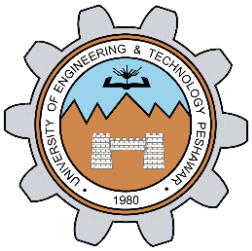
INFERENTIAL STATISTICS

LAB # 03



Fall 2021

Data Analytics Lab

Submitted by: **Shah Raza**

Registration No.: **18PWCSE1658**

Class Section: **B**

"On my honor, as student of University of Engineering and Technology, I have neither given nor received unauthorized assistance on this academic work."

Student Signature: _____

Submitted to:

Engr. Mian Ibad Ali Shah

November 28, 2021

Department of Computer Systems Engineering

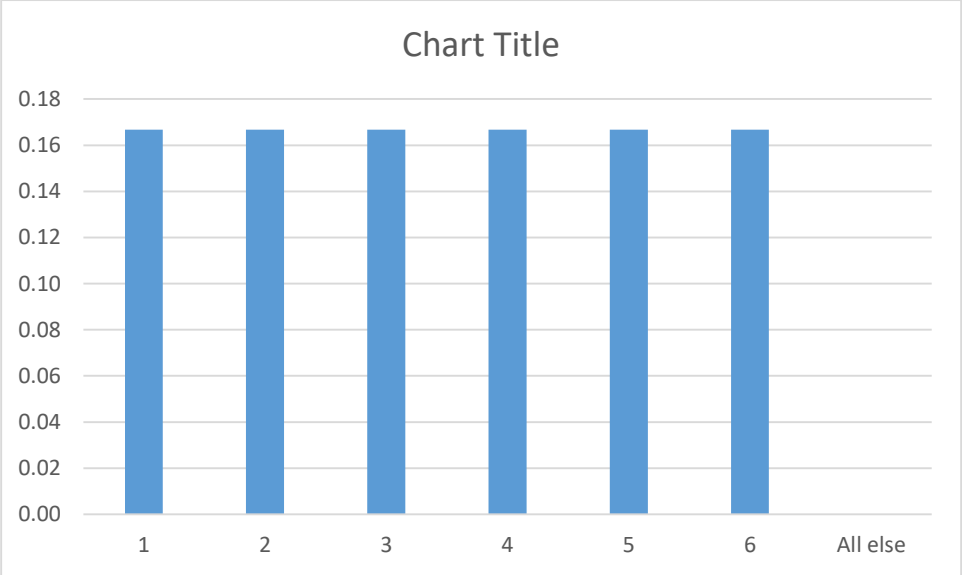
University of Engineering and Technology, Peshawar

Question: What is a distribution?

Answer:

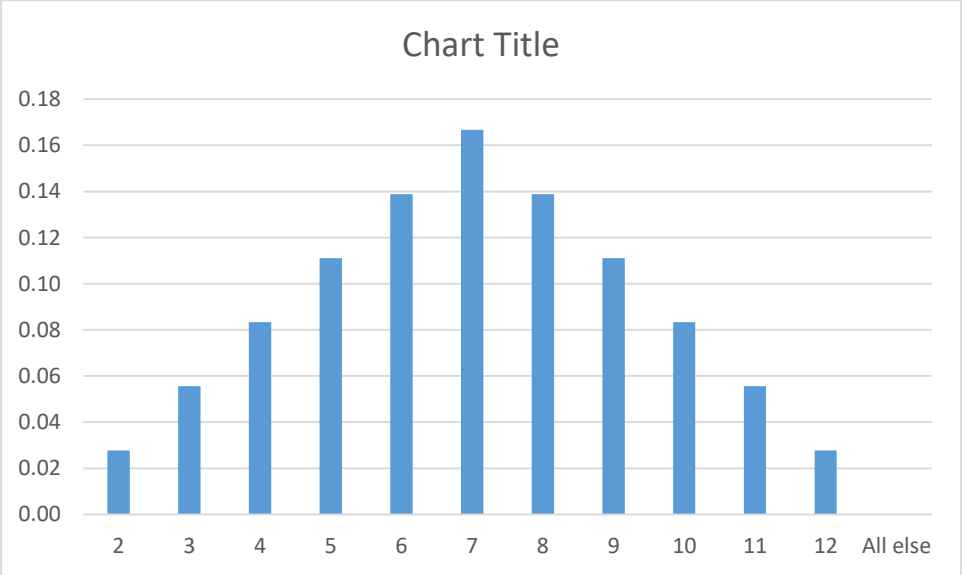
A distribution is a function that shows the possible values for a variable and how often they occur.

