Chapter 8. Inferences on a Population Mean

8.1 Confidence Intervals

8.2 Hypothesis Testing

8.1 Confidence Intervals
8.1.1 Confidence Interval Construction

- Confidence Intervals
  - A ***confidence interval*** for an unknown parameter ϑ is an interval that contains a set of plausible values of the parameter.
  - It is associated with a ***confidence level*** **1-** α, which measures the probability that the confidence interval actually **contains the unknown parameter value**.
  - Confidence levels of 90%, 95%, and 99% are typically used.

8.1.1 Confidence Interval Construction

- Inferences on a Population Mean
  - Inference methods on a population mean based upon the *t-procedure* are appropriate for large sample sizes $n \geq 30$ and also for small sample sizes as long as the data can reasonably be taken to be approximately normally distributed.
  - Nonparametric techniques (Chapter 15) can be employed for small sample sizes with data that are clearly not normally distributed.

- Two-Sided t-Interval
  - A confidence interval with confidence level $1-\alpha$ for a population mean μ based upon a sample of n continuous data observations with a sample mean $\overline{x}$ and a sample standard deviation s is

$$(\overline{x} - \frac{t_{\alpha/2,n-1}s}{\sqrt{n}}, \overline{x} + \frac{t_{\alpha/2,n-1}s}{\sqrt{n}})$$

  - The interval is known as a ***two-sided t-interval.***

- Technically speaking, $\frac{\sqrt{n}(\overline{X}-\mu)}{S}$ has a t-distribution only when the random sample is from a Normal distribution.

- Nevertheless, **the central limit theorem ensures** that the distribution of $\overline{X}$ is **approximately normal** for reasonably large sample sizes, and in such cases it is sensible to construct t-intervals **regardless of the actual distribution of the data** observations.

Example 14 : Metal Cylinder Production (p.340)

- Data : 60 metal cylinder diameters (page 271, Figure 6.5).
- Summary statistics:

n = 60	Median = 50.01	Max. = 50.36
$\bar{x}$ = 49.999	Upper quartile = 50.07	Min. = 49.74
s = 0.134	Lower quartile = 49.91	

- Critical points:

Sample size n = 60
Confidence level 90%: $t_{0.05,59}$ = 1.671
Confidence level 95%: $t_{0.025,59}$ = 2.001
Confidence level 99%: $t_{0.005,59}$ = 2.662

- Confidence interval with confidence level 90%:

$(49.999 - 1.671 \times \frac{0.134}{\sqrt{60}}, 49.999 + 1.671 \times \frac{0.134}{\sqrt{60}})$=(49.970, 50.028)

- Confidence interval with confidence level 95%:

$(49.999 - 2.001 \times \frac{0.134}{\sqrt{60}}, 49.999 + 2.001 \times \frac{0.134}{\sqrt{60}})$=(49.964, 50.033)

- Confidence interval with confidence level 99%:

$(49.999 - 2.662 \times \frac{0.134}{\sqrt{60}}, 49.999 + 2.662 \times \frac{0.134}{\sqrt{60}})$=(49.953, 50.045)

- Conclusion with confidence interval:

  With over 99% certainty, the average cylinder diameter lies within 0.05 mm of 50.00mm, that is, within the interval (49.95, 50.05).

- Comment:

  It is important to remember that this confidence interval is for the ***mean*** cylinder diameter, and not for the actual diameter of a randomly selected cylinder.

8.1.2 Effect of the Sample Size on Confidence Intervals

- Interval length(L)

$$L = 2 \times \frac{t_{\alpha/2,n-1}S}{\sqrt{n}}$$

- If a confidence interval with a length no large than L_0 is required, then the desired sample size n must satisfy

$$2 \times \frac{t_{\alpha/2,n-1}S}{\sqrt{n}} \leq L_0.$$

8.1.4 Simulation Experiment

- In practice, **an experimenter observes just one data set**, and it has a probability of 0.95 of providing a 95% confidence interval that does indeed straddle the true value μ.

8.1.5 One-Sided Confidence Intervals

- **One-Sided *t*-Interval**: One-sided confidence intervals with confidence levels 1-α for a population mean μ based on a sample of n continuous data observations with a sample mean $\overline{x}$ and a sample standard deviation s are

$$(-\infty,\ \overline{x} + \frac{t_{\alpha,n-1}s}{\sqrt{n}})$$

which provides an upper bound on the population mean μ, and

$$(\overline{x} - \frac{t_{\alpha,n-1}s}{\sqrt{n}}, \infty)$$

which provides a lower bound on the population mean μ.

