Statistics for the Sciences

Interval Estimation of Population Means

Xuemao Zhang
East Stroudsburg University

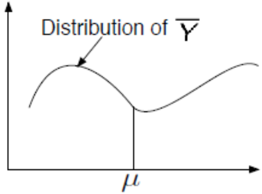
January 18, 2025

Outline

- Point Estimation
  - The Method of Moments
  - Maximum Likelihood Estimation
- Central Limit Theorem
- Confidence Intervals of μ
- Confidence Intervals of $\mu_1 - \mu_2$
- Lab 1: Confidence Intervals of μ
- Lab 2: Confidence Intervals of $\mu_1 - \mu_2$

Point Estimation

- Parameters are numerical descriptive measures for populations.
- Numerical descriptive measures calculated from the sample are called statistics.
  - We use statistics as estimates of population parameters.
- Statistics vary from sample to sample if we repeat the sampling procedure, hence Statistics are random variables.
- It is impossible to get the exact parameter of the population from a sample.



The Method of Moments

- When a random sample is selected, there are two ways to **point estimate** the parameters
  - The Method of Moments
  - The Method of Maximum Likelihood
- The Method of Moments is a technique used to estimate population parameters by equating sample moments (i.e., sample statistics such as the mean, variance, etc.) to theoretical moments (i.e., expected values of these statistics under a specified distribution).
- Examples:
  - $\overline{Y}$ is an estimate of μ by method of moments
  - Sample proportion $\hat{p}$ is an estimate of p by method of moments
  - Sample variance s^2 is an estimate of σ^2 by method of moments

Maximum Likelihood Estimation

- The method of maximum likelihood was first introduced by R. A. Fisher, a geneticist and statistician, in the 1920s.
- Most statisticians recommend this method, at least when the sample size is large, since the resulting estimators have certain desirable properties.
- The **likelihood function** tells us how likely the observed sample is as a function of the possible parameter values.
- **Maximizing the likelihood** gives the parameter values for which the observed sample is most likely to have been generated—that is, the parameter values that `agree most closely` with the observed data.
  - Let's use an example to illustrate the idea.

Maximum Likelihood Estimation

Example A sample of ten seedlings from a particular plant species is obtained. Upon observation after a month, it is found that the first, third, and tenth seedlings did not survive, whereas the others did.

- Let $p = P(\text{not survive})$, i.e., p is the proportion of all seedlings that did not survive.
- Define (Bernoulli) random variables $Y_1, Y_2, \ldots, Y_{10}$ by

$$Y_1 = \begin{cases} 1, & \text{if 1st seedling is dead} \\ 0, & \text{if 1st seedling survived} \end{cases}$$

$$\vdots$$

$$Y_{10} = \begin{cases} 1, & \text{if 10th seedling is dead} \\ 0, & \text{if 10th seedling survived} \end{cases}$$

Maximum Likelihood Estimation

- Then for the obtained sample, $Y_1 = Y_3 = Y_{10} = 1$ and the other seven Y_i's are all zero.
- The probability mass function of any particular Y_i is $p^{y_i}(1-p)^{1-y_i}$.
- Suppose that the conditions of various seedlings are independent of one another. This implies that the Y_i's are independent, so their joint probability mass function is the product of the individual pmf's:

$$f(y_1, \ldots, y_{10}; p) = \cdots = p^3(1-p)^7$$

Maximum Likelihood Estimation

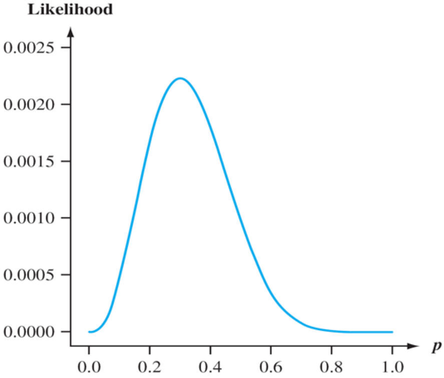
- Suppose that $p = 0.25$. Then the probability of observing the sample that we actually obtained is $(0.25)^3(0.75)^7 = 0.002086$.
- If instead $p = 0.50$, then this probability is $(0.50)^3(0.50)^7 = 0.000977$.

$$f(p|y_1, \ldots, y_{10}) = f(y_1, \ldots, y_{10}; p) = f(y_1; p) \cdot f(y_2; p) \cdots f(y_{10}; p) = p^3(1-p)^7$$

- It is called the **likelihood function** of p
- For what value of p is the obtained sample most likely to have occurred? That is, for what value of p is the above joint pmf as large as it can be? What value of p maximizes the joint pmf?

