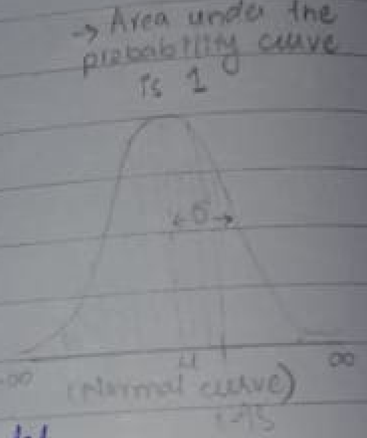
Monday 8/Apr/19 After Mid 2 "Continuous Probability Distributions"

(108)

Topics:

① Normal distribution, std. normal distribution
② Uniform.
③ Hypothesis testing.
④ Correlation & Regression.

① Normal Distribution:=



Density of normal random variable X

pdf $\Rightarrow f(x) = \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} dx$

- Bell-shaped, symmetric & unimodal
- Mean = Median = Mode
- Curve never touches x-axis.
- Area under the curve & above horizontal axis is equal to 1.

② Standard Normal Distribution:=

Distribution of normal random variable with mean '0' & variance '1'.

$$y = \frac{e^{-z^2/2}}{\sqrt{2\pi}} \quad \text{where} \quad z = \frac{X-\mu}{\sigma}$$

z-table → pg# 760

Eg#01 1(a) Area to the left of z = 2.06

$P(z < 2.06) = \boxed{2.0} + \boxed{0.06}$

$\boxed{z = 0.9803}$

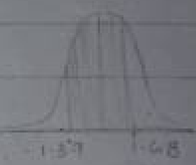
(1b) Area to the right of z = -1.19

$P(z > -1.19) = 1 - P(z < -1.19)$

$= 1 - (-1.1 - 0.09)$

$= 1 - 0.1170$

$\boxed{z = 0.883}$



(C) Area between z = 1.68 & z = -1.37

$P(-1.37 < z < 1.68) = P(z < 1.68) - P(z > -1.37)$

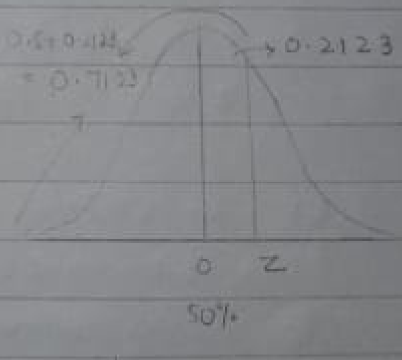
$= (1.6 + 0.08) - (-1.3 - 0.07)$

$= 0.9535 - 0.0853$

$\boxed{z = 0.8682}$

Eg#04 B/w 0 to z = 0.2123

z = ?



$z = 0.5 + 0.06$

$\boxed{z = 0.56}$

z	.06
0.5	0.7123

(from z-table)

Eg#02 z value to the right of the mean.

(a) 54.78%

$= 0.5478$

$z = 0.1 + 0.02$

$\boxed{z = 0.12}$

(b) 69.85%

$= 0.6985$

$z = 0.5 + 0.02$

$\boxed{z = 0.52}$

(c)

Eg#03 z value to the left of the mean.

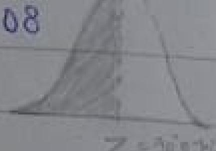
(a) 98.87%

$= 1 - 0.9887$

$= 0.0113$

$z = -2.2 - 0.08$

$\boxed{z = -2.28}$



(b) 82.12%

$= 1 - 0.8212$

$= 0.1788$

$z = -0.9 - 0.02$

$\boxed{z = -0.92}$

(c) 60.64%

$= 1 - 0.6064$

$= 0.3936$

$z = -0.2 - 0.07$

$\boxed{z = -0.27}$

(11)

Eg#04 $z = ?$

(i) 75 th = 0.75 ⇒ $z = 0.67$

(ii) 80 th = 0.80 ⇒ $z = 0.84$

(iii) 92nd = 0.92 ⇒ $z = 1.41$.

Eg#05 * Summer Spending: $\mu = 146.21$, $\sigma = 29.44$

$P(x < 160) = ?$

$$P(X < 160) = P\left(z < \frac{160 - 146.21}{29.44}\right) \qquad z = \frac{x - \mu}{\sigma}$$

$$P(z < 0.47) = 0.4 + 0.07$$

$$\boxed{z = 0.6808}$$

Eg#06 * Monthly Newspaper Recycling: $\mu = 28$ $\sigma = 2$

(a) B/W 27 & 31 pounds per month - ?

$$P(27 < X < 31) = P\left(z < \frac{31 - 28}{2}\right) - P\left(z > \frac{27 - 28}{2}\right)$$

$$= P(z < 1.5) - P(z > -0.6)$$

$$= 0.9332 - 0.2743$$

$$\boxed{z = 0.6589}$$

(b) $P(x > 30.2) = 1 - P(X < 30.2)$

$$= 1 - P\left(\frac{30.2 - 28}{2}\right)$$

$$= 1 - P(x < 1.1)$$

$$z = 1 - 0.8643 = 0.1357$$

Eg#07 * Coffee Consumption: $\mu = 1.64$, $\sigma = 0.24$.

