1. Find the Karl Pearson Correlation coefficient

Method 1:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X | Y | X\*Y | X^2 | Y^2 |
|  | 57 | 10 | 570 | 3249 | 100 |
|  | 42 | 60 | 2520 | 1764 | 3600 |
|  | 40 | 30 | 1200 | 1600 | 900 |
|  | 33 | 41 | 1353 | 1089 | 1681 |
|  | 42 | 29 | 1218 | 1764 | 841 |
|  | 45 | 27 | 1215 | 2025 | 729 |
|  | 42 | 27 | 1134 | 1764 | 729 |
|  | 44 | 19 | 836 | 1936 | 361 |
|  | 40 | 18 | 720 | 1600 | 324 |
|  | 56 | 19 | 1064 | 3136 | 361 |
|  | 44 | 31 | 1364 | 1936 | 961 |
|  | 43 | 29 | 1247 | 1849 | 841 |
| Sum | 528 | 340 | 14441 | 23712 | 11428 |

r=(∑XY-(∑X\*∑Y)/n)/ sqrt(∑X^2(-∑(X)^2/n)\* sqrt(∑Y^2(-∑(Y)^2/n)

r=(14441-(528\*340)/12)/ sqrt(23712 -(528^2/12))\*sqrt(11428 -(340^2/12))

r=14441-14960/sqrt(23712-23232)\* sqrt(11428-9633.33)

r=-519/21.90\*42.36 = **0.5594**

Method 2:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | X | Y | x-x̄ | y- ȳ | (x-x̄)^2 | (y- ȳ)^2 | x\*y |
|  | 57 | 10 | 13 | -18.33 | 169 | 336.11 | -238.333 |
|  | 42 | 60 | -2 | 31.67 | 4 | 1002.78 | -63.3333 |
|  | 40 | 30 | -4 | 1.67 | 16 | 2.78 | -6.66667 |
|  | 33 | 41 | -11 | 12.67 | 121 | 160.44 | -139.333 |
|  | 42 | 29 | -2 | 0.67 | 4 | 0.44 | -1.33333 |
|  | 45 | 27 | 1 | -1.33 | 1 | 1.78 | -1.33333 |
|  | 42 | 27 | -2 | -1.33 | 4 | 1.78 | 2.666667 |
|  | 44 | 19 | 0 | -9.33 | 0 | 87.11 | 0 |
|  | 40 | 18 | -4 | -10.33 | 16 | 106.78 | 41.33333 |
|  | 56 | 19 | 12 | -9.33 | 144 | 87.11 | -112 |
|  | 44 | 31 | 0 | 2.67 | 0 | 7.11 | 0 |
|  | 43 | 29 | -1 | 0.67 | 1 | 0.44 | -0.66667 |
| Sum | 528 | 340 |  |  | 480 | 1794.6667 | -519 |
| x̄ or ȳ | 44 | 28.33333 |  |  |  |  |  |

sd(x)= sqrt((x-x̄)^2)=sqrt(480)=21.90

sd(y)= sqrt((y- ȳ)^2)= sqrt(1794.66)=42.36

r=∑XY/sd(x)\*sd(y)

=-519/21.90\*42.36= **0.5594**

1. Find the Spearman’s rank correlation coefficient.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X | Y | r1 | r2 | d=r1-r2 | d^2 |
| 97.8 | 73.2 | 5 | 8 | 3 | 9 |
| 99.2 | 85.5 | 1 | 2 | 1 | 1 |
| 98.8 | 78.9 | 2 | 5 | 3 | 9 |
| 98.3 | 75.8 | 4 | 7 | 3 | 9 |
| 98.4 | 77.2 | 3 | 6 | 3 | 9 |
| 96.7 | 87.2 | 7 | 1 | -6 | 36 |
| 97.1 | 83.8 | 6 | 4 | -2 | 4 |
| 80 | 85 | 8 | 3 | -5 | 25 |
|  |  |  |  |  | 102 |

r=1-((6∑d^2)/(n^3-n))

r=1-((6\*102)/(8^3-8))

r=1-(612/504)

r=**0.2145**

1. Obtain a regression line for the following:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | y | x-x̄ | y- ȳ | x^2 | y^2 | x\*y |
| 50 | 30 | -10 | -20 | 100 | 400 | 200 |
| 60 | 60 | 0 | 10 | 0 | 100 | 0 |
| 50 | 40 | -10 | -10 | 100 | 100 | 100 |
| 60 | 50 | 0 | 0 | 0 | 0 | 0 |
| 80 | 60 | 20 | 10 | 400 | 100 | 200 |
| 50 | 30 | -10 | -20 | 100 | 400 | 200 |
| 80 | 70 | 20 | 20 | 400 | 400 | 400 |
| 40 | 50 | -20 | 0 | 400 | 0 | 0 |
| 70 | 60 | 10 | 10 | 100 | 100 | 100 |
| 540 | 450 | 0 | 0 | 1600 | 1600 | 1200 |
| 60 | 50 |  |  |  |  |  |

byx = (n∑xy-∑x∑y)/(n∑x^2-(∑x)^2)

