CSE 544 A5 Solution

Spring 2021

1 Q1

x	$F_Y(x)$	$\hat{F}_x^-$	$\hat{F}_x^+$	$\lvert\hat{F}_x^- - F_Y(x)\rvert$	$\lvert\hat{F}_x^+ - F_Y(x)\rvert$
0.04	0.0134	0.0	0.1	0.0134	0.0866
0.74	0.2463	0.1	0.2	0.1463	0.0463
0.84	0.2815	0.2	0.3	0.0815	0.0185
1.19	0.3964	0.3	0.4	0.0964	0.0036
1.88	0.6269	0.4	0.5	**0.2269**	0.1269
1.99	0.6639	0.5	0.6	0.1639	0.0639
2.23	0.7432	0.6	0.7	0.1432	0.0432
2.57	0.8572	0.7	0.8	0.1572	0.0572
2.65	0.8823	0.8	0.9	0.0823	0.0177
2.78	0.9282	0.9	1.0	0.0282	0.0718

$0.2269 < c = 0.25$, accept.

2 Q2

$H_0 : X \equiv Y$, calculate the difference of means for given dataset $T_{obs} = |5 - 4.5| = 0.5$. Calculate the difference of means for the 6 possible permutations:

No.	Permutation	$T_i = \lvert\bar{X} - \bar{Y}\rvert$
1	52,7	0.5
2	57,2	0.5
3	25,7	4
4	27,5	4
5	75,2	3.5
6	72,5	3.5

$$p_{val} = \frac{1}{N!}\sum_{i=1}^{N!} I(T_i > T_{obs}) \tag{1}$$
$$= \frac{4}{6} \tag{2}$$
$$= 0.66 > 0.05 \tag{3}$$

Reject.

3 Q3

Solution:

(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, we get

$$P(Win) = \frac{\text{Total Wins}}{\text{Grand Total}} = \frac{121}{267}$$
$$P(\text{Dealer A}) = \frac{\text{Total for Dealer A}}{GrandTotal} = \frac{110}{267}$$

If the outcomes are independent of dealers

$$\text{Expected frequency of Win and Dealer A} = (\text{Grand Total}) \; * \text{P(Win)} \; * \text{P(Dealer A)}$$

Similarly, we populate the table below with the expected frequencies:

	Dealer A	Dealer B	Dealer C	Total
Win	49.85	49.40	21.75	121
Draw	6.59	6.53	2.88	16
Lose	53.56	53.07	23.37	130
Total	110	109	48	267

Evaluating the χ^2 test:

Observed	Expected	$\frac{(E-O)^2}{E}$
48	49.85	0.0686
54	49.40	0.4283
19	21.75	0.3477
7	6.59	0.0255
5	6.53	0.3585
4	2.88	0.4356
55	53.56	0.0387
50	53.07	0.1776
25	23.37	0.1137
$\sum O = 267$	$\sum E = 267$	$\frac{(E-O)^2}{E} = 1.9942$

We now have a Q-statistics of 1.9942

$$\text{df (Degree of Freedom)} = (3-1)*(3-1) = 4$$
$$\text{p-value} = Pr(\chi^2_4 > 1.9942) = 1 - 0.2632 = 0.7368$$

Since $0.7368 > 0.05$, we fail to reject H_0.

(b)

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Dealer A	48	40	58	53	65	25	52	34	30	45
Dealer B	54	48	51	47	62	35	70	20	25	40
Dealer C	19	40	35	41	38	32	32	37	37	15

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

Dealer A: Mean = 45.0
Dealer B: Mean = 45.2
Dealer C: Mean = 32.6

$$\rho_{A,B} = \frac{1396}{1790.8219341966974} = 0.77953 \text{ (Positive linear correlation)}$$
$$\rho_{B,C} = \frac{-96.20}{1234.1944093213192} = -0.07795 \text{ (No linear correlation)}$$
$$\rho_{A,C} = \frac{22.0}{1007.5776893123428} = 0.02183 \text{ (No linear correlation)}$$

As the probability of winning each game is the 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.

4 Q4

(a) The p-value is 0.0 for both $n = 200$ and $n = 1000$, reject.

