Piece 1

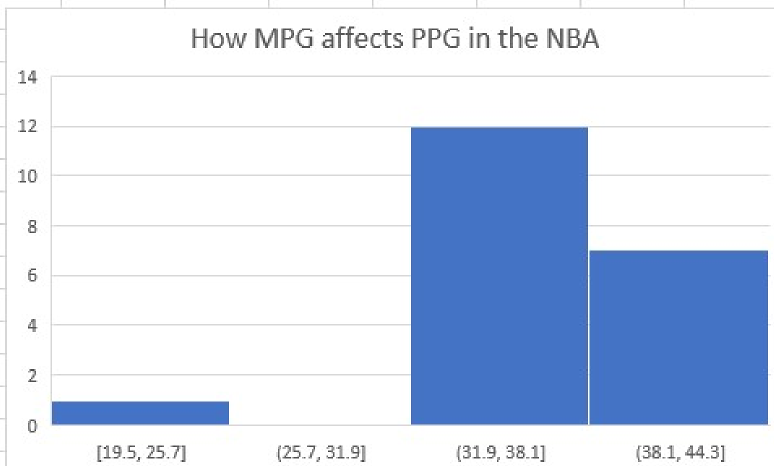
Dataset Problems from Top 20 NBA Players

In this piece of my final project, I will be taking problems from each section of the textbook and changing a question from each section to correlate with my personal dataset. My dataset contains statistics for my favorite twenty NBA players from the 2021-2022 season. The textbook that I am using the question from is Mathematical Statistics with Applications 7th Edition. Within the textbook, I will be taking examples from chapters one through five. The dataset was found online from ESPN's public database. This database is fact checked constantly by ESPN employees as well as NBA employees to ensure that these statistics are accurate.

<u>Section 1.2 Characterizing a Set of Measurements: Graphical Methods:</u>

For this question, I decided that I would create a relative frequency histogram that would highlight the points per game of the player with how many minutes they play. I believe that this graph would be useful because there are many NBA players that play a variety of different minutes however, I believe that the more minutes you play, the more points you would score. Therefore, this relative frequency histogram would show that with more minutes, you score more buckets!

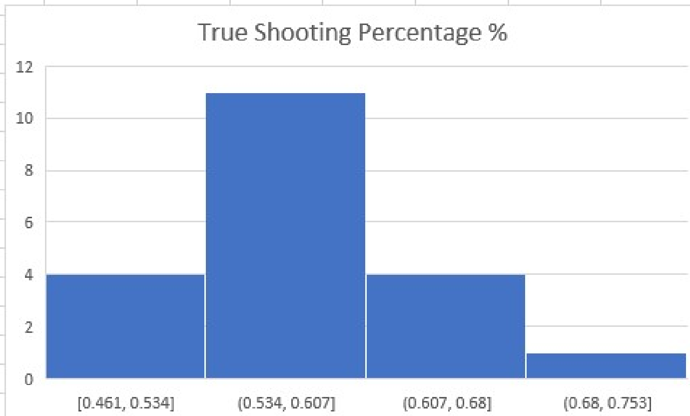
MPG	PPGPointsPoints per game.
37.3	31.7
44	26.3
34.7	27.4
34.2	31
38.5	23.6
41	25.6
36.8	31.7
37.5	27.1
37	27.4
39.9	18.6
37	21.8
37.3	15.4
38.3	25.5
37.8	25.2
34.1	14.8
19.5	8
40.7	20.8
32.8	12
39.3	27
37.7	10.7



<u>Section 1.3 Characterizing a Set of Measurements: Numerical Methods:</u>

When we take a look at the stats of an NBA player, we tend to overlook the true shooting percentage of a player. True shooting percentage is the percent value of all the players attempted shots from the free throw line, any two-point attempts, and three-point attempts. If you take all their attempts and divide it by the number of attempts made, you receive the true shooting percentage. This is something that a lot of NBA players care about because the better the percentage, the better the paycheck. Therefore, to make this histogram, I took the true shooting percentage, highlighted it, and inserting a histogram. However, to complete this section, we also must solve for the standard deviation. In order to solve for the standard deviation, we have to insert our data into the following formula $\sqrt{(\sum (x-x1)^2/ n-1)}$. After solving for the standard deviation with our data, the answer is 0.055331184697239 which means that there is not a lot of variances between these players true shooting percentage. Since these players are some of the best players in the league, it does not surprise me that they are close to each other's true shooting percentage.

