Paired T test and One-sample T test

Miaoyan Wang

Department of Statistics
UW Madison

Two-Sample Studies

Two-sample studies aim to compare two populations. For example, the goals are to:

A. Compare milk yield of cows on two different diets.

B. Compare timber volumes of two species of trees.

C. Compare heart rates of patients before and after a drug treatment.

D. Compare test scores of 7th graders before and after the summer break.

Paired vs. Independent Two Samples

- There are two types of two-sample studies:
  - Two samples are **paired**.
  - Two samples are **independent** or **unpaired**.
- A **paired two-sample study** is a study with two levels of a treatment, where each observation on one treatment is naturally paired with an observation on the other treatment. (order matters.)
- An **independent two-sample study** is a study with two levels of a treatment, where there is no order constraint between the observations on the two treatments. (order does not matter.)

Paired vs. Independent Two Samples

- When to use which study?
- For example, consider
  - Heart rates of 10 patients before and after a drug treatment.
  - Heart rates of 10 patients before the drug treatment vs. heart rates of another 10 patients after the drug treatment.
- Which study would be better for detecting the drug effect?
  Paired studies are usually preferred, because of increased precision (i.e. reduced variability) in estimating population mean difference.
- The method we use for data analysis should follow the study design.

Example: Lake Clarity 1980 vs. 1990

	Wisconsin	
Lake	1980	1990
1	2.11	3.67
2	1.79	1.72
3	2.71	3.46
4	1.89	2.60
5	1.69	2.03
6	1.71	2.10
7	2.01	3.01
...	...	...
17	1.47	2.43
18	1.67	1.91
19	2.31	3.06
20	1.76	2.26
21	1.58	1.48
22	2.55	2.35
sample mean	1.854	2.351
sample variance	0.168	0.354
sample sd	0.410	0.595

- Question of interest: Are the population mean in 1990 the same as that in 1980?
- In general, how to perform hypothesis testing on two population means?

Hypothesis testing is a form of statistical procedure that uses sampled data to draw conclusions about a population parameter.

Null Hypothesis vs. Alternative Hypothesis

- Y_{1i}: Random variable of Secchi depth of the ith lake in 1990 for $i = 1, \ldots, n$.
- Y_{2i}: Random variable of Secchi depth of the ith lake in 1980 for $i = 1, \ldots, n$.
- $\mu_1 = E(Y_{1i})$: Population mean Secchi depth in 1990.
- $\mu_2 = E(Y_{2i})$: Population mean Secchi depth in 1980.
- Our goal is to test $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$.
- Under the **null hypothesis** $H_0 : \mu_1 = \mu_2$
- The null hypothesis H_0 is generally the claim initially favored or believed to be true.
- Under the **alternative hypothesis** $H_A : \mu_1 \neq \mu_2$
- The alternative hypothesis H_A is generally the departure from H_0 that one wishes to be able to detect.

Null Hypothesis vs. Alternative Hypothesis

- $D_i = Y_{1i} - Y_{2i}$: difference of the ith observation between two treatments.
- $\mu_D = E(D_i) = \mu_1 - \mu_2$: Population mean difference between two treatments.
- Equivalent to testing $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$, we now consider testing
$$H_0 : \mu_D = 0 \text{ vs. } H_A : \mu_D \neq 0.$$
- Want to find a point estimation for μ_D. Candidate: sample mean difference $\bar{D}$ based on an i.i.d. sample of size $n = 22$ $(D_1, D_2, \ldots, D_{22})$.
- Sample average $\bar{D} = \frac{1}{n} \sum_i D_i = 0.497$ and sample variance $\frac{1}{n-1} \sum_i (D_i - \bar{D}_i)^2 = 0.19$.

Test Statistic

- **Hypothesis**: Assume that the $H_0 : \mu_D = 0$ holds.
- **Model**: Assume that $D_i \sim_{\text{i.i.d.}} N(0, \sigma_D^2)$.
- What is the distribution of $\bar{D}$?