Python codes

- import numpy as np
- import pandas as pd
- from scipy import stats
- import statsmodels.stats.weightstats as sms
- data = pd.read_csv('E:/data/taxi.txt', sep='\t', index_col=0)

  # The data is of tire life times in kilometers.
- print(data)   # output → next sheet
- print(data['BrandA'])   # output → next sheet
- print(data.describe())   # produces basic statistics. → next sheet
- ds1=sms.DescrStatsW(data['BrandA'])
- print("confidence interval=(%.4f,%.4f)" %ds1.tconfint_mean(0.05, 'two-sided'))

  confidence interval=(31776.7759,42723.2241)

Python outputs

print(data) →

Taxi	BrandA	BrandB
1	34400	36700
2	45500	46800
3	36700	37700
4	32000	31100
5	48400	47800
6	32800	36400
7	38100	38900
8	30100	31500

print(data['BrandA']) →

Taxi	
1	34400
2	45500
3	36700
4	32000
5	48400
6	32800
7	38100
8	30100

print(data.describe()) →

	BrandA	BrandB
count	8.000000	8.000000
mean	37250.000000	38362.500000
std	6546.754921	6181.062669
min	30100.000000	31100.000000
25%	32600.000000	35175.000000
50%	35550.000000	37200.000000
75%	39950.000000	40875.000000
max	48400.000000	47800.000000

8.1.6 z-Intervals

- Two-Sided z-Interval

  If an experimenter wishes to construct a confidence interval for a population mean μ based on a sample of size n with a sample mean $\overline{X}$ and using an assumed ***known*** value for the population standard deviation σ, then the appropriate confidence interval is

  $$(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

  which is known as a ***two-sided z-interval.***

- One-sided z-intervals are constructed analogously to the one-sided t-intervals with the z-quantile and σ replacing the t-quantile and s.

8.2 Hypothesis Testing
8.2.1 Hypotheses

- Hypothesis Tests of a Population Mean
  - A ***null hypothesis*** $\boldsymbol{H_0}$ for a population mean μ is a statement that designates possible values for the population mean.
  - It is associated with an ***alternative hypothesis*** $\boldsymbol{H_A}$, which is the "opposite" of the null hypothesis.
  - A ***two-sided*** set of hypotheses is

    $H_0 : \mu = \mu_0$ versus $H_A : \mu \neq \mu_0$

    for a specified value μ_0 of μ.

- A ***one-sided*** set of hypotheses is either

  $H_0 : \mu \leq \mu_0$ versus $H_A : \mu > \mu_0$

  or

  $H_0 : \mu \geq \mu_0$ versus $H_A : \mu < \mu_0$

Example 14 : Metal Cylinder Production

- The machine that produces metal cylinders is **set to make cylinders with a diameter of 50 mm.**
- The two-sided hypotheses of interest are

$$H_0 : \mu = 50 \quad \text{versus} \quad H_A : \mu \neq 50$$

where the null hypothesis states that the machine is calibrated correctly.

Example 48 : Car Fuel Efficiency

- A manufacturer claim : its cars achieve an average of **at least 35 miles per gallon** in highway driving.
- The one-sided hypotheses of interest are

$$H_0 : \mu \geq 35 \quad \text{versus} \quad H_A : \mu < 35$$

- The null hypothesis states that the manufacturer's claim regarding the fuel efficiency of its cars is correct.

8.2.2 Interpretation of p-values

- Types of error
  - Type I error: An error committed by rejecting the null hypothesis when it is true.
  - Type II error: An error committed by accepting the null hypothesis when it is false.
- Significance level
  - is specified as the upper bound of the probability of type I error.

- *p*-value of a test
  - Definition: The p-value of a test is the **probability of obtaining a given data set or worse when the null hypothesis is true**.
  - A data set can be used to measure the plausibility of null hypothesis H_0 through the construction of a *p*-value.
  - The smaller the *p*-value, the less plausible is the null hypothesis.

8.2.3 Calculation of p-values

Example 14 : Metal Cylinder Production

- For testing $H_0: \mu = \mu_0$ vs $H_A: \mu \neq \mu_0$
- The data set of metal cylinder diameters:
- $n = 60, \quad \bar{x} = 49.99856, \quad s = 0.1334$
- $\mu_0 = 50.0$ --→ $t = \frac{49.99856 - 50.0}{0.1334/\sqrt{60}} = -0.0836$
- $p\text{-value} = 2 \times P(T \geq 0.0836)$ where $T \sim t_{59}$.
- $p\text{-value} = 2 \times 0.467 = 0.934$

P-value for two-sided t-test

- Consider testing

$$H_0: \mu = \mu_0 \text{ vs } H_A: \mu \neq \mu_0$$

- $p - \text{value} = 2 \times P(T \geq |t|)$

  where $\mathrm{T} = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$.

