**Task 1**

**Question 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 corn starch 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 corn starch had an effect or not.

Answer :Ho (null hypothesis): raw corn starch diet has no effect

Ha (alternate hypothesis): raw corn starch diet has positive effect

I have to perform one tailed test because only positive effect is to be tested and by default the Level of Significance(α) value is 0.05

Referring left Z table for probability percentage as (1-0.05 = 0.95) which falls in the non-critical region and the corresponding z score critical value is found as 1.645, so if calculated z score value is more than z score critical value than reject null hypothesis otherwise accept null hypothesis.

**Standard error of sampling distribution (SE) = population standard deviation/sqrt(number of samples) = 15/sqrt(36) = 15/6 = 2.5**

**Z score calculated value of standard normal distribution = (sample mean - population mean)/SE = 108-100/2.5 = 8/2.5 = 3.2**

**So, I can infer that a diet high in raw corn starch will have a positive effect on blood glucose levels.**

**Question 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?

Answer : **Mean of the difference in sample proportion of Republicans of state1 and state2 = 52% - 47% = 5% = 0.05**

**Standard deviation obetween state1 and state2 = sqrt((0.52\*0.48)/100 + (0.47\*0.53)/100) = 0.0706**

**To find that percentage of republicans in state2 to be more than that of state1 is equivalent of finding the probability that state1 - state2 is less than 0.**

**, Zscore = (0 - 0.05)/0.0706 = -0.7082 and the corresponding percentage probability is 0.24**

**Question 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?

Answer : **Individual score = 1100**

**mean = 1026**

**std deviation = 209**

**Z score is (1100-1026)/209 = 0.3540 which is equivalent to 63.68%.**

**Hence, the individual scored 63.68% more than rest of the test taker and 36.32% less than rest of the test taker.**

**Task 2**

**Question 1** : Is gender independent of education level? A random sample of 395 people were

surveyed and each person was asked to report the highest education level they

obtained. The data that resulted from the survey is summarized in the following table:

High School Bachelors Masters Ph.d. Total

Female 60 54 46 41 201

Male 40 44 53 57 194

Total 100 98 99 98 395

Are gender and education level dependent at 5% level of significance? In

other words, given the data collected above, is there a relationship between the gender

of an individual and the level of education that they have obtained?

So that

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gender/Education** | **High School** | **Bachelors** | **Masters** | **Ph.d.** | **Total** |
| **Female** | **60** | **54** | **46** | **41** | **201** |
| **Male** | **40** | **44** | **53** | **57** | **194** |
| **Total** | **100** | **98** | **99** | **98** | **395** |

**Here, I use Chi-square test to test the independence of the categorical features.**

**Ho(Null Hypothesis) : Gender and Education are Independent.**

**Ha(Alternative Hypothesis) : Gender and Education are dependent.**

**χ2=∑(O−E)^2/E where χ2 is Chi-square, O is Observed Frequency and E is Expected Frequency**

**E = (Total Row \* Total Column)/Size of Sample where Size of Sample is given as 395**

**Hence,**

* **1>Observation value is 60 and it's related Expected value is (201\*100)/395 = 50.886**
  + **2>Observation value is 40 and it's related Expected value is (194\*100)/395 = 49.113**
  + **3> Observation value is 54 and it's related Expected value is (201\*98)/395 = 49.868**
  + **4> Observation value is 44 and it's related Expected value is (194\*98)/395 = 48.131**
  + **5> Observation value is 46 and it's related Expected value is (201\*99)/395 = 50.377**
  + **6> Observation value is 53 and it's related Expected value is (194\*99)/395 = 48.622**
  + **7> Observation value is 41 and it's related Expected value is (201\*98)/395 = 49.868**
  + **8> Observation value is 57 and it's related Expected value is (194\*98)/395 = 48.131**

**So, χ2= (60-50.886)^2/50.886 + (40-49.113)^2/49.113 + (54-49.868)^2/49.868 + (44-48.131)^2/48.131 + (46-50.377)^2/50.377 + (53-48.622)^2/48.622 + (41-49.868)^2/49.868 + (57-48.131)^2/48.131 = 1.632 + 1.690 + 0.342 + 0.354 + 0.380 + 0.394 + 1.576 + 1.634 = 8.002**

