Statistics

Lecture 10

Chou

Review

- Sampling Methods
- The Sampling Distribution of the Sample Mean
- Central Limit Theorem

Review

Ex1: Tell true or false of the following statements with regard to the sample mean.
(a) Sample mean is equal to the population mean.

(b) The mean of sample mean is equal to the population mean.

(c) The standard deviation of sample mean is equal to the population standard deviation.

(d) If the sample size is large, sample mean is approximately normally distributed.

(e) If the population is normal, sample mean is normally distributed no matter what is the sample size.

Review

Ex2: Michael eats at the same fast food restaurant every day. Suppose the time X between the moment Michael enters the restaurant and the moment he is served his food is normally distributed with mean 4.2 minutes and standard deviation 1.3 minutes.

(a) Find the probability that when he enters the restaurant today it will be at least 5 minutes until he is served.

(b) Find the probability that average time until he is served in eight randomly selected visits to the restaurant will be at least 5 minutes.

Review

Ex3: X $\sim$ N(2,9), sample size $=$ 3

$P((X-3)>6)=$

$P((\bar{X}-3)>6)=$

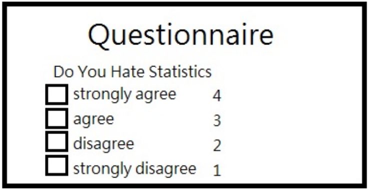
$P(|X-3|>6)=$

Statistical Inference

- Numerical descriptive measures calculated from the **population** are called **parameters**.
    - For quantitative variables, the population is described by mean (location parameter) μ and standard deviation (scale parameter) σ.
    - For a binomial population, location and scale are described by the population proportion p.
- If the values of the parameters are unknown we make inference on them based on sampled measurements.
- Types of inference: estimation, hypothesis testing.

Estimation

I would like to know if students at NCCU hate statistics :(


Questionnaire

Do You Hate Statistics

☐ strongly agree	4
☐ agree	3
☐ disagree	2
☐ strongly disagree	1

We estimate the population mean scores, μ, of all students by the sample mean scores, $\bar{x}$, of the 30 students sampled.

An estimate of this kind is called a **point estimate** for μ because it consists of a single number, or point.

Estimation

Point estimator: A point estimator is the value of a statistic used to estimate the parameter.

$$\text{statistic} \rightarrow \text{parameter}$$

$$\bar{x} \rightarrow \mu$$

$$s \rightarrow \sigma$$

$$s^2 \rightarrow \sigma^2$$

Estimation

However, a sample mean is usually not equal to the population mean; generally, there is **sampling error**. Therefore, we should accompany any point estimate of μ with information that indicates the accuracy of that estimate. This information is called a **confidence-interval estimate** for μ.

Estimation

- **Interval estimator**: A confidence interval (or interval estimate) is a range (or an interval) of values used to estimate the true value of a population parameter. A confidence interval is sometimes abbreviated as C.I..
  - The **confidence level** is the probability (often expressed as the equivalent percentage value) that is the proportion of times that the confidence interval actually does contain the population parameter, assuming that the estimation process is repeated a large number of times.

    C.I. $=$ point estimate $\pm$ margin of error

Point estimator for a Population mean

- Point estimator: The sample mean $\bar{X}$ is the best point estimate of the population mean μ.

Interval estimator for a Population mean

- Interval estimator:
  - Notation $z_{\frac{\alpha}{2}}$: The $z_{\frac{\alpha}{2}}$ is a value that satisfy

$$P(z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$$

  - Useful $z_{\frac{\alpha}{2}}$ values:

$100(1-\alpha)\%$	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
90%	0.05	1.645
95%	0.025	1.96
99%	0.005	2.58

Interval estimator for a Population mean

We need to consider two situations to compute a confidence interval:
We use sample data to estimate μ with $\bar{x}$, and

- σ known
- σ unknown $\rightarrow$ use s

σ known

Requirement:

- The value of the population standard deviation is known.
- Either or both of these conditions is satisfied: The population is normally distributed or $n \geq 30$.

If x is a normally distributed variable with mean μ and standard deviation σ, then, for samples of size n, the variable $\bar{x}$ is also normally distributed and has mean μ and standard deviation $\sigma/\sqrt{n}$

C.I. $=$ point estimate $\pm$ margin of error

margin of error $= z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

σ known

σ known

Step1 For a confidence level of $100(1-\alpha)\%$, use standard normal table to find $z_{\frac{\alpha}{2}}$.

Step2 The confidence interval for μ is from

$$\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ to } \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

n is the sample size, $\bar{x}$ is the sample mean

Example

Ex1: Suppose we have $n = 40$ and $\bar{X} = 76.3$. Assume σ is known to be 12.5. Construct a 95% confidence interval for population mean μ.

