**Name = Vishal Kumar**

**Email = kumarvisha28@gmail.com**

**Assignment = Peer-graded Assignment: Analysing Student Performance**

**Course: Coursera: Pre-MBA Statistics IIMA - IIM Ahmedabad**

**In an MBA class, the probability that a student is from an engineering background is 60%, a commerce background is 30%, or some other background is 10%. Given that a student has an engineering background, their scores in component 1 are distributed (approximately) as a normal random variable with a mean 75 and a standard deviation 3. Given that a student has a commerce background, their scores in component 1 are distributed (approximately) as a normal random variable with a mean 76 and a standard deviation of 5. Given that a student has neither an engineering background nor a commerce background, their component 1 scores are distributed (approximately) as a normal random variable with a mean 85 and a standard deviation of 4**.

**1. Given that a student has an engineering background, what is the probability they score 70 or less in component 1?**

Given:

Mean score for engineering background (M) = **75**

Standard deviation for engineering background (Sd) **= 3**

We need to find P (X <= 70 |E), where X represents the component 1 score

Score = **70**

Standardize the score of 70 using the formula for z-score:

Z = X-M/Sd = **-1.666666667**

We need to find P (Z<= -1.667), where Z is a standard normal random variable with a mean of 0 and a standard deviation of 1.

Consulting the standard normal distribution table or using a calculator:

P (Z<= -1.667) = **0.0475**

**2. In the given data set, what fraction of students with an engineering background have scored 70 or less in component 1?**

|  |  |  |
| --- | --- | --- |
| **Using data set to calculate** | | |
| **Total Number of students** | **300** | **100%** |
| **students from engineering** | **184** | **61.3333%** |
| **Students from engineering who got less than or equal to 70 marks in Component 1** | **8** | **4.3478%** |

**3. Given the distributions, what is the expected value of the class score in component 1?**

Probability of engineering background P(E) 0.6

Probability of commerce background P(C) 0.3

Probability of Others background P(C) 0.1

**Expected scores and their distributions:**

|  |  |  |
| --- | --- | --- |
| **Background** | **Mean(M)** | **Standard Deviation (Sd)** |
| **Engineering (Eng)** | **75** | **3** |
| **Commerce (Com)** | **76** | **5** |
| **Others (Oth)** | **85** | **4** |

Calculate the expected value of the class score:

**E(class) = P(E)\*Eng(M)+P (C)\*Com (M)+P(O)\*Oth(M)**

**E(Class) = 76.3**

**4. In the given data set, what is the average score by students in component 1?**

Average score of students in Component 1 = **75.47333**

Mean = Sum of individual numbers

Number of observations

= 22642/300

= **75.473333**

**5. Given that a student scored 80 or more in component 1, what is the probability that this student is neither from an engineering background nor a commerce background?**

To solve this problem, we will use Bayes' theorem and the law of total probability

Let's denote:

**E**: Event that a student is from an engineering background.

**C**: Event that a student is from a commerce background.

**O**: Event that a student is from some other background (neither engineering nor commerce).

Given probabilities:

**P(E) = 0.6**

**P (C) =0.3**

**P(O) = 0.1**

We need to find P (O∣X≥80), where X is the score in component 1.

Using Bayes' theorem:

**P(O∣X≥80) = [ P(X≥80∣O) ⋅ P(O)] / P(X≥80)**

Calculate P(X≥80) using the law of total probability:

**​ P(X≥80) =P(X≥80∣E) ⋅ P(E) + P(X≥80∣C) ⋅ P(C) + P(X≥80∣O) ⋅ P(O)**

calculate each term:

|  |  |  |
| --- | --- | --- |
| **Background** | **Mean(M)** | **Standard Deviation (Sd)** |
| **Engineering (Eng)** | **75** | **3** |
| **Commerce (Com)** | **76** | **5** |
| **Others (Oth)** | **85** | **4** |

For engineering background (E):

**P(X≥80∣E) = P (Z ≥ (80−75)/ 3)**

**P (Z≥ 1.67) = 0.0475**.

For commerce background (C):

