Parametric Statistical Inference: Estimation

Two Areas of Statistical Inference

I. Estimation of Parameters

II. Hypothesis Testing

I. Estimation of Parameters – involves the estimation of unknown population values (parameters) by the known sample values (statistic).

Remark: Statistic are used to estimate the unknown parameters.

Parametric Statistical Inference: Estimation

Types of Estimation

I. **Point Estimate** – consists of a single value used to estimate population parameters.

Example: $\overline{X}$ can be used to estimate μ

s can be used to estimate σ

s^2 can be used to estimate σ^2

Parametric Statistical Inference: Estimation

II. **A confidence – Interval Estimate** – consists of an interval of numbers obtained from a point estimate, together with a percentage specifying how confident we are that the parameters lies in the interval.

Example: Consider the following statement.

A 95% confidence interval for the mean grade of graduate students is (1.0, 1.5).

Remark: The number 95% or 0.95 is called the confidence coefficient or the degree of confidence. The end points (1.0, 1.5) that is 1.50 and 1.0 are respectively called the lower and upper confidence limits.

Parametric Statistical Inference: Estimation

- **Remark:** In general, we can always construct a 100% confidence interval. The Greek letter is referred as the level of significance and a fraction is called the confidence coefficient which is interpreted as the probability that the interval encloses the true value of the parameter.
- A level of confidence equal to 95% means that the probability is 0.95 that the parameter value being estimated is contained, & 0.05 that is not contained, within the interval we obtain.
- $\alpha = 0.05$
- $\alpha \rightarrow$ is the probability of error indicating that the parameter will not be included in the interval estimate
- $1 - \alpha \rightarrow$ level of confidence

The following table presents the most commonly used confidence coefficients and the corresponding z – values.

Confidence Coefficient	α	Z_{α}	$\alpha/2$	$Z_{\alpha/2}$
90%	0.10	1.282	.05	1.645
95%	0.05	1.645	0.025	1.960
99%	0.01	2.326	0.005	2.576

Estimating the Population Mean

I. The Point Estimator of μ is $\overline{X}$.

II. The Interval Estimator of μ is the $(1-\alpha)100\%$ confidence interval given by:

1. $\left(\overline{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$ when σ is known.
2. $\left(\overline{X} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right)$ when σ is unknown,

where $t_{\frac{\alpha}{2}}$ is the t-value with $v = n-1$ degrees of freedom.

TABLE A2 | *t*-Distribution Critical Values

	Level of Significance for One-Tailed Test					
	0.10	0.05	0.025	0.01	0.005	0.0005
	Level of Significance for Two-Tailed Test					
df	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Estimating the Population Mean

Remark: If σ is unknown but for as long as $n > 30$, we still use (1) instead of (2). This explains the notion that the *t* is used only for small sample cases $(n \leq 30)$.

The Nature of *t*-Distribution

- develop by William Sealy Gosset (1896 – 1937), an employee of the Guinness Brewery in Dublin, where he interpreted data and planned barley experiments.
- his findings was under the pseudonym “student” because of the Guinness Company’s restrictive policy on publication by its employees.

The Nature of *t*-Distribution

The Sampling Distribution William Sealy Gosset studied are then called Student's *t*-distributions which is given by

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

The Nature of *t*-Distribution

Properties of the *t*-distribution

1. unimodal;
2. asymptotic to the horizontal axis;
3. symmetrical about zero;
4. dependent on *v*, the degrees of freedom (for the statistic under discussion, $v = n - 1$).
5. more variable than the standard normal distribution, $V(t) = \frac{v}{v-2}$ for $n > 2$;
6. approximately standard normal if *v* is large.

Definition: Degrees of freedom – the number of values that are free to vary after a sample statistic has been computed, and they tell the researcher which specific curve to use when a distribution consists of a family of curves.

Parametric Statistical Inference: Hypothesis Testing

- **Statistical hypothesis testing** – used in making decisions in the face of uncertainty in the context of choosing between two competing statements about a population parameter of interest.

- **Remark:** Statistical hypothesis testing involves two competing claims, that is, statements regarding a population parameter, and making a decision to accept one of these claims on the basis of evidence (and uncertainty in the evidence).

Parametric Statistical Inference: Hypothesis Testing

Definition: A statistical hypothesis is any statement or assumption about the population.