->Rolling 1 dice (Probability of 1,2,3,4,5,6 is 1/6 or 0.16 and 0 for every other number)



Discrete Uniform Distribution

->Rolling 2 dice (Probability of 1 is 0, 2 is 1/36 or 0.003 and so on)



Binomial Distribution

So Distribution shows frequency at which all the possible values occur.

Task #1: Given an approximately normally distributed dataset, calculate the mean and standard deviation of the dataset and standardize the dataset.

Original dataset	
567.45	743.62
572.45	747.20
572.45	748.20
589.12	748.28
613.87	748.53
615.78	750.03
628.45	752.12
644.87	754.70
650.45	755.03
652.20	758.37
656.87	760.53
661.45	764.03
666.45	769.28
667.70	775.45
668.95	781.20
675.28	781.70
675.78	785.62
685.53	792.78
694.28	793.37
697.62	795.28
705.78	797.62
705.87	798.95
708.12	799.70
711.03	799.95
714.03	810.87
716.03	811.53
722.28	813.62
728.12	814.03
728.70	814.78
729.03	817.87
730.12	818.87
731.95	820.70
735.03	821.12
736.95	825.62
737.37	828.62
738.28	841.45
739.78	842.03
740.62	842.87

849.62	878.78
874.70	897.45

Mean and Standard deviation:

Mean	743.03
STD	73.95306

Standardized dataset:

Standardized Data	
-2.374169264	-0.149785327
-2.30655881	-0.108092213
-2.30655881	-0.082174873
-2.08119063	-0.076540668
-1.746518883	-0.064145418
-1.720601542	-0.043862282
-1.549321725	-0.032593873
-1.327334068	0.007972399
-1.251835727	0.056426558
-1.228172068	0.069948649
-1.165068978	0.07107549
-1.103092728	0.074456012
-1.035482274	0.094739149
-1.018579661	0.122910171
-1.001677047	0.157842239
-0.916037139	0.162349603
-0.909276094	0.207423239
-0.777435708	0.236721102
-0.659117414	0.28404842
-0.614043778	0.355039397
-0.503613369	0.438425623
-0.502486529	0.516177645
-0.472061824	0.522938691
-0.432622393	0.575900213
-0.39205612	0.672808531
-0.365011939	0.680696417
-0.280498871	0.706613758
-0.201620008	0.738165303
-0.193732122	0.756194757
-0.189224758	0.766336325
-0.174575826	0.769716848

0.917333006	1.157350118
0.926347733	1.330883616
0.954518756	1.338771503
0.96015296	1.350039912
0.970294528	1.441314025
1.011987642	1.780493136
1.025509732	1.83570834
1.050300232	2.088120701
1.055934437	-10.04727969
1.116783845	

Task #2: Given is the same dataset from the lesson. The population standard deviation is known to be Rs 15000. Calculate the mean and the standard error. Find the appropriate z-score for calculating a 90% confidence interval and Find the 90% confidence interval.

Dataset	
Rs 117,313	Rs 112,276
Rs 104,002	Rs 108,637
Rs 113,038	Rs 96,818
	Rs 92,307
Rs 101,936	Rs 114,564
Rs 84,560	Rs 109,714
Rs 113,136	Rs 108,833
Rs 80,740	Rs 115,295
Rs 100,536	Rs 89,279
Rs 105,052	Rs 81,720
Rs 87,201	Rs 89,344
Rs 91,986	Rs 114,426
Rs 94,868	Rs 90,410
Rs 90,745	Rs 95,118
Rs 102,848	Rs 113,382
Rs 85,927	

Mean and Standard Error:

Mean		Standard Error
Rs 100,200		Rs 2,738.61

Z-score for 90% CI:

Step 1: 90% = 0.90 = (1-α)
α = 0.10

Step 2: Since we are dealing with confidence intervals (or two-sided tests) we are interested in α/2, or better: (1-α/2)
Find (1-α/2)

(1-α/2) = (1-0.05) = 0.95

Step 3: Search for the closest value to 0.95 (or the exact one) in the table.
In this case it 0.9505

Step 4: The critical value ($z_{\alpha/2}$), is given by the sum of the corresponding row and column headers
1.6+0.05 = 1.65

Answer: $z_{0.05} = 1.65$

90% Confidence Interval:

Confidence Interval
[Rs 95,681.66, Rs 104,719.08]