P-value for one-sided t-test

- Consider testing

$$H_0: \mu \leq \mu_0 \text{ vs } H_A: \mu > \mu_0$$

Then

$$p - \text{value} = P(T \geq t)$$

- Consider testing

$$H_0: \mu \geq \mu_0 \text{ vs } H_A: \mu < \mu_0$$

Then

$$p - \text{value} = P(T \leq t)$$

Making conclusions using p-values

- Rejection of the Null Hypothesis

  If a *p*-value is smaller than the significance level, then the hypothesis H_0 is **rejected** in favor of the alternative hypothesis H_A.

- Acceptance of the Null Hypothesis

  A p-value **larger than 0.10** is generally taken to indicate that the null hypothesis H_0 is a plausible statement. The null hypothesis H_0 is therefore **accepted**.

  However, this does **not mean** that the null hypothesis H_0 has been **proven to be true**.

Python codes for one-sample tests concerning a mean

- import numpy as np
- import pandas as pd
- import statsmodels.stats.weightstats as sms
- data = pd.read_csv(“taxi.txt",sep='\t',index_col=0)
- dat=data/1000
- print(dat['BrandA'])
- print(dat.describe())
- dat_A= dat['BrandA']
- ds=sms.DescrStatsW(dat_A)
- print("One Sample Two-sided t-test")
- print("alternative hypothesis: true mean is not equal to 40")
- print("t, p-value, df: %.4f %.4f %.1f" %ds.ttest_mean(40, 'two-sided'))
- print("mean: %.4f" %np.mean(dat_A))

Python output

Taxi

1	*34.4*
2	*45.5*
3	*36.7*
4	*32.0*
5	*48.4*
6	*32.8*
7	*38.1*
8	*30.1*

	BrandA	BrandB
Count	8.000000	8.000000
mean	37.250000	38.362500
std	6.546755	6.181063
min	30.100000	31.100000
25%	32.600000	35.175000
50%	35.550000	37.200000
75%	39.950000	40.875000
max	48.400000	47.800000

[Output]

One Sample Two-sided t-test

alternative hypothesis: true mean is not equal to 40

t, p-value, df: -1.1881 0.2735 7.0

mean: 37.2500

8.2.4 Significance Levels

- Significance Level of a Hypothesis Test
  - A hypothesis test with a ***significance level*** of ***size*** α

    rejects the null hypothesis H_0 if a p-value *smaller* than α is obtained and

    accepts the null hypothesis H_0 if a p-value *larger* than α is obtained.

Two-Sided Hypothesis Test for a Population Mean with sig. level $\boldsymbol{\alpha}$

- A size α test for the two-sided hypotheses

$$H_0 : \mu = \mu_0 \quad vs \quad HA : \mu \neq \mu_0$$

rejects the null hypothesis H_0 if the ***test statistic*** $|\boldsymbol{t}|$ falls in the ***rejection region,***

$$\{\boldsymbol{t}: \ |\boldsymbol{t}| > \boldsymbol{t}_{\boldsymbol{\alpha/2,n-1}}\}$$

and accepts the null hypothesis H_0 if the ***test statistic*** $|\boldsymbol{t}|$ falls in the ***acceptance region,***

$$\boldsymbol{A} = \{\boldsymbol{t}: \ |\boldsymbol{t}| \leq \boldsymbol{t}_{\boldsymbol{\alpha/2,n-1}}\}$$

- Recall that $\mathrm{T} = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$

One-Sided Hypothesis Test for a Population Mean with sig. level α

- A size α test for the one-sided hypotheses

$$H_0 : \mu \leq \mu_0 \quad vs \quad H_A : \mu > \mu_0$$

rejects the null hypothesis H_0 if the ***test statistic t*** falls in the ***rejection region,***

$$\{\boldsymbol{t}\text{:}\ \ \boldsymbol{t} > \boldsymbol{t_{\alpha,n-1}}\}$$

and accepts the null hypothesis H_0 if the ***test statistic t*** falls in the ***acceptance region,***

$$\boldsymbol{A} = \{\boldsymbol{t}\text{:}\ \ \boldsymbol{t} \leq \boldsymbol{t_{\alpha,n-1}}\}$$