$-2.67 \rightarrow 0.0038(500) \simeq 2.4$

?

idk

Eg#08 * Police Academy Qualification:

$P(z) = 0.90$, $x = ?$, $\sigma = 20$, $\mu = 200$

$$z = \frac{x - \mu}{\sigma} \rightarrow ①$$

For z: $z = 1.2 + 0.08 = 1.28$

eq① $1.28 = \frac{x - 200}{20} \Rightarrow x = (1.28)(20) + 200$

$$\boxed{x = 225.6}$$

Eg#09

$K = ?$

(a) $P(z > k) = 0.3015$

Area to the left $= 0.6985$

$k = 0.52$

(b) $P(k < z < -0.18) = 0.4197$

Area to the left of $-0.18 = 0.4286$

$= 0.4286 - 0.4197$

$= 0.0089$

$k = -2.37$

9/Apr/19

Eg#11: Diameter $= 3.0 \pm 0.01$ cm

$= (2.99 - 3.01)$

$\mu = 3.0 \quad \sigma = 0.005$.

$$P(2.99 < x < 3.01) = P\left(\frac{2.99-3}{0.005} < z < \frac{3.01-3}{0.005}\right).$$

$$= P(-2.0 < z < 2.0) \text{ (OR) } P(z<2.0) + P(z<-2$$

since symmetric $\leftarrow = 2(0.0228)$

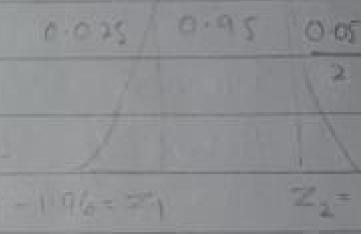
$$= 0.0456$$

$P(2.99 < x < 3.01) = 4.56\%$

Eg#12 $\mu = 1.50$ & $\sigma = 0.2$.

$P(z) = 95\%$

$P(-1.96 < z < 1.96) = 0.95$.

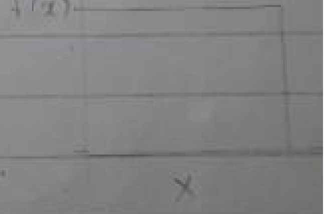


$$1.96 = \frac{1.50 + d - 1.50}{0.2}$$

$d = 0.392$.

③ Continuous Uniform Distribution:

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \leq x \leq B. \\ 0, & \text{elsewhere} \end{cases}$$



Mean $= (a+b)/2$

Variance = $(b-a)^2/12$ $\left(\frac{x-\bar{x}}{n}\right)$

→ no mode

→ Odd order moments = 0. → $\mu_1, \mu_3, \mu_5, \ldots = 0$

→ Skewness = 0 & Kurtosis = 9/5, therefore distribution is platykurtic

Eg 13 $[0, 4]$, reserved for no more than 4 hours.

(a) Density function

$$f(x) = 1/4$$

(b) $P(X \geq 3) = \int_3^4 1/4 \, dx \Rightarrow \frac{1}{4}$

→ Chap # 8

④ Sampling Distribution: "Fundamental Sampling Distribution"

– Probability distribution of statistics.

(only mean).

Sampling Distribution of Sample Means:

① Mean of sample means is same as the population mean.

$$\mu_{\bar{x}} = \mu.$$

② S.D of sample means < S.D of population.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \text{(Standard Error)}$$

Eg#01 class of four obtaining 2, 6, 4 and 8 marks out of 8.

(a) Population Mean & standard deviation?

$$\mu = (2+6+4+8)/4 = 5$$

$$\sigma = \sqrt{\Sigma(x-\mu)^2/n} = 2.236$$

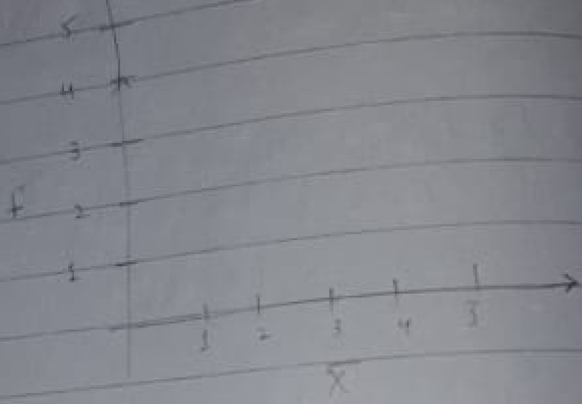
(b) Graph?