Ex2: A sample of size 49 has sample mean 35 and sample standard deviation 14. Construct a 98% confidence interval for the population mean using this information. Interpret its meaning.

Factors Affecting Confidence Interval Estimates

The width of a confidence interval are determined by:

1. The sample size, n.
2. The variability in the population, σ (usually estimated by s).
3. The desired level of confidence.

Determining the Required Sample Size

If the margin of error and confidence level are given, then we must determine the sample size needed to meet those specifications. To find the formula for the required sample size, we solve the margin of error formula, $E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, for n

Sample size for estimating μ

The sample of size required for a $100(1-\alpha)\%$ level confidence interval for μ with a specified margin of error, E, is given by the formula

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$$

rounded up to the nearest whole number.

Example

Estimating the mean age, μ, of all students in the school.

(a) Determine the sample size needed in order to be 95% confident that μ is within 0.5 year of the estimate, $\bar{x}$. Given that $\sigma = 12.1$ years.

(b) Find a 95% confidence interval for μ if a sample of the size determined in part (a) has a mean age of 20.8 years.

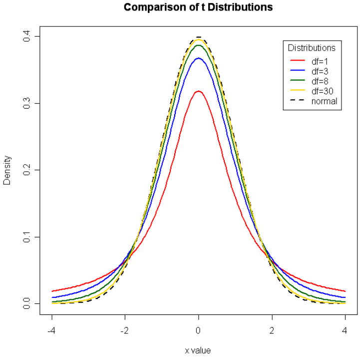
σ unknown

- estimate σ by s
- use "t"
  - **Student's t-distribution**: If a population has a normal distribution, then the distribution of

    $$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

    is a t distribution with **degree of freedom** $n - 1$.
  - There is a different t-distribution for each sample size. We identify a particular t-distribution by its number of degrees of freedom (df). **degree of freedom (df)** $= n - 1$.

t-distribution



A variable with a t-distribution has an associated curve, called a t-curve. Although there is a different t-curve for each number of degrees of freedom, all t-curves are similar and resemble the standard normal curve, as illustrated in the Figure above.

t-table (Appendix B5)

σ unknown

Step1 For a confidence level of $100(1-\alpha)\%$, use t table to find $t_{\frac{\alpha}{2}}$ with $df = n-1$

Step2 The confidence interval for μ is from

$$\bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \text{ to } \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

n is the sample size, $\bar{x}$ is the sample mean

Important Properties of the t distribution

- The t distribution is different for different sample sizes.
- The t distribution has the same general symmetric bell shape as the standard normal distribution, but it reflects the greater variability that is expected with small samples.
- The t distribution has mean 0.
- The standard deviation of the t distribution varies with degree of freedom, but it is greater than 1.
- As the sample size n gets lager, the t distribution gets closer to the standard normal distribution.

When to use z or t?

Example

Finding confidence intervals, assuming the population has a normal distribution:

- Weight lost on Weight Watchers diet: 95% confidence; $n = 40$, $\bar{x} = 3kg$, $s = 4.9kg$.

- Weight lost on Weight Watchers diet: 95% confidence; $n = 10$, $\bar{x} = 3kg$, $s = 4.9kg$.

- Life span of desktop PC: 90% confidence; $n = 21$, $\bar{x} = 6.8$ *years*, $\sigma = 2.4$ *years*.

Estimation of a Population Proportion

$np \geq 5$ and $n(1-p) \geq 5$

- Point estimator: $p = x/n$
- Interval estimator: $p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$

Example

To estimate the proportion of students at a large college who are female, a random sample of 120 students is selected. There are 69 female students in the sample. Construct a 90% confidence interval for the proportion of all students at the college who are female.

Example

In a random sample of 900 adults, 42 defined themselves as vegetarians. Of these 42, 29 were women.

1. Give a point estimate of the proportion of all self-described vegetarians who are women.

2. Verify that the sample is sufficiently large to use it to construct a confidence interval for that proportion.

3. Construct a 99% confidence interval for the proportion of all all self-described vegetarians who are women.

Determining the Required Sample Size

The estimated minimum sample size n needed to estimate a population proportion p to within E at $100(1-\alpha)\%$ confidence is

$$n = p(1-p)\left(\frac{z}{E}\right)^2$$

Determining the Required Sample Size

There are 3 factors that determine the size of a sample, none of which has any direct relationship to the size of the population.

1. The level of confidence desired.
2. The margin of error the researcher will tolerate.
3. The variation in the population being studied.

Example

Find the necessary minimum sample size to construct a 98% confidence interval for p with a margin of error $E = 0.05$,

1. assuming that no prior knowledge about p is available; and
2. assuming that prior studies suggest that p is about 0.1.

Example

The administration at a college wishes to estimate, to within two percentage points, the proportion of all its entering freshmen who graduate within four years, with 90% confidence. Estimate the minimum size sample required.