ASSIGNMENT

Piyush Soni [SBU ID # 113273375]

Abhishek Bhattad [SBUID# 113277131]

Aditya Nandan Bhide [SBU ID# 113216565]

Niket Shah [SBU ID# 113274949]

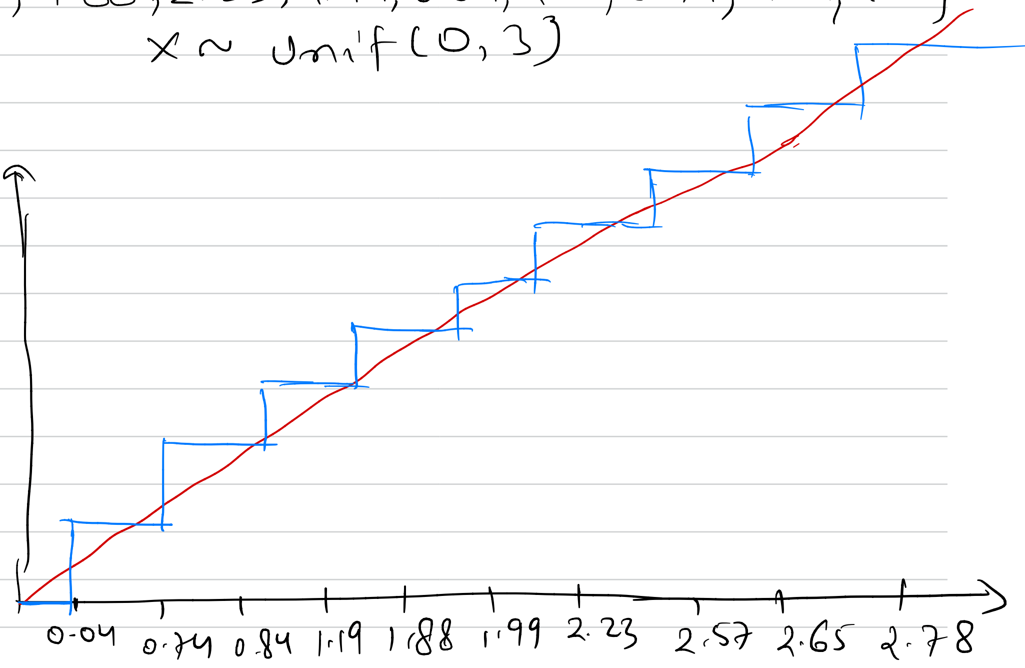
A-5

(1) $D = \{2.78, 0.84, 1.88, 2.23, 1.99, 0.04, 2.65, 0.74, 1.19, 2.57\}$

$X \sim Unif(0, 3)$

$H_0: F_D \overset{d}{=} Unif(0, 3)$

$H_1: F_D \overset{d}{\neq} Unif(0, 3)$



x	$F_x(x)$	$\hat{F}_D^-(x)$	$\hat{F}_D^+(x)$	d_1: $\lvert F_x(x) - \hat{F}_D^-(x)\rvert$	d_2: $\lvert F_x(x) - \hat{F}_D^+(x)\rvert$	max(d_1, d_2)
0.04	$0.04/3 = 0.013$	0	0.1	$\lvert 0.013 - 0\rvert = 0.013$	0.087	0.087
0.74	$0.74/3 = 0.246$	0.1	0.2	$\lvert 0.246 - 0.1\rvert = 0.146$	0.046	0.146
0.84	$0.84/3 = 0.28$	0.2	0.3	$\lvert 0.28 - 0.2\rvert = 0.08$	0.02	0.08
1.19	$1.19/3 = 0.396$	0.3	0.4	$\lvert 0.396 - 0.3\rvert = 0.096$	0.004	0.096
1.88	$1.88/3 = 0.626$	0.4	0.5	$\lvert 0.626 - 0.4\rvert = 0.226$	0.126	0.226
1.99	$1.99/3 = 0.66$	0.5	0.6	$\lvert 0.66 - 0.5\rvert = 0.16$	0.06	0.16
2.23	$2.23/3 = 0.743$	0.6	0.7	$\lvert 0.743 - 0.6\rvert = 0.143$	0.043	0.143
2.57	$2.57/3 = 0.856$	0.7	0.8	$\lvert 0.856 - 0.7\rvert = 0.156$	0.056	0.156
2.65	$2.65/3 = 0.883$	0.8	0.9	$\lvert 0.883 - 0.8\rvert = 0.083$	0.017	0.083
2.78	$2.78/3 = 0.926$	0.9	1.0	$\lvert 0.926 - 0.9\rvert = 0.026$	0.074	0.074

