A1 $D = \{1.68, 1.34, 1.98, 0.97, 1.09, 2.65, 1.23, 1.78, 0.6, 0.85\}$

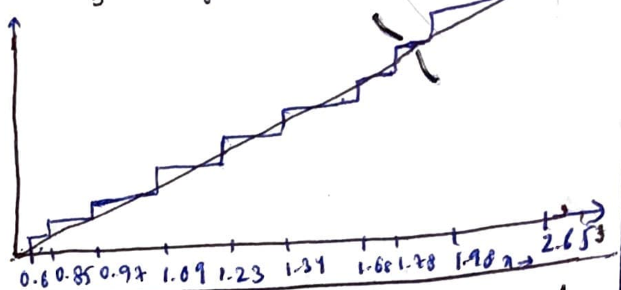
7 6 9 3 4 10 5 8 1 2

$Y = \text{Unif}(0, 3)$

critical threshold $c = 0.37$

$H_0 : F_D \equiv F_Y$ vs. $H_1 : F_D \not\equiv F_Y$

$F_{unif(0,3)}(\alpha) = \dfrac{\alpha - 0}{3-0} = \dfrac{\alpha}{3}$



x	$F_Y(x)$	$\hat{F}_X^-(x)$	$\hat{F}_X^+(x)$	$\lvert\hat{F}_X^-(x) - F_Y(x)\rvert$	$\lvert\hat{F}_X^+(x) - F_Y(x)\rvert$
0.6	0.2	0	0.1	0.2	0.1
0.85	0.2833	0.1	0.2	0.1833	0.0833
0.97	0.3233	0.2	0.3	0.1233	0.0233
1.09	0.3633	0.3	0.4	0.0633	0.0367
1.23	0.41	0.4	0.5	0.01	0.09
1.34	0.4467	0.5	0.6	0.0533	0.1533
1.68	0.56	0.6	0.7	0.04	0.14
1.78	0.5933	0.7	0.8	0.1067	0.2067
1.98	0.66	0.8	0.9	0.14	$\boxed{0.24}$
2.65	0.8833	0.9	1.0	0.0167	0.1167

$D(F_X, F_Y) = \max_{\alpha} \lvert \hat{F}_X(\alpha) - F_Y(\alpha) \rvert = 0.24 < 0.37\,(c)$

Since $D < c$

Therefore, we fail to reject or accept the H_0 (null hypothesis)

Ans2 $H_0 : X \equiv Y$ vs $H_1 : X \not\equiv Y$ $[X = \{2, 9\}$ and $Y = \{4\}]$

Total number of samples $N = |X| + |Y| = 2 + 1 = 3$

Step 1 $T_{obs} = |\bar{X} - \bar{Y}| = \left|\frac{2+9}{2} - \frac{4}{1}\right| = \left|\frac{11}{2} - 4\right| = 1.5$

Step 2: Permute $X \cup Y$ in all $N! = 3! = 6$ ways to get X_i of size $|X|$ and Y_i of size $|Y|$.

↳ All possible permutations are listed in the table below.

Step 3: Compute $T_i = |\bar{X_i} - \bar{Y_i}|$ for $i = 1$ to 6.

(Also tabulated below)

Step 4: Compute $\frac{1}{N!} \sum_{i=1}^{N!} I(T_i > T_{obs}) = \frac{1}{6} \sum_{i=1}^{6} I(T_i > T_{obs})$ and

reject H_0 if this computed value is less than or equal to α

Here the threshold value given $= 0.05$

I will compute the value corresponding to $I(T_i > T_{obs})$ in the table as follows:-

i	X_i	Y_i	$\bar{X_i}$	$\bar{Y_i}$	$T_i = \lvert\bar{X_i} - \bar{Y_i}\rvert$	$I(T_i > T_{obs})$
1	$\{2, 9\}$	$\{4\}$	5.5	4	1.5	0
2	$\{9, 2\}$	$\{4\}$	5.5	4	1.5	0
3	$\{2, 4\}$	$\{9\}$	3	9	6	1
4	$\{4, 2\}$	$\{9\}$	3	9	6	1
5	$\{4, 9\}$	$\{2\}$	6.5	2	4.5	1
6	$\{9, 4\}$	$\{2\}$	6.5	2	4.5	1

Step 4 $\frac{1}{6} \sum_{i=1}^{6} I(T_i > T_{obs}) = \frac{1}{6}[1+1+1+1+0+0] = \frac{4}{6} = \frac{2}{3} = 0.6667 > 0.05 (\alpha)$

Therefore we fail to reject H_0 / i.e. Accept $H_0 : X \equiv Y$

3. a) In this question, we wish to check for independence between the X-[outcome of the tables] and the Y-[dealer].

