The 2021 Football Season

Making Statistics Question about Football

Patrick Niederhauser

Probability and Applied Statistics, CSIS 3327

Byron Hoy

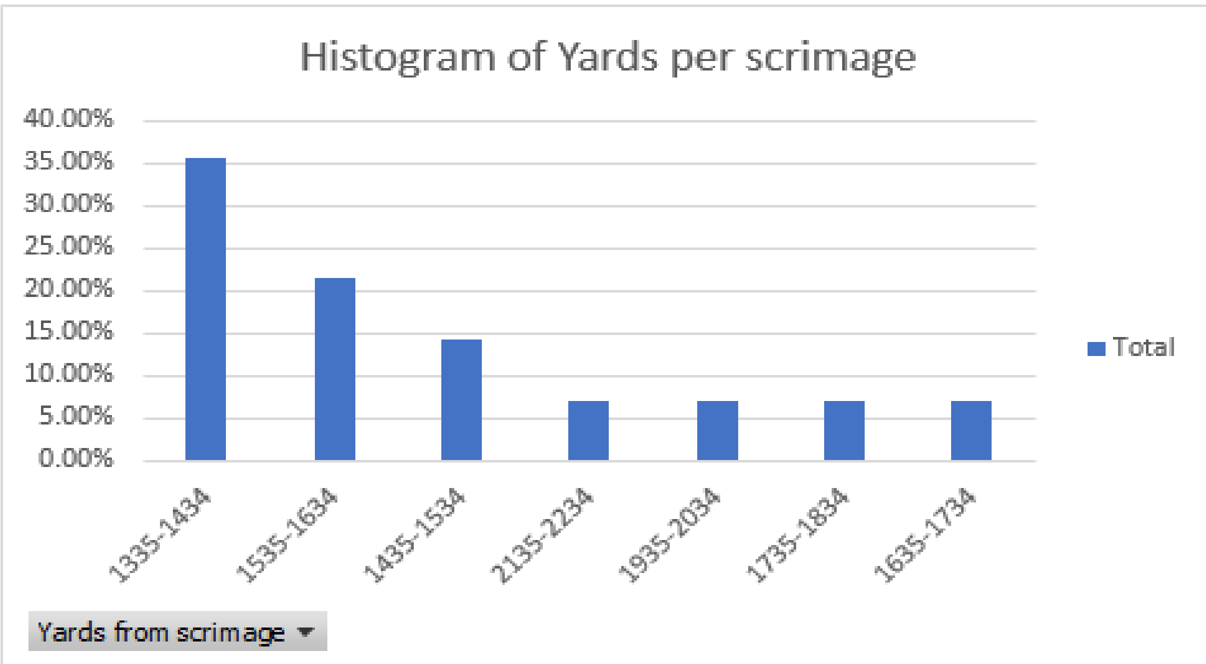
December 7th ,2022

Table of Contents

Section 1.2: Characterizing a Set of Measurements Graphical Methods

The most important statistic when talking about football is all-purpose yardage. The stat represents both passing yards and rushing yards and is the main statistic looked at when judging offensive efficiency. Offensive weapons such as the running back and wide river are mainly judged on total all-purpose yardage. Given below is the data for 15 of the top-performing players of the 2021 NFL season. Construct a relative frequency histogram for this data.

2,171	1,965	1,770	1,667	1,630
1,558	1,553	1,519	1,476	1,433
1,383	1,361	1,337	1,335	1,331