$\Rightarrow d = \max_{x} \left(|F_X(x) - \hat{F}_D(x)| \right)$

$= 0.226$

$\Rightarrow$ Given threshold, $\underline{c = 0.25}$

Since, $\underline{d < c}$

thus, H_0 is true

$\therefore \boxed{F_D = Unif(0, 3)}$

(2) $X = \{5\}$, $Y = \{2, 7\}$
(D_1) (D_2)

$H_0 : X \overset{d}{=} Y$ $\quad P_{threshold} = \underline{0.05}$

$H_1 : X \overset{d}{\neq} Y$; $N = |X| + |Y| = 1 + 2 = \underline{3}$

$\Rightarrow T_{obs} = |\bar{D}_1 - \bar{D}_2|$

$= |5 - 9/2| = \underline{0.5}$

i	D_1^i	D_2^i	$\lvert\bar{D_1^i} - \bar{D_2^i}\rvert = T_i$	$I(T_i > T_{obs})$
1	$\{5\}$	$\{2, 7\}$	$\lvert 5 - 9/2\rvert = 0.5$	0
2	$\{5\}$	$\{7, 2\}$	$\lvert 5 - 9/2\rvert = 0.5$	0
3	$\{7\}$	$\{2, 5\}$	$\lvert 7 - 7/2\rvert = 3.5$	1
4	$\{7\}$	$\{5, 2\}$	$\lvert 7 - 7/2\rvert = 3.5$	1
5	$\{2\}$	$\{5, 7\}$	$\lvert 2 - 6\rvert = 4$	1
6	$\{2\}$	$\{7, 5\}$	$\lvert 2 - 6\rvert = 4$	1

$\Rightarrow P = \dfrac{\sum I(T_i > T_{obs})}{N!} = \dfrac{4}{6} = \dfrac{2}{3} = \underline{0.67}$

Since, $P > P_{th}$ thus H_0 is ACCEPTED!!

(3) H_0: Outcome $\perp$ dealer

(a) $\alpha = 0.05$

	A	B	C	Total
Win	48	54	19	121
Draw	7	5	4	16
Loose	55	50	25	130
Total	110	109	48	267

$$E_{11} = \frac{121}{267} \times 110 = 49.85$$

$$E_{12} = \frac{121}{267} \times 109 = 49.39$$

$$E_{13} = \frac{121}{267} \times 48 = 21.75$$

$$E_{21} = \frac{16}{267} \times 110 = 6.59$$

$$E_{22} = \frac{16}{267} \times 109 = 6.53$$

$$E_{23} = \frac{16}{267} \times 48 = 2.88$$

$$E_{31} = \frac{130}{267} \times 110 = 53.55$$

$$E_{32} = \frac{130}{267} \times 109 = 53.07$$

$$E_{33} = \frac{130}{267} \times 48 = 23.37$$

$$Q_{obs} = \frac{(E_{11}-O_{11})^2}{E_{11}} + \frac{(E_{12}-O_{12})^2}{E_{12}} + \frac{(E_{13}-O_{13})^2}{E_{13}}$$

$$+ \frac{(E_{21}-O_{21})^2}{E_{21}} + \frac{(E_{22}-O_{22})^2}{E_{22}} + \frac{(E_{23}-O_{23})^2}{E_{23}}$$

$$+ \frac{(E_{31}-O_{31})^2}{E_{31}} + \frac{(E_{32}-O_{32})^2}{E_{32}} + \frac{(E_{33}-O_{33})^2}{E_{33}}$$