TS%True Shooting PercentageTrue shooting percentage is a measure of shooting efficiency that takes into account field goals, 3-point field goals, and free throws.
0.552
0.526
0.606
0.643
0.59
0.562
0.577
0.549
0.604
0.584
0.642
0.461
0.505
0.604
0.638
0.62
0.481
0.688
0.584
0.596



Section 2.3: A Review of Set Notation

In street basketball (3 versus 3) there are typically three different positions that can play on the court at the same time. You can have a point guard (otherwise known as the ball handler PG), a forward (typically the bucket getter F), and a big (larger player otherwise known as a center B). However, the game of street basketball is not locked to have one point guard, one forward, or one big. You can have three bigs, three forwards, or any combination of positions in the three-on-three game. Define the super set and the subsets.

A = if there is one point guard in the game.

B = if one big is in the game.

Find the Union and the Intersection of A and B:

S = {(PG, F, B), (PG, F, F), (PG, B, B), (PG, PG, PG), (PG, PG, F), (PG, PG, B), (F, B, B), (F, F, B), (F, F, F), (B, B, B)

A = {(PG, F, F), (PG, B, B), (PG, B, F)}

B = {(B, F, F), (Pg, Pg, B), (B, F, PG)}

A ∪ B = {(F, F, PG), (PG, B, B), B, F, PG), (B, F, F), (PG, PG, B)}

A ∩ B = {(B, F, PG)

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

On a team in the NBA, you need to always have five players on the court. Generally, there are specifically five general positions that most teams follow. Each basketball team's starting five positions tend to be point guard, shooting guard, small forward, power forward, and center. However, these positions are either offensive (denoted by a +) or defensive (denoted by a *). List the sample space for the starting line up of an NBA game.

S = {(PG+), (SG+), (SF+), (PF+), (C+), (PG*), (SG*), (SF*), (PF*), (C*)}

What this sample space is telling us is that although most of the time, the power forward and the center are the defensive players, any of the positions on the court can be an offensive or defensive player. An example of a defensive point guard would be Jose Alvarado because although he doesn't score the most points, he is able to guard players very well. However, an example of an offensive center would be Joel Embiid or Giannis.

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

In the NBA, the starting line up for the point guard and shooting guard positions tend to be very valuable and earn a lot of praise. Most NBA teams tend to have up to five players with a point/shooting guard position. Since there are only two starting spots, what is the probability that the point guard is a defensive point guard?

The guards that we are looking at are:

1. Stephen Curry (+)
2. Jose Alvarado (*)
3. Trae Young (+)
4. Luka Doncic (+)
5. Jayson Tatum (*)

S = {(SC, JA), (SC, TY), (SC, LD), (SC, JT), (JA, TY), (JA, LD), (JA, JT), (TY, LD), (TY, JT), (LD, JT)

Because we know that all the five players have an equal chance of being selected, we know that the probability of them getting selected is 1/5. We have six different sets that have a mainly defense guard. Therefore, there is a 6/10 or 60% chance of a defensive guard.

Section 2.6 Tools for Counting Sample Points

For a home game, the New York Knicks are going to get picked up by the team shuttle before the game so that they can practice. The last eight players need to get picked up locally from around Madison Square Garden. Four of the players are in the Bronx. Another two players are in the heart of New York City, and the rest are in Manhattan. How many different ways does the team shuttle have to go to pick up all of the players.

$$\frac{8!}{(4!\,2!\,2!)} = \frac{40320}{96} = 420$$

Section 2.8 Two Laws of Probability

A study on Kevin Durant's 2021-2022 statistics determined that he had a free throw percentage of 89.5% and a two-point percentage of 40.3%. Since free throws and two-pointers are independent of each other, find $P(A \cup B)$, $P(A' \cap B')$, $P(A' \cup B')$.