=(9(1200)-(60)(50))/9(1600)-(60)^2

=(10800-3000)/(14400-3600)

=7800/10800=**0.722**

bxy = (n∑xy-∑x∑y)/(n∑y^2-(∑y)^2)

=(9(1200)-(60)(50))/9(1600)-(50)^2

=(10800-3000)/(14400-2500)

=7800/11900=**0.655**

1. Find the least squares line for the data (10,14)(12,14)(15,23)(23,25) and (20,21) and compute the coefficient of determination.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | Y | x-x̄ | y- ȳ | x^2 | y^2 | x\*y |
| 10 | 14 | -6 | -6 | 36 | 36 | 36 |
| 12 | 17 | -4 | -3 | 16 | 9 | 12 |
| 15 | 23 | -1 | 3 | 1 | 9 | -3 |
| 23 | 25 | 7 | 5 | 49 | 25 | 35 |
| 20 | 21 | 4 | 1 | 16 | 1 | 4 |
| 80 | 100 | 0 | 0 | 118 | 80 | 84 |
| 16 | 20 |  |  |  |  |  |

r= (n(∑xy )- (∑x)(∑y)) / sqrt((n∑x^2 – (∑x)^2)( n∑y^2 – (∑y)^2))

=(5(84)-(16)(20))/sqrt((5(118)-(16)^2))(5(80)-(20)^2)

=(420-320)/sqrt((590-256)(400-400))

=100

intercept of the least squares line:

m = Σ(XY) / Σ(X)²=84/118=0.711

intercept (b):

b = Ȳ - m \* X̄

b = 20 - 0.711 \* 16 = 8.614

equation of the least squares line is: Y = 0.711X + 8.614

sum of squares total (SST) = Σ(Y)² = 80

sum of squares regression (SSR)= Σ(XY)² / Σ(X)² = 84² / 118 = 58.0322

sum of squares error (SSE) = SST – SSR = 80 - 58.0322 = 21.9678

coefficient of determination (R²) is:

R² = SSR / SST

= 58.0322 / 80= 0.725

1. For which type of data Chi square test is used?

The chi-square test is used for categorical data. It is used to compare the distribution of categorical variables to a hypothesized distribution

1. State the formula of test statistic for Chi square test of goodness of fit

χ2 = Σ(O - E)^2 / E

where:

O is the observed frequency in each category

E is the expected frequency in each category

Σ is the sum of all values

1. The table below gives the number of accidents that occurred in a certain factory on the various days of particular week,

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Days | Sun | Mon | Tue | Wed | Thu | Fri | Sat |
| No. of accidents | 6 | 4 | 9 | 7 | 8 | 10 | 12 |

Test 5% level whether the accidents are uniformly distributed over the different days.

(H0): Accidents are equally distributed over all the days of week

(Ha) : Accidents do hot occur equally

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Days | No. of accidents | E | (O-E)^2 | X^2=((O-E)^2)/E |
| Sun | 6 | 8 | 4 | 0.5 |
| Mon | 4 | 8 | 16 | 2 |
| Tue | 9 | 8 | 1 | 0.125 |
| Wed | 7 | 8 | 1 | 0.125 |
| Thu | 8 | 8 | 0 | 0 |
| Fri | 10 | 8 | 4 | 0.5 |
| Sat | 12 | 8 | 16 | 2 |
| Total |  |  |  | 5.25 |

Level of significance: α= 0.05

Degree of freedom = h-1=7-1=6

Critical value ⇒For 6 degrees of freedom at 5% level of significance table value.

x2 is 12.59

Decision ⇒ Since the calculated value of x2 is less than the table value. The hypothesis is accepted.

∴The accidents occur equally on all working days.

1. A) State the name of non-parametric test which is equivalent to one way ANOVA

* The non-parametric test which is equivalent to one way ANOVA is the Kruskal-Wallis test.

B) List the names of non-parametric tests

* Mann-Whitney U test
* Wilcoxon signed-rank test
* Kruskal-Wallis test
* Friedman test
* Spearman's rank correlation test
* Kendall's rank correlation test
* Chi-square test

C) When do we use non-parametric tests instead of parametric tests?

* When the data does not meet the assumptions of parametric tests, such as normality or equal variances.
* When the data is measured on an ordinal or nominal scale rather than a continuous scale.
* When the sample size is small, and the population distribution is unknown or non-normal.

D) State the name of non-parametric test which is equivalent to paired t-test.

* The non-parametric test which is equivalent to paired t-test is the Wilcoxon signed-rank test.

E) Which nonparametric test is analogous to one sample t-test?

* The non-parametric test which is analogous to one sample t-test is the sign test.

F) Which nonparametric test is analogous to paired(dependent) t-test?

* The non-parametric test which is analogous to paired(dependent) t-test is the Wilcoxon signed-rank test.

G) Which nonparametric test is analogous to unpaired(independent) t-test?

* The non-parametric test which is analogous to unpaired(independent) t-test is the Mann-Whitney U test.