(c) All possible samples of size 2 with replacement & calculate mean?

Samples $= 2^4 = 16$.

Samples	$\bar{X}$	Samples	$\bar{X}$	Samples	$\bar{X}$
(2, 2)	2	(6, 2)	4	(8, 4)	6
(2, 6)	4	(4, 2)	3	(8, 8)	8
(2, 8)	5	(4, 6)	5		
(2, 4)	3	(4, 4)	4		
(6, 4)	5	(4, 8)	6		
(6, 8)	7	(8, 2)	5		
(6, 6)	6	(8, 6)	7		

$\bar{x}$	f
2	1
3	2
4	3
5	4
6	3
7	2
8	1
Total	16



11/Apr/19

→ Central Limit Theorem:

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

Eg#02 $\mu = 25$, $\sigma = 3$, $n = 20$.

$$P(\bar{x} > 26.3) = 1 - P(\bar{x} < 26.3)$$

$$= 1 - P\left(z < \frac{26.3 - 25}{3/\sqrt{20}}\right)$$

$$= 1 - P(z < 1.94)$$

$$= 1 - 0.9738$$

$$P(\bar{x} > 26.3) = 0.0262$$

Eg#

$\mu = 96$ months / 8 year , $\sigma = 16$.

$n = 36$

$$P(90 < \bar{x} < 100) = P\left(\frac{90-96}{16/\sqrt{36}} < z < \frac{100-96}{16/\sqrt{36}}\right)$$

$$= P(-2.25 < z < 1.5).$$

$$= P(1.5) - P(-2.25)$$

$$P(90 < \bar{x} < 100) = 0.9210.$$

→ Finite Population Correction Factor:

(Sampling without replacement)

$$\sqrt{\frac{N-n}{N-1}}$$

where n = sample size.

N = population size.

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

→ Continuous Prob Dist → continuous → discrete

→ Normal Approximation to the Binomial Distribution:

If $\bar{X} = np \geq 5$ and $\bar{X} = nq \geq 5$ and n is very large and p is small then Apply normal distribution.

Eg #01 6% read ⇒ $P = 0.06$

$n = 300$, $x = 25$.

* Case of binomial solved using normal distribution

$np = 300(0.06)$
$= 18$

$nq = 300(0.94)$
$= 282$

Discrete → continuous

$P(X = 25) = P(24.5 < x < 25.5)$

$= P\left(z < \frac{25.5 - 18}{\sqrt{300(0.06)(0.94)}}\right) - P\left(z < \frac{24.5 - 18}{\sqrt{300(0.06)(0.94)}}\right)$

$= P(z < 1.823) - P(z < 1.580)$
$= 0.0227.$

Eg #02 $n = 100$ $\sigma = \sqrt{npq} = 4.665$

$p = 0.320$ $\mu = np = 32$

discrete → continuous

$$P(x \leq 26) = P(x \leq 26.5)$$
$$= P\left(z \leq \frac{26.5 - 32}{4.665}\right)$$
$$= P(z < -1.18)$$
$$= 0.1190.$$

12/Apr/19

Hypothesis Testing :-

Hypothesis is assumption. Statistic decision making rule is hypothesis rule

Decision making techniques for evaluating claims about population.

Three Method of Testing Statistical Hypothesis:-

1. Traditional Method
2. P-value Method.
3. Confidence Interval Method.

① Traditional Method:-

A statistical hypothesis is a conjecture about population parameter. The conjecture may or may not be true.

→ Null Hypothesis → Alternative Hypothesis.

- H_0 : Null Hypothesis :
- H_1 : Alternative Hypothesis : Statement opposite to null hypothesis.

The null parameter symbolizes by H_0 is statistical hypothesis states there is no difference between parameter and a specific value or there is no difference between two parameter (population and sample parameter, mean, variance, SD)

Assume there is no difference between sample mean and population mean i.e nullifying so it is called null hypothesis

→ To state hypothesis reaserchers must translate conjecture or claim from words to mathematical symbol.

$=, >, \neq, <$

→ The null hypothesis contains the equal sign.

Two-Tailed test	Right-tailed test	Left tailed test
H_0, $u = K$	H_0, $u = K$	H_0, $u = K$
H_1, $u \neq K$	H_1, $u > K$	H, $u < K$

Level of Significance (α) : After stating hypothesis select correct statistical test choosing an appropriate α.

(0.05, 0.10, 0.01)

(1.21) Types of Decision:

P (type I) = α

	H_0 (true)	H_0 (false)
Reject H_0	Error	Correct
Donot reject H_0	Correct	Error

→ P (type II) = β

Types of Test:

1. z - test.
2. t - test
3. χ^2 - test of independence
4. F - Test of ANOVA analysis variance.

Acceptance & Critical Regions:

15/Apr/19

1. z - test for mean.