Two Types of Hypotheses Involved in a Hypothesis Testing Procedure

I. The Null Hypothesis H_0

- a statement that will involve specifying an educated guess about the value of the population parameter.

- the hypothesis of no effect and non-significance in which the researcher wants to reject.

II. The Alternative Hypothesis H_a

- The statement to be accepted, in case, we reject the null hypothesis.
- the contradiction of the null hypothesis H_0

Example: If the null hypothesis says that the average grade of the graduate students is 50, then we write,

$$\mu = 50$$

Parametric Statistical Inference: Hypothesis Testing

Example: There are three possible alternative hypothesis which may be formulated from the preceding null hypothesis .

a. $H_a : \mu < 50$ (the average grade of the graduate students is less than 50)

b. $H_a : \mu > 50$ (the average grade of the graduate students is greater than 50)

c. $H_a : \mu \neq 50$ (the average grade of the graduate students is not equal to 50)

- For every hypothesis test, a pair of hypotheses is set up
  - A null hypothesis H0
  - An alternate hypothesis H1
- The null hypothesis is always the one to be tested
- If evidence from sample is sufficient to reject H0 then we accept H1
- Otherwise H0 is not rejected & accepted to be true

Parametric Statistical Inference: Hypothesis Testing

Remarks: 1. The first two alternative hypotheses are called one-tailed or a directional test.

2. The third alternative hypothesis is called two-tailed or a non-directional test.

Decision Rule –

- if the computed value of the test statistic is more extreme than its critical value/s, then reject H0, else accept H1.
- The critical values are obtain having reference to appropriate area tables

Parametric Statistical Inference: Hypothesis Testing

Remark: The decision is based on the value of a test statistic, the value of which is determined from sample measurements.

Critical Region –

- the area under sampling distribution that includes unlikely sample outcomes which is also known as the rejection region.
- the area where the null hypothesis is rejected.

Parametric Statistical Inference: Hypothesis Testing

Acceptance Region –the region where the null hypothesis is accepted.

Critical Value – the value between the critical region and the acceptance region

Parametric Statistical Inference: Hypothesis Testing

Two Types of Errors that may be committed in rejecting or accepting the null hypothesis

I. Type I Error – occurs when we reject the null hypothesis when it is true.

- denoted by α.

II. Type II Error – occurs when we accept the null hypothesis when it is false.

- denoted by β.

Parametric Statistical Inference: Hypothesis Testing

The following table displays the possible consequences in the decision to accept or reject the null hypothesis.

State of affairs	**Hypothesis test conclusion**	
	H0 is accepted	H0 is rejected
H0 is True	Correct Decision	Type I Error Probability = α
H0 is False	Type II Error Probability = β	Correct Decision

Parametric Statistical Inference: Hypothesis Testing

Remark: α is called the level of significance which is interpreted as the maximum probability that the researcher is willing to commit a Type I Error.

Remark: The acceptance of the null hypothesis H_0 does not mean that it is true but it is a result of insufficient evidence to reject it.

Parametric Statistical Inference: Hypothesis Testing

Remark: α and β errors are related. For a fixed sample size *n*, an increase of α results to a decrease in β and a decrease in α results to an increase in β. However, decreasing the two errors simultaneously can only be achieved by increasing the sample size *n*. As α increases, the size of the critical region also increases. Thus, if H_0 is rejected at α, then H_0 will also be rejected at a level of significance higher than α.

Parametric Statistical Inference: Hypothesis Testing

- **In hypothesis testing procedure, the following steps are suggested:**

1. State the null hypothesis and the alternative hypothesis.
2. Decide on the level of significance.
3. Determine the decision rule, the appropriate test statistic and the critical region.
4. Gather the given data and compute the value of the test statistic. Check the computed

   value if it falls inside the critical region or in the acceptance region.
5. Make the decision and state the conclusion in words.

Parametric Statistical Inference: Hypothesis Testing

Remark: Alternatively, the *p*-value can also be used to make decision about the population of interest.

Definition: The *p*-value represents the chance of generating a value as extreme as the observed value of the test statistic or something more extreme if the null hypothesis were true.

Remark: The p-value serves to measure how much evidence we have against the null hypothesis. The smaller the p-value, the more evidence we have.

Remark: If the p-value is less than the level of significance, then the null hypothesis is rejected, otherwise, the null hypothesis is accepted.