H_0: outcome of the tables $\perp$ dealer vs H_1: outcome of tables $\not\perp$ dealer

	1 Dealer A	2 Dealer B	3 Dealer C	Total
1 Win	48 (49.85)	54 (49.397)	19 (21.7528)	121
2 Draw	7 (6.59176)	5 (6.5318)	4 (2.8764)	16
3 Lose	55 (53.558)	50 (53.0712)	25 (23.3708)	130
Total	110	109	48	267

The above table captures the observed data. We need to compute the corresponding expected data (mentioned in brackets)

$E_{11} = \frac{110}{267} \times 121$, $E_{12} = \frac{109}{267} \times 121$, $E_{13} = \frac{48}{267} \times 121$

$\boxed{E_{11} = 49.85, \quad E_{12} = 49.397, \quad E_{13} = 21.7528}$

$E_{21} = \frac{110}{267} \times 16$, $E_{22} = \frac{109}{267} \times 16$, $E_{23} = \frac{48}{267} \times 16$

$\boxed{E_{21} = 6.59176, \quad E_{22} = 6.5318, \quad E_{23} = 2.8764}$

$E_{31} = \frac{110}{267} \times 130$, $E_{32} = \frac{109}{267} \times 130$, $E_{33} = \frac{48}{267} \times 130$

$\boxed{E_{31} = 53.558, \quad E_{32} = 53.0712 \quad E_{33} = 23.3708}$

Step 1

1) Compute $q_{obs} = \sum_r \sum_c \frac{(E_{rc} - O_{rc})^2}{E_{rc}}$

$$q_{obs} = \frac{(49.85 - 48)^2}{49.85} + \frac{(49.397 - 54)^2}{49.397} + \frac{(21.7528 - 19)^2}{21.7528} + \frac{(6.59176 - 7)^2}{6.59176}$$

$$+ \frac{(6.5318 - 5)^2}{6.5318} + \frac{(2.8764 - 4)^2}{2.8764} + \frac{(53.558 - 55)^2}{53.558} + \frac{(53.0712 - 50)^2}{53.0712} + \frac{(23.3708 - 25)^2}{23.3708}$$

$q_{obs} = 1.99949$

Step 2 Compute degrees of freedom

$df = (\#rows - 1) \times (\#cols - 1)$

$= (3-1) \times (3-1) = 4.$

Step 3, $p\text{-value} = Pr(\chi^2_{df=4} > q_{obs})$

$= Pr(\chi^2_4 > 1.99949)$

$= 1 - Pr(\chi^2_4 \leq 1.99949)$

$= 1 - F_{\chi^2_4}(1.99949)$

$= 1 - 0.26414731$ [from cdf calculator for chi-square distribution)

$= 0.73585$

Step 4, $p\text{-value} = 0.73585 > \underbrace{0.05 (\alpha)}_{\text{given in question}}$

∴ We fail to reject H_0

(Accept H_0 : outcome of tables ⊥ dealers)

Claim is correct.

A3.(b)

Pearson correlation coefficient for dealer A and dealer B. :

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

Now, $\bar{A} = \frac{48+40+58+53+65+25+52+34+30+45}{10} = \frac{450}{10}$

$\bar{A} = 45$

$\bar{B} = \frac{54+48+51+47+62+35+70+20+25+40}{10} = \frac{452}{10}$

$\bar{B} = 45.2$

$$\hat{\rho}_{A,B} = \frac{\begin{matrix}(48-45)(54-45.2)+(40-45)(48-45.2)+(58-45)(51-45.2)+ \\ (53-45)(47-45.2)+(65-45)(62-45.2)+(25-45)(35-45.2)+(52-45)(70-45.2) \\ +(34-45)(20-45.2)+(30-45)(25-45.2)+(45-45)(40-45.2)\end{matrix}}{\sqrt{3^2+(-5)^2+13^2+8^2+20^2+(-20)^2+7^2+(-11)^2+(-15)^2+0^2} \times \sqrt{(8.8)^2+(2.8)^2+(5.8)^2+(1.8)^2+(16.8)^2+(-10.2)^2+(24.8)^2+(-25.2)^2+(-20.2)^2+(-5.2)^2}}$$

