**Assignment 5 – Bivariate Analysis**

**Dataset: bullriders**

**Research Question:** In 2012, what variable had the strongest linear relationship with Earnings:  Percentage rides or Cup points?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** |  | **Question** | **Answer**  **write the R code in answer column**  **Output in the Conclusion.** | **Marks** |
| 1. |  | Reading the dataset into a dataframe named bike in R using commands | bull\_riders = read.csv('BullRiders.csv') | 3 |
| 2. |  | **Examine the Dataset** |  |  |
|  | a) | Select the question that can be answered with correlation.  Is there a difference between the earnings of professional and non-professional bull riders?  Is there a relationship between the name of the bull and the number of times he has bucked a rider?  Which variable has the strongest linear relationship with earnings: successful ride percentage or Cup points?  On average, how much does a professional bull rider earn each year? | Which variable has the strongest linear relationship with earnings: successful ride percentage or Cup points?  **Ans: Cup Points**  cor(bull\_riders$CupPoints12, bull\_riders$Earnings12)  cor(bull\_riders$RidePer12, bull\_riders$Earnings12) | 2 |
| 3. |  | Make a Scatterplot of Earnings and Ride Percentage | plot(bull\_riders$RidePer12, bull\_riders$Earnings12) | 3 |
|  | b) | Does the scatterplot show a **linear** relationship? | **Yes**  plot(bull\_riders$RidePer12, bull\_riders$Earnings12) | 1 |
|  | c) | What is the **correlation** of Earnings with Ride Percentage for 2012? | **0.6481714** | 2 |
|  |  | Does the scatterplot show a **linear** relationship? | **Yes** |  |
| 4. |  | Make a Scatterplot of Earnings and Cup points | plot(bull\_riders$CupPoints12, bull\_riders$Earnings12) |  |
|  |  | Does the scatterplot show a **linear** relationship? | **Yes** |  |
|  |  | What is the **correlation** of Earnings with Cup Points for 2012? (report to three decimal places) | **0.743** |  |
| 5. |  | Find any outlier is there? If so remove it and re run to find the new correlation coefficient | **0.94**  bull\_riders\_12\_no\_outliers = bull\_riders[!(bull\_riders$Earnings12 == max\_value),] |  |
| 6. |  | **Conclusion** | | 5 |
|  |  | An initial examination of the relationships between Ride Percentage (RidePer) and Earnings, and Cup Points (CupPoints) and Earnings showed that Cup Points had the **stronger** [stronger/weaker] relationship to Earnings. Ride Percentage and Earnings showed a correlation value of **0.64** while Cup Points and Earnings had a correlation value of **0.743**. Visual examination showed an **outlier**. Removal of this increased the initial relationship: Ride Percentage and Earnings now had correlation value of **0.87** and Cup Points and Earnings had a correlation value of **0.94**. **Cupoints** [cup Points/ RidePer] had the higher relationship to Earnings.. | | |

**DataSet: AustinCityLimits**

**Research Question:** Among male artists, is there an association between winning a Grammy and the genre of music that he plays?

**Ans Research:**

**Conclusion**

There **is** [is/isnot] an association between winning a Grammy and the Genre of music an artist plays. The probability of winning a Grammy, regardless of Genre, is **42.2%**. However, examination of a **contingency** **table** containing both Grammy and Genre showed **variation** saying that the probability of winning a Grammy changes by genre. If an artist is in the Country Genre, the conditional probability of winning a Grammy is **10.3%**, while if an artist is in the **Rock/Folk/Indie Genre**, the conditional probability of winning a Grammy is **28.6%**. Visual examination of the barplot shows the conditional probabilities of winning a Grammy are **associated/related** across Genres.