- A size α test for the one-sided hypotheses

$$H_0 : \mu \geq \mu_0 \quad vs \quad H_A : \mu < \mu_0$$

rejects the null hypothesis H_0 if the ***test statistic*** $\boldsymbol{t}$ falls in the ***rejection region,***

$$\{\boldsymbol{t}\text{:}\ \ \boldsymbol{t} < -\boldsymbol{t}_{\boldsymbol{\alpha},\boldsymbol{n}-\mathbf{1}}\}$$

and accepts the null hypothesis H_0 if the ***test statistic*** $\boldsymbol{t}$ falls in the ***acceptance region,***

$$\boldsymbol{A} = \{\boldsymbol{t}\text{:}\ \ \boldsymbol{t} \geq -\boldsymbol{t}_{\boldsymbol{\alpha},\boldsymbol{n}-\mathbf{1}}\}$$

- **Power of a hypothesis Test**
  - The ***power*** of a hypothesis test

    power = 1 – P(Type II error | H_A)

    which is the probability that the null hypothesis is rejected when it is false.

8.2.5 z-Tests

- Two-Sided z-test

The p-value for the two-sided hypothesis testing problem

$$H_0\colon \mu = \mu_0 \quad \text{versus} \quad H_A\colon \mu \neq \mu_0$$

based upon a data set of size n from $N(\mu, \sigma^2)$ with σ known.

The p-value is given by

$$\text{p-value} = 2 \times P(Z > |z|)$$

where $Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$.

- As for the two-sided test,

  a size α test rejects the null hypothesis H_0 if the ***test statistic*** z falls in the ***rejection region,***

$$\{\boldsymbol{z}: |\boldsymbol{z}| > \boldsymbol{z}_{\boldsymbol{\alpha/2}}\},$$

  and accepts the null hypothesis H_0 if the ***test statistic*** z falls in the ***acceptance region,***

$$A = \{z: |\boldsymbol{z}| \leq \boldsymbol{z}_{\boldsymbol{\alpha/2}}\}.$$

- The only difference between t-test and z-test is that the t-statistic is used for the t-test while z-statistic is used for the z-test instead.

Computation of the power of a hypothesis test

- Consider testing $H_0: \mu = \mu_0$ vs $H_A: \mu \neq \mu_0$ with significance level α. Assume that we have a random sample of size n from $N(\mu,\ \sigma^2)$. For computational convenience, assume σ is known.
- Power of test when $\mu = \mu^* > \mu_0$, $\beta(\mu^*)$.

$$\beta(\mu^*) = 1 - P_{\mu=\mu^*}(Accept\ H_0)$$
$$= 1 - P_{\mu=\mu^*}(|Z| \leq z_{\alpha/2})$$
$$= 1 - P_{\mu=\mu^*}(\frac{|\overline{X}-\mu_0|}{\sigma/\sqrt{n}} \leq z_{\alpha/2}) = 1 - P_{\mu=\mu^*}(-z_{\alpha/2} \leq \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2})$$

- (Continued)

$$1 - P_{\mu=\mu^*}\left(-z_{\alpha/2} \le \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right)$$

$$=1 - P_{\mu=\mu^*}\left(-z_{\alpha/2} - \frac{\mu^*-\mu_0}{\sigma/\sqrt{n}} \le \frac{\overline{X}-\mu^*}{\sigma/\sqrt{n}} \le z_{\alpha/2} - \frac{\mu^*-\mu_0}{\sigma/\sqrt{n}}\right)$$

$$= 1 - \Phi\left(z_{\alpha/2} - \frac{\mu^* - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\mu^* - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$= \beta(\mu^*)$$

Determination of sample size in hypotheses testing

- Find n for which $\beta(\mu^*) = \beta^*$ with $\mu^* > \mu_0$.

$$\beta(\mu^*) = 1 - \Phi\left(z_{\alpha/2} - \frac{\mu^* - \mu_0}{\sigma/\sqrt{n}}\right) + \Phi\left(-z_{\alpha/2} - \frac{\mu^* - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$\approx 1 - \Phi\left(z_{\alpha/2} - \frac{\mu^* - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$z_{\alpha/2} - \frac{\mu^* - \mu_0}{\sigma/\sqrt{n}} \approx z_{\beta^*}$$

So, $\sqrt{n} \approx \sigma \frac{z_{\alpha/2} - z_{\beta^*}}{\mu^* - \mu_0}$

Chapter summary

8.1 Confidence Intervals for Mean

- **t-intervals**
- **z-intervals**

8.2 Hypothesis Testing about Mean

- **p-value**
- **significance level**
- **acceptance region**
- **power of test**