$$\hat{\rho}_{A,B} = \frac{26.4-14+75.4+14.4+336+204+173.6+277.2+303+0}{\sqrt{1462} \times \sqrt{2193.6}}$$

$$\hat{\rho}_{A,B} = \frac{1396}{1790.8219} = 0.77953$$

$$\boxed{\hat{\rho}_{A,B} = 0.77953}$$

$\hat{\rho}_{A,B} > 0.5 \Rightarrow$ +ve linear correlation between wins of dealer A and dealer B

Pearson correlation coefficient for dealer B and dealer C,

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

Now, $\bar{B} = 45.2$

$$\bar{C} = \frac{19+40+35+41+38+32+32+37+37+15}{10} = \frac{326}{10}$$

$$\bar{C} = 32.6$$

$$\hat{\rho}_{B,C} = \frac{(54-45.2)(19-32.6) + (48-45.2)(40-32.6) + (51-45.2)(35-32.6) + (47-45.2)(41-32.6) + (62-45.2)(38-32.6) + (35-45.2)(32-32.6) + (70-45.2) \times (32-32.6) + (20-45.2)(37-32.6) + (25-45.2)(37-32.6) + (40-45.2)(15-32.6)}{\sqrt{(8.8)^2 + (2.8)^2 + (5.8)^2 + (1.8)^2 + (16.8)^2 + (-10.2)^2 + (24.8)^2 + (-25.2)^2 + (-20.2)^2 + (-5.2)^2} \times \sqrt{(-13.6)^2 + (7.4)^2 + (2.4)^2 + (8.4)^2 + (5.4)^2 + (-0.6)^2 + (-0.6)^2 + (4.4)^2 + (4.4)^2 + (-17.6)^2}}$$

$$\hat{\rho}_{B,C} = \frac{-119.68 + 20.72 + 13.92 + 15.12 + 90.72 + 6.12 - 14.88 - 110.88 - 88.88 + 91.52}{\sqrt{2193.6} \times \sqrt{694.4}}$$

$$= \frac{-96.2}{1234.19441} = -0.0779$$

$$\boxed{\hat{\rho}_{B,C} = -0.0779}$$

Its a negative correlation but relationship b/w the two variables is very weak

$\Rightarrow$ no correlation between wins of dealer B and dealer C. $[\because |\hat{\rho}_{B,C}| \leq 0.5]$

Pearson correlation coefficient for C and A

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

Now, $\bar{A} = 45$

$\bar{C} = 32.6$

$$\hat{\rho}_{CA} = \frac{(19-32.6)(48-45)+(40-32.6)(40-45)+(35-32.6)(58-45)+(41-32.6)(53-45)+(38-32.6)(65-45)+(32-32.6)(25-45)+(32-32.6)(52-45)+(37-32.6)(34-45)+(37-32.6)(30-45)+(15-32.6)(45-45)}{\sqrt{(-13.6)^2+(7.4)^2+(2.4)^2+(2.4)^2+(8.4)^2+(5.4)^2+(-0.6)^2+(-0.6)^2+(4.4)^2+(4.4)^2+(-17.6)^2} \times \sqrt{3^2+(-5)^2+13^2+8^2+20^2+(-20)^2+7^2+(-11)^2+(-15)^2+0}}$$

$$\hat{\rho}_{CA} = \frac{-40.8-37+31.2+67.2+108+12-4.2-48.4-66+0}{\sqrt{694.4} \times \sqrt{1462}}$$

$$= \frac{22}{1007.577689} = 0.0218345$$

$$\boxed{\hat{\rho}_{CA} = 0.0218}$$

+ve but weak relation. $\hat{\rho}_{CA} = 0.0218 \Rightarrow$ no correlation b/w wins of dealer A and dealer C.