$n>30$ and σ is known.

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

eg #01

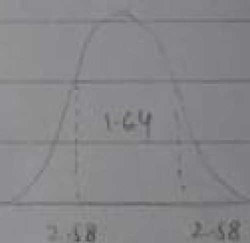
$H_0 \Rightarrow \mu = 1600$

$H_1 \Rightarrow \mu \neq 1600$ (Two-tail test)

$\alpha = 0.01$

$\sigma = 80$ hrs

From Table



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

$$= \frac{1576 - 1600}{80/\sqrt{30}}$$

$z = -1.64$ Donot reject H_0.

eg #02

$\sigma = 30$ $\alpha = 0.05$

$\bar{x} = 110$

$H_0 \Rightarrow \mu = 100$ $H_1 \Rightarrow \mu > 100$

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

$$= \frac{110 - 100}{30/\sqrt{16}}$$

$z = 1.333$

Do not reject H_0.

eg #03

$n = 36$. $H_0 : \mu \geq 80$

$\alpha = 0.10$ $H_1 : \mu < 80$

$\sigma = 19.2$

$\bar{x} = 75$

$$z = \frac{75 - 80}{19.2/\sqrt{36}}$$

$z = -1.56$

Rejected.

	0.10	0.05	0.01
$\mu > \mu_0 : \mu < \mu_0$	+1.28, −1.28	+1.64, −1.64	+2.33, −2.33
$\mu_1 \neq \mu_0$	+1.64, −1.64	+1.96, −1.96	+2.58, −2.58

eg#04 $\bar{x} = 5950$ $H_1: \mu > 5700$

$n = 36$ $H_0: \mu \leq 5700$

$\alpha = 0.05$

$\sigma = 659$

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

$$= \frac{5950 - 5700}{659 / \sqrt{36}}$$

$$z = 2.27$$

$P(z = 2.27) = 0.9884$

$P(z > 2.27) = 1 - 0.9884$

P-value $= 0.0116$. Rejected.

NOTE: If p-value $\geq \alpha \Rightarrow$ Donot Reject H_0.

If p-value $< \alpha \Rightarrow$ Reject H_0

16/Apr/19

eg#05 $\bar{x} = 8.2$, $n = 32$, $\alpha = 0.05$, $\sigma = 0.6$

$H_0 \Rightarrow \mu = 8$ (Two tailed test).

$H_1 \Rightarrow \mu \neq 8$

$$z = \frac{8.2 - 8}{0.6/\sqrt{32}} = 1.8856.$$

$P(z = 1.88) = 1 - 0.9699 = 0.0301$

$P(-z) = 2(0.0301) = 0.0602$

P-value $> \alpha$ $\Rightarrow$ Do not reject H_0.

→ Confidence Interval on μ when σ is known:

statistics $\pm$ $Z_{\alpha/2}$ (S.E)

$\boxed{\sigma/\sqrt{n}}$

eg#5 (continued) $8.2 - (1.96)\boxed{(0.1060)} < \mu < 8.2 + 1.96(0.1060)$

$7.99224 < \mu < 8.40776$.

(1-0.05) 95% confident that μ lie between this interval.

Accept.

→ Test of Difference between two means

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- sample size < 30
  & standard deviation known.

$$(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

9/11/06

$H_0 : \mu_1 = \mu_2$ (There is no significant difference)

$H_1 : \mu \neq \mu_2$ (There is significant difference).

$\alpha = 0.05$. $n = 50$.

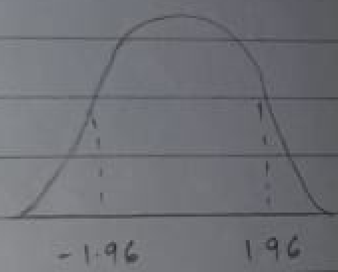
$\sigma_1 = 5.62$ $\sigma_2 = 4.83$.

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{(88.42 - 80.61) - 0}{\sqrt{\frac{(5.62)^2}{50} + \frac{(4.83)^2}{50}}}$$

$= 7.45$.

Rejected.



7.45 is greater

$$(88.42 - 80.61) - 1.96\sqrt{\frac{(5.62)^2}{50} + \frac{(4.83)^2}{50}} < \mu_1 - \mu_2 <$$

$$(88.42 - 80.61) + 1.96\sqrt{\frac{(5.62)^2}{50} + \frac{(4.83)^2}{50}}$$

$7.81 - 1.96(1.047) < \mu_1 - \mu_2 < 7.81 + 1.96(1.047)$

$5.755 < \mu_1 - \mu_2 < 9.864$.