$$= \frac{(49.85-48)^2}{49.85} + \frac{(49.39-54)^2}{49.39} + \frac{(21.75-19)^2}{21.75}$$

$$+ \frac{(6.59-7)^2}{6.59} + \frac{(6.53-5)^2}{6.53} + \frac{(2.88-4)^2}{2.88}$$

$$+ \frac{(53.55-55)^2}{53.55} + \frac{(53.07-50)^2}{53.07} + \frac{(23.37-25)^2}{23.37}$$

$$= 0.0686 + 0.430 + 0.3477 + 0.0255$$
$$+ 0.358 + 0.4356 + 0.039$$
$$+ 0.177 + 0.113$$

$$\boxed{Q_{obs} = 1.995}$$

$\Rightarrow$ Degree of freedom (df) $= (3-1)(3-1)$

$= 2 \times 2$

$= \underline{\underline{4}}$

$\Rightarrow$ p-value $= Pr(\chi^2_{df} > Q_{obs})$

$= Pr(\chi^2_4 > 1.995)$

$= 1 - cdf(1.995)$

$= 1 - 0.2635$

$= \underline{\underline{0.7365}}$

Since, P-value $> 0.05\ (\alpha)$

Accept Ho, i.e. outcome of tables and the dealer are independent

(b) Dealer -

$A = \{48, 40, 58, 53, 65, 25, 52, 34, 30, 45\}$

$B = \{54, 48, 51, 47, 62, 35, 70, 20, 25, 40\}$

$C = \{19, 40, 35, 41, 38, 32, 32, 37, 37, 15\}$

$\bar{A} = 45$

$\bar{B} = 45.2$

$\bar{C} = 32.6$

$$\Rightarrow \hat{\rho}_{A,B} = \frac{\Sigma(A-\bar{A})(B-\bar{B})}{\sqrt{\left(\Sigma(A-\bar{A})^2\right)\left(\Sigma(B-\bar{B})^2\right)}}$$

$$= \frac{1396}{\sqrt{1462 \times 2193.6}}$$

$$= \underline{\underline{0.7795}}$$

$$\therefore \hat{\rho}_{A,B} > 0.5$$

Thus, dealers A and B are positive linearly correlated.

$$\Rightarrow \hat{\rho}_{B,C} = \frac{\Sigma(B-\bar{B})(C-\bar{C})}{\sqrt{\left(\Sigma(B-\bar{B})^2\right)\left(\Sigma(C-\bar{C})^2\right)}}$$

$$= \frac{-96.2}{\sqrt{2193.6 \times 694.4}}$$

$$= -0.07794$$

$$\Rightarrow -0.5 < \rho_{B,C} < 0.5$$

B and C dealers are uncorrelated

$$\hat{\rho}_{C,A} = \frac{\Sigma(A-\bar{A})(C-\bar{C})}{\sqrt{(\Sigma(A-\bar{A})^2)(\Sigma(C-\bar{C})^2)}}$$

$$= \frac{22}{\sqrt{694.4 \times 1462}} = \underline{\underline{0.02183}}$$

$$\therefore \quad -0.5 < \hat{\rho}_{C,A} < 0.5$$

→ Thus, dealers C and A are uncorrelated

→ Since, dealer C is uncorrelated to both A and B, If there is no bias than the result should not depend on the dealers i.e., A, B and C should linearly correlated to each other.