$[\,|\hat{\rho}_{CA}| \le 0.5\,]$

5. $\{X_1, X_2, \ldots X_n\} \sim Nor(\mu_1, \sigma_1^2)$ ... (1)

$\{Y_1, Y_2, \ldots Y_m\} \sim Nor(\mu_2, \sigma_2^2)$ ... (2)

$X \perp Y$

$H_0 : \mu_1 > \mu_2$ vs $H_1 : \mu_1 <= \mu_2$

$H_0 : \mu_1 - \mu_2 > 0$ vs $H_1 : \mu_1 - \mu_2 <= 0$

T statistic for unpaired t-test $T = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} = \dfrac{\bar{D}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$

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

S_x : sample std deviation for $\{X_1 \ldots X_n\}$

S_y : sample std deviation for $\{Y_1 \ldots Y_m\}$

For one sided test with $H_0 : \mu_1 > \mu_2$, we have :

Reject H_0 if $T < -\delta$

Type I error: Pr(Test rejects H_0 | H_0 is true)

$= Pr(T < -\delta \mid H_0 \text{ is true})$

$= Pr\left(\dfrac{\bar{D} - 0}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} < -\delta \right)$

Denominator is +ve, Multiplying both sides by $\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}$

Type I error $= Pr\left(\bar{D} < -\delta \sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}} \right)$ ... (3)

Now, we know that :

$\bar{D} = \bar{X} - \bar{Y} = \bar{X} + (-1)\bar{Y} \sim Nor\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$

Transforming D to a std normal to get result in terms of Φ, we have after using transformation property :

$Z = \dfrac{\bar{D} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$

Note since n and m are large

∴ sample std deviation is same as actual std deviation

and
$$Z = \frac{\bar{D} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} = \frac{\bar{D} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \quad \text{--- (4)}$$

From (3) eq:

$$\text{Type I error} = Pr\left(\bar{D} < -\delta\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}\right)$$

A) Subtracting $(\mu_1 - \mu_2)$ from both sides :-

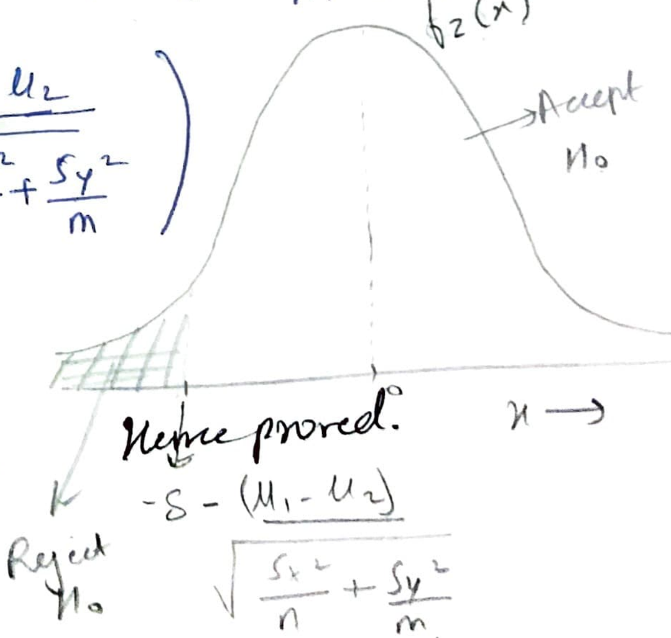
$$= Pr\left(\bar{D} - (\mu_1 - \mu_2) < -\delta\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}} - (\mu_1 - \mu_2)\right)$$

B) Dividing both sides by $\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}$:

$$= Pr\left(\frac{\bar{D} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} < -\delta - \left(\frac{\mu_1 - \mu_2}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}\right)\right)$$

$$= Pr\left(Z < -\delta - \frac{\mu_1 - \mu_2}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}\right)$$

$$\boxed{\text{Type-I error} = \phi\left(-\delta - \frac{\mu_1 - \mu_2}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}\right)}$$



Hence proved.

