CSE544: Assignment 5 Solutions

From the table, the maximum difference is 0.24.

Q1.

x	$F_Y(x)$	$\hat{F_x^-}$	$\hat{F_x^+}$	$\lvert\hat{F_x^-} - F_Y(x)\rvert$	$\lvert\hat{F_x^+} - F_Y(x)\rvert$
0.6	0.2	0.0	0.1	0.2	0.1
0.85	0.283	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	0.0366
1.23	0.41	0.4	0.5	0.01	0.09
1.34	0.4466	0.5	0.6	0.0533	0.1533
1.68	0.56	0.6	0.7	0.04	0.14
1.78	0.593	0.7	0.8	0.106	0.206
1.98	0.66	0.8	0.9	0.14	**0.24**
2.65	0.8833	0.9	1.0	0.0166	0.116.

$0.24 < C(0.37)$.

Therefore, we fail to reject H_0 (Null Hypothesis).

Q2

$H_0 : X \equiv Y$

X = {2, 9} Y = {4}

Calculating the difference of Means for the given Dataset

$T_{Obs} = |5.5 - 4| = 1.5$

Calculating Difference of Means for the 6 possible permutations:

S.No	Permutation	$T_i = \lvert\overline{X} - \overline{Y}\rvert$
1	{2, 9} {4}	1.5
2	{9, 2} {4}	1.5
3	{2, 4} {9}	6
4	{4, 9} {2}	4.5
5	{4, 2} {9}	6
6	{9, 4} {2}	4.5

$$p_{val} = \frac{1}{N!} \sum_{i=1}^{N!} I(T_i > T_Obs)$$

$$p_{val} = \frac{4}{6}$$

$$p_{val} = 0.66$$

Here p-value = 0.66; (> 0.05 threshold). Therefore, we fail to reject H_0.

Q3

(a)

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

From the above table,
$Grand\ Total = 267$

$$P(Win) = \frac{TotalWins}{GrandTotal} = \frac{121}{267}$$
$$P(DealerA) = \frac{TotalA}{GrandTotal} = \frac{110}{267}$$

If the outcomes are independent of dealer;
$Expected\ frequency\ of\ Win\ and\ Dealer\ A\ =\ (Grand\ Total) * P(Win) * P(DealerA)$

Similarly, we populate the below table with expected frequencies:

	Dealer A	**Dealer B**	**Dealer C**
Win	49.85	49.4	21.75
Draw	6.59	6.53	2.88
Lose	53.56	53.07	23.37

Evaluating the χ^2 test:

Observed	**Expected**	$\frac{(O-E)^2}{E}$
48	49.85	0.0686
54	49.4	0.428
19	21.75	0.3477
7	6.59	0.0255
5	6.53	0.358
4	2.88	0.435
55	53.56	0.0387
50	53.07	0.178
25	23.37	0.113
$\sum O = 267$	$\sum E = 267$	$\sum \frac{(O-E)^2}{E} = 1.993$

We now have the $Q - statistic(Q) = 1.993$

$$df(Degree\ of\ Freedom) = (3-1)*(3-1)$$
$$p - value = Pr(\chi^2_4 > 1.993) = 1 - 0.2629 = 0.737$$

As $p - value(0.737) > 0.05$, we fail to reject H_0.

(b)

$\rho_{X,Y} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{(X_i-\overline{X})^2}\sqrt{(Y_i-\overline{Y})^2}}$

Dealer A; $\overline{X} = 45$
Dealer B; $\overline{X} = 45.2$
Dealer C; $\overline{X} = 32.6$

$\rho_{A,B} = \frac{1396}{\sqrt{14622193.6}} = 0.779(> 0.5)$ – Positive linear correlation
$\rho_{B,C} = \frac{-96.2}{\sqrt{2193.6694.4}} = -0.779$ – No linear correlation
$\rho_{A,C} = \frac{1396}{\sqrt{1462694.4}} = 0.0218$ – No linear correlation

As the probability of winning each game is same; the results for each of the dealers should be correlated. We observe from above that results from Dealer C are not linearly correlated with Dealer A and Dealer B. We could infer that Dealer C is responsible for loss of money.

Q4.

(a)

$$p - value : 1999vs.2009 = 0.0$$
$$p - value : 2009vs.2019 = 0.0$$

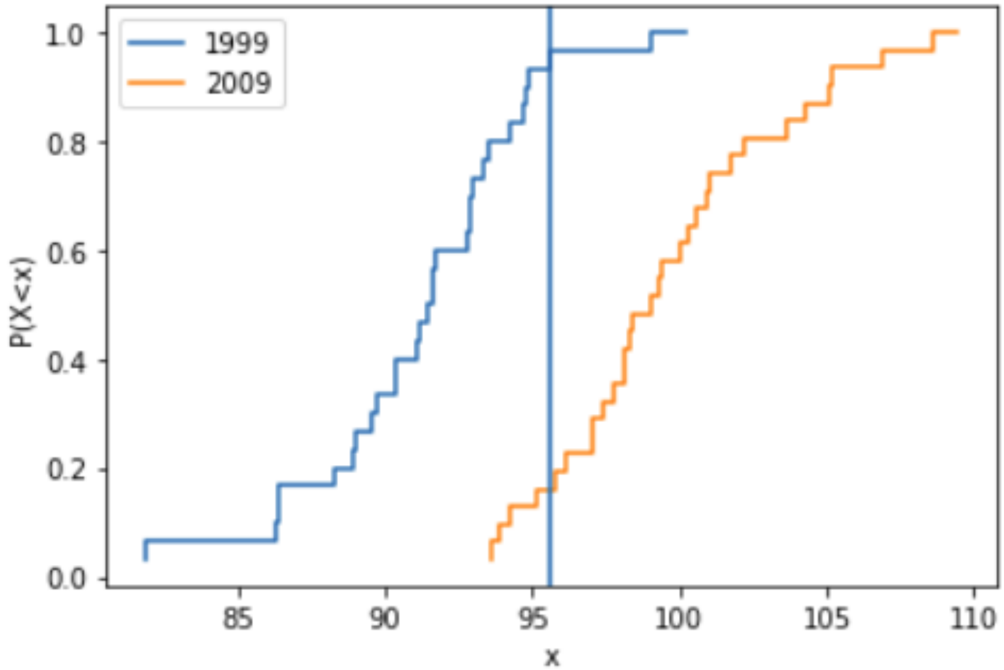
Reject null hypothesis for both scenarios as $0.0 < 0.05$

