For left tailed test ⇒ If ~~[illegible]~~ $Z_{cal} < Z_{tab}$, then H_0 is rejected

σ - S.D. of population
s - " — sample

PAGE:
DATE: / /

SDM - cheatsheet

Module - 1

	2 TAILED TEST	RIGHT ~~TEST~~ TAILED TEST	LEFT TAILED TEST
NH	$H_0: M = M_0$	$H_0: M \le M_0$	$H_0: M \ge M_0$
AH	$H_a: M \ne M_0$	$H_a: M > M_0$	$H_a: M < M_0$

CONFIDENCE INTERVAL $= \bar{x} \pm Z\left(\frac{\sigma}{\sqrt{n}}\right)$

Test statistic,

$$z = \frac{\bar{x} - \bar{M}}{\sigma/\sqrt{n}}$$

$\bar{x}$ = mean of given sample
$\bar{M}$ - mean of req. sample (population)

Level of significance & Critical Value:

CRITICAL VALUE	LEVEL OF SIGNIFICANCE		
$Z\alpha$	1%	5%	10%
2 TAILED TEST	$\lvert Z\alpha \rvert = 2.58$	$\lvert Z\alpha \rvert = 1.96$	$\lvert Z\alpha \rvert = 1.645$
RIGHT TAILED	$Z\alpha = 2.33$	$Z\alpha = 1.645$	$Z\alpha = 1.28$
LEFT TAILED	$Z\alpha = -2.33$	$Z\alpha = -1.645$	$Z\alpha = -1.28$

PREREQUISITES:

1) Mean: Avg. of Data

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

2) Variance: Sum of squares of differences between all numbers & mean.

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$$

3) Standard Deviation: Square root of variance. Measure of extent to which data varies from the mean. (σ)

Type I & II errors:

	Accept H_0	Reject H_0
H_0 is true	✓	Type I error
H_0 is false	Type II error	✓

Type I: Hypothesis is true but our test REJECTS it (FALSE NEGATIVE)

Type II: Hypothesis is false but our test ACCEPTS it (FALSE POSITIVES)

Module - 2 - Large Sample Test

LARGE SAMPLE TESTS:

$$z = \frac{t - E(t)}{S.E.(t)} \sim N(0,1)$$

z = std. normal variate
$E(t)$ - Mean
$S.E.(t)$ - std. error of t
t - test statistic

1) Test of significance for Single Mean:

Here, $S.E.(t) = \sigma/\sqrt{n}$

$$\therefore z = \frac{\bar{x} - M}{\sigma/\sqrt{n}}$$

2) Test of significance of difference of Mean:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

3) ~~Test of significance for single proportion~~

Remarks:

1) If: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

then - $z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{(1/n_1)^2 + (1/n_2)^2}}$

2) If σ is not known, ~~&~~ sample variances (s) are used ($\sigma_1 = \sigma_2 = \sigma$)

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

3) — " — (But $\sigma_1 \ne \sigma_2 \ne \sigma$)
$\sigma_1^2 = s_1^2$; $\sigma_2^2 = s_2^2$

Test of difference of standard deviations -

$$z = \frac{s_1 - s_2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}}$$

3) Test of significance for Single Proportion:

$$z = \frac{X - nP}{\sqrt{nPQ}}$$

X - no. of successes in 'n' independent trials

P - constant prob. of success in each trial

$Q = 1 - P$

$$= \frac{np - nP}{\sqrt{nPQ}}$$

$p = \text{obs. prob. of success} = \frac{X}{n}$

$q = 1 - p$

If $|z| > 3 \Rightarrow H_0$ is rejected

Probable limits for proportion in population $= p \pm 3\sqrt{\frac{pq}{n}}$

The limits at α level of significance $= p + Z_\alpha \sqrt{\frac{pq}{n}}$

4) Test of significance for difference of proportions:

$$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

If $P_1 \neq P_2$; $Q_1 \neq Q_2$

$$z = \frac{p_1 - p_2}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}}$$

If P & Q are not known,

$$\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}$$

$e = 2.718$

Module 5 - Curve fitting

1) FITTING A STRAIGHT LINE:

$$y = a + bx$$

Normal Eqn: $\Sigma y = na + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

2) FITTING OF EXPONENTIAL CURVES

a) $y = ae^{bx}$

$\log_{10} y = \log_{10} a + bx \log_{10} e$

$Y = \log_{10} y$; $A = \log_{10} a$; $B = b\log_{10} e$

$\therefore Y = A + BX$

Normal Eqn: $\Sigma Y = nA + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$a = \text{Antilog } A$; $b = B/\log_e 10$

b) $y = ax^b$

$\log_{10} y = \log_{10} a + b\log_{10} x$

$Y = \log_{10} y$; $A = \log_{10} a$; $X = \log_{10} x$; $B = b$

$\therefore Y = A + BX$

Normal Eqn: $\Sigma Y = nA + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$a = \text{Antilog } A$, $b = B$

c) $y = ab^x$

$\log_{10} y = \log_{10} a + x\log_{10} b$

$Y = \log_{10} y$; $A = \log_{10} a$; $B = \log_{10} b$

$\therefore Y = A + BX$

Normal Eq: $\Sigma Y = nA + B\Sigma X$

$\Sigma XY = A\Sigma X + B\Sigma X^2$

$a = \text{Antilog } A$; $b = \text{Antilog } B$

σ - S.D of population
s - " - sample

PAGE:
DATE: / /

Module 6 - Correlation & Regression

Karl Pearson's coefficient of correlation:

$$r = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

$$cov(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\sqrt{\Sigma(y-\bar{y})^2}}$$

$$\sigma_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$r = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}\sqrt{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}}$$

$$r = \frac{\Sigma dx dy - \frac{\Sigma dx \Sigma dy}{n}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{n}}\sqrt{\Sigma dy^2 - \frac{(\Sigma dy)^2}{n}}}$$