(5) (10)

Date 18/Apr/19

eg#05 $n_1 = 50$, $\sigma_1 = 7.35$, $\bar{x}_1 = 181$
$n_2 = 72$, $\sigma_2 = 4.81$, $\bar{x}_2 = 176$
$\alpha = 0.05$

$H_0: \mu_0 = \mu_1$ $H_1: \mu_0 \neq \mu_1$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n}} = \frac{(181 - 176) - 0}{\sqrt{(7.35)^2/50 + (4.81)^2/72}}$$

$$z = 4.223$$

Now compare with z-critical
H_0 rejected, since $z = 4.223$ doesn't lie between -1.96 & 1.96

eg#06 $n_1 = 50$, $\sigma_1 = 6.28$, $\bar{x}_1 = 86.7$
$n_2 = 50$, $\sigma_2 = 5.61$, $\bar{x}_2 = 77.8$
$\alpha = 0.01$

$H_0: \mu_0 - \mu_1 \geq 12$ $H_1: \mu_0 - \mu_1 < 12$

$z = -2.603$ H_0, rejected.

Confidence Interval:
$(86.7 - 77.8) - (2.33)(1.190) < \mu_0 - \mu_1 < (86.7 - 77.8) + (2.33)(1.190)$
$6.127 < \mu_0 - \mu_1 < 11.67$
hence H_0 rejected.

Date

→ T-test for Mean: (t-distribution)

* Similarity between T-test & normal distribution:
  1. Bellshaped
  2. Symmetric about mean
  3. Mean = median = mode
  4. Never touches x-axis.

$$h(t) = \frac{\Gamma(\nu+1)/2}{\Gamma\nu/2\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right) \qquad -\infty < t < \infty$$

ν → Degree of freedom = No. of independent − restricted

* Difference:
  1. Variance > 1
  2. Depends on sample size.
  3. Population variance unknown.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \text{and} \quad \nu = n - 1$$

Confidence Intervals

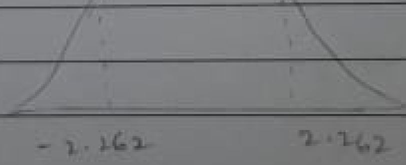
$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

eg #08 $H_0: \mu = 16.3$ $\quad H_1: \mu \neq 16.3$

$\alpha = 0.05$ $\quad \bar{x} = 17.7$ $\quad s = 1.8$ $\quad n = 10$

$\nu = n - 1 = 10 - 1 = 9.$

$$t = \frac{17.7 - 16.3}{1.8/\sqrt{10}} = 2.46$$



Date

t - critical = 2.262 → from T-table

H_0 Rejected.

eg#09

$\alpha = 0.10$, $n = 8$, $\bar{X} = 58.875$, $S = ?$, $\nu = 8 - 1 = 7$.

$H_0 \Rightarrow \mu \geq 60$, $H_1 \Rightarrow \mu < 60$

$S = \sqrt{\Sigma(x-\bar{x})^2/n-1} = \sqrt{180.875/7} = 5.08$

$t = \dfrac{58.875 - 60}{5.08/\sqrt{8}} = -0.626 \Rightarrow$ t - critical $= -1.415$

Do not reject H_0.

eg#10 $\alpha = 0.05$, $n = 15$, $\bar{X} = 40.6$, $\sigma = 6$

H_0: $\mu \leq 36.7$ H_1: $\mu > 36.7$

$\nu = n - 1 = 15 - 1 = 14$

$t = \dfrac{40.6 - 36.7}{6/\sqrt{15}} = 2.517$

t - critical $= 1.76 \Rightarrow H_0$ Rejected.

19/Apr/19 Case I: Two Mean Independent T-test:

Testing difference between two mean (independent sample test):

① Variances are unequal & independent:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

Assumptions:-

① Samples are random samples.

② Independent sample

③ Sample size < 30, population must be normally or approximately normally distributed.

Confidence interval (Independent Sample)

$$(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

ν = smaller of $n_1 - 1$ and $n_2 - 1$.

eg # 11 $n_1 = 8$, $S_1 = 38$, $\bar{X}_1 = 191$

$n_2 = 10$, $S_2 = 12$, $\bar{X}_2 = 199$

$\nu = 8 - 1 = 7$ (b/c 7 is smaller than 9).

$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$.

(130)

$$t = \frac{(191 - 199) - 0}{\sqrt{38^2/8 + 12^2/10}} = -0.573$$

t critical $= 2.365$

Do not reject H_0.

Date

Confidence interval:

$\alpha = 0.05$

$(191-199) - (2.365)(13.96) < \mu_1 - \mu_2 < (191-199) + (2.365)(13.96)$

$-41.0154 < \mu_1 - \mu_2 < 25.0154$

Since 0 lies between this interval do not reject H_0.

eg#12 $\bar{x}_1 = 10.26$, $s_1 = 8.56$

$\bar{x}_2 = 6.93$, $s_2 = 4.93$

$\nu = 56$

Confidence interval $= -0.18979, 6.84979$

$t = 1.89566$

t-critical $= -2.0037, 2.0037$

p-value $= 0.06317$

$\alpha = 0.05$

① Are sample dependent or independent?

