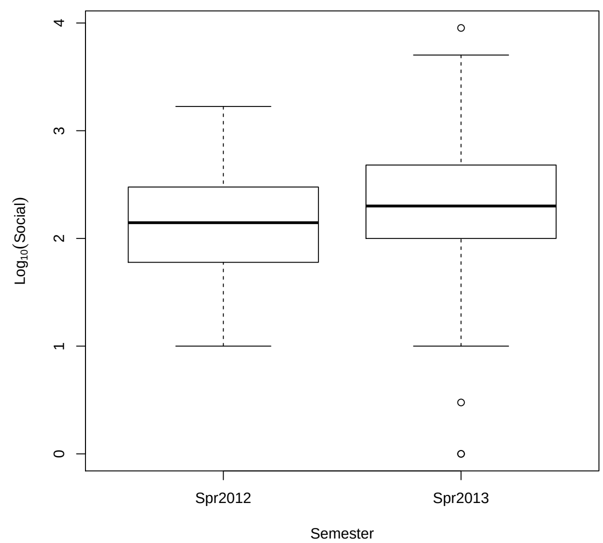
Comparing Paired and Unpaired Populations – Solutions
STAT-UB.0103 – Statistics for Business Control and Regression Models

Comparing Two Populations (Unpaired)

1. There were 44 students in the Spring 2012 version of this course who reported using social media. The mean reported log (base 10) usage per week was 2.15 ($\log_{10} 140$ minutes) with a standard deviation of 0.503. In the Spring 2013 version of the course, 60 students reported reported using social media. The mean reported log (base 10) usage per week was 2.32 ($\log_{10} 210$ minutes) with a standard deviation of 0.734.



Is there evidence that social media usage in one year was higher than in another year?

(a) What are the populations?

Solution: There are many plausible answers here. For example: population 1 is reported nonzero usages of all Spring 2012 NYU freshmen; population 2 is reported nonzero usages of all Spring 2013 NYU freshmen.

(b) What are the null and alternative hypotheses?

Solution: Let μ_1 be the mean log usage for the first population, and μ_2 be the mean log usage for the second population.

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2.$$

(c) What are the samples?

Solution: The reported usages for students in the two semesters. We have

$$n_1 = 44 \qquad n_2 = 60$$
$$\bar{x}_1 = 2.15 \qquad \bar{x}_2 = 2.32$$
$$s_1 = 0.503 \qquad s_2 = 0.734$$

(d) What is the test statistic?

Solution: First, we compute

$$\bar{x}_1 - \bar{x}_2 = 2.15 - 2.32 = -0.17.$$

Next,

$$\mathrm{SE}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{(0.503)^2}{44} + \frac{(0.734)^2}{60}} = 0.121.$$

The test statistic is

$$z = \frac{-0.17}{0.121} = -1.40.$$

(e) What is the rejection region? What is the result of the test?

Solution: Reject H_0 if $|z| > z_{\alpha/2}$. With $\alpha = .05$, we take $z_{\alpha/2} = 1.96 \approx 2$. We do not reject H_0.

(f) What assumptions do you need for the test to be valid?

Solution: We need that the samples are independent unbiased draws from the populations.

Confidence Intervals (Unpaired)

2. In Spring 2012, the reported log (base 10) social media usages for $n_1 = 44$ students had mean $\bar{x}_1 = 2.15$ and standard deviation $s_1 = 0.503$. In Spring 2013, the reported log (base 10) social media usages for $n_2 = 60$ students had mean $\bar{x}_2 = 2.32$ and standard deviation $s_2 = 0.734$. Find a 95% confidence interval for the difference in the mean social media usages for all Stern underclassmen between Spring 2012 and Spring 2013.

Solution: We first compute

$$\bar{x}_1 - \bar{x}_2 = 2.15 - 2.32 = -0.17.$$

Next,

$$\mathrm{SE}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{(0.503)^2}{44} + \frac{(0.734)^2}{60}} - 0.121.$$

An approximate 95% confidence interval for $\mu_1 - \mu_2$ is

$$\begin{aligned} -0.17 \pm 2(0.121) &= -0.17 \pm 0.242 \\ &= (-0.412, 0.072). \end{aligned}$$

3. Suppose you take a sample of size $n_1 = 49$ from one population and observe a sample mean of $\bar{x}_1 = 100$ and a sample standard deviation of $s_1 = 14$. Take a sample of size $n_2 = 36$ from another population and observe a sample mean of $\bar{x}_2 = 95$ and a sample standard deviation of $s_2 = 18$. Find a 95% confidence interval for the difference in the population means.

Solution: The estimated difference is

$$\bar{x}_1 - \bar{x}_2 = 100 - 95 = 5.$$

The standard error of the estimated difference is

$$\mathrm{SE}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{14^2}{49} + \frac{18^2}{36}} = 3.6.$$

Thus, the 95% confidence interval is 5 ± 2(3.6), or 5 ± 7.2. With 95% confidence, the difference in the population means is between -2.2 and 12.2.

Paired and unpaired samples

4. Suppose That you want to compare the mean daily rates of return for two stocks. For $n = 30$ consecutive trading days, you record the daily returns of the two stocks.

   (a) Consider two samples. The first sample is $X_{1,1}, X_{1,2}, \ldots, X_{1,30}$, the consecutive returns of the first stock. The second sample is $X_{2,1}, X_{2,2}, \ldots, X_{2,30}$, the consecutive returns of the second stock. Are these samples independent of each other? Why or why not?

   Solution: Dependent. There could be daily shocks (market events) affecting both stocks; these shocks would make the performances of the two stocks dependent.

   (b) How could we test whether or not one stock performs better than the other on average?

   Solution: Look at the differences $D_i = X_{1,i} - X_{2,i}$, and test whether or not $\mu_D = 0$ using a test on a population mean (the usual z or t test).

5. In the following situations, are the samples paired or unpaired?

   (a) You want to compare the performances of two restaurants. You measure the weekly profits of both restaurants for 10 consecutive weeks.

   Solution: Paired.

   (b) You want to compare expected starting salaries between males and females using the class survey data.

   Solution: Unpaired.

   (c) Your company can use one of two possible advertisements. You show one ad to one group of people, and ask them to rate the likelihood of buying your product after seeing the ad. You show the second ad to a second group of people, and ask them the same question.

   Solution: Unpaired.

   (d) Your company can use one of two possible advertisements. You show both ads to a group of people, and ask them to rate their opinions of both ads.

   Solution: Paired.