$P(A \cup B) = .895 + .403 – (.895 * .403) = 1.298 - 0.360685 = 0.937315$ or 93.7%

$P(A' \cap B') = P(A \cup B)' = 0.062685$

$P(A' \cup B') = .105 + .597 - .062685 = 0.639315$

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

In the NBA, 84% of the jerseys are made by the organization themselves and 16% are outsourced by another company. 1.1% of the NBA made jerseys are defective and 8.9% of the third party jerseys are defective. If a jersey is selected at random, what is that chance that the jersey will not be defective.

A = jerseys created by the NBA

B = jerseys created by a third party

X = not defective

$P(X) = P(X \cap (A \cup B)) = P(X \cap A) + P(X \cap B) = .89(.84) + .911(.84) = 1.51284$

Section 2.10 The Law of Total Probability and Bayes' Rule

From the majority of basketball fans, 52% of women like to watch mainstream teams such as the Los Angeles Laker or the Miami Heat. While 39% of men like watching mainstream teams like the Boston Celtics. Out of a group of 10 people, 7 females and 3 males, the response was negative towards mainstream teams. What is the probability that it was a female?

P(Female) = 5.2/10 = .52

P(Male) = 3/10 = .3

P(Negative | Female) = 1 - .72 = 0.28

P(Negative | Male) = 1 – 0.39 = 0.61

P(Female | Negative) = .48 * .3 / .28 * .7 + .48 * .3 = 4.235%

Section 3.2 The Probability Distribution for a Discrete Random Variable

A fan of the sport wants to select their favorite three players. There are four forwards and six centers. If the fan selects players at random, what are the odds that the fan selects no centers, one center, two centers, or three centers?

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

$$P(3) = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} = \frac{20}{120} = 0.1\overline{6}$$

$$P(2) = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = \frac{60}{120} = 0.50$$

$$P(1) = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{36}{120} = 0.30$$

$$P(0) = \frac{\binom{4}{3}\binom{6}{0}}{\binom{10}{3}} = \frac{4}{120} = 0.0\overline{33}$$

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

Using the previous answers from above, what is the mean, variance, and the standard deviation of Y?

Mean $= 0(.033) + 1(.3) + 2(.5) + 3(.16) = \boxed{1.78}$

Variance $= (0-1.78)^2(0.033) + (1-1.78)^2(.30) + (2-1.78)^2(.5) + (3-1.78)^2(.16)$

$\Rightarrow -0.861 + 0.22 + 1.065 \boxed{= 0.424}$

Standard Deviation: $\sqrt{\text{Variance}} = \sqrt{0.424} = \boxed{0.6511}$

Section 3.4 The Binomial Probability Distribution

Stephen Currey is the greatest three point scorer of all time. He has a 39.7% chance to make every three-point shot he takes. Identify the event favor that Steph's shot as a success S. The probability of S on trial 1 is .397. Consider the event B that S occurs on the second trial. B can occur in two ways, either he makes both shots or the first shot is a make and the second shot is a miss. Show that P(B) = .397. What is P(B|trial 1 in S)? Does this conditional probability differ markedly from P(B)?

The probability of first trial being a failure is 1 - .397 = 0.603

The probability of the second trial becoming a success is .397.

The probability of back-to-back successes = 0.157609

The probability of a failure then back-to-back is = .603 * .397 = 0.239391

Total probability of two trials becoming both successes or the first trial is a failure and the second is a success = 0.157609 + 0.239391 = 0.397

P(B) does equal .397 because the conditional doesn't differ at all from P(B).

Section 3.5 Geometric Probability Distribution

Jose Alvarado has a true shooting percentage of 62%. If we took four NBA fans and told them the true percentage of Jose, what is the probability that the first person thinks Jose would make the three and the fourth player also agrees that he would make his fourth three.