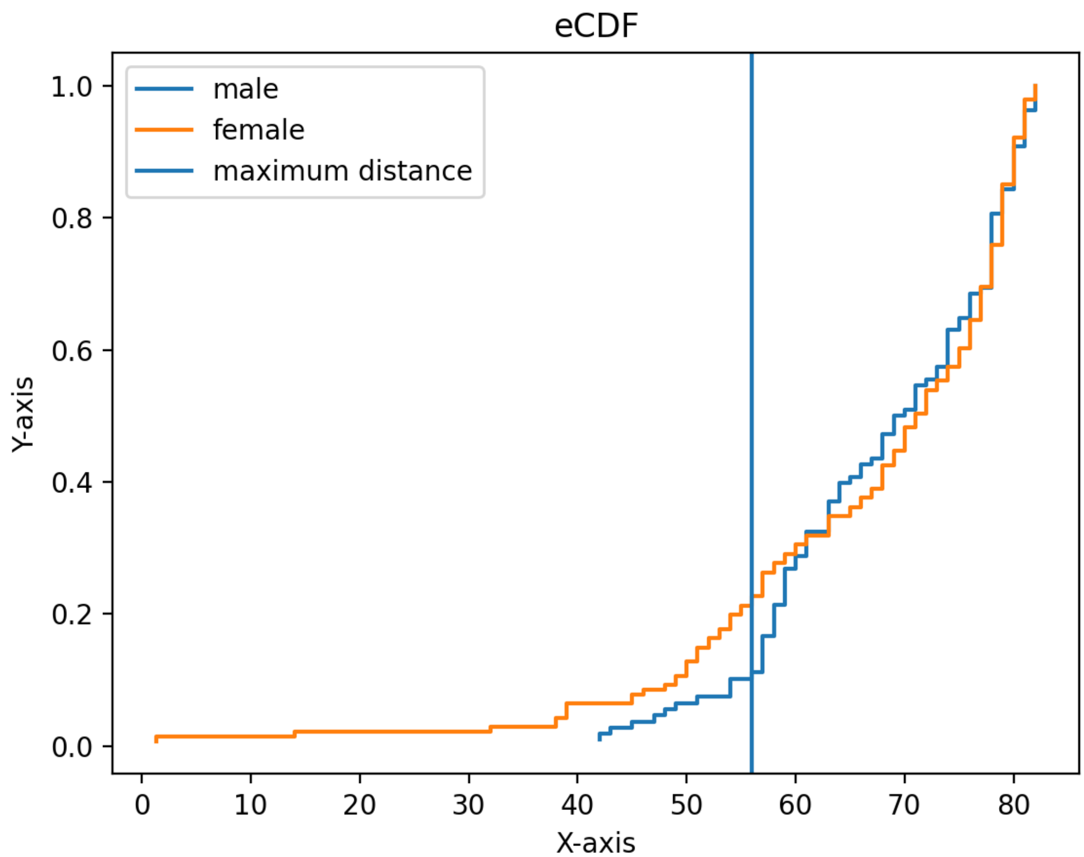
Thus, dealer 'C' is not loyal and could be the reason for loss of money.

4 a] n = 200
p-value = 0.0
Reject H_0
People getting stroke will not have the same glucose level as people who do not get stroke

n = 1000
p-value = 0.0
Reject H_0
People getting stroke will not have the same glucose level as people who do not get stroke

4 b] p-value = 0.404
Accept H_0
Female patients get a stroke at the same age as male patients

4 c]Maximum distance: 0.11800630417651693
Age at maximum distance: 56.0



Accept H_0
Female patients get a stroke at the same age as male patients

⑤ ⓐ $\{X_1, \ldots X_n\} \overset{iid}{\sim} Nor(\mu_1, \sigma_1^2) \rightarrow S_1$ (std. dev.)

$\{Y_1, \ldots Y_m\} \overset{iid}{\sim} Nor(\mu_2, \sigma_2^2) \rightarrow S_2$

$\Rightarrow X \perp Y$

$H_0 : \mu_1 > \mu_2$ $\qquad$ $\delta > 0 \rightarrow$ critical value

$H_1 : \mu_1 \leq \mu_2$

$\Rightarrow$ Let, $D = \mu_1 - \mu_2$

Then, $H_0 : D > 0$

$H_1 : D \leq 0$

Now, $\bar{D} = \bar{X} - \bar{Y}$

$$\Rightarrow E[\bar{D}] = E[\bar{X} - \bar{Y}]$$

$$\overset{LOE}{=} E[\bar{X}] - E[\bar{Y}] = E\left[\frac{\sum X_i}{n}\right] - E\left[\frac{\sum Y_i}{m}\right]$$

$$\overset{LOE}{=} \frac{\sum E[X_i]}{n} - \frac{\sum E[Y_i]}{m}$$

$$= \frac{\cancel{n} E[X_1]}{\cancel{n}} - \frac{\cancel{m} E[Y_1]}{\cancel{m}}$$

