STAT 2593

Lecture 036 - Analysis of Paired Data

Dylan Spicker

Analysis of Paired Data

Learning Objectives

1. Construct hypothesis tests and confidence intervals for two sample data, when the data are paired.

Paired Data

- Sometimes, there is a natural pairing of data between two samples.

Paired Data

- Sometimes, there is a natural pairing of data between two samples.
  - Two treatments given to the same individual.

Paired Data

- Sometimes, there is a natural pairing of data between two samples.
  - Two treatments given to the same individual.
  - Before-and-after testing.

Paired Data

- Sometimes, there is a natural pairing of data between two samples.
  - Two treatments given to the same individual.
  - Before-and-after testing.
  - Twin studies.

Paired Data

- Sometimes, there is a natural pairing of data between two samples.
  - Two treatments given to the same individual.
  - Before-and-after testing.
  - Twin studies.
- In this case we will have

Paired Data

- Sometimes, there is a natural pairing of data between two samples.
  - Two treatments given to the same individual.
  - Before-and-after testing.
  - Twin studies.
- In this case we will have
  - $X_1, \ldots, X_n$ with mean μ_1 and variance σ_1^2.

Paired Data

- Sometimes, there is a natural pairing of data between two samples.
  - Two treatments given to the same individual.
  - Before-and-after testing.
  - Twin studies.
- In this case we will have
  - $X_1, \ldots, X_n$ with mean μ_1 and variance σ_1^2.
  - $Y_1, \ldots, Y_n$ with mean μ_2 and variance σ_2^2.

Paired Data

- Sometimes, there is a natural pairing of data between two samples.
  - Two treatments given to the same individual.
  - Before-and-after testing.
  - Twin studies.
- In this case we will have
  - $X_1, \ldots, X_n$ with mean μ_1 and variance σ_1^2.
  - $Y_1, \ldots, Y_n$ with mean μ_2 and variance σ_2^2.
  - $X_j \not\perp Y_j$, for the same j.

Comparing Means in Paired Data

- Since the data are naturally paired, we can consider the differences.

Comparing Means in Paired Data

- Since the data are naturally paired, we can consider the differences.
  - Take $D_i = X_i - Y_i$ to denote the i-th difference.

Comparing Means in Paired Data

- Since the data are naturally paired, we can consider the differences.
  - Take $D_i = X_i - Y_i$ to denote the i-th difference.
  - We can treat $D_1, \ldots, D_n$ as a sample from a single population.

Comparing Means in Paired Data

- Since the data are naturally paired, we can consider the differences.
  - Take $D_i = X_i - Y_i$ to denote the i-th difference.
  - We can treat $D_1, \ldots, D_n$ as a sample from a single population.
  - If X and Y are both normally distributed, D will be as well.

Comparing Means in Paired Data

- Since the data are naturally paired, we can consider the differences.
  - Take $D_i = X_i - Y_i$ to denote the i-th difference.
  - We can treat $D_1, \ldots, D_n$ as a sample from a single population.
  - If X and Y are both normally distributed, D will be as well.
  - The variance of D_i will be given by $\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$.

Comparing Means in Paired Data

- Since the data are naturally paired, we can consider the differences.
  - Take $D_i = X_i - Y_i$ to denote the i-th difference.
  - We can treat $D_1, \ldots, D_n$ as a sample from a single population.
  - If X and Y are both normally distributed, D will be as well.
  - The variance of D_i will be given by $\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$.
  - σ_{12} is the **covariance** between X and Y.

Using Paired Data for Comparisons

- The mean difference, $\overline{D}$ will have variance

$$\sigma_D^2 = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{n}.$$

Using Paired Data for Comparisons

- The mean difference, $\overline{D}$ will have variance

$$\sigma_D^2 = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{n}.$$

- We can then treat this as any standard one population sample we have seen.

Using Paired Data for Comparisons

- The mean difference, $\overline{D}$ will have variance
$$\sigma_D^2 = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{n}.$$
- We can then treat this as any standard one population sample we have seen.
  - If the variances are known, and normality is assumed, we can use the standard normal distribution.

Using Paired Data for Comparisons

- The mean difference, $\overline{D}$ will have variance
$$\sigma_D^2 = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{n}.$$
- We can then treat this as any standard one population sample we have seen.
  - If the variances are known, and normality is assumed, we can use the standard normal distribution.
  - If the variances are unknown, normality is assumed, and the sample size is large, use the normal approximation.

Using Paired Data for Comparisons

- The mean difference, $\overline{D}$ will have variance
$$\sigma_D^2 = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{n}.$$
- We can then treat this as any standard one population sample we have seen.
  - If the variances are known, and normality is assumed, we can use the standard normal distribution.
  - If the variances are unknown, normality is assumed, and the sample size is large, use the normal approximation.
  - If the variances are unknown, normality is assumed, and the sample size is small, use the t distribution.

Using Paired Data for Comparisons

- The mean difference, $\overline{D}$ will have variance
$$\sigma_D^2 = \frac{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}{n}.$$
- We can then treat this as any standard one population sample we have seen.
  - If the variances are known, and normality is assumed, we can use the standard normal distribution.
  - If the variances are unknown, normality is assumed, and the sample size is large, use the normal approximation.
  - If the variances are unknown, normality is assumed, and the sample size is small, use the t distribution.
- Paired data should be used with caution. The results are **less conservative**.

Summary

- When the data are paired, we can consider the differences between observations directly.
- This results in a single sample, for which all of the previous procedures apply.
- The variance of the samples depends on the individual variances as well as the covariance between the two samples.