Independent.

② Which value is compared with α?

p-value.

③ Type 1 error?

P-(type I) $= \alpha$ (level of significance)

④ Right, left or two tailed test?

Two tailed test since two critical values are given.

Date

5) Conclusion:
Do not reject Ho b/c p-value > α and conclude that there is difference in

6) If initially $\alpha = 0.10$
then reject Ho => pvalue < α

Q#13 $H_0: \mu_1 = \mu_2$ $\alpha = 0.05$ $n = 10$
$H_1: \mu_1 \neq \mu_2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \Rightarrow S = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n-1}}$$

$\bar{x}_1 = 36.7$
$\bar{x}_2 = 33.78$

$(x_1 - \bar{x}_1)^2$	$(x_2 - \bar{x}_2)^2$	
22.09		$S_1^2 = 7.344$
1.69		$S_2^2 = 5.943$
0.09		
0.49		$t = \frac{(36.7 - 33.78) - 0}{\sqrt{7.344/10 + 5.943/9}}$
0.49		
7.29		$= 2.48$
5.29		
0.49		
0.09		t-critical = ± 2.306
28.09		
66.1		Reject Ho

Date

$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(S.E)$

$2.92 - (2.306)(1.1809) < \mu_1 - \mu_2 < 2.92 + (2.306)(1.1809)$

$0.197 < \mu_1 - \mu_2 < 5.643$

- Testing difference between two mean when $\sigma_1 = \sigma_2$. (Independent Sample t-test).

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

S.D $\longrightarrow$ d

eg#14 $H_0: \mu_1 - \mu_2 \leq 3 \qquad H_1: \mu_1 - \mu_2 > 3$

$\alpha = 0.10 \quad , \quad \sigma_1 = \sigma_2$

$\bar{x}_1 = 50.272 \quad , \quad \bar{x}_2 = 37.833$

$s_1^2 = \cancel{2720.6656} \quad , \quad s_2^2 = 3815.1211$

52.16 61.766

$$t = \frac{(50.272 - 37.833) - 3}{\sqrt{\frac{(11-1)(\cancel{2720.6656}) + (6-1)(3815.1211)}{11 + 6 - 2}}\sqrt{\frac{1}{11} + \frac{1}{6}}}$$

52.16 61.766

$= 0.2385$

$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(S.E)$

$(12.439) - (0.10)(\quad) < \mu_1 - \mu_2 < (12.439) + (0.10)($

$< \mu_1 - \mu_2 <$

Date

→ Testing Difference between two means (Dependent Samples)

Two-tailed	Left-tailed	Right-tailed
$H_0: \mu_D = 0$	$H_0: \mu_D = 0$	$H_0: \mu_D = 0$
$H_1: \mu_D \neq 0$	$H_1: \mu_D < 0$	$H_1: \mu_D > 0$

eg#15 $n = 9$, $\alpha = 0.05$

X_1	X_2	$(X_1 - X_2)^2$
11.42	16.69	27.772
8.41	9.44	1.060
3.98	6.53	6.502
7.37	5.58	3.204
2.28	2.92	0.409
1.10	1.88	0.608
1.00	1.78	0.608
0.90	1.50	0.36
1.35	1.22	0.0169
		40.5399

$$\bar{D} = \frac{\Sigma(x_1 - x_2)}{n} = \frac{-9.73}{9} =$$

$$S_2^2 = \frac{\Sigma(D - \bar{D})^2}{n_d - 1}$$

Date

eg#16 $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$

$\alpha = 0.10$

x_1	x_2	$d = x_1 - x_2$
210	190	
235	170	
208	210	
190	188	
172	173	
244	228	

$\bar{d} = 16.67$

$s_d = 25.39$

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n}}$$

$$= \frac{16.67 - 0}{25.39/\sqrt{6}}$$

$t = 1.608$

$\nu = 6 - 1 = 5$

$t_{(5)} = \pm 2.015$

Do not reject H_0.

Correlation And Regression

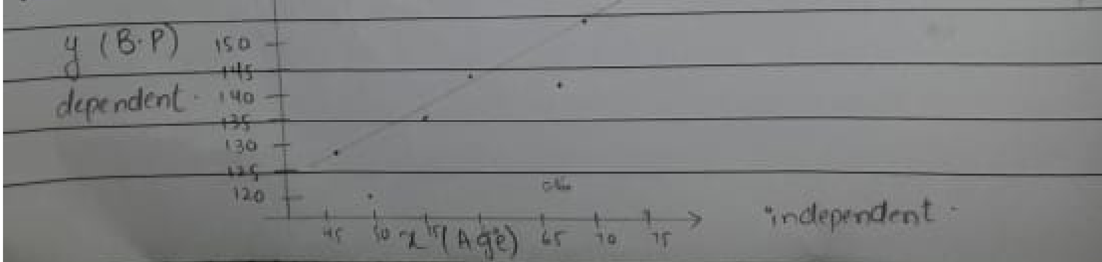
→ Determining whether a relationship between two or more numerical or quantitative variable exists.