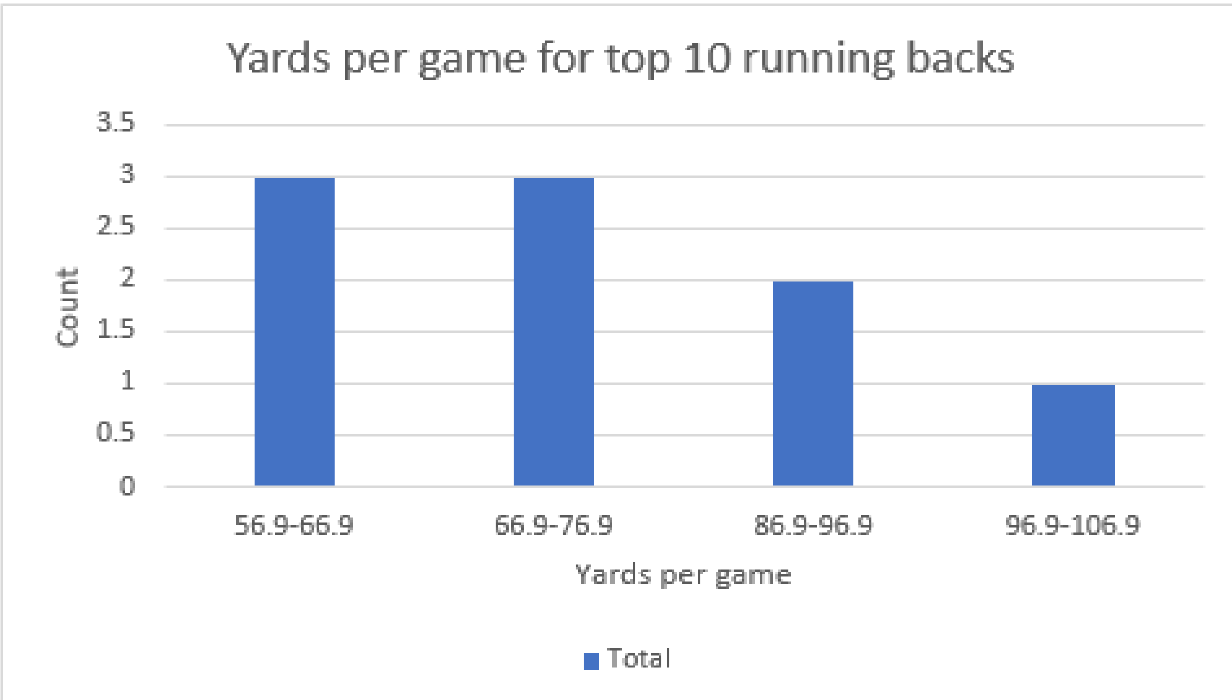


The above solution shows us the relative frequency of the yards of scrimmage for the 2021 NLF season. By looking at the graph we can identify that of the top 15 performers most of them are in the 1335-1434 yards range, this shows there is a steep decline in the top 15 of players. Notice how the top performers which fell into the 2135 and 2234 range don't have the lowest frequency. This is interesting because according to our graph people who fall into the 2 thousand plus yards range have the same frequency and several other lower ranges.

Section 1.3: Characterizing a Set of Measurements: Numerical Methods

The running back is one of the most important positions on a football team. When considering a running backs worth lots of teams will often look at their yards per game stat. This statistic shows how many yards per average a running back has gotten over the course of an NFL season. Below are the top 9 performing running backs of the 2021 NFL season. Construct a relative frequency histogram on this data and solve for the standard deviation of this data?

106.5	70.6	56.9	75.3	89.9
89.2	69.1	64.8	58.9	



Our above histogram is very interesting and represents lots of important data. When talking about a running back it almost sounds trivial for them to average a high number of rushing yards per game, but this graph shows how difficult it truly is. Out of the top 9 performing running backs only one of the average 100 yards per game. To solve for out standard deviation, we have to plug our data into the standard deviation formula, which is $\sqrt{\frac{\sum(x-x_1)^2}{n-1}}$. Once we plug in and solve, we get that our standard deviation is 15.48. Standard deviation is a representation of variance in data, so since our deviation is on the higher end, we can say that there is lots of variance in our data. This means that the player who averaged 106.5 yards per game last year is more than likely skewing our data. This always shows that this player is better than all the rest by a decent margin.

Section 2.3: A review of Set Notation

In football there are several different formations that an offense can run. A skill group is defined by your runbacks, wide receivers, and tight ends. If the offense is running a west coast offense, then they are only going to have 3 players from a skill group out at one time. Define the super set for this problem and then define the following subsets. A: if one wide receiver is on the field B: if one running back is on the field. Then find the union intercept and complement for a and b.