Maximum Likelihood Estimation

- Using knowledge of mathematics, it can be shown that the graph of the likelihood reaches its peak when $p = 0.3$.



Maximum Likelihood Estimation

- **Definition** Denote the parameter of interest by θ. For a random sample $y_1, \ldots, y_n$, the **maximum likelihood estimator** (MLE) of the parameter θ based on the sample is

$$\widehat{\theta}(y_1, \ldots, y_n)$$

  such that $L(\theta | y_1, \ldots, y_n)$ attains its maximum as a function of θ at $\theta = \widehat{\theta}(y_1, \ldots, y_n)$.
- Note that the likelihood function $L(\theta | y_1, \ldots, y_n) = f(y_1; \theta) \cdot f(y_2; \theta) \cdots f(y_n; \theta)$ is the joint probability (or probability density) of observing the sample data given the parameter θ.
- Maximizing a likelihood function is equivalent to maximizing the log-likelihood (natural log of the likelihood function) function.

Maximum Likelihood Estimation

Example Let $Y_1, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Find the MLEs of μ and σ^2.

Answer

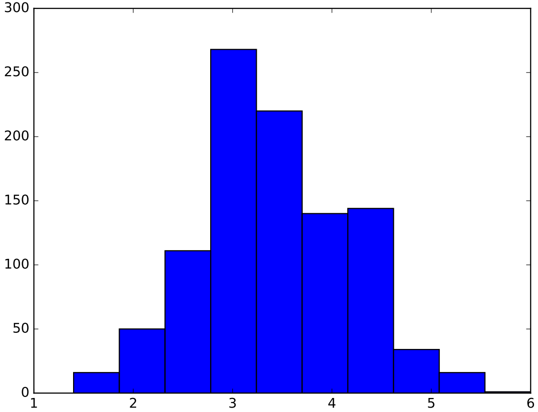
$$\begin{aligned}\hat{\mu}_{mle} =& \overline{Y}\\ \hat{\sigma^2}_{mle} =& \sum (Y_i - \overline{Y})^2/n\\ =& (n-1)s^2/n\end{aligned}$$

Central Limit Theorem

- Recall that Statistics are random variables in repeated sampling.
- The probability distributions for statistics are called **sampling distributions**.
- The sampling distribution of the sample mean $\overline{Y}$ is the distribution of all possible sample means, with all samples having the same sample size n taken from the same population.

Central Limit Theorem

- A simulation study: Consider repeating this process: Roll a balanced-die 5 times. Find the sample mean. What do we know about the behavior of all sample means that are generated as this process continues indefinitely?



Central Limit Theorem

- The Central Limit Theorem: Let $Y_1, \ldots, Y_n$ be a sequence of iid random variables. Let $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2 < \infty$. Define $\overline{Y}_n = \frac{\sum_{i=1}^{n} Y_i}{n}$. Then, $\frac{\overline{Y}_n - \mu}{\sigma/\sqrt{n}}$ has a limiting standard normal distribution or

$$\overline{Y}_n \sim N(\mu_{\overline{Y}_n} = \mu, \sigma_{\overline{Y}_n} = \frac{\sigma}{\sqrt{n}})$$

- Note: The subscript $_n$ specifies the sample size n.

