**Inferential Statistics – Capstone 1 Project**

In order to understand how to slice our data and find patterns/relationships between different variables, we really have to dig deep into the key question we want to ask and what the stakeholders find important from a heuristic level. By integrating both the domain knowledge of both football experts and inferential techniques, we can begin to extract more relevant insights.

The key questions we are trying to answer are the following:

1. How strong is the relationship between athletic ability, measured by the drills at the NFL Combine, and how high the player is drafted?
2. What is this relationship, and can it be modeled?

Any football expert can attest that context is incredibly important in analyzing the Combine measurable. For example, running a 4.8 second 40-yard dash is either extremely bad or extremely good, depending on the context. With that in mind, the most important dimensions for providing context to the combine performances are the player’s position (QB, RB, DB, etc) and his weight. It makes intuitive sense to group players by their positions, since each group usually comprises of players that carry the same role for their team, so their physical attributes, abilities, and skillsets group together. In addition, even for the same players within the same position, a very important consideration is the player’s weight. Expectations for speed, agility, and strength drills are vastly different players with weight differences of 20 pounds, so the burning question is, by how much?

To address this question, we split this exercise into two parts. First, we run a regression for each drill against weight of all the players to get a good sense of the relationship. Next, we can investigate further into each position group and run the same regression so that we can extract more detailed insights into each, and possibly find new trends or patterns. Furthermore, we can understand the spread of this data by finding the R-squared of each of these regressions. Lastly, as a point of reference, we will find the average weights of each position, then use these to find the average measurable for each combine drill for each group (Rounds 1-3, 4-7, and Undrafted) by plugging this value into the regression equations. This will serve as a general sense of where each of these groups stands.

The results confirm the original hypotheses that vertical leap and broad jump have negative correlations with weight, while bench press, 40-yard dash, shuttle, and 3Cone are positively correlated with weight. The ranges of the R-squared for each exercise, in descending order, are:

1. 40-Yard Dash: 0.68 – 0.75
2. 3Cone: 0.53 – 0.61
3. Shuttle: 0.52 – 0.59
4. Broad Jump: 0.40 – 0.48
5. Bench Press: 0.36 – 0.42
6. Vertical Leap: 0.32 – 0.41

Furthermore, we can also confirm our intuition that performance at the NFL Combine has a fairly strong relationship with which draft group that the prospect gets selected in (Rounds 1-3, 4-7, or Undrafted). We can see this by observing the slopes and orders of the regression lines. For exercises where a lower value is considered a better performance (40-Yard Dash, 3Cone, and Shuttle), the regression line for Rounds 1-3 is predictably below the other two, with Rounds 4-7 above it, then Undrafted on top. These regression lines form the reverse order when the higher value is considered a better performance (Broad Jump, Bench Press, and Vertical Leap).

The position with the highest R-Squares are following for each exercise:

1. 40-Yard Dash: Defensive Linemen
2. Bench Press: Wide Receivers (we excluded Quarterbacks because of the very low sample size)
3. Vertical Leap: Defensive Linemen
4. Broad Jump: Defensive Linemen
5. Shuttle: Defensive Linemen
6. 3Cone: Defensive Linemen