S={(wr,wr,wr), (wr,wr,rb),(wr,rb,rb),(rb,rb,rb,),(rb,rb,te),(rb,te,te),(te,te,te),(te,te,wr),(te,wr,wr)}

A= {(wr, brr,), (te,te,wr)}

B={(wr,wr,rb),(rb,te,te)}

$A \cup B$ = {(wr,rb,rb,),(te,te,wr),(wr,wr,rb),(rb,te,te)}

$A \cap B$ =∅

Á $= = \{(wr, wr, wr), (wr, wr, rb), (rb, rb, rb), (rb, rb, te), (rb, te, te), (te, te, te), (te, wr, wr)\}$

$B^{`} = \{(wr, wr, wr), (wr, rb, rb), (rb, rb, rb), (rb, rb, te), (te, te, te), (te, te, wr), (te, wr, wr)\}$

This information, although it seems like it has little to no importance for football, can help a lot. As a defense if I know a team is running a west coast offense then I know there are only 9 possibilities in which different skill groups can be active. This can help a lot with planning in preparation of a matchup against certain opponents.

Section 2.4: A Probabilistic Model for an Experiment: The Discrete Case

Every player in the NFL that plays defense is either a defensive lineman, linebacker, cornerback, or a safety. These positions make a defense whole and are required for a football team. Each position is either strong side (+) or weak side (-) depending on the orientation of the defenses. List the sample space for the defense.

$$S = \{(dl^+), (dl^-), (lb^+), (lb^-), (cb^+), (cb^-), (s^+), (s^-)\}$$

This set notation can help us identify different formations that are possible on the football field. This means that as an offense if you identify the weak side then you can run or throw the ball to that side easier. Although this doesn't directly identify that it gives us all the possibilities.

Section 2.5: Calculating the Probability of an Event: The Sample-Point Method

Four players are fighting in training camp to be the starting quarterback, all these players have an equal opportunity to get the job. Two players are listed as mobile quarterbacks, while the other two players are listed as pocket passers. The team can only roster two quarterbacks at a time, what is the probability that one of the rostered quarterbacks is a mobile one?

$$S = \{(pq_1pq_2), (pq_1mq_1), (pq_1mq_2), (pq_2mq_1), (pq_2mq_2), (pq_1mq_2)\}$$

Since all players are given an equal chance to obtain the roster spots, we assume that all of them share a similar probability of$\frac{1}{6}$. We can then pick out all the sets in which we have at least one mobile quarterback on the roaster, in this case, we have 5 sets. So, the P(mq) = 5/6 or 83.33% chance. This indicates that the likelihood of rostering a mobile quarterback in this instance would be highly likely.

The Buffalo Bills game is almost sold out and tickets are in high demand, because of this all tickets are equal in value. There are only two types of tickets left box seats, or front row stadium seating. 3 Bills fans buy these last tickets in secession, before buying the tickets, ticket master asks the user their preference. Display the set for all the possibilities of preferences that can be displayed and give the probability that no one prefers box seats.

$$S = \{(B,B,B), (B,B,S), (B,S,B), (B,S,S), (S,S,S), (S,B,B), (S,S,B), (S,B,S)\}$$

Above we have the preferences for the fans waiting in line, since there is no unfair probability, all sets have a 1/8 chance of occurring. So, the probability of no one preferring box seats would be P(S,S,S). Since all sets have equal probability, we can assume that the probability that no one prefers box seats would be 1/8 or 12.5% chance.

Section 2.6: Tools for Counting Sample Points

An NFL team is looking for new equipment for the upcoming 2022 NFL season. There are 5 different options for pads, 4 different options for helmets and 3 different options for their team jersey. If a team uniform consists of one set of pads, one helmet and one jersey, how many different uniforms are available?

N = (n1) (n2) (n3) = (5)(4)(3) = 60 possible ways

By using the multinomial rule and multiplying all our possible variables together, we get that there is a total of 60 possible combinations of uniforms. This may seem like an unimportant statics question, but it has real world applications. This problem might be applicable to equipment managers that are trying to find all the different unique sets of uniforms they can have their team were.