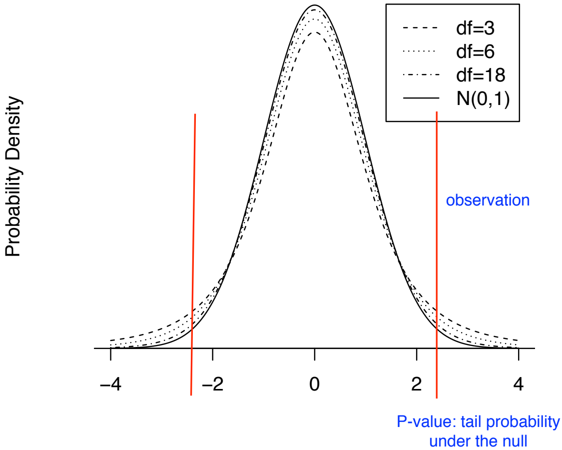
$$\bar{D} \sim N\left(0, \frac{\sigma_D^2}{n}\right).$$

- Because σ_D^2 is unknown, we have to plug in the estimator $S_D^2 := \frac{1}{n-1}\sum_i (D_i - \bar{D}_i)^2$ in place σ_D^2.
- Cautious. It does not hold $\bar{D} \sim N(0, 0.190/22)$!
- The **test statistic** is a function of the data whose sampling distribution can be mathematically well characterized. How about rescaled $\bar{D}$? We propose test statistics

$$T = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}},$$

T Distribution

- The test statistics T follows a T-distribution with degree of freedom n under the null hypothesis.



Example: Lake Clarity 1980 vs. 1990

- From the summary statistics, we have $n = 22$, $\bar{d} = 0.497$, and $s_D = 0.435$.
- The standard error is:
  $s_d/\sqrt{n} = 0.435/\sqrt{22} = 0.0927$
- The observed test statistic is:
  $t = \frac{\bar{d}-0}{s_d/\sqrt{n}} = \frac{0.497-0}{0.0927} = 5.357$
- Compute a **p-value** defined as the probability of observing a value as extreme or more extreme than what we observed, *if the H_0 is true*.
  $2 \times P(T_{21} \geq 5.357)$ which is less than 0.002.
- Interpretation: If H_0 is true, then we observed a very rare event. In other words, we have strong evidence to reject H_0.

Interpretation of the p-value

- The p-value can be interpreted as evidence again H_0. The smaller the p-value, the greater the evidence.
- In the classical hypothesis testing, a threshold value α is determined and the p-value is compared against it.
  - If the p-value is less than α, then we **reject** the H_0.
  - If the p-value is greater than α, then we **do not reject** the H_0.
- Lake clarity 1980 vs. 1990 example: Reject H_0 at the 5% level. There is very strong evidence that the mean Secchi depths in 1980 and 1990 are different.

Another Example: Lake Clarity 1980 vs. 1990

- How about testing

$$H_0 : \mu_1 = \mu_2 + 0.5 \text{ vs. } H_A : \mu_1 > \mu_2 + 0.5$$

- The test statistic is:

$$T = \frac{\bar{D} - 0.5}{S_D/\sqrt{n}} \sim T_{n-1}$$

- The standard error is: $s_d/\sqrt{n} = 0.435/\sqrt{22} = 0.0927$
- The observed test statistic is: $t = \frac{\bar{d}-0.5}{s_d/\sqrt{n}} = \frac{0.497-0.5}{0.435/\sqrt{22}} = -0.0294$
- The p-value is: $P(T_{21} \geq -0.0294) = 1 - pt(-0.0294, df = 21)$ which is more than 0.5 from calculator, R, or T-table.
- The conclusion is: Do not reject H_0 at 5% level. There is no evidence against that the H_0 that the mean Secchi depths differ by 0.5 m between 1990 and 1980.