Experience the adrenaline-packed world of professional bull riding (PBR), where the spotlight shines not only on the daring riders but also on the powerful bulls. In this worksheet, we'll dive into the data from the 2023 season of the Touring Pro Division to understand how points are achieved by bulls themselves.

1. Fit a logistic regression model where we want to predict the Avg BullScore that are greater than the mean `Avg BullScore` using orld champion average score, average Buckoff time, highest bull score and 45Pt rides.

A. Create new variable to implement this model and record the R-code below here.

bulls <- bulls %>%

mutate(betterThanAverage = ifelse(Avg\_BullScore > mean(Avg\_BullScore), 1, 0) )

B. Interpret the Coefficient of Highest BullScore predictor in both odds and log-odds form.

Beta-Hat\_{Highest BullScore} = -0.51984, which is in ***log-odds*** can be interpreted as for every one unit increase in the Highest BullScore, the ***log odds*** will decrease by 0.52, given all other variables remain constant.

And

Beta-Hat\_{Highest BullScore} = 0.5946157, which means for every one unit increase in the Highest BullScore, the ***odds*** of avg BullScore being better than mean-scores is decreased by a factor of 0.595.

C. If the bull has a World champ average of 46, an average Buckoff time of 2.5, highest BullScore of 44.5, and has one over 45Pt ride, what would his probability of having an average bull score that is higher than mean avg score.

newdata <- data.frame(World\_Champ\_Avg\_Score = 46,

Avg\_Buckoff\_Time = 2.5,

Highest\_BullScore = 44.5,

fortyFivePt\_Rides = 1)

predict(mod, newdata, type = "response")

\*Output\*---> 1.834293

exp(1.834293)/(1+exp(1.834293)) ----> 0.8622724 is the probability of this bull averaging a score greater than the mean-scores

D. Check the overall efficiency of the model, are there any concerning outliers or influential points we should be concerned about? What would we hypothetically do to fix these issues?

mod <- glm(betterThanAverage ~ World\_Champ\_Avg\_Score + Avg\_Buckoff\_Time + Highest\_BullScore + fortyFivePt\_Rides, data = bulls)

plot(mod, pch = 16)

Could improve the linearity of the residuals as well as the linearity on the normal q-q plot as there is significant fraying off the line. We could fix this by removing those points that are outliers or we could take a different route and use a transformation in the model to produce better variability and linearity.

2. Using the same model from question 1…

A. Produce 95% confidence intervals for the two of model’s population means.

confint(mod)

lwr upr

World\_Champ\_Avg\_Score 0.41983784 0.77721152

Avg\_Buckoff\_Time -0.03836348 0.01652574

Highest\_BullScore -0.73162062 -0.30805514

fortyFivePt\_Rides 0.71637235 1.04522278

B. Produce a 95% prediction interval, given that the bull’s World champ avg score is 45.5, avg buckoff time is 3.73, highest bull score 45.5, and has had a 45-point ride.

i. Record R-code here.

mod <- glm(`betterThanAverage` ~ World\_Champ\_Avg\_Score + Avg\_Buckoff\_Time + Highest\_BullScore + fortyFivePt\_Rides, data = bulls)

newdata <- data.frame(World\_Champ\_Avg\_Score = 45.5,

Avg\_Buckoff\_Time = 3.73,

Highest\_BullScore = 45.5,

fortyFivePt\_Rides = 1)

predict.glm(mod, interval = "prediction", newdata = newdata)

\*Output\* -----> 1.001762

ii. What is the probability off this of this Bull’s buckoff time being better than mean average?

exp(1.001762 )/(1+exp(1.001762)) ----> 0.7314049, 73.1% chance of this specific bull scoring better than the mean-scores.