Testing of Hypothesis. 10.04.2022

Sample — $x_1, x_2, \ldots, x_n$ — a obsn. which we draw from a popn of size (N)

obsn

$n << N$.

We can calculate mean and s.d. i.e. $\bar{x}$ & s.

$\bar{x} = \frac{1}{n}\sum x_i$ $\quad s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$

$X_1 \quad X_2 \quad \ldots \quad X_N$ → mean $= \mu$

$\mu = \frac{1}{N}\sum X_i$

$\sigma^2 =$ popn variance.
$= \frac{1}{N}\sum(X_i - \mu)^2$

μ is not known to us

obj — How we can test the population parameter on the basis of sample statistic

philips — Electric bulbs
Avg. Burning Hrs. $\geq$ 1000 hrs.

Customer. Avg. Burning hrs $<$ 1000 hr.

on the basis of the above statement

$\mu \geq 1000$ (Producer)

$\mu < 1000$ (Consumer).

avg.

1000

Lower limit.

Initially the statement made by Producer in the form of composite (not exact)

On the basis of composite statement the consumer we can form the statement.

$\mu \geq 1000$ → Composite hyp.

$\mu < 1000$ → Alternative hyp.

Null hyp → which will specify the population parameter completely exactly

$$H_0: \mu = 1000$$

$$H_1: \mu < 1000$$

obj.. From producer. Do not reject H_0

obj From consumer. Reject H_0

Trade off between Producer & Consumer.

Decision	Accept H_0	Reject H_0
H_0 is True	Correct	Wrong
H_0 is false	Wrong	Correct

In any decision making

Correct - Correct

Correct - Wrong

Wrong - Correct

Wrong - Wrong.

If Producer's Produce is good but there is a chance (Prob) of rejecting the lot = Type I Error.

P[of rejecting the good lot] = Type I Error.

Prob of accepting the bad lot = Type II Error.

Type I Error ~ Producer's risk.

Type II ,, ,, Consumer's risk.

Obj to reduce Type I Error.

Type I Error
level of significance
or Critical region
i.e. region of rejection

	No Covid	Covid	
No Covid	55	~~10~~ 5	60
Covid	10	~~25~~/30	40
	65	35	100

In testing of hyp. Type I Error is denoted by α

Generally $\alpha = 5\%$ | $\alpha = 1\%$

out of 100 times, 5 times can be rejected

In testing of hyp. there is an assumption. follow normal distribution

1. Continuous
2. Bell shaped
3. Symmetrical: mean = median = mode
4. Skewness = 0
5. Kurtosis = 3

Prob. density function (P.d.f)

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$X \sim ND(\mu, \sigma)$$

Standard normal variate (SNV).

$$z = \frac{x-\mu}{\sigma}$$

mean of $z = 0$.

variance of $z = 1$.

$z \sim ND(0, 1)$ → S.N.V.

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

$-\infty < z < +\infty$

Normal → Chi-square, t

Chi-square → F

Application:
- Univariate
- Bivariate
- Multivariate

Univariate → mean, s.d. — z, t, χ^2, F.

Bivariate → Correlation, Regression } t, F, χ^2.

Multivariate → Covariance test.

Application of t:

Pop^n 1. ✓ Stock: 1 — Stock 2 → Pop^n 2

10 years — $\bar{x}_1 = 10$ — $\bar{x}_1$, s_1^2 → μ_1

10 years — $\bar{x}_2 = 10$ — $\bar{x}_2$, s_2^2 → μ_2

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

Alternative hyp

- Left sided
- Both sided
- Right sided

Step ① Collection of sample (data) from some source.

Step ② ~~...~~ check the obs^n follows normal distribution. ✓

Step ③ Construct the hypothesis.

1. Null hypo. (H_0)
2. Alt "

univariate. Bivariate multivariate

Testing

Reject H_0

Do not H_0

H_0: $\mu_1 = \mu_2$ | $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 \neq \mu_2$

Step (4) Test of Statistic

t — derived from normal dist^n & chi-square dist.

Chi-square dist^n derived from normal dist.