$$= E[X_1] - E[Y_1]$$

$$\boxed{E[\bar{D}] = \mu_1 - \mu_2}$$

$$\Rightarrow Var(\bar{D}) = Var(\bar{X} - \bar{Y})$$

$$\overset{LOV\ I}{=} Var(\bar{X}) + Var(\bar{Y})$$

$$= Var\left(\frac{\sum X_i}{n}\right) + Var\left(\frac{\sum Y_i}{n}\right)$$

$$\Rightarrow \text{var}(\bar{D}) = \text{var}\left(\frac{\sum x_i}{n}\right) + \text{var}\left(\frac{\sum y_i}{m}\right)$$

$$= \frac{1}{n^2}\text{var}\left(\sum x_i\right) + \frac{1}{m^2}\text{var}\left(\sum y_i\right)$$

$$\overset{\text{cov}}{=} \frac{1}{n^2}\left(\sum \text{var}(x_i)\right) + \frac{1}{m^2}\left(\sum \text{var}(y_i)\right)$$

$$= \frac{1}{n^2} \times \cancel{n} \times \text{var}(x_1) + \frac{1}{m^2} \times \cancel{m} \times \text{var}(y_1)$$

$$= \frac{\text{var}(X_1)}{n} + \frac{\text{var}(Y_1)}{m}$$

$$= \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$$

So,

$$\frac{D}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim \text{Nor}\left(\frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}, 1\right)$$

$$\sim \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} + \text{Nor}(0, 1)$$

Since, σ_x & σ_y are not known, thus we will replace them by sample variances (S_x, S_y)

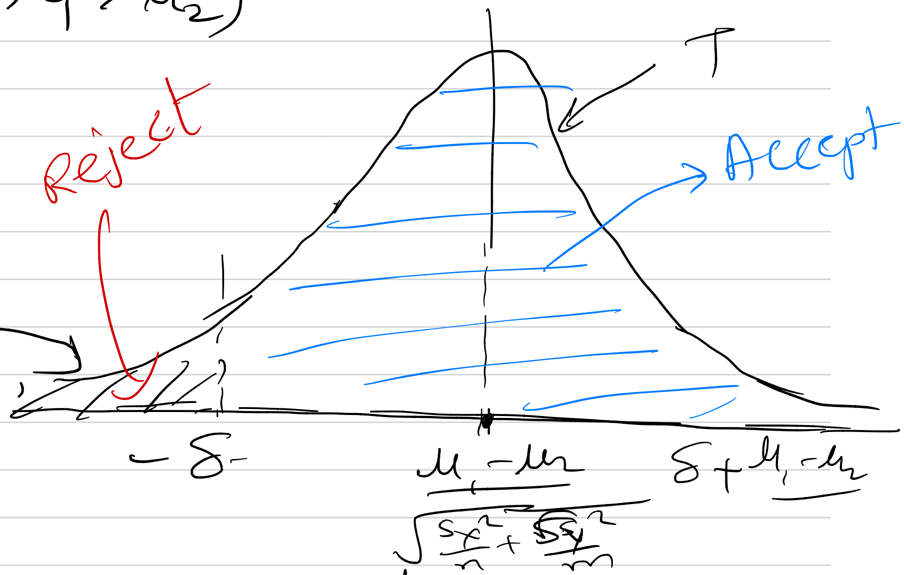
Thus, $$\text{var}(\bar{D}) = \frac{S_x^2}{n} + \frac{S_y^2}{m}$$

$\mu_1 > \mu_2$

$\Rightarrow Pr(\text{Type-I error}) = Pr(\text{reject } H_0 \mid H_0 \text{ true})$

$\Rightarrow \boxed{T < -\delta}$ $\{\text{for } \mu_1 > \mu_2\}$

To reject H_0



$\therefore Pr(\text{Type-I error})$

$= Pr(T < -\delta)$

$= F_T(-\delta)$

$= \Phi\left(-\delta - \dfrac{\mu_x - \mu_y}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}}\right)$