**Degree of freedom = (Number of Columns -1)\*(Number of Rows -1) = (4-1)\*(2-1) = 3\*1 = 3**

**Now, referring Chi Squared table with input params (Degree of freedom = 3 and level of significance = 0.05), the critical value for χ2 is found to be 7.814.**

**It's found that the calculated χ2 value(8.002) is greater than Critical χ2 value(7.814).**

**Hence, I can infer that Alternative Hypothesis is true here which means Gender and Education are dependent each other.**

**Question 2**: Using the following data, perform a one way analysis of variance using α=.05. Write up

the results in APA format.

[Group1: 51, 45, 33, 45, 67]

[Group2: 23, 43, 23, 43, 45]

[Group3: 56, 76, 74, 87, 56]

Answer :

**Using one way ANOVA, we conduct Hypothesis Testing.**

**N is total number of data points across all groups**

**n is total number of data points within a individual group**

**a is total number of levels of factor**

**N = 15, n = 5, a = 3 and α=.05**

**df-between = a-1 = 3-1 = 2**

**df-within = N-a = 15-3 = 12**

**df-total = N-1 = 15-1 =14**

**Now, we have to calculate as follows:-**

**Group1 Mean : (51+45+33+45+67)/5 = 241/5 = 48.2**

**Group2 Mean: (23+43+23+43+45)/5 = 177/5 = 35.4**

**Group3 Mean: (56+76+74+87+56)/5 = 349/5 = 69.8**

**Group1 Variance = (51-48.2)^2+(45-48.2)^2+(33-48.2)^2+(45-48.2)^2+(67-48.2)^2/5-1 = 612.8/4 = 153.2**

**Group2 Variance = (23-35.4)^2+(43-35.4)^2+(23-35.4)^2+(43-35.4)^2+(45-35.4)^2/5-1 = 515.2/4 = 128.8**

**Group3 Variance = (56-69.8)^2+(76-69.8)^2+(74-69.8)^2+(87-69.8)^2+(56-69.8)^2/5-1 = 732.8/4 = 183.2**

**Mean Square-within = (153.2+128.8+183.2)/3 = 155.07**

**SumofSquares-within = Mean Square-within \* df-within = 155.07\*12 = 1860.8**

**Mean of Group Means = (48.2+35.4+69.8)/3 = 153.4/3 = 51.13**

**Variance of Groups Means = (48.2-51.13)^2+(35.4-51.13)^2+(69.8-51.13)^2/3-1 = 604.58/2 = 302.29**

**Mean Square-between = 302.29\*5 = 1511.45**

**SumofSquares-between = Mean Square-between\* df-between = 1511.45\*2 = 3022.9**

**F = Mean Square-between/Mean Square-within = 1511.45/155.07 = 9.75**

**Referring F table for α=.05, the corresponding value for df-between and df-within is 3.8853 which means if the calculated F is greater than F critical(2,12) i.e- 3.8853 then we reject Null Hypothesis.**

**Inference : Reject Null Hypothesis**

**Question 3** : Calculate F Test for given 10, 20, 30, 40, 50 and 5,10,15, 20, 25.

For 10, 20, 30, 40, 50:

Answer : **F test is known as ratio of variance of set of values.**

**Set 1 : 10, 20, 30, 40, 50**

**Mean of Set 1 = (10+20+30+40+50)/5= 30**

**Variance of Set 1 =((10-30)^2+(20-30)^2+(30-30)^2+(40-30)^2+(50-30)^2)/5-1 = 250.25**

**Set 2 : 5,10,15, 20, 25**

**Mean of Set 2 = (5+10+15+20+25)/5= 15**

**Variance of Set 2 =((5-15)^2+(10-15)^2+(15-15)^2+(20-15)^2+(25-15)^2)/5-1 = 62.75**

**F Test for 10, 20, 30, 40, 50 = Variance of Set 1/ Variance of Set 2 = 250.25/62.75 = 3.988**