Section 2.7: Conditional probability and the Independence of Events

The table below shows the data gathered from the 2021 NFL season that represents scrimmage yards for wide receivers and running backs. This chart determines if the top 100 players in terms of yards per scrimmage for the NFL 202 scored above or below 1000 yards of scrimmage and if they are a running back or a wide receiver. Of the 100 top players 28 were wide receivers that scored over 1000 yards and 25 were wider receivers that scored under 1000 yards. 24 were running backs that scored over 1000 yards and 23 were running backs that scored under 1000 yards. If event A is over 1000 yards from Scrimmage, and events b and c are wide receivers and running backs respectively. Explain why events A and B are independent or dependent.

	Wide Receivers	Running Backs	Total
Over 1000 scrimmage yards	28	24	52
under 1000 scrimmage yards	25	23	48
Total	53	47	100

To prove these events to be independent we must prove that any of the following probability statements are true: P(a|b)=P(A), P (B|A) =P(B), or P (A∩ B) = P(A)P(B). Frist we can solve for P (A∩ B) =P(A)P(B), we first need to solve for the P(B) which we can do by just plugging in and solving which gives us $\frac{.28}{.52}$ = .54. We than can see that the P(A) is 0.52 based on the chart above we can now plug in our values P (A∩ B) =P (.52) P (.54), which gives us .28 meaning that our events are independent, because it satisfies one of the three truth statements.

Section 2.8: Two Laws of Probability

Making it to the NFL is a huge accomplishment, most experts say there is about a 1% chance that if you play pro in high school, you will make it to the NFL. They also say that there is about a 7% chance that if you play division one football in college you will make it to the NFL. Given the probability of event A and event B, assuming that they are independent events, find the P (A∪ B), P ($A^{\sim} \cap B^{\sim}$) P ($A^{\sim} \cup B^{\sim}$).

We can first solve for P (A ∪ B) which is equal to P(A)+P(B)- (P∩ B). We can then plug in our probability and get P (0.01) +P (0.07)-P (0.01* 0.07) = 0.01 so, the p (A ∪ B) =0.01. Next, we can solve for the P ($A^{\sim} \cap B^{\sim}$), by using de-Morgan's law we get $P(A \cup B)^{\sim}$, and since we just solved for $P(A \cup B)$ we get $(0.01)^{\sim}$ which equals 0.99 or 99%. We finally can solve for the P ($A^{\sim} \cup B^{\sim}$) Which equals P ($A^{\sim}$) +P ($B^{\sim}$)-P ($A^{\sim} \cup B^{\sim}$). We already know all the probabilities we need to solve this problem is simply just a plug and chug. $P(0.01)^{\sim}+P(0.07)^{\sim}-(0.01)^{\sim}$ = 0.99+0.93-0.99 = 0.93, this means that we have a 93% chance.

Section 2.9: Calculating the Probability of an Event: The Event-Composition Method

In the 2021 NFL season the Chiefs had a 12 and 5 record meaning that they had about a 70% chance of winning a football game on any given day. In an NFL season there are 17 regular season games, what is the probability that the chiefs will go undefeated in a 17-game season?

To solve this problem, we first must understand that the probability for each game is 70%. So, we can solve this problem by simply multiplying our 70% chance 17 times. For example, the P(a)=$(.70)^{17}$ which gives us 0.002 meaning the chiefs have about a 0.2% chance to go 17-0. This would be assuming that every week no matter how hard the opponent is the chiefs have about a 70% chance of winning.

Suppose that that the NFL is no longer using Microsoft surface tablets, and instead are going with Apple iPads. Although this turns out to be, because out of every shipment of apple iPads two are turning up broken. If the San Francisco 49'ers received an order of 10 iPads what is the probability that the IT person who is sorting them finds the last defective iPad on the 5th attempt. This assumes that he is only going through the iPads one at a time.

$\frac{\frac{5}{1}}{\frac{10}{2}} = \frac{5}{45}$= .11

To solve this problem, we can use the combination formula to get our probability of event a and event b. In this case the probability of event a is the probability that we find no defective iPads on 5 attempts and the probability of event b is the total outcomes of our events. Once we divide them, we get .11 which means there is an 11% chance that by the fifth iPad we find the defective ones.