$T \sim \dfrac{\mu_1-\mu_2}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} + Z$

$$\Rightarrow \boxed{Pr(\text{Type-I error}) = \Phi\left(-\delta - \dfrac{\mu_1-\mu_2}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}}\right)}$$

$\Rightarrow Pr(\text{Type-II error}) = Pr(\text{Accept } H_0 \mid H_0 \text{ is false})$

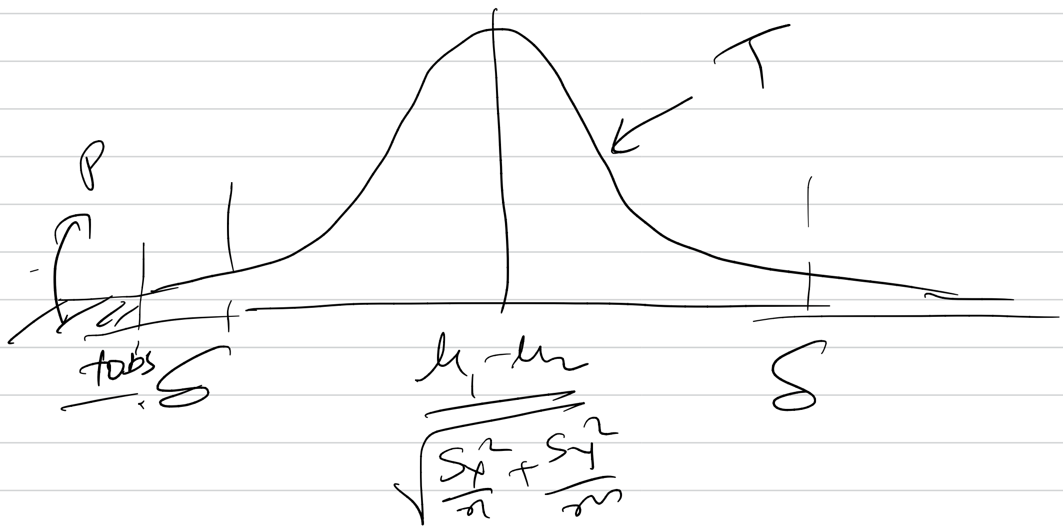
$= Pr(T > -\delta \mid \mu_1 \le \mu_2)$

$= 1 - F_T(-\delta)$

$$\boxed{Pr(\text{Type-II error}) = 1 - \Phi\left(-\delta - \dfrac{\mu_1-\mu_2}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}}\right)}$$

5. (b) p-value = Pr (more extreme statistic than t_0 | H_0 is true)

$$\Rightarrow t_{obs} = \frac{\bar{D}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$



$$\Rightarrow \text{p-value} = Pr(T < t_{obs} \mid \mu_1 > \mu_2)$$

$$= F_T(t_{obs}) \qquad T \sim \frac{\mu_1 - \mu_2}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} + Z$$

$$= \Phi\left(t_{obs} - \frac{\mu_1 - \mu_2}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}\right)$$

$$\Rightarrow \boxed{\text{p-value} = \Phi\left(\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}\right)}$$

6. (a) $X \sim N(1.5, 1)$

$Y \sim N(1, 1)$ $\qquad X \perp Y$

$\Rightarrow \sigma_x = 1, \quad \sigma_y = 1$

$\Rightarrow H_0 : \mu_x = \mu_y$ $\qquad | \quad \alpha = 0.05$

$H_1 : \mu_x \neq \mu_y$

$\Rightarrow$ Z-test

Let, $D = \mu_x - \mu_y$

$\bar{D} = \bar{X} - \bar{Y}$

New,

$H_0 : D = 0$

$H_1 : D \neq 0$

$$\Rightarrow Z = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\text{var}(\bar{X} - \bar{Y})}}$$

