.

   a. Display the boxplot for HDL.

   b. Calculate the summary statistics for HDL.

   c. Use the one-sample t-test and the one-sample Wilcoxon sign rank test to test whether the mean of HDL cholesterol is equal to 50 (< 50).

   d. Are the results the same? If not, which test is appropriate and why?

   e. Log-transform HDL (lHDL=lHDL) and compare the mean of lHDL to log(50) using the one-sample t-test and one sample Wilcoxon test. Are the results similar?

2. **Paired test:** Compare the BMI at baseline and at 24 month (BMI24) to see if there is a difference between two means.

   a. Use the visual display to see if the mean of BMI at baseline is similar to the mean of BMI24 two years later.

   b. Calculate the summary statistics of BMI and BMI24 and compare the results.

   c. Use one-sample test on the difference (BMI24-BMI) to test H_0: Mean of BMI = Mean of BMI24.

   d. Use the paired test by using the "paired=T" option. Compare the results from c) and d).

   e. Which test is the appropriate test to use, the paired t-test or the paired Wilcoxon test?

3. **<u>Two-Sample test:</u>** Comparing of HDL between the treatment group and the Placebo group.

   a. Use the visual display to see if the means of HDL are similar between the treatment group and the placebo group. Use side-by-side boxplot with notch=T option

   b. Look at the summary statistics by group and compare the results.

   c. Use the two-sample t-test H_0: Mean of HDL1 = Mean of HDL2.

      i. Use t.test() with unequal variance between groups (default)

      ii. Use t.test() with equal variance between groups

      iii. Choose between i.) vs. ii.) by testing if two groups have equal variance

   d. Use Wilcoxon rank sum test to test H_0: Mean of HDL1 = Mean of HDL2.

4. Power Analysis

a. Calculate the power for one-sample t-test to detect a difference of delta=10 ($\mu - \mu_0$), when n=100, and sd=25. Use Type I error equal to 0.05 and 0.01. What is the power for each Type I error?

b. Calculate the power for paired t-test to detect a difference of delta=10 ($\mu_1 - \mu_2$), when n=100, and sd=25. Use Type I error equal to 0.05 and 0.01. What is the power for each Type I error?

c. Calculate the power for two-sample t-test to detect a difference of delta=10 ($\mu_1 - \mu_2$),when n=100, and sd=25. Use Type I error equal to 0.05 and 0.01.

d. Repeat c), but using one-sided alternative, e.g. H_A: delta $\geq$ 10.

5. **Sample Size Calculations**

   a. Calculate the required samples size, n, for one-sample t-test to detect a difference of delta=10, with 80% power and 90% power, when the sd=20 and Type I error is equal to 0.05.

   b. Calculate the required samples size, n, for paired t-test to detect a difference of delta=10, with 80% power and 90% power, when the sd=10 and Type I error is equal to 0.05.

   c. Calculate the required samples size, n, for two-sample t-test to detect a difference of delta=10, with 80% power and 90% power, when the sd=20 and Type I error is equal to 0.05.

   d. Repeat c), but for one-sided alternative.

6. Please complete the **muddy point assessment** of this lab available at: http://tinyurl.com/bioinf525-lab2-2