**Statistics**

**Q-1. A university wants to understand the relationship between the SAT scores of its**

**applicants and their college GPA. They collect data on 500 students, including their SAT**

**scores (out of 1600) and their college GPA (on a 4.0 scale). They find that the correlation**

**coefficient between SAT scores and college GPA is 0.7. What does this correlation**

**coefficient indicate about the relationship between SAT scores and college GPA?**

Answer: a correlation coefficient 0.7 shows that strong relationship between SAT scores and college GPA.as correlation coefficient range lies between -1 to +1

**Q-2. Consider a dataset containing the heights (in centimeters) of 1000 individuals. The**

**mean height is 170 cm with a standard deviation of 10 cm. The dataset is approximately**

**normally distributed, and its skewness is approximately zero. Based on this information,**

**answer the following questions:**

**a. What percentage of individuals in the dataset have heights between 160 cm**

**and 180 cm?**

**b. If we randomly select 100 individuals from the dataset, what is the probability**

**that their average height is greater than 175 cm?**

**c. Assuming the dataset follows a normal distribution, what is the z-score**

**corresponding to a height of 185 cm?**

**d. We know that 5% of the dataset has heights below a certain value. What is**

**the approximate height corresponding to this threshold?**

**e. Calculate the coefficient of variation (CV) for the dataset.**

Answer:given

Population mean=170

n=1000

std=10

1. z=mean-population mean/std

For 160

z=160-170/10

z=-1(0.8413)

For 180

z=180-170/10

z=+1(0.84213)

168%

b)p(z>175)

p(z>175-170/10/sqrt(100)

p(z>5)

c)z=185-170/10

z=1.5

e) cv=std/mean\*100

Cv = (10/170) \*100

cv=5.88

**Q-3. Consider the ‘Blood Pressure Before’ and ‘Blood Pressure After’ columns from the**

**data and calculate the following**

answer: a) for blood pressure before

range=max-min value

=148-120

=28

For bp after:

Range=max-min value

=141-118

=23

Hence blood pressure before is having the more spread

**Q-4. A group of 20 friends decide to play a game in which they each write a number**

**between 1 and 20 on a slip of paper and put it into a hat. They then draw one slip of paper**

**at random. What is the probability that the number on the slip of paper is a perfect square?**

**(i.e., 1, 4, 9, or 16)?**

Answer: perfect square between 1 to 20 is 1,4,9,16

Total outcome=20

Probability =favorable/total outcome

=4/20

=0.2

**Q-5. A certain city has two taxi companies: Company A has 80% of the taxis and**

**Company B has 20% of the taxis. Company A's taxis have a 95% success rate for picking**

**up passengers on time, while Company B's taxis have a 90% success rate. If a randomly**

**selected taxi is late, what is the probability that it belongs to Company A?**

Answer: a: selecting taxi belongs to company A

B: selecting the taxi late

p(a/b) =p(a)\*p(b/a)/p(b)

P(B) = P(B|A) \* P(A) + P (B|not A) \* P (not A)

p(not a)=1-p(a)

=1-0.8

p(b)=p(b/a)\*p(a) +p(b/nota)\*p(not a)

p((b)=0.95\*0.8+0.90\*0.2

p(b)=0.94

So finally

p(a/b)=p(a)\*p(b/a)/p(b)

=0.95\*0.8/0.94

=0.8085

**Q-7. The equations of two lines of regression, obtained in a correlation analysis**

**between variables X and Y are as follows:**

**and . 2X + 3 − 8 = 0 2Y + X − 5 = 0 The variance of X = 4 Find the**

**a. Variance of Y**

**b. Coefficient of determination of C and Y**

**c. Standard error of estimate of X on Y and of Y on X.**

a) variance of Y =16

b) 1

c)0

**Q-8. The anxiety levels of 10 participants were measured before and after a new therapy.**

**The scores are not normally distributed. Use the Wilcoxon signed-rank test to test whether**

**the therapy had a significant effect on anxiety levels. The data is given below: Participant**

**Before therapy After Therapy Difference**

Answer:

Given n=10

H0=there is no difference between means

H1=there is difference between means

Level of significance =0.05

Test static:

T=sqrt((1/n-1) \*Ed^^2-(Ed) ^^2)

T=0.098

Table value=(n-1) df

=10-1 df

=2.262

Conclusion:

Table value is greater than calculated value

Accept the null hypothesis

**Q-9. Given the score of students in multiple exams**

**Test the hypothesis that the mean scores of all the students are the same. If not, name the**

**student with the highest score.**

Answer:

Null hypothesis=all means are equal

H1=at least two means are equal

Level of significance=0.05

Test statistics:

Correction factor= 107188.3

Total sum square= 1049.733333

F1= 53.73333

F2= 869.0667

Error=126.93

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| source | df | ss | mss | f ratio |
| f1 | 4 | 53.73333 | 13.43333 | 0.846661 |
| f2 | 2 | 869.0667 | 434.5333 | 27.38727 |
| error | 8 | 126.93 | 15.86625 |  |
|  |  |  |  |  |

Table value =f1=0.06

F2=0.010

Conclusion:

Table value is less than the calculated value hence reject null hypothesis

**Q-10. A factory produces light bulbs, and the probability of a bulb being defective is 0.05.**

**The factory produces a large batch of 500 light bulbs.**

**a. What is the probability that exactly 20 bulbs are defective?**

**b. What is the probability that at least 10 bulbs are defective?**

**c. What is the probability that at max 15 bulbs are defective?**

**d. On average, how many defective bulbs would you expect in a batch of 500?**

**answer:** a) p(x=20) =ncx p^x q^n-x

p(x=20) =0.1397

b) p(x>=10) =1-p(x<10)

p(x<10) =0.0236

p(x>=10) =0.9764

c)p(x<=15) =0.9997

d)e(x)=n\*p

=500\*0.05

=25

**Q-12. A pharmaceutical company develops a new drug and wants to compare its**

**effectiveness against a standard drug for treating a particular condition. They conduct a**

**study with two groups: Group A receives the new drug, and Group B receives the standard**

**drug. The company measures the improvement in a specific symptom for both groups after**

**a 4-week treatment period.**

**a. The company collects data from 30 patients in each group and calculates the**

**mean improvement score and the standard deviation of improvement for each**

**group. The mean improvement score for Group A is 2.5 with a standard**

**deviation of 0.8, while the mean improvement score for Group B is 2.2 with a**

**standard deviation of 0.6. Conduct a t-test to determine if there is a significant**

**difference in the mean improvement scores between the two groups. Use a**

**significance level of 0.05.**

**b. Based on the t-test results, state whether the null hypothesis should be**

**rejected or not. Provide a conclusion in the context of the study.**

Answer: given

Group a: mean=2.5,std=0.8,n1=30

Group B:mean=2.2,std=0.6,n2=30

H0:two means are not equal

H1: two means are equal

Level of significance: 0.05

Test statistic:

t=mean1-mean2/sqrt(s1^2/n1)+sqrt(s2^2/n2)

t=1.64

Table value=t follows (n1+n2-2)degree of freedom

=2.00

Table value is greater than tested value

b)reject the null hypothesis

Group a has different.measures the improvement in a specific symptom of group b