$\therefore$ for H_0 to be true $\rightarrow \mu_x = \mu_y \Rightarrow$

$$Z = \frac{\bar{X} - \bar{Y}}{\text{var}(\bar{X} - \bar{Y})}$$

$$\Rightarrow \text{var}(\bar{X} - \bar{Y}) \overset{\text{LOV}}{\underset{\perp}{=}} \text{var}(\bar{X}) + \text{var}(\bar{Y})$$

$$= \frac{1}{n^2}\text{var}(\Sigma X_i) + \frac{1}{m^2}\text{var}(\Sigma Y_i)$$

$$\overset{\text{LOV}}{=} \frac{1}{n^2}\Sigma\,\text{var}(X_i) + \frac{1}{m^2}\Sigma\,\text{var}(Y_i)$$

$$\Rightarrow \mathrm{var}(X_i - Y_i) = \frac{1}{n^2}\sum \mathrm{var}(X_i) + \frac{1}{m^2}\sum \mathrm{var}(Y_i)$$

$$\overset{iid}{=} \frac{1}{n^2} \times n \times \mathrm{var}(X_1) + \frac{1}{m^2} \times m \times \mathrm{var}(Y_1)$$

$$= \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}$$

$$\therefore \quad Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

$$= \frac{1.598739 - 1.062504}{\sqrt{\frac{1}{20} + \frac{1}{20}}} \quad \left\{ \text{Given, } n = m = 20 \right\}$$

$$= \frac{0.533699}{\sqrt{1/10}}$$

$$= 1.68770422$$

$$\therefore \quad |Z| < Z_{\alpha/2}$$

$$\Rightarrow 1.6877 < 1.96$$

Thus, z-test will accept H_0 i.e. $\mu_x = \mu_y$ which would be false as given $\mu_x = 1.5$, $\mu_y = 1 \Rightarrow \boxed{\mu_x \neq \mu_y}$

it didn't perform well.

p-value of z-test –

$$= Pr\left(\begin{matrix}\text{more extreme} \\ \text{statistic than } z_0\end{matrix} \;\middle|\; H_0 \text{ is true}\right)$$

$$= Pr\left(|z| > z_0 \mid \mu_1 = \mu_2 \right)$$

$$= Pr\left(|z| > 1.68 \right)$$

$$= 2\left(1 - cdf(1.68)\right)$$

$$\because \quad z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim Nor(0,1)$$

thus,

$$(p\text{-value})_z = 2\left(1 - \Phi(1.68)\right)$$

$$= 2\left(1 - 0.9535\right)$$

$$= \underline{\underline{0.093}} \quad > 0.05$$

Since, p-value > 0.05, we can confidently accept H_0.

T-test :-

$$D = \mu_x - \mu_y$$

$$\bar{D} = \bar{X} - \bar{Y}$$

$\Rightarrow H_0 : D = 0$ | C (Threshold) $= 2.086$

$H_1 : D \neq 0$

$$\Rightarrow T = \frac{(\bar{X} - \bar{Y}) - D_0}{\sqrt{var(\bar{D})}}$$

$\because H_0$ is $D = 0$, thus $\underline{D_0 = 0}$

$$\Rightarrow T = \frac{(\bar{X} - \bar{Y})}{\sqrt{var(\bar{D})}}$$

$$\Rightarrow var(\bar{D}) = var(\bar{X} - \bar{Y})$$

$$\overset{cov}{\underset{+}{=}} var(\bar{X}) + var(\bar{Y})$$

$$= \frac{1}{n^2} var(\Sigma X_i) + \frac{1}{m^2} var(\Sigma Y_i)$$

$$\overset{cov}{=} \frac{1}{n^2} \Sigma \, var(X_i) + \frac{1}{m^2} \Sigma \, var(Y_i)$$

$$\overset{iid}{=} \frac{n}{n^2} var(X_1) + \frac{m}{m^2} var(Y_1)$$

$$= \frac{var(X_1)}{n} + \frac{var(Y_1)}{m} = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}$$

Since, for T-test σ_x and σ_y are not known thus, we would use plugin estimator. (sample variance)

$$\bar{S}_x^2 = \frac{1}{n-1} \sum (x-\bar{x})^2$$

$$\Rightarrow T = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \quad \left\{ \begin{array}{c} \text{Given} \\ n = m = 20 \end{array} \right\}$$