**P(X≥80∣C) = P (Z≥ (80−76)/5)**

**​ P(Z≥0.8) = 0.2119.**

For other background(O):

**P (X≥80∣O) =P (Z≥ (80−85)/4)**

**​=P(Z≥−1.25) =P (Z≤1.25) = 0.8944.**

Now compute:

**P(X≥80) =P(X≥80∣E) ⋅ P(E) + P(X≥80∣C) ⋅ P(C) + P(X≥80∣O) ⋅ P(O)**

**P(X≥80) =0.0475⋅0.60+0.2119⋅0.30+0.8944⋅0.10**​

**P(X≥80) = 0.0285 + 0.06357 + 0.08944 = 0.18151**

Using Bayes' theorem:

**P(O∣X≥80) = [ P(X≥80∣O) ⋅ P(O)] / P(X≥80)**

**P(O∣X≥80) = [ 0.8944 ⋅ 0.1]/ 0.18151**

**P(O∣X≥80) = 0.4929**

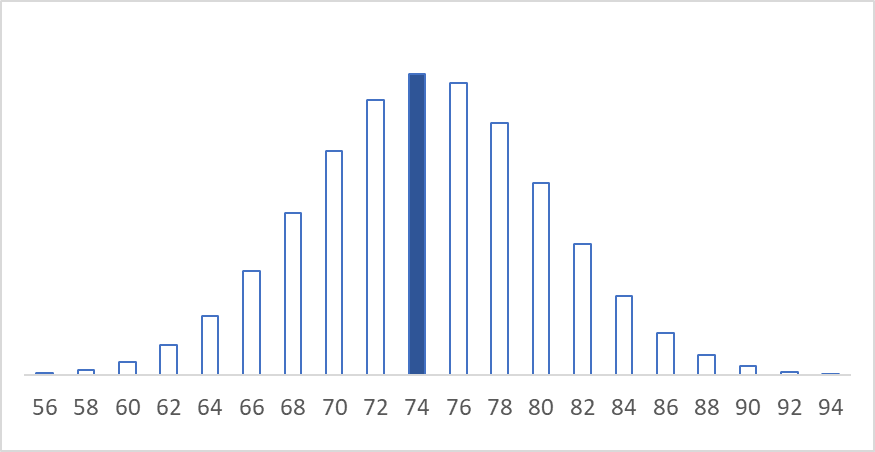
Therefore, the probability that a student is neither from an engineering background nor a commerce background given that they scored 80 or more in component 1 is approximately **0.4929.**

**6. What percentage of the students who have scored over 80 in component 1 are neither from an engineering background nor a commerce background?**

|  |  |  |
| --- | --- | --- |
| **Total number of students** | **300** | **100%** |
| **Total number of students from other background** | **13** | **4.3333%** |
| **Students who have scored over 80 in component 1 are neither from engineering nor commerce** | **11** | **3.6667%** |

**7.** **The final score obtained by a student is the average of the scores in the three components. Draw a sample of the students by choosing students with serial numbers 1, 11, 21, … 291. Assume this to be a random sample.**

**a) Based on this sample, what is a point estimate of the mean score of students taking this course?**

****

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Serial no.** | **Mean** | **Serial no.** | **Mean** | **Serial no.** | **Mean** |
| 1 | 71.000 | 101 | 72.000 | 201 | 67.000 |
| 11 | 82.667 | 111 | 77.333 | 211 | 64.333 |
| 21 | 71.000 | 121 | 77.000 | 221 | 72.333 |
| 31 | 88.667 | 131 | 69.667 | 231 | 83.333 |
| 41 | 74.667 | 141 | 79.333 | 241 | 75.000 |
| 51 | 77.333 | 151 | 72.667 | 251 | 66.333 |
| 61 | 77.000 | 161 | 78.000 | 261 | 71.333 |
| 71 | 66.000 | 171 | 72.333 | 271 | 75.333 |
| 81 | 79.667 | 181 | 71.667 | 281 | 70.333 |
| 91 | 68.333 | 191 | 82.667 | 291 | 80.333 |
| **Total** | **Xi =75.633** | **Total** | **Xi = 75.267** | **Total** | **Xi = 72.567** |

Sample size = 30

Sample mean = 74.489

Sample standard deviation = 5.804

**The point estimate of the mean score of this course is 74.489**

**b) Based on this sample, what is a 95% confidence interval of the mean scores of students taking this course?**

Sample size = 30

Sample mean = 74.489

Sample standard deviation = 5.903

t α/2 is the critical value of the t-distribution with n−1 degrees of freedom and confidence level 𝛼 =0.05 (for a 95% confidence interval).