Confidence Intervals of μ

- Let $\overline{Y}$ be the sample mean of a random sample of size n from a **normal distribution** with mean μ, the random variable

  $$T = \frac{\overline{Y} - \mu}{S/\sqrt{n}}$$

  has a t-distribution with $n - 1$ degrees of freedom (df).
- If the population is not normally distributed, but the sample size n is large (> 30), then the statistics T above is approximately t-distributed with $n - 1$ df.
  - The t-distribution of T is robust to small or even moderate departures from normality unless the sample size n is quite small.
- If the population is not normally distributed, but the sample size n is very large, then the distribution of statistics T is very close to standard normal.

Confidence Intervals of μ

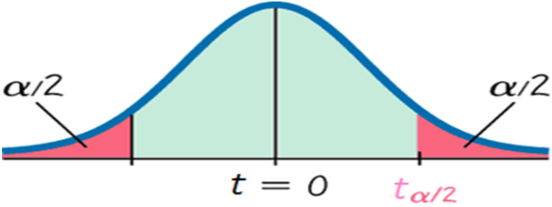
- A confidence interval (or interval estimate) is a an interval (a, b) of values so that you are fairly sure (with high probability) that the population parameter lies between these two values.
- Let α be a small probability like 0.01, 0.025 and 0.05.
- We want to find a 100(1-α)% confidence interval $(L(\overline{Y}), U(\overline{Y}))$ of μ such that
  - It is a random interval

$$P(L < \mu < U) = 1 - \alpha$$

Confidence Intervals of μ

- We start with the sampling distribution

$$P\left(-t_{\alpha/2} < \frac{\overline{Y}-\mu}{S/\sqrt{n}} < t_{\alpha/2}\right) = 1-\alpha$$



Confidence Intervals of μ

- Results:

Let $Y_1, \ldots, Y_n$ be a random sample from a normal population with mean μ.

(1) A two-sided $1-\alpha$ CI of μ is $\overline{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$.

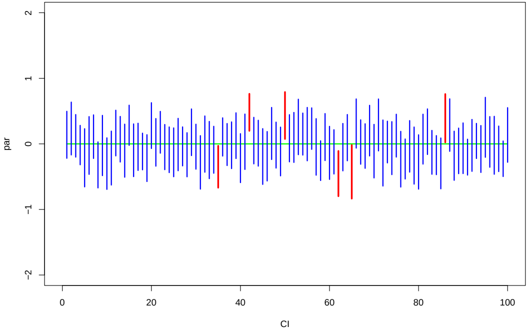
(2) A one-sided $1-\alpha$ CI of μ is $[\overline{Y} - t_{\alpha}\left(\frac{S}{\sqrt{n}}\right), \infty)$.

(3) A one-sided $1-\alpha$ CI of μ is $(-\infty, \overline{Y} + t_{\alpha}\left(\frac{S}{\sqrt{n}}\right)]$.

where $t_{\alpha/2}$ is determined from the t distribution with $df = n-1$.

Interpreting a Confidence Interval

- A simulation study:
  - Population (standard normal) mean μ=0
  - Sample size n=30
  - Take 100 samples from the population
  - Construct a 95% symmetric two-tailed confidence interval of μ
  - Count the number of confidence intervals containing μ



```
## [1] 94
```

Interpreting a Confidence Interval

- There are a huge number of confidence intervals that could be drawn.
  - In theory, all confidence intervals could be listed for a finite population.
  - 100(1-α)% (in the long run) will `work` (capture the true mean).
  - 100 α% (in the long run) will be `duds` (not capture the true mean).
- For a fixed confidence interval, randomness disappears.
- We simply say: `We are 00(1-$\alpha$)%  confident that the confidence interval contains the population mean.`

Confidence Intervals of $\mu_1 - \mu_2$

- Two samples are said to be paired or matched samples when for each data value collected from one sample there is a corresponding data value collected from the second sample, and both these data values are collected from the same source.
  - $(y_{1i}, y_{2i}), i = 1, \ldots, n$
- We can eliminate unwanted variability in the experiment by analyzing only the differences $d_i = y_{1i} - y_{2i}, i = 1, \ldots, n$ to see if there is a difference in the two population means, $\mu_1 - \mu_2$.
- Thus, the **paired two-sample inference** problem is reduced to **one-sample inference** problem.

