Data Science Masters :Assignment 17

Problem Statement 1:
Blood glucose levels for obese patients have a mean of 100 with a standard deviation of 15. A researcher thinks that a diet high in raw cornstarch will have a positive effect on blood glucose levels. A sample of 36 patients who have tried the raw cornstarch diet have a mean glucose level of 108. Test the hypothesis that the raw cornstarch had an effect or not.

Solution:

Given Data:
Mean = 100, std.dev = 15
Sample size = 36, Sample Mean = 108

Possible hypothesis :

H0: μ= 100 --- NULL Hypothesis
H1: μ > 100 --- Alternate Hypothesis

Significance level 5% (0.05) (Assumed Value - since the level of confidence is not mentioned in the given problem)

Calculating the random chance probability using z score and z-table

z= (108-100) / (15/√36)=3.20

With the reference of z- table and p-value associated with 3.20 is 0.9993 i.e. probability of having value less than 108 is 0.9993 and more than or equals to 108 is (1-0.9993)=0.0007.

Calculated Probability is less than 0.05 (Significance level)
Hence, We will reject the NULL hypothesis (H0) i.e. there is raw cornstarch effect.

Problem Statement 2:
In one state, 52% of the voters are Republicans, and 48% are Democrats. In a second state, 47% of the voters are Republicans, and 53% are Democrats. Suppose a simple random sample of 100 voters are surveyed from each state.
What is the probability that the survey will show a greater percentage of Republican voters in the second state than in the first state?

Solution:

Given Data:
P1 = the proportion of Republican voters in the first state i.e. 52% = 0.52
P2 = the proportion of Republican voters in the second state i.e. 47% = 0.47
p1 = the proportion of Republican voters in the sample from the first state
p2 = the proportion of Republican voters in the sample from the second state

Value of Expected Mean of Sample propotions => E(p1 - p2) = P1 - P2 = 0.52 - 0.47 = 0.05.
Value of Std. dev of the difference =>
σd = sqrt{ [P1(1 - P1) / n1] + [P2(1 - P2) / n2] }
= sqrt{ [(0.52)(0.48) / 100] + [(0.47)(0.53) / 100] }
= sqrt (0.002496 + 0.002491) = sqrt(0.004987) = 0.0706
We need to find the probability that p1 is less than p2. This is equivalent to finding the probability that p1 - p2 is less than 0.
We need to transform the random variable (p1 - p2) into a z-score.

z = X - E(p1-p2) / Std.dev (p1-p2)

p1-p2
= (0 - 0.05)/0.0706 = -0.7082

With the reference of z- table, probability of a z-score being -0.7082 or less is 0.24.

Hence,probability that the survey will show a greater percentage of Republican voters in the second state than in the first state is 0.24

Problem Statement 3:
You take the SAT and score 1100. The mean score for the SAT is 1026 and the standard deviation is 209. How well did you score on the test compared to the average test taker?

Solution:

Given Data:
Mean = 1026
Std. dev = 209
Score of a guy (myself) = 1100

Find the z value (1100) => z= 1100-Mean / Std.dev
Substitute the given values :
z = 1100 - 1026 / 209
= 0.354

Hence, this means that my score was 0.354 std devs above the mean. i.e. With the help of Z table, we can find the percentage of test-takers scored below my score which is equal to 0.6368 or 63.68%