→ Independent variable: Can be controlled or manipulated.

→ Dependent variable: Cannot be controlled or manipulated.

Scatter Plot

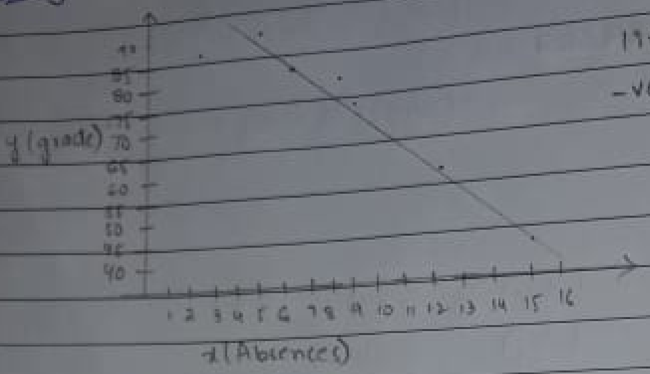
eg#01



136

Date ________

eg Scatter Plot



linear

-ve → indirect relationship

→ Correlation Coefficient:

Pearson Product moment correlation coefficient.

- Symmetric, no unit, range -1 to 1.

* Sample correlation coefficient

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

Where $r = \frac{Cov(x,y)}{\sigma x \, \sigma y}$ and $Cov(x,y) = E(x,y) - E(x)E($

$\sigma^2 x = E(x^2) - [E(x)]^2$.

eg#03

Subject	xy	x^2	y^2
A	5504	1849	16384
B	5760	2304	14400
C	7560	3136	18225
D	8723	3721	20449
E	9447	4489	19881
F	10640	4900	23104
	47634	20399	112443.

Date

$$r = \frac{6(47634) - (345)(819)}{\sqrt{[6(20399) - (345)][6(112443) - (819)]}}$$

$r = 0.89$

near to 1 $\Rightarrow$ strong relationship.

25/April/19

eg#04

$\Sigma(xy) = 3745$

$\Sigma x = 57$ $(\Sigma x)^2 = 3249$

$\Sigma y = 511$ $(\Sigma y)^2 = 261121$

$n = 7$

$\Sigma x^2 = 579$ $\Sigma y^2 = 38993$.

$$r = \frac{7(3745) - (57)(511)}{\sqrt{[7(579) - 3249][7(38993) - 261121]}} = -0.9442.$$

strong inverse relationship

→ Significance of Correlation:

Hypothesis testing by t-test.

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

$\nu = n - 2$.

where $H_0: \rho = 0$

$H_1: \rho \neq 0$.

Date

eg #05 $n = 6$, $r = 0.89$, $\nu = 4$

$$t = 0.89\sqrt{\frac{6-2}{1-(0.89)^2}} = 3.90.$$

let $\alpha = 0.05$, t-critical $= \pm 2.776$

Reject the null hypothesis

hence result is significant.

→ Possible relationship between variables

When null hypothesis has been rejected, then

1. Direct cause & effect relationship. (x causes y)
2. Reverse cause & effect relationship. (y causes x)
3. Relationship between variables may be caused by 3rd variable.

→ Regression:

- Applied when correlation is significant.
- Regression line is line equation.

"Dependence of one variable over another variable is called Regression."

- Estimates the value of dependent variable from unknown independent variable.

→ Simple linear Regression model

$$Y = \beta_0 + \beta_1 x_i + \epsilon_i$$ → then mathematical model / statistical model.

ϵ = error / residual variance / disturbance term.

$E(\epsilon) = 0$

139

Date

Y - random , X = regressor variable (negligible error)

26/Apr/19

→ Method of Least square

The estimation of α and β should have minimum error

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

$$y = \alpha + \beta x + \epsilon$$

$$a = \bar{Y} - b\bar{X}$$

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

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

→ Method

Estimate β_0 and β_1

SSE - Sum of square of error

$$SE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 =$$

$$\sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$$

Date

Differentiate SSE w.r.t b_0 & b_1

$$\partial(SSE)/\partial b_0 = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)$$

$$\partial(SSE)/\partial b_1 = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)\,x_i$$

// partial derivative

$$n b_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i$$

eq = 0

$$\beta_0 = b_0 = a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2} \quad - \text{ intercept}$$

$$\beta_1 = b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} \quad = \text{ slope}$$

eg # 06 (Blood pressure) → eg # 3

(iv) $\hat{y} = \beta_0 + \beta_1 x + e$

$P = \alpha + \beta(\text{Ages}) + \epsilon$.