(b)

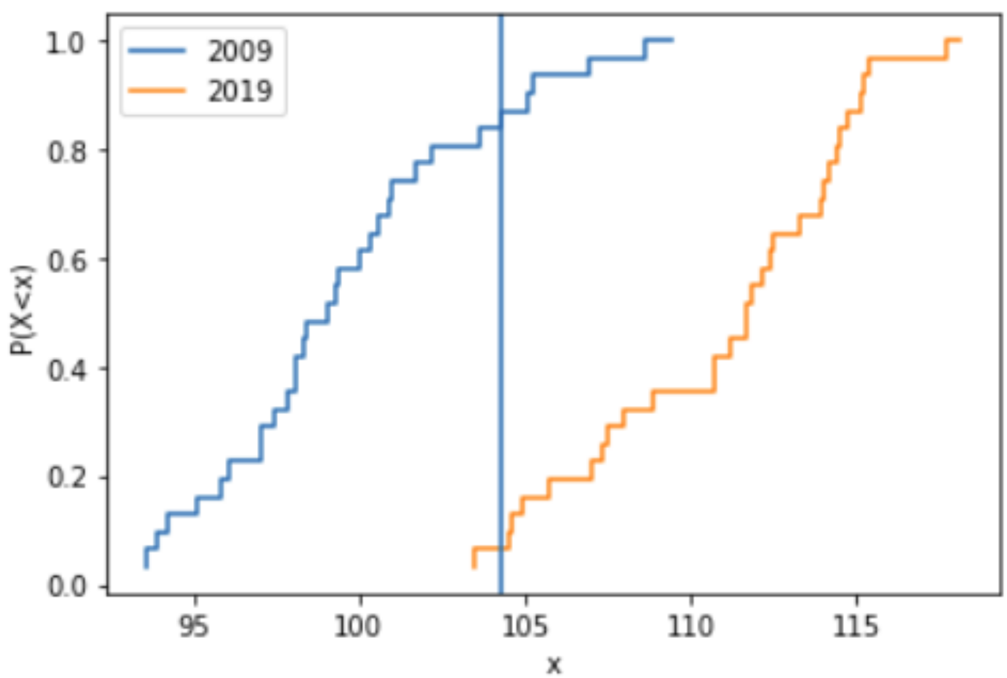
$$p - value : 1999vs.2009 = 0.0$$
$$p - value : 2009vs.2019 = 0.0$$

Reject null hypothesis for both scenarios as $0.0 < 0.05$

(c)




```
Point of maximum difference is 95.6
KS statistic is 0.804
```




```
Point of maximum difference is 104.3
KS statistic is 0.806
```

Based on above plots:
for the period 1999-2009;

$$D = 0.804$$
$$D > 0.05 (max\ difference\ threshold)$$

Thus, for 1999-2009, we reject the null hypothesis.

Similarly, the period 2009-2019;

$$D = 0.806$$
$$D > 0.05 (max\ difference\ threshold)$$

Thus, for 2009-2019, we reject the null hypothesis.

Note that in Q5 below, since we assumed n and m are large, σ_1 and σ_2 are well approximated by S, and by consistency of sample variance, we can replace σ_1 & σ_2 with S_x & S_y respectively.

Q5. $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\mu_1, \sigma_1^2)$

$Y_1 \dots Y_m \overset{i.i.d}{\sim} N(\mu_2, \sigma_2^2)$

$X's \perp Y's.$

Note that $\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right)$

$\bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right)$

Since $X's \perp Y's$,

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$$

(a) $H_0: \mu_1 > \mu_2 \quad H_1: \mu_1 \leq \mu_2$

Type I error:

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

$$= P\left(\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} < -\delta \;\middle|\; \mu_1 > \mu_2\right)$$

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

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

Sine $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}} \sim N(0,1)$

Type 2 error:

$P(T > -\delta \mid H_0 \text{ is false})$

$$= P\left(\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}} > -\delta \;\middle|\; \mu_1 \le \mu_2\right)$$

$$= 1 - P\left(\frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}} \le -\delta\right)$$

$$= 1 - P\left(\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}} \le -\delta - \frac{\mu_1-\mu_2}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}}\right)$$

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

(b) Let $t_{obs} = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$ be the observed statistic.

$$\text{p-value} = P(T < t_{obs})$$

$$= P\left(\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} < t_{obs}\right)$$

$$= P\left(\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} < t_{obs} - \frac{(\mu_1 - \mu_2)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}\right)$$

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

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