Section 2.10: The Law of Total Probability and Bayes' Rule

The top 100 players in the NFL according to scrimmage yards are represented as 53% wide receivers and 47% running backs. The 2021 NFL season Stats indicate that 51% of running backs and 53% of wide receiver broke 1000 scrimmage yards. If a player from the list was selected at random what is the conditional probability that they would be a running back that had over 1000 scrimmage yards in the 2021 NFL season?

Since we are given all the probabilities, we need for bays theorem this problem is not very challenging. According to the textbook we can use the formula P(A|B) P(B)+P (A|$B^{\sim}$) P ($B^{\sim}$), to solve for total probabilities. Once we plug in our probabilities given to us into our formula, we get P (.53) P (.53) +P (.47) P (.51) =0.5206. This means we have about a 52% chance when picking from the top 100 NFL players to randomly select a Running back that had over one thousand scrimmage yards in the 2021 NFL season.

Section 3.2: The Probability Distribution for a Discrete Random Variable

In the NFL players who do not get drafted still have a chance to get signed by an NFL team by attending try outs. The New York Giants are hosting try outs for undrafted free agents. There are 6 defensive players and 9 offensive players trying out. The head coach decided that they only have 3 roaster spots remaining, and to be fair they are going to select who the sign at random. What is the probability that one offensive player will get singled? What is the probability that two offensive players will get signed? What is the probability that three offensive players will get signed?

$$\binom{n}{r} = \frac{n!}{(r!(n-r)!)}$$

$$P\ (3) = \frac{\binom{6}{0}\binom{9}{3}}{\frac{15}{3}} = \frac{(1)(84)}{(455)} = .18 \text{ or } 18\%$$

$$P\ (2) = \frac{\binom{6}{1}\binom{9}{2}}{\frac{15}{3}} = \frac{(6)(36)}{(455)} = .47 \text{ or } 47\%$$

$$P\ (1) = \frac{\binom{6}{2}\binom{9}{1}}{\frac{15}{3}} = \frac{(15)(9)}{455} = .29 \text{ or } 29\%$$

$$P\ (0) = \frac{\binom{6}{3}\binom{9}{0}}{\frac{15}{3}} = \frac{(20)(1)}{(455)} = 0.04 \text{ or } 4\%$$

When calculating this problem, we first must calculate all the combinations in which we could have. We do this by plugging in 15 and 3 into the combination formula of $\binom{n}{r} = \frac{n!}{(r!(n-r)!)}$, and we got 455. This is going to be our denominator for all our calculations. We then can solve for each individual probability case, in this problem there are 4 of them. Each probability we subtract one from the offensive probability and add it to the deserve probability. The above work is what we get when plugging in all our offensive probabilities.

Section 3.3: The Expected Value of a Random Variable or a Function of a Random Variable

The above answer gave us a probability distribution for Y, which is the probability that an offensive free agent gets signed. Find the mean, variance, and standard deviation of Y.

Mean $\mu = \sum_{y=0}^{3} P(y)$= (0) (0.04) +(1) (.29) +(2) (.47) +(3) (.18) = **1.77**

Variance $\sigma^2 = \sum_{y=0}^{3}(y-\mu)^2$p(y)

$=(0-1.77)^2(0.04) + (1-1.77)^2(0.29) + (2-1.77)^2(0.47) + (3-1.77)^2(0.18) =$ **0.622**

Standard deviation = $\sqrt{0.622}$ = **.79**

Section 3.4 The Binomial Probability Distribution

In the top 100 of NFL players according to scrimmage yards of the 2021 NFL season 24% were Running backs that rushed for over 1000 yards from scrimmage. Suppose that 10 players are selected from random, what is the probability that exactly 3 players are running backs that have over 1000 yards?

P(y) = $\binom{n}{y} P^y (1-p)^{n-y}$

$\binom{10}{3} .23^3 (1-.23)^{10-3} = (120).012(.77)^7$ = **0.23112**

By using the binomial distribution formula, we can plug in our n and y values to our formula and solve. Our n is 10, because that is our sample size, and our y is 3 because that is our variable, we are also given the probability of .24 in the problem. After plugging in the problem becomes a simple plug and chug and finally, we get an answer of .23 or 23%.