F derived from chi-square dist.

$$t = \frac{\text{Observed} - \text{Expected}}{\text{Standard error}} = (O - E)/S.E.$$

H_1	H_2
$\bar{x}_1$	$\bar{x}_2$
s_1^2	s_2^2

$n = n_1 + n_2$

$$= \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma/\sqrt{n}}$$

$$s = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

Under H_0 we can calculate the value of 't' calculated t.

't' is calculated

We have to check the calculated value of t with given value of tabulated 't' (given) depending on level of significance (α) & degrees of freedom.

If the alternative hyp (H_1) is both sided then level of significance $\alpha/2$ & $d.f = n_1 + n_2 - 2$.

Decision If the calculated value of t < Tabulated value.
⇒ H_0 is accepted
i.e., Do not reject H_0

Otherwise reject H_0.

When we apply any software like SPSS, Jamovi, R-studio, python, we apply p-value criterion. Concept: p-value gives the degree of acceptance.

If $p > 0.05$ ⇒ Do not reject H_0

$p = 0.06$ — Low acceptance

$p = 0.60$ — Medium

$p = 0.90$ — High

$\alpha/2 = 0.05/2 = 0.025$

Normal Table.

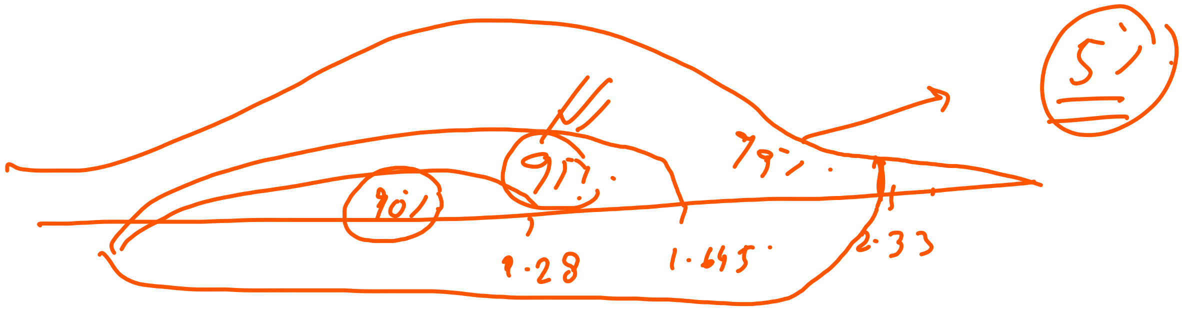
$z_{0.05} = 1.645$

$z_{0.01} = 2.33$

$z_{0.10} = 1.28$

$z_{0.025} = 1.96$

$z_{0.005} = 2.575$



When we apply chi-square (Test)

then chi-square = Sum of square of standardised normal variate.

$$0 < \sum\left(\frac{x_i-\mu}{\sigma}\right)^2 = \chi^2 < \infty$$

(1) Goodness of fit statistic
(2) To test pop^n s.d.
(3) To test the attributes.

Education having different levels. Income having.

1) < UG. 4) > PG.
2) UG.
3) PG.

Low
Avg.
High

Attributes Levels

obj

Attribute

Levels

Education \ Income	L_1	L_2	L_3	
L1	50	25	25	100 R_1
L2	30	15	5	50 R_2
L3	20	20	10	50 R_3
L4	10	20	20	50 R_4
	150 C_1	50 C_2	50 C_3	250

H_0: Education is independent of Income

H_1

n : given

Test of statistic (χ^2)

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

f_o = observed frequencies

f_e = expected " [Based on prob]

AH1 \ AH2	L1	L2	L3	
L1	R_1C_1/N	R_1C_2/N	R_1C_3/N	R_1
L2	R_2C_1/N	R_2C_2/N	R_2C_3/N	R_2
L3	$\frac{R_3C_1}{N}$	$\frac{R_3C_2}{N}$	$\frac{R_3C_3}{N}$	R_3
	C_1	C_2	C_3	N

follow a χ^2 dist with $(R-1)\times(C-1)$ d.f at level of significance (α)

R = no of rows. C = no of cols.