$\Rightarrow \bar{X} = 1.598739$

$\Rightarrow \bar{S}_x^2 = 1.177191 \Rightarrow S_x = 1.33113109$

$\Rightarrow \bar{Y} = 1.062504$

$\Rightarrow \bar{S}_y^2 = 1.141242 \rightarrow S_y = 1.06828928$

$$\Rightarrow T = \frac{1.598739 - 1.062504}{\sqrt{\frac{1.177191}{20} + \frac{1.141242}{20}}}$$

$$= \frac{0.533699}{0.340472} = \underline{1.567523} = \underline{t_{obs}}$$

$\Rightarrow \therefore \quad |T| < C(2.086)$

$\therefore$ T-test will also accept the null hypothesis, i.e. $\boxed{\mu_x = \mu_y}$ which is not same as ground truth.

p-value for t-test = $Pr\left(\begin{matrix}\text{statistic more} \\ \text{extreme} \\ \text{than } t_0\end{matrix} \middle| H_0 \text{ is true}\right)$

$$(\text{p-value})_{t\text{ test}} = Pr(|t| > t_{obs} \mid \mu_1 = \mu_2)$$

$$= Pr(|t| > t_{obs}) = Pr(|t| > 1.567)$$

$$\text{p-value} = 2(1 - cdf(t_{obs}))$$
$$= 2(1 - cdf_T(1.567))$$
$$= 2(1 - 0.9332)$$
$$= 0.0668 \times 2$$
$$\Rightarrow \quad = \underline{\underline{0.1336}}$$

Since, p-value > 0.05, we can confidently accept H_0.

$cdf_T(1.567) = 0.9332$ from 'T' table with df = 19

using online calculator

(b) $n = 1000$ $\quad x \sim Nor(1.5, 1)$, $y \sim Nor(1, 1)$

$\boxed{\sigma_x = \sigma_y = 1}$

$$\Rightarrow Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

$\bar{X} = 1.461259$

$\bar{Y} = 0.98352$

$$\Rightarrow Z = \frac{1.461259 - 0.98352}{\sqrt{\frac{1}{1000} + \frac{1}{1000}}} = \frac{0.477739}{\sqrt{1/500}}$$

$$= \underline{10.6825687}$$

$\therefore \quad |Z| > Z_{\alpha/2} (1.96)$

$\therefore$ Z-test will reject the H_0. i.e. $(\mu_x \neq \mu_y)$ which is same as ground truth.

$$\text{p-value} = 2(1 - \Phi(10.68))$$

$$= 0$$

Since, p-value of Z-test is '0', we can confidently reject H_0.

$$\Rightarrow T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$

$$\Rightarrow \bar{X} = 1.461259, \quad \bar{Y} = 0.98352$$

$$\Rightarrow \bar{S}_x^2 = \frac{1}{n-1} \Sigma (X - \bar{X})^2 = 1.029887$$

$$\Rightarrow \bar{S}_y^2 = 0.958606$$

$$\Rightarrow T = \frac{1.461259 - 0.98352}{\sqrt{\frac{(1.029887)}{1000} + \frac{(0.958606)}{1000}}}$$

$$= \frac{0.477739}{\sqrt{0.00198849}}$$

$$= \frac{0.477739}{0.04459252} = \underline{\underline{10.71}}$$

$$\Rightarrow |T| > C \; (2.086)$$

Thus, t-test will also reject the Ho.

$\therefore$ $\underline{\mu_x \neq \mu_y}$ which is same as ground truth.

As compared to ⓐ part, both test performed better with larger sample size.

p-value of T-test $= 2(1 - cdf(t_{obs}))$

$= 2(1 - cdf(10.71))$

$= 0$

Since, p-value of t-test is '0', we can confidently reject H_0.

⇒ For small sample size, 20, $H_0 (\mu_x = \mu_y)$ is accepted by both z-test and t-test which is wrong as H_0 is not true $(\mu_x \neq \mu_y)$.

With large sample size (1000), both test results in rejection of H_0 and provide the true result i.e. $\mu_x \neq \mu_y$.

⇒ For small sample size, p-value of z test (0.093) is much lesser than the p-value of t-test (0.1336). Thus, confidence in acceptance of H_0 is much more in t-test than in z-test.

Apart from that, there is no other significant advantage of using z-test over t-test as observed with this example.