Properties:

1) $-1 \le r \le 1$
2) If x & y are independent, $cov(x,y) = 0$

Ranks correlation:

$$\rho = 1 - \frac{6\Sigma d^2}{n^3 - n}$$

$d = x - y$

$\Sigma d = \Sigma(x-y) = \Sigma x - \Sigma y = n(\bar{x} - \bar{y})$

Related Ranks:

$$\rho = 1 - \frac{6\left[\Sigma d^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \dots\right]}{n^3 - n}$$

m - no. of items having equal ranks.

• REGRESSION •

1) Line of regression of y on x:

$y = a + bx$

$$y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$\therefore y - \bar{y} = b_{yx}(x - \bar{x})$

2) Line of regression of x on y:

$x = a + by$

$$x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})$$

$\therefore x - \bar{x} = b_{xy}(y - \bar{y})$

Expressions for Regression Co-eff:

1) $$b_{yx} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$$

$$b_{xy} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(y-\bar{y})^2}$$

2) $$b_{yx} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}$$

$$b_{xy} = \frac{\Sigma xy - \frac{\Sigma x \Sigma y}{n}}{\Sigma y^2 - \frac{(\Sigma y)^2}{n}}$$

Interval estimation $= \bar{x} \pm t \frac{s}{\sqrt{n}}$

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

3) $b_{yx} = \dfrac{\Sigma d_x d_y - \frac{\Sigma d_x \Sigma d_y}{n}}{\sqrt{\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{n}}}$

$b_{xy} = \dfrac{\Sigma d_x d_y - \frac{\Sigma d_x \Sigma d_y}{n}}{\sqrt{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{n}}}$

Properties :

1) $r = \sqrt{b_{yx} . b_{xy}}$
2) If $b_{xy} > 1$ & then $b_{yx} < 1$
3) $r \geq \frac{1}{2}(b_{xy} + b_{yx})$
4) Regression coeff are independent of change of origin but not change of scale.
5) Both reg. coff have the same sign.

~~Module 3 – Small Sample Test~~

~~1) t–TEST~~

~~curve : $y = c\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}}$~~

~~teststatistic, $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$~~

~~where, s = sample std. devi~~

~~$= \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$~~

~~$v = n-1$ = degree of freedom~~

Module 3 – Small Sample Test

1) t – test : (used when $n<30$ & σ is not known)

test statistic,

$t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ — (if σ s.d. of population is not given)

$t = \dfrac{\bar{x} - \mu}{s/\sqrt{n-1}}$ — (if σ is given)

$v = n - 1$ = degree of freedom

curve: $y = c\left(1 + \frac{t^2}{v}\right)^{-\frac{(v+1)}{2}}$

$\bar{x}$ – mean of sample ; μ – mean of population

s – s.d. of sample ; σ – s.d. of "

n – no. of observat^{n} ; v = degree of freedom

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

2) CHI – SQUARE TEST :

a) To test the goodness of fit :

H_0 : No difference in given & expected values

H_1 : Significant difference between given & expected values

Test statistic,

$\chi^2_{cal} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$

If $\chi^2_{cal} < \chi^2_{tab} \Rightarrow$ accept H_0

degree of freedom = $n-1$

b) For independence of attributes

			TOTAL
	a	b	(a+b)
	c	d	(c+d)
TOTAL	(a+c)	(b+d)	(a+b+c+d)

$\frac{(a+b)(a+c)}{N}$ $\frac{(a+b)(b+d)}{N}$

E(a)	E(b)	(a+b)
E(c)	E(d)	(c+d)
(a+c)	b+d	a+b+c+d

$\frac{(a+c)(c+d)}{N}$ $\frac{(c+d)(b+d)}{N}$

To generalize,

$~~E(~~ E = \frac{(\text{row total})(\text{column total})}{N}$

degree of freedom $= (r-1)(c-1)$

If $\chi^2_{cal} < \chi^2_{tab} \Rightarrow$ Accept H_0

3) SNEDECOR'S F - TEST $[F(v_1, v_2)$

$v_1 = n_1 - 1$; $v_2 = n_2 - 1$

$$F_{cal} = \frac{\text{Greater variance}}{\text{smaller variance}}$$

$$= \frac{S_1^2}{S_2^2} \text{ or } \frac{S_2^2}{S_1^2}$$

S_1^2 is the unbiased estimator of σ_1^2

$$\therefore S_1^2 = \frac{1}{n_1 - 1}(\Sigma x_1 - \bar{x})^2 = \frac{n_1 s_1^2}{n_1 - 1}$$

~~Similly~~ $~~S_2^2~~$ $S_2^2 = \frac{1}{n_2-1}(\Sigma x_2 - \bar{x}_2)^2 = \frac{n_2 s_2^2}{n_2 - 1}$

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

$F_{cal} < F_{tab} \Rightarrow H_0$ is accepted

t-test (continued)

1) For two independent means (?)

$$~~t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}~~ \qquad t = \frac{\bar{x}_1 - \bar{x}_2}{S.E.}$$

S.E. (standard error)

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

$$= \sqrt{\frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2}}$$

$v = n_1 + n_2 - 2$ //

~~2) Paired t-test~~

$$s_1^2 = \frac{n_1}{n_1 - 1} s_1^2$$