Lookup t α/2 for α/2= 0.025 and 𝑑𝑓=29 in the t-distribution table or using a calculator:

t α/2 ​ = **2.045** (for a two-tailed test)

Now calculate the margin of error:

Margin of Error=t α/2 × n/ sqrt(n)

Margin of Error = **2.045 \* 5.903/sqrt (30)**

Margin of Error = **2.045 \* 1.077**

Margin of Error = **2.204**

Confidence Interval = mean +- Margin of error

Confidence Interval = **72.285 +- 76.693**

Therefore, the 95% confidence interval for the mean scores of students taking the course is approximately (**72.285,76.693**).

**8. Suppose that the data set here represents a sample from a population of similar students. Based on the data set, would you conclude that students with engineering backgrounds have an average score of 75 in component 1? (Use α=0.05).**

Assuming the data set represents similar students of engineering

**Null Hypothesis (𝐻0):** The average score (𝜇) of students with engineering backgrounds in component 1 is 75.

**𝐻0: 𝜇 = 75**

**Alternative Hypothesis (𝐻1):** The average score (𝜇) of students with engineering backgrounds in component 1 is not equal to 75.

**𝐻1: 𝜇 ≠ 75**

Given:

From the data sample for the score of Component 1:

**Sample Size = 30**

**Sample Mean (Xi) = 74.8667**

**Sample Standard Deviation (s) = 4.1167**

Unbiased sample standard deviation = Sqrt(n/n-1) \* s

​ Unbiased sample standard deviation= sqrt (30/29) \* 4.1167

**Unbiased sample standard deviation = 4.1871**

**Significance level: α = 0.05**

Standard error = 4.1871/sqrt (30)

Standard error =4.1871/5.4772

**Standard error =0.7644**

For a two-tailed test with α = 0.05 and degrees of freedom 𝑑𝑓 = n−1 =29, the critical values are approximately ±2.045 (from t-distribution table).

**Based on the critical value approach:**

The region of acceptance is [74.8867 - 2.045/0.7644, 74.8867+ 2.045/0.7644]

The region of acceptance is [74.8867 - 1.5633, 74.8867+ 1.5633]

The region of acceptance is **[76.4300, 73.3034]**

**Conclusion:** The mean 74.8867 lies in the acceptance region, so we cannot reject the null hypothesis.

**Based on the p-value approach:**

The sample mean is (74.8867-75) = **-0.1333** away from the hypothesised mean.

The probability that the mean of the sample size 30 is at least **0.1333** away from the hypothesised mean is:

2\* *Prob* [T < -0.1333/0.7644] = **0.8628**

**P value > 0.05, fails to reject the null hypothesis.**

Conclusion: **Yes, we conclude that students with engineering backgrounds have an average score of 75 in component 1.**

**9. Suppose we choose two random samples of 30 students each from the class. The first sample contains students who have commerce backgrounds, and the second sample contains students who have engineering backgrounds. The average of the scores of the first sample is 75.8333, with a sample standard deviation of 5.7813. The average of the scores of the second sample is 74.7444, with a sample standard deviation of 3.4416. Based on these samples, would you conclude that students with commerce backgrounds scored better than students with engineering backgrounds? (Use 𝛼 = 0.05).**

To determine if students with commerce backgrounds scored better than students with engineering backgrounds based on the given samples, we will conduct a hypothesis test.

**Null hypothesis (H₀):** The mean score of students with commerce backgrounds (𝜇1) is equal to or less than the mean score of students with engineering backgrounds (𝜇2).

