**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 | bike = data.frame(BullRiders) | 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? | Yes Cup Points | 2 |
| 3. |  | Make a Scatterplot of Earnings and Ride Percentage | plot(bike$Earnings12,bike$RidePer12) | 3 |
|  | b) | Does the scatterplot show a **linear** relationship? | Yes | 1 |
|  | c) | What is the **correlation** of Earnings with Ride Percentage for 2012? | cor(bike$Earnings12,bike$Rides12)  0.7392811 | 2 |
|  |  | Does the scatterplot show a **linear** relationship? | Yes |  |
| 4. |  | Make a Scatterplot of Earnings and Cup points | plot(bike$Earnings12,bike$CupPoints12)  abline(lm(bike$CupPoints12~bike$Earnings12), col="red") |  |
|  |  | Does the scatterplot show a **linear** relationship? | Yes |  |
|  |  | What is the **correlation** of Earnings with Cup Points for 2012? (report to three decimal places) | round(cor(bike$Earnings12,bike$CupPoints12), digits = 3)  0.744 |  |
| 5. |  | Find any outlier is there? If so remove it and re run to find the new correlation coefficient | nooutlier <- bike[bike$Earnings12 < 1000000 ,]  vars <-c("Earnings12", "CupPoints12")  round(cor(nooutlier[,vars]), digits = 3)  Earnings12 CupPoints12  Earnings12 1.000 0.948  CupPoints12 0.948 1.000 |  |
| 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 \_\_\_\_**0.7392811**\_\_\_\_ while Cup Points and Earnings had a correlation value of  \_\_**0.744**\_\_\_\_. Visual examination showed an \_**outlier**\_\_\_\_\_\_. Removal of this increased the initial relationship: Ride Percentage and Earnings now had correlation value of  \_\_\_**0.876**  \_\_\_ and Cup Points and Earnings had a correlation value of \_\_**0.948**  \_\_\_\_\_. \_\_\_\_\_**Cup Points**\_\_\_ [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?

**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  **42.21** \_\_\_\_ %. However, examination of a **contingency\_**\_\_\_\_\_\_\_\_ table containing both Grammy and Genre showed that the **conditional** \_\_\_\_\_\_\_ 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 **Singer-Songwriter**\_\_\_\_\_\_ 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 **not equal**\_\_\_\_\_\_\_\_ across Genres.