Section 3.5: The Geometric Probability Distribution

In the State of New Jersey people are normally an eagle's fan or a giant's fan. In my house 80% of my family prefers the Giants over the Eagles. If asked at random what is the probability that exactly 4 people are asked before we find one Eagles fan?

$P(y) = (1-p)^{y-1}P$

$P(0.80) = (1-.80)^{4-1}0.80$

P (0.80) = $P(0.80) = (.20)^3 0.80$

P (0.80) = 0.0064 or .64%

We can use are Geometric Distribution formula, which is $P(y) = (1-p)^{y-1}P$. We are given our main probability which is 80% and we are also given our y value, which is 4. We are looking for the probability that we will get a Giants fan 4 times in a row before we get one Eagle's fan. This problem is a simple plug and chug, which looks like the work I showed above.

Section 3.7: The Hypergeometric Probability Distribution

NFL teams relies heavily on Microsoft surface tablets throughout their seasons. During the 2021 NFL season Tom Brady Broke 5 surface tablets. As a team the Buccaneers are only given 30 tablets. The buccaneers I.T. staff collect all surface tablets at the end of the season. What are the odds that the first five tablets they collect will all be broken by Tom Brady.

P(y) = $\frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$

P(y)= $\frac{\binom{25}{5}\binom{5}{0}}{\binom{30}{5}} = \frac{(53130)(1)}{(142506)}$ = 0.37 = 37%

Section 3.8: The Poison Probability Distribution

According to the data the top twenty-five players have a mean of 89.7 receiving yards pers game. Use the Poisson distribution to prove the probability that the player has an average receiving yardage of fifty.

P(y) = $\frac{\lambda^y}{y!} e^{-\lambda}$

P (50) = $\frac{89.7^{50}}{50!} e^{-89.7}$ = 1.58610616 x 10−6

Section 3.11: TchebySheff's Theorem

During the NFL draft process, teams will make players take exams to verify their skills, intellect, and football intelligence. The exam is multiple choice and consists of 50 questions and each question has 4 answers. This test has been labeled as overwhelmingly difficult by many former players. If each problem is 2 points and you have zero football knowledge and guess on each question, what is the expected outcome of correct guesses? What is the variance? What is the Standard deviation?

$E(y) = np$ = 50 $\left(\frac{1}{4}\right)$

E(y) = 12.5

V(y) = np(1-p)

V(y) = 50(¼)(1-¼)

V(y) = 9.375

Standard deviation = $\sqrt{V(y)}$

Standard deviation = 3.06

The textbook gives us our formulas for expected variance and standard deviation in the textbook. To find the expected we just plug in the values given from our problem into the formula E(y) = np. To find the variance we use the formula np(1-p) and to find the Standard deviation we use the formula $\sqrt{V(y)}$.

Section 4.2: The Probability Distribution for a Continuous Random Variable

Referring to section 3.2, in the chart below is the probability in which offensive players are signed over defensive player. This is assuming that a team only has 3 positions to fill, and the positions are filled at random. If Y is a random variable give the distribution for F(y) and sketch the function.

y	0	1	2	3
P(y)	0.06	0.29	0.47	0.18

F (0) = P (Y≤0)

F (0) = 0.06

F (1) = P (Y≤1)