Confidence Intervals of $\mu_1 - \mu_2$

- Suppose there are two **independent** samples selected from two populations. The two data sets are of the form $x_1, \ldots, x_{n1}$ and $y_1, \ldots, y_{n2}$
- Let $X_1, \ldots, X_{n_1}$ be a random sample from a population with mean μ_1.Let $Y_1, \ldots, Y_{n_2}$ be a random sample from a population with mean μ_2. Consider the statistics

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}}$$

- If the two populations are **normal** and $\sigma_1 = \sigma_2$. then T follows the t-distribution with $df = n_1 + n_2 - 2$.
- If the two populations are **normal** but $\sigma_1 \neq \sigma_2$ then T follows the t-distribution with df determined by Welch-Scatterthwaite equation

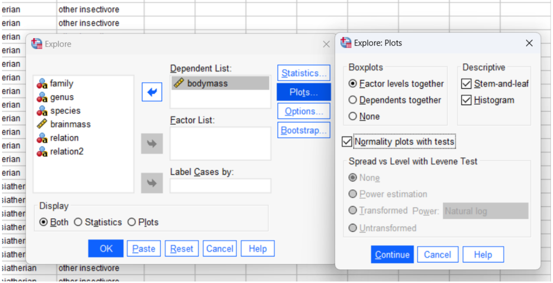
$$df = \frac{\left(\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_x^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_y^2}{n_2}\right)^2}$$

Confidence Intervals of $\mu_1 - \mu_2$

- If either populations is not **normal**, but the sample sizes are large, then T is very close to standard normal.
- Confidence intervals of $\mu_1 - \mu_2$ is based on any of the above sampling distributions of T.

Lab 1: Confidence Intervals of μ

- Consider the variable `bodymass` in data `kaufman.csv`
- Click on `Analyze` → `Descriptive Statistics` → `Explore...`
  - We need to check normality assumption



Explore

Dependent List: bodymass

Factor List:

Label Cases by:

Display: ● Both ○ Statistics ○ Plots

Explore: Plots

Boxplots: ● Factor levels together ○ Dependents together ○ None

Descriptive: ☑ Stem-and-leaf ☑ Histogram

☑ Normality plots with tests

Spread vs Level with Levene Test: ● None ○ Power estimation ○ Transformed Power: Natural log ○ Untransformed

Lab 1: Confidence Intervals of μ

- We can see that normality assumption is not valid.
- `T-CI` can be used due to the large sample size.
- Or we can use `Z-CI` following the formula

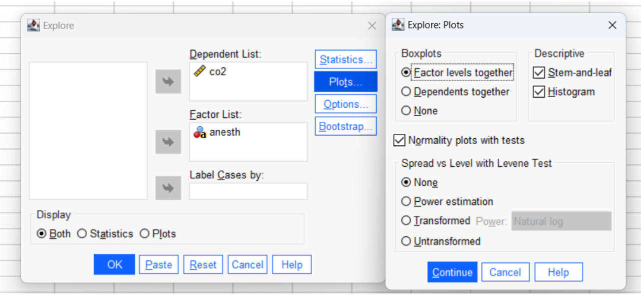
$$\overline{Y} \pm z_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

Lab 2: Confidence Intervals of $\mu_1 - \mu_2$

- Consider data `lowco2.csv`: Low et al (2016) examined the effects of two different anesthetics on aspects of the physiology of the mouse. Twelve mice were anesthetized with isoflurane and eleven mice were anesthetized with alpha chloralose and blood CO_2 levels were recorded after 120 minutes. We want to check if there is any difference between the anesthetics in the mean blood CO2 level. This is an independent comparison because individual mice were only given one of the two anesthetics.

Lab 2: Confidence Intervals of $\mu_1 - \mu_2$

- Check normality assumption: Click on `Analyze` → `Descriptive Statistics` → `Explore...`
  - It seems that normality is ok.