**𝐻0: 𝜇1 ≤ 𝜇2**

**Alternative hypothesis (H₁):** The mean score of students with commerce backgrounds (𝜇1) is greater than the mean score of students with engineering backgrounds (𝜇2).

**𝐻1: 𝜇1 > 𝜇2**

The significance level (𝛼) is given as **0.05.**

**Given:**

**Commerce Background:**

Sample Size = 30

Sample Mean (𝜇1) = 75.8333

sample standard deviation (*s1*) = 5.7813.

unbiased sample standard deviation (*s2*) = 5.880

**Engineering Background:**

Sample Size = 30

sample Mean (𝜇2) = 74.7444

sample standard deviation (*s2*) = 3.4416.

Unbiased sample standard deviation (*s2*) = 3.500

**The estimate of the statistic = 75.8333 - 74.7444 = 1.0899**

The estimator follows the t distribution with the standard error:

**√ (5.8802 + 3.5002)/30 = 1.2493**

**Now, compare t with the critical value from the t-distribution with 𝑑𝑓= 58 (degrees of freedom), at 𝛼 = 0.05 (one-tailed test).**

**T =** 1.0899/1.2494 **= 0.8724**

The test statistic T equals **0.8724**, which is in the 95% region of acceptance:

[-∞ : 1.6716].

**P value approach:**

*Prob* [T < 0.8724] = 0.1933

**The p-value equals 0.1933, (p (x ≤ T) = 0.1933). It means that the chance of a type I error, rejecting a correct H0, is too high: 0.1933 (19.33%).**

The larger the p-value the more it supports *H*0.

**Conclusion: Since p-value > α, H0 cannot be rejected.**

The average of Group-1's population is assumed to be less than or equal to the average of Group-2's population.

In other words, the average of Sample -1 is greater than the average of Sample-2, but not big enough to be statistically significant and for the null assumption to be rejected.

**10. Based on the average of all three components, grades are awarded to students. The rulebook says that no more than 20% of the students can score the highest grade: A, and together, no more than 60% of the students can score the highest and second highest grade i.e., A and B. Students who score neither A nor B get a C grade. Now, the instructor wants to give as many A and B grades as the rule book permits. In this context, answer the following questions:**

Total number of students: **300**

Calculate the maximum number of students who can receive **grade A** (highest grade). According to the rulebook, no more than 20% of the students can receive grade A.

**20% of 300=0.2×300=60**

Calculate the maximum number of students who can receive **grade B** (second highest grade).

**60% of 300 = 0.6×300 =180**

Since we already have 60 students who can receive grade A, the maximum number of additional students who can receive grade B is:

**180 − 60= 120**

So, the maximum number of students who can receive grade B is 120.

Calculate the number of students who will receive **grade C** (neither A nor B). Students who do not receive grades A or B will receive grade C.

Total students− (Students receiving A+ Students receiving B) =

**300−(60+120) = 300−180 = 120**

Therefore, the number of students who will receive grade C is **120.**

**Summary of grade distribution:**

* Grade A: 60 students
* Grade B: 120 students
* Grade C: 120 students

**a) What is the average score obtained by students with A grades?**

By using a data

Average score of A grade students = 4945/60 = **82.4167**

**b) What is the average score obtained by students with B grades?**

By using a data

Average score of B grade students = 9146/120 = **76.2167**

**c) What fraction of students with an engineering background have scored A grades?**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Student grade** | **A** | **B** | **C** | **Total** |
| **Commerce** | **29** | **36** | **38** | **103** |
| **Others** | **11** | **1** | **1** | **13** |
| **Engineering** | **20** | **83** | **81** | **184** |
| **Total** | **60** | **120** | **120** | **300** |

Total number of engineering students = 184

Students who got an A from the engineering background = 20

The fraction = 20/184 **= 0.10869 or 10.869%**

**d) Among students who scored A grade, what fraction had engineering backgrounds?**

The total number of students who got an A = 60

Students who got an A from the engineering background = 20

The fraction = 20/60 **= 0.33333 or 33.3333%**