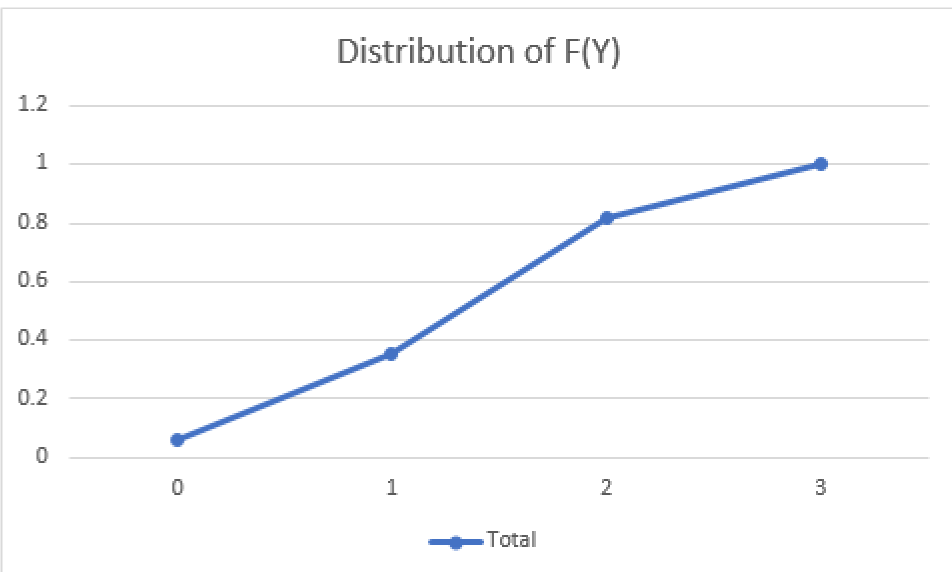
F (1) = 0.06+0.29 = 0.35

F (2) = P (Y≤2)

F (2) = 0.06+0.29+0.47 = 0.82

F (3) = P (Y≤4)

F (2) = 0.06+0.29+0.47 + 0.18= 1



This problem expands on one of our previous questions distribution results. The graph gives a visual representation of our data.

Section 4.3: Expected Values for Continuous Random Variables

In the above section we covered a density function for offensive players being signed over defensive players. Given the following, solve for the mean and variance.

$f(y) = \vdash_0^{(0.06)+(0.29)y}, 0 \leq y \leq 1$

$\int_0^1 0.06 + 0.29y = \int_0^1 0.06y + \frac{0.29y^2}{2}$

$$\int_0^1 0.06(1) + \frac{0.29(1)^2}{2} - 0.06(0) + \frac{0.29(0)^2}{2}$$

=0.0205

First, we need to solve for the interval which is $\int_0^1 0.06y + \frac{0.29y^2}{2}$. We then need to plug in our limits and solve what looks like this $\int_0^1 0.06(1) + \frac{0.29(1)^2}{2} - 0.06(0) + \frac{0.29(0)^2}{2}$. Once we solve it, we get 0.0205 which is equal to 2.05%.

Section 4.4: The Uniform Probability Distribution

Kickers in the NLF are for the most part fairly accurate, however on really windy days kick offs can be random. A kick will often aim for the end zone on kick offs, but due to increased wind the ball can land anywhere. If the football lands at a random spot between the end zone and the 20-yard line of the football field, find the probability that it lands closer to the end zone than the 20-yard line.

This question is more theoretical than practical work. If we have two balls and we are measuring the distance between given points, then in theory there is always going to be a halfway mark. This means that if we through a ball there is an equal probability that it lands closer to point a than b and vice versa. So, to answer the question above the probability that the ball lands closer to the endzone then the 20-yard line is 50%.

Section 5.2: Bivariate and Multivariate Probability Distributions

The NFL has a system for distributing players that are cut from one team, this system is called waivers. If a player is cut a team can submit a waiver claim for said player to enter a raffle to be rewarded that player. If two players are cut and the Bill, Jets, and Patriots all submit wavier claims for both players, find the join probability for the Jets and patriots. Each team can either be awarded one player, both players, or no players.

Awarded player(s)	Jets 0	Jets 1	Jets 2
Patriots 0	1/9	2/9	1/9
Patriots 1	2/9	2/9	0
Patriots 2	1/9	0	0

We can create this table of probability by looking at each individual case and assigning a proper probability to each one. We can check our answer by adding each case up and hopefully getting 1. If we get one, then we can confirm our probabilty is correct. After adding we get 9/9 which means our probability chart is correct.

Section 5.3: Marginal and Conditional Probability Distribution

The above question left us with the join distribution displayed in the chart below. Find the marginal probability distribution of the Jets.

P (0) = 1/9+2/9+1/9 = 4/9

P (1) = 2/9+2/9+0/9 = 4/9

P (2) = 1/9+0/9+0/9 = 1/9

Since we already know our probabilty values for the Jets to find the marginal probabilty we can add each column up to determine our new answer. The first column is for the probability that one player is signed to the jets, so we can add 1/9 2/9 and 1/9 to find our new probabilty. We can then do this for the next two columns and find the probabilities for all test cases.