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

$$\beta = \frac{6(47634) - (345)(819)}{6(20399) - (345)^2}$$

$\beta = 0.964$ (Sign of β & correlation should be same).

141

Date

$a = \alpha = \bar{y} - \beta \bar{x}$

$= 136.5 - 0.96(57.5)$

$a = \alpha = 81.07$

$P = 81.07 + (0.964)(Ages) + \epsilon$

Another formula of α

$$\alpha = \frac{(819)(20399) - (345)(47634)}{6(20399) - (345)^2} = 81.04$$

Estimated regression is

$\hat{P} = 81.07 + (0.964)(Ages) + \epsilon$

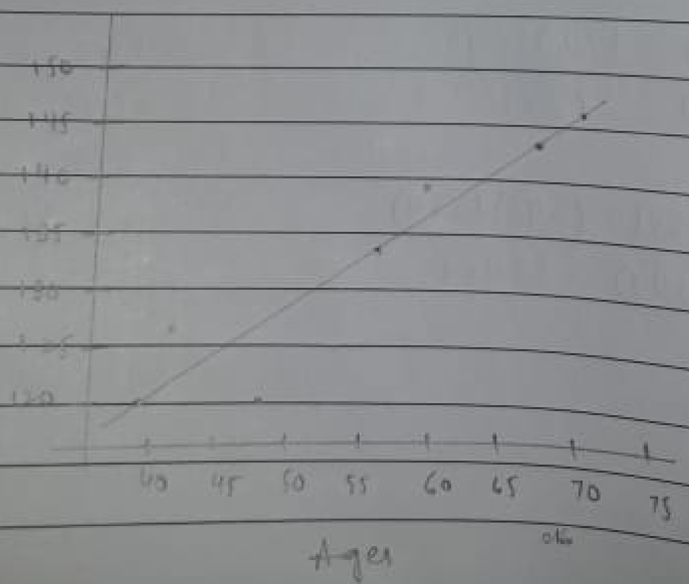
if $x = 71$ & ignore ϵ

$\hat{P} = 149.5$

If x increase by one year than on avg. pressure will increase by 0.964 units.

Scatter Plot:

$\hat{P} = 81.07 + 0.964(Ages)$



Date 29/Apr/19

Q#07 $y = \alpha + \beta x + \epsilon$

$\Sigma x = 1104$ $\Sigma y = 1124$

$n = 33$ $\Sigma xy = 41355$ $\Sigma x^2 = 41,086$

$\bar{y} = 34.06$ $\bar{x} = 33.45$

$\alpha = \bar{y} - \beta\bar{x}$

$$\beta = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2} = \frac{(33)(41355) - (1104)(1124)}{(33)(41086) - (1104)^2}$$

$\beta = 0.9036$

$\alpha = 34.06 - 0.9036(33.45)$

$\alpha = 3.834$

$y = 3.834 + 0.903x + \epsilon$

Multiple Linear Regression.

Date

→ Several independent & one dependent.

$$y' = a + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k.$$

→ Normal equations for Regression coefficient for two independent variables.

$$\Sigma y = na + b_1 \Sigma x_1 + b_2 \Sigma x_2$$

$$\Sigma x_1 y = a \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2.$$

$$\Sigma x_2 y = a \Sigma x_2 + b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2.$$

$$b_1 = \frac{(\Sigma x_1 y)(\Sigma x_2^2) - (\Sigma x_2 y)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

$$b_2 = \frac{(\Sigma x_2 y)(\Sigma x_1^2) - (\Sigma y x_1)(\Sigma x_1 x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1 x_2)^2}$$

eg#01

$n = 5$

$y = \alpha + \beta_1 x_1 + \beta_2 x_2$

$\Sigma x_1 = 14$ $\Sigma x_1^2 = 40.18$ $\Sigma x_1 y = 7967.5$

$\Sigma x_2 = 124$ $\Sigma x_2^2 = 3102$ $\Sigma x_2 y = 70120$

$\Sigma x_1 x_2 = 349.1$

$$b_1 = \frac{(7967.5)(3102) - (70120)(349.1)}{(40.18)(3102) - (349.1)^2}$$

$$= 85.379.$$

Date

$$b_2 = \frac{(70120)(40.18) - (7967.5)(349.1)}{(40.18)(3102) - (349.1)^2}$$

$= 12.99$

$a = \alpha = \frac{\Sigma y}{} =$

$y' = -44.572 + 87.679x_1 + 14.519x_2$

If $x_1 \uparrow$ by one unit than on avg $y \uparrow$ by b_1 on it while keeping the effect of x_2 constant.

→ Multiple correlation

Range → 0 - 1

Co. of determination → $r^2 \times 100$.

if value is closer to 0, weak relationship.

$$R = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}}}$$