P(X=4) = .62 * (1-.62)^4-1

= .62 * 0.054872 = **0.034**

Section 3.7 Hypergeometric Probability Distribution

Five NBA fans were selected by the NBA to select their favorite NBA player from a group of 15 total fans. Of the total fans, 10 of them selected Stephen Curry to be the best player in the NBA and 5 fans selected Giannis Antetokounmpo. The lot was randomly selected and only one Giannis fan was selected? Is there any doubt on the randomness of this selection?

$$P(Y\leq 1) = \frac{\binom{10}{5}\binom{5}{0}}{\binom{15}{5}} + \frac{\binom{10}{4}\binom{5}{1}}{\binom{15}{5}} = \frac{252}{3003} + \frac{210\cdot 5}{3003} = \frac{252+1050}{3003} = \boxed{0.4335}$$

Section 3.8 Poisson Probability Distribution

Luka Doncic scores 31.7 points-per game on average. Use the Poisson Distribution to prove the probability that the player has at least 25 points per game.

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

$$P(25) = \frac{31.7^{25}}{25!}e^{-31.7} = \boxed{0.037}$$

Section 3.11 Tchebysheff's Theorem

14.8% of NBA players wear Kobe Bryant's Kobe shoes. If the same proportion of 90 G-League players are interviewed on what shoes they preferred,

a. What is the expected number of Americans who prefer Kobes?

P = .148, n = 90

E(x) = np = 90(.148) = 13.32

b. What is the standard deviation of the number Y who would prefer Kobes?

$\sigma = \sqrt{V(x)} = \sqrt{npq} = \sqrt{90(.148)(.852)} = 3.3687$

c. Is it likely that the number of Americans who preferred Kobes exceeds 13 people?

It is likely because 13.32 + 3.3687 = 16.6887

Section 4.2 Probability Distribution of a Continuous Random Variable

In the past 5 years, the NBA has sold over 110 (in thousands) jerseys in the United States. The total amount of jerseys sold in a 5 year span is a random variable Y with a probability density function given by:

$$\begin{cases} y, & 0 \le y \le 1 \\ 5 - y & 1 \le y \le 2 \\ 0, & elsewhere \end{cases}$$

a. Find F(y): For 0 < y < 1 $F(y) = \int_0^y t dt = \frac{y^2}{2}$
   For 1 ≤ y < 2, $F(y) = \int_0^1 t dt + \int_1^y (5 - t) dt = 5y - \frac{y^2}{2} - 1.$
b. Find the probability that the sales will be between 100 and 120 copies in a 5 year span.

$$P(.92 \le 1.1) = F(1.1) - F(.92) = 4.445$$

Section 4.4 Uniform Probability Distribution

Giannis Antetokounmpo dribble twice to get to the end of full court. If he dribbles the ball one time, which half of the court would he be on? (A is the half with the opponents basket, B is his basket that he is defending). Find the probability that he is closer to the opponent's basket than his defending basket.

If Giannis is in the opponent's half of the court, he is in the interval (A, $\frac{(A+B)}{2}$). This would be half of the total interval length; therefore, the probability of this happening is 0.50 or 50%.

Section 5.2 Bivariate and Mulitvariate Probability Distribution (was unable to use database, wrote example from textbook and solved it for Chapter 5)

Suppose that a radioactive particle is randomly located in a square with sides of unit length. That is, if two regions within the unit square and of equal area are considered, the particle is equally likely to be in either. Let X and Y denote the coordinates of the particle's location. A reasonable model for the relative frequency histogram for X and Y is the bivariate analogue of the univariate uniform density function:

$$f(x, y) = \begin{cases} 1, & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & elsewhere \end{cases}$$

a. Find(.3, .4)

$$F(.3, .4) = \int_{\infty}^{.4}\int_{\infty}^{.3} f(x,y)\,dx\,dy = \int_0^{.4}\int_0^{.3} (1)\,dx\,dy$$

$$\Rightarrow \int_0^{.4} \left(x \Big]_0^{.3} \right) dy = \int_0^{.4} .3\,dy = .3y \Big|_0^{.4} \boxed{= .12}$$

Section 5.3 Marginal and Conditional Probability Distribution

From a group of three Republicans, two Democrats, and one independent, a committee of two people is to be randomly selected. Let X denote the number of Republicans and Y denote the number of Democrats on the committee. Find the joint probability function of X and Y and then find the marginal probability function of X.

$$P(x=1, y=1) = p(1,1) = \frac{\binom{3}{1}\binom{2}{1}\binom{1}{0}}{\binom{6}{2}} = \frac{3(2)}{15} \boxed{= \frac{6}{15}}$$