UnShaded region : Acceptance region , Shaded : Rejection of H_0

Type II error = Pr(Accept H_0 | H_0 is false)

$T \geq -\delta \rightarrow$ Accept H_0:

$$= Pr\left(\frac{\bar{D}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \geq -\delta \right)$$

$$= Pr\left(\bar{D} \geq -\delta \sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}} \right)$$

Using the steps marked A. and B in Type I error calculation, we get:-

$$= Pr\left(\frac{\bar{D} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \geq -\delta - \left(\frac{(\mu_1 - \mu_2)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \right) \right)$$

$$= 1 - Pr\left(Z < -\delta - \frac{(\mu_1 - \mu_2)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \right)$$

$\because$ n and m are large.
sample std dev ≈ actual std dev

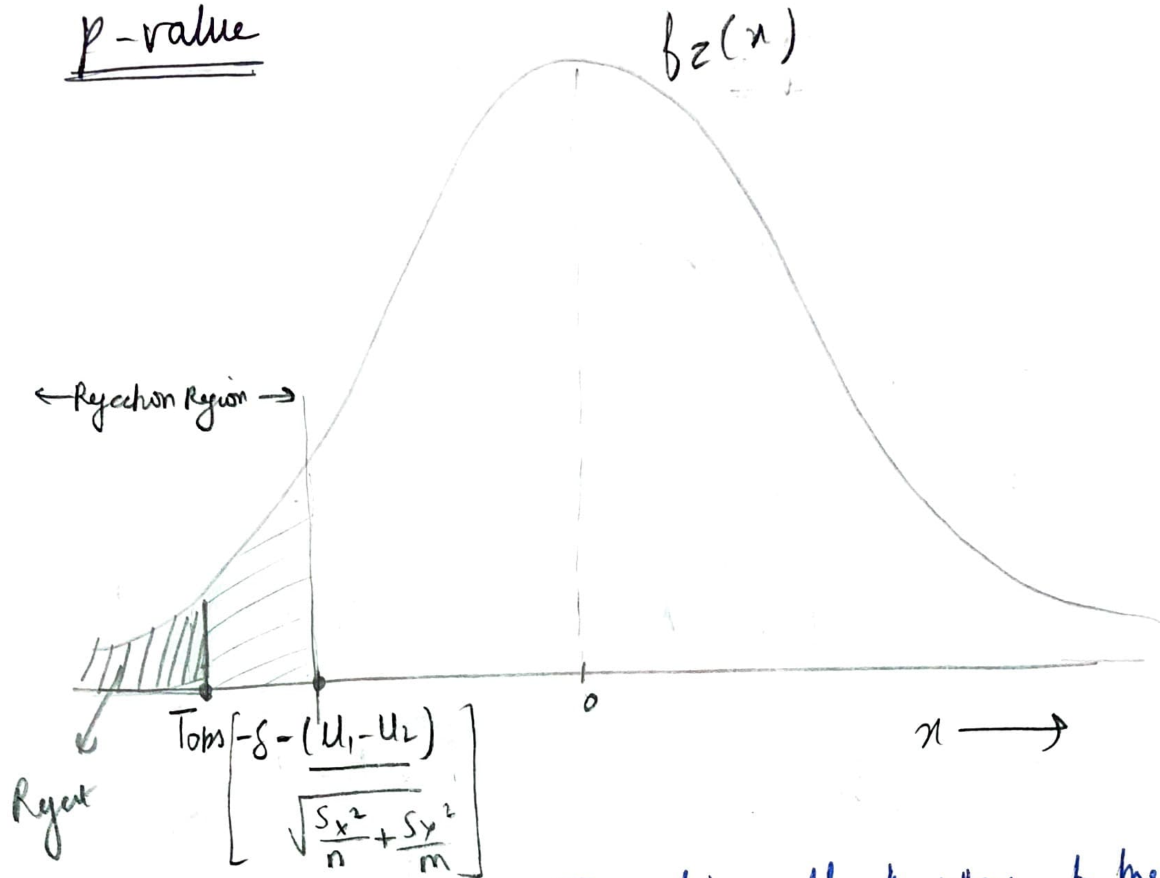
$$= 1 - \Phi\left(-\delta - \frac{\mu_1 - \mu_2}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \right)$$

[This part is answered based on the facts and reasoning developed during the Type I error computation]

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

Hence proved.

(b) p-value



We will reject H_0 if T_{obs} lies all the way to the left in the graph above :

p-value = Area to the left of T_{obs} = $\Phi(T_{obs})$

The above graph is of the std normal distribution and we had already computed it in eq (4) of part a)

$$Z = \frac{\bar{D} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$

[Note n and m are Large
∴ sample std deviation = true std dev]

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

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

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

Hence proved

14.8

a) Permutation Test

Let Null Hypothesis : H_0 = Scoring distribution remains same, i.e., $X \equiv Y$.