Explore

Dependent List: co2

Factor List: anesth

Label Cases by:

Display: ● Both ○ Statistics ○ Plots

Explore: Plots

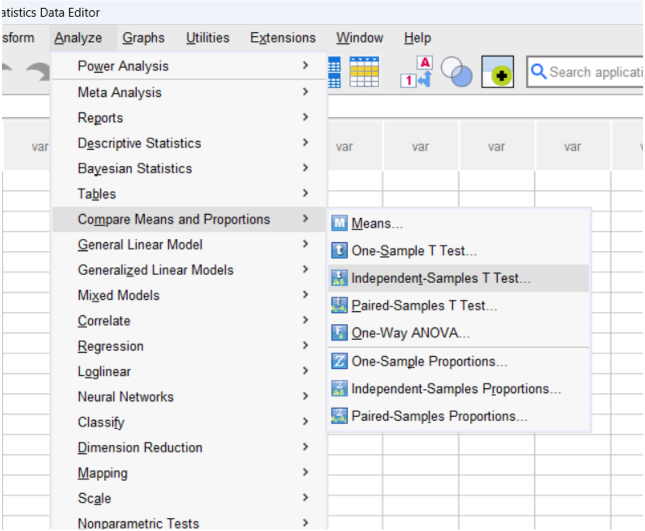
Boxplots: ● Factor levels together ○ Dependents together ○ None

Descriptive: ☑ Stem-and-leaf ☑ Histogram

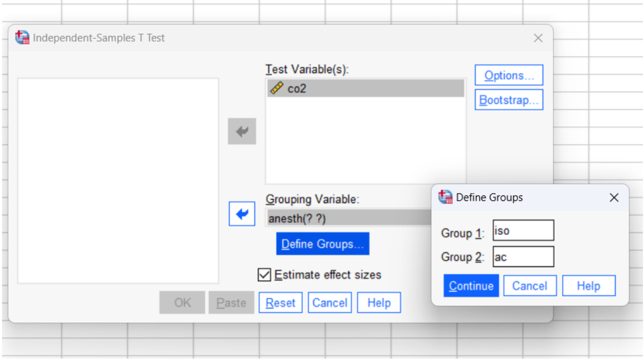
☑ Normality plots with tests

Spread vs Level with Levene Test: ● None ○ Power estimation ○ Transformed Power: Natural log ○ Untransformed

Lab 2: Confidence Intervals of $\mu_1 - \mu_2$



Lab 2: Confidence Intervals of $\mu_1 - \mu_2$



Test Variable(s): co2

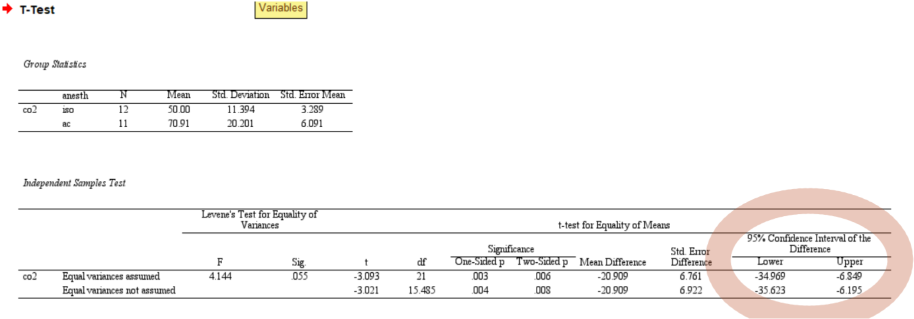
Grouping Variable: anesth(? ?)

☑ Estimate effect sizes

Group 1: iso

Group 2: ac

Lab 2: Confidence Intervals of $\mu_1 - \mu_2$



Group Statistics

	anesth	N	Mean	Std. Deviation	Std. Error Mean
co2	iso	12	50.00	11.394	3.289
	ac	11	70.91	20.201	6.091

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
co2	Equal variances assumed	4.144	.055	-3.093	21	.003	.006	-20.909	6.761	-34.969	-6.849
	Equal variances not assumed			-3.021	15.485	.004	.008	-20.909	6.922	-35.623	-6.195

License