Estimation

Properties of Good Estimate

(i) A good estimator is one which is as close to the value of the parameter as possible

(ii) unbiased: $\hat{\theta}$ is an unbaised estimator of θ if the ~~estimator~~ expected value of the estimator is equal to the parameter being estimated.
eg: when sample mean $(\bar{X})$ and population mean μ

$$E(\bar{x}) = \mu$$

(iii) consistent: estimator is said to be consistent if the estimator approaches ~~popup~~ population parameter as sample size increases.

(iv) efficient: generally defined by comparing it with another and an estimate is ~~to~~ efficient if its variance is less than the other

(v) sufficient: An estimator is said to be sufficient if it contains all the information in the sample regarding parameter

Statistical Inference

Estimation

Testing of Hypothesis

⇒ **Parameters**: characteristics of population

⇒ **Statistics**: characteristics of sample

⇒ **Estimation**: prediction of characteristics of population with the help of sample

- → **Point Estimation**: specific value of a population which is assumed to be parameter
- → **Interval Estimation**: A range b/w which a parameter is ~~are~~ expected to lie

⇒ **Estimator**: Any function of random sample used to estimate unknown parameter of population.

Confidence Interval Estimation

⇒ Decision makers prefer to use interval-estimate that is likely to contain the population parameter value.

⇒ Interval estimations is calculating a and b which makes interval which contains parameters.

⇒ this interval is known as confidence coefficient denoted by $(1-\alpha)$

Obtained by:

Point Estimate ± Margin of error

Margin Error = Z_c × Standard error of a particular static

Z_c = variable that expresses confidence value i.e. 0.90

~~Interval Estimation (σ Known)~~

Confidence Intervals

- Population Mean
  - σ Known
  - σ Unknown
- Population Proportion

Confidence Interval for μ (σ known)

$$\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$\bar{x}$ is point estimate

$Z_{\alpha/2}$: normal distribution critical value for a probability of $\alpha/2$ in each tail

$\frac{\sigma}{\sqrt{n}}$: is the standard error

Finding Critical Value:

Consider confidence 95%

at this $Z_{\alpha/2} = \pm 1.96$

after finding put all values in eq^n

$$\bar{x} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

if confidence is 90% then $\alpha = 1 - 0.90 = 0.10$

$\alpha/2$ = 0.05

check on which z-score do this number 0.05 come

Confidence Interval for μ (σ unknown)

⇒ if σ is unknown, we can use sample S.D.

⇒ it introduces extra uncertainty, so we use t-distribution.

$$\overline{X} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

⇒ $t_{\alpha/2}$ depends on degrees of freedom (d.f.)

$$d.f. = n-1$$

In t-table check for df and $\alpha/2$ to get $t_{\alpha/2}$

Confidence Interval for Population Proportion, π

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

The distribution of sample proportion is approx normal if sample size is large with above standard deviation

We will estimate this with sample data: $\sqrt{\frac{p(1-p)}{n}}$

$$p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

$np > 5$ and $n(1-p) > 5$

p is sample proportion

n is sample size

Q A random sample of 100 people shows that 25 are left handed. Form a 95% confidence interval for the true proportion

$p = \frac{25}{100} = 0.25$

$np = 100 \times 0.25 = 25 > 5$

$n(1-p) = 100 \times (0.75) = 75 > 5$

$p \pm Z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$

$\Rightarrow \frac{25}{100} \pm 1.96 \sqrt{\frac{0.25 \times 0.75}{100}}$

$0.1651 \leq \pi \leq 0.3349$

Hypothesis Testing

Hypothesis testing begins with an assumption called a hypothesis that we make about a population parameter.

Procedure

(1) Set up Hypothesis

First thing is to set up a hypothesis about a population parameter.
Then we collect sample data, produce sample statistics and use this information to decide how likely is our hypothesis population parameter is correct.
The 2 hypothesis are:
1) Null Hypothesis

It is a claim ~~at~~ about population parameter that is assumed to be true until proven false (H_0)

2) Alternative Hypothesis

Any hypothesis which is complementary to null hypothesis (H_1)

$$H_0 : \mu = 500\,ml$$

$$H_1 : \mu \neq 500\,ml$$

② Set up Significance Level

The next step is to test the validity of H_0 against H_1 at a certain level of significance, which is the confidence with which an experiment rejects or retain a null hypothesis.

explain how percentage works i.e. 95% significance

③ Setting a Test Criterion

This is constructing a test criterion. This involves selecting an appropriate probability distribution for the particular test. Some commonly used are t, F and χ^2 (Chi-square)

④ Doing Computations

Performance of various computations - from a random sample of size n, necessary for test. These include testing statistic and standard error of testing statistic.

⑤ Making Decisions

We may draw statistical conclusions and take decision i.e. either to accept or reject null hypothesis.

2 types of error:

1. Hypothesis is true but test reject it (Type I)

2. Hypothesis is false but test accept it (Type II)

The Probability of committing a type I error is ~~same as level of significance~~ α = Prob (Type I)
whereas for type II is β = Prob (Type II)

Z-test

$$Z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$$

σ = S.D of population

Large sample

t-test

small sample

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

$\bar{x}$ = mean of sample
μ = mean of population
s = std. deviation

$$= \sqrt{\frac{\Sigma x^2}{n-1}}, \quad x = X - \bar{X}$$

If value less than table value then hypothesis is true.