Alternate Hypothesis : H_1 :- $X \not\equiv Y$

where X, Y are distributions of different years. For $n = 200$,

Now, for 1999 v/s 2009,

$X = 1999$, $Y = 2009$

p-value (from code) = 0.0

$\because$ p-value $\leq \alpha = 0.05$

we reject Null hypothesis.

For 2009 v/s 2019 data,

$X = 2009$ $Y = 2019$

p-value = 0.0

$\because$ p-value $\leq \alpha$, we reject H_0

b) For $n = 2000$,

$X = 1999$, $Y = 2009$, p-value = 0.0

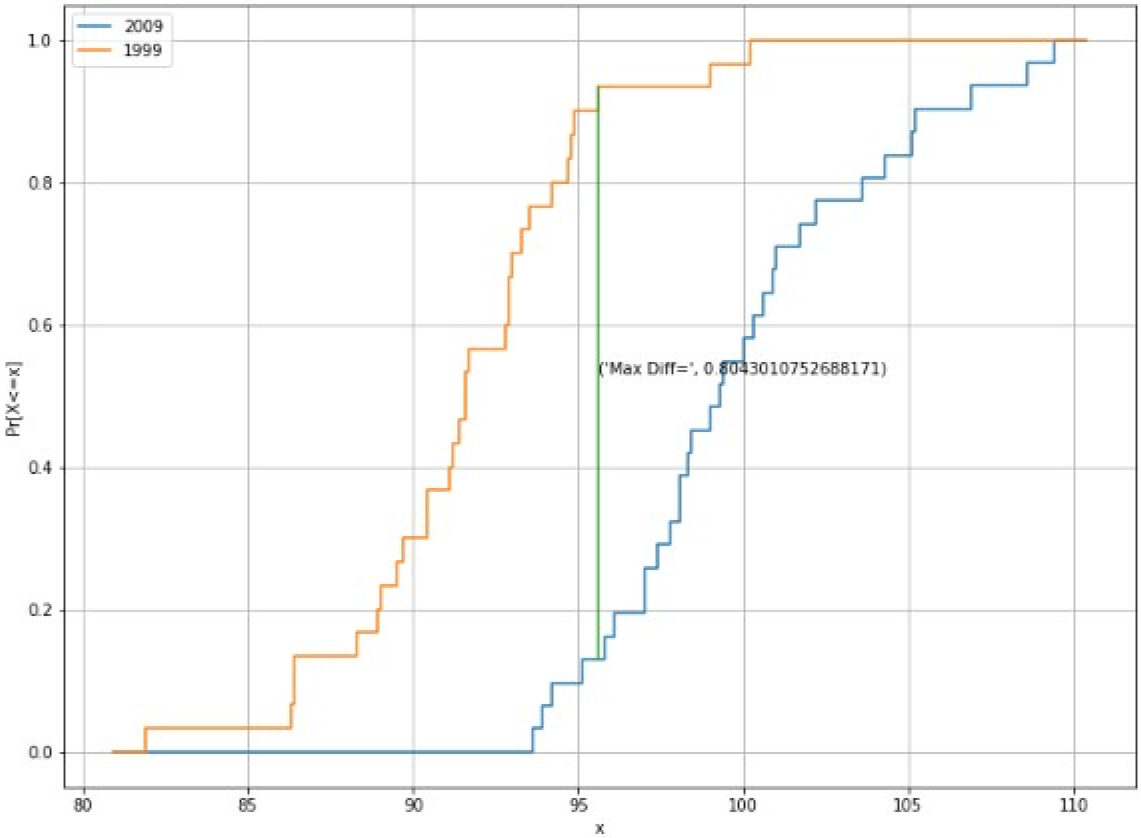
$\therefore$ We reject H_0 ($\because$ p-value $\leq \alpha$)

For $X = 2009$, $Y = 2019$, p-value = 0.0

$\therefore$ We reject H_0 ($\because$ p-value $\leq \alpha$)

```
1 k_s_test(pts_1999, pts_2009, 1999)
```

```
Max value is 0.8043010752688171 at X=[95.6]
D > C, We fail to reject Null Hypothesis
```



```
1 k_s_test(pts_2009, pts_2019, 2009)
```

```
Max value is 0.8064516129032253 at X=[104.5]
D > C, We fail to reject Null Hypothesis
```