(b) The p-value is 0.417(within a small range) > 0.05, accept.

(c) The maximum difference is at 57.0(or 56.0) and the K-S statistic is $0.118 > 0.05$ for Female vs. Male. Reject.

5 Q5

First we know $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n})$ and $\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{m})$. Since X and Y are independent, $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_1, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m})$.

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

(a) Type-1 error is shown by

$$P(T < -\delta|H_0) = P(\frac{\bar{X}-\bar{Y}}{\sqrt{S_X^2/n + S_Y^2/m}} < -\delta|\mu_1 > \mu_2) \tag{4}$$

$$= P(\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S_X^2/n + S_Y^2/m}} < -\delta - \frac{\mu_1-\mu_2}{\sqrt{S_X^2/n + S_Y^2/m}}) \tag{5}$$

$$= \Phi(-\delta - \frac{\mu_1-\mu_2}{\sqrt{S_X^2/n + S_Y^2/m}}) \tag{6}$$

(5) $\rightarrow$ (6) because $\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S_X^2/n+S_Y^2/m}} \sim N(0,1)$.
Type-2 error is

$$P(T > -\delta|H_1) = P(\frac{\bar{X}-\bar{Y}}{\sqrt{S_X^2/n + S_Y^2/m}} > -\delta|\mu_1 \leq \mu_2) \tag{7}$$

$$= 1 - P(\frac{\bar{X}-\bar{Y}}{\sqrt{S_X^2/n + S_Y^2/m}} \leq -\delta|\mu_1 \leq \mu_2) \tag{8}$$

$$= 1 - P(\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S_X^2/n + S_Y^2/m}} \leq -\delta - \frac{\mu_1-\mu_2}{\sqrt{S_X^2/n + S_Y^2/m}}) \tag{9}$$

$$= 1 - \Phi(-\delta - \frac{\mu_1-\mu_2}{\sqrt{S_X^2/n + S_Y^2/m}}) \tag{10}$$

(b) Let $t_{obs} = \frac{\bar{X}-\bar{Y}}{\sqrt{S_X^2/n+S_Y^2/m}}$ be the observed statistics.

$$p - value = P(T < t_{obs}) \tag{11}$$

$$= P(\frac{\bar{X}-\bar{Y}}{\sqrt{S_X^2/n + S_Y^2/m}} < t_{obs}) \tag{12}$$

$$= P(\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S_X^2/n + S_Y^2/m}} < t_{obs} - \frac{\mu_1-\mu_2}{\sqrt{S_X^2/n + S_Y^2/m}}) \tag{13}$$

$$= \Phi(\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S_X^2/n + S_Y^2/m}}) \tag{14}$$

6 Q6

(a)

Data	Mean	Deviation(dof=n)	**Deviation(dof=n-1)**
X	1.5987385648483647	1.0575119993886528	1.0849844184430466
Y	1.06250437094748	1.0412396178159737	1.0682893072126083

$$T = \frac{\bar{X}_1 - \bar{Y}_1}{\sqrt{S_{X_1}^2/30 - S_{Y_1}^2/30}} = 1.574969707990326$$

$$Z = \frac{\bar{X}_1 - \bar{Y}_1}{\sqrt{\sigma_X^2/30 - \sigma_Y^2/30}} = 1.695721411991167$$

$|T| < 2.086$,$|Z| < 1.962$, accept in both tests.

(b)

Data	Mean	Deviation(dof=n)	**Deviation(dof=n-1)**
X	1.461258989664249	1.0143257450006955	1.0148332885626046
Y	0.983520204716034	0.9785947052535909	0.9790843698853109

$$T = \frac{\bar{X_2} - \bar{Y_2}}{\sqrt{S_{X_2}^2/1000 - S_{Y_2}^2/1000}} = -10.713428819788566$$

$$Z = \frac{\bar{X_2} - \bar{Y_2}}{\sqrt{\sigma_X^2/1000 - \sigma_Y^2/1000}} = -10.68256398632362$$

$|T| > 1.962$,$|Z| > 1.962$, reject in both tests.