Sardar Vallabhbhai Patel Institute of Technology, Vasad

Subject: Probability and Statistics Subject Code: 310006

**Tutorial 4: Basic Statistics: Correlation and regression**

**Note: Example numbers 1, 2, 4, 6, 8, 9, 11, 12, 14, 16 will be solved in the tutorial. The remaining examples are for self-practice.**

**Q.1** The coefficient of rank correlation of marks obtained by 10 students in English and Economics was found to be 0.6. It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 7 instead of 1. Find the correct coefficient of rank correlation.

**Q.2** Compute Karl Pearson's coefficient of correlation between $X$ and $Y$ for the following data:

| X: | 100 | 98 | 78 | 85 | 110 | 93 | 80 |
|---|---|---|---|---|---|---|---|
| Y: | 85 | 90 | 70 | 72 | 95 | 81 | 74 |

**Q.3** Compute the correlation coefficients between X and Y using following data:

| X: | 2 | 4 | 5 | 6 | 8 | 11 |
|---|---|---|---|---|---|---|
| Y: | 18 | 12 | 10 | 8 | 7 | 5 |

**Q.4** Ten competitors in a test are ranked by three judges in the following order:

| Rank by First Judge: | 6 | 10 | 2 | 9 | 8 | 1 | 5 | 3 | 4 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| Rank by Second Judge: | 5 | 4 | 10 | 1 | 9 | 3 | 8 | 7 | 2 | 6 |
| Rank by Third Judge: | 4 | 8 | 2 | 10 | 7 | 5 | 9 | 1 | 3 | 6 |

Use the method of rank correlation to gauge which pairs of judges has nearest common approach.

**Q.5** Compute the coefficient of correlation between X and Y using the following data:

| X: | 2 | 4 | 5 | 6 | 8 | 11 |
|---|---|---|---|---|---|---|
| Y: | 18 | 12 | 10 | 8 | 7 | 5 |

**Q.6** Psychological tests of intelligence and of engineering ability were applied to 10 students as per the following data. Find the coefficient of correlation.

| Intelligence ration | 105 | 104 | 102 | 101 | 100 | 99 | 98 | 96 | 93 | 92 |
|---|---|---|---|---|---|---|---|---|---|---|
| Engineering ability | 101 | 103 | 100 | 98 | 95 | 96 | 104 | 92 | 97 | 94 |

**Q.7** Find the correlation coefficient for the following data:

| X: | -3 | -2 | -1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|
| Y: | 9 | 4 | 1 | 0.5 | 1 | 4 | 9 |

**Q.8** From the following data of the marks obtained by 8 students in Computer Networking (CN) and Complier Design (CD) papers, compute rank coefficient of correlation

| CN: | 15 | 20 | 28 | 12 | 40 | 60 | 20 | 80 |
|---|---|---|---|---|---|---|---|---|
| CD: | 40 | 30 | 50 | 30 | 20 | 10 | 30 | 60 |

**Q.9** Given that $n$ = 25, $\sum X$ = 125, $\sum X^2$ = 650, $\sum Y$ = 100, $\sum Y^2$ = 460 and $\sum XY$ = 508. Later, it was found that two of the points (8, 12) and (6, 8) were wrongly entered as (6, 14) and (8, 6). Prove that $r$ = 2/3 .

**Q.10** Obtain the two lines of regression for the following data:

| Sales (No of tablets) | 190 | 240 | 250 | 300 | 310 | 335 | 300 |
|---|---|---|---|---|---|---|---|
| Advertising expenditure (Rs) | 5 | 10 | 15 | 20 | 20 | 30 | 30 |

**Q.11** Find the regression equation showing the capacity utilization on production from the following data:

| | Average | Standard Deviation |
|---|---|---|
| Production (in lakh unit) | 35.6 | 10.5 |

<table>
<tr><td></td><td>
<table>
<tr><td>Capacity utilization (in %)</td><td>84.8</td><td>8,5</td></tr>
<tr><td>Correlation coefficient</td><td colspan="2">r = 0.62</td></tr>
</table>
Estimate the production when capacity utilization is 70%.
</td></tr>
<tr><td>Q.12</td><td>Calculate the regression coefficients and find the two lines of regression for the following data:
<table>
<tr><td>X:</td><td>57</td><td>58</td><td>59</td><td>59</td><td>60</td><td>61</td><td>62</td><td>64</td></tr>
<tr><td>Y:</td><td>67</td><td>68</td><td>65</td><td>68</td><td>72</td><td>72</td><td>69</td><td>71</td></tr>
</table>
Find the value of y when x=65.
</td></tr>
<tr><td>Q.13</td><td>The following are the lines of regression $9y = x + 288$ $and$ $4y = x + 38$. Estimate y when x = 99 and x when y = 30. Also, find the means of x and y</td></tr>
<tr><td>Q.14</td><td>Obtain correlation coefficient between x and y if two regression lines are 4x - 5y + 33 = 0 and 20x - 9y – 107 = 0.</td></tr>
<tr><td>Q.15</td><td>Obtain the regression line of y on x for the following data:
<table>
<tr><td>X:</td><td>100</td><td>98</td><td>78</td><td>85</td><td>110</td><td>93</td><td>80</td></tr>
<tr><td>Y:</td><td>85</td><td>90</td><td>70</td><td>72</td><td>95</td><td>81</td><td>74</td></tr>
</table>
</td></tr>
<tr><td>Q.16</td><td>The population ($p$) of a small community on the outskirts of a city grows rapidly over a 20 year period:
<table>
<tr><td>t:</td><td>0</td><td>5</td><td>10</td><td>15</td><td>20</td></tr>
<tr><td>p:</td><td>100</td><td>200</td><td>450</td><td>950</td><td>2000</td></tr>
</table>
As an engineer working for a utility company, you must forecast the population 5 years into the future to anticipate the demand for power. Employ an exponential model and linear regression to make this prediction.
</td></tr>
<tr><td>Q.17</td><td>In partially destroyed laboratory record of an analysis of correlation data, the following results are eligible.<br>• Variance of x, $\sigma^2 = 9$<br>• Two lines of regressions: 8x - 10y + 66 =0, 40x - 18y = 214.<br>From the above obtain mean values of x and y, the standard deviation of y and correlation coefficient.</td></tr>
</table>