Cole Martin

Dr. Jacob

MATH-1223-01

August 17, 2021

**NFL Quarterback Passing Stats**

The data set I chose is the NFL quarterback passing stats from 2020. The data set includes all of the regular season games played, and the data can be found at <https://www.nfl.com/stats/player-stats/>. I only picked the top 25 quarterbacks to be a part of this project because, if I kept going down the list then I would start getting into some of the backup quarterbacks that barely got to see any playing time during the 2020 season. There are so many talented quarterbacks in the NFL and it will be very interesting to see who has the best stats. The categorical variables I will be using are going to be the team the quarterback plays for and the conference they are in. The quantitative variables I will be using will be passing yards (Pass Yds) and passing touchdowns (TD). I am interested in this data because I love fantasy football, and it is very important to have a good quarterback that can put up really good stats. I hope to find out if passing yards (Pass Yds) has an impact on the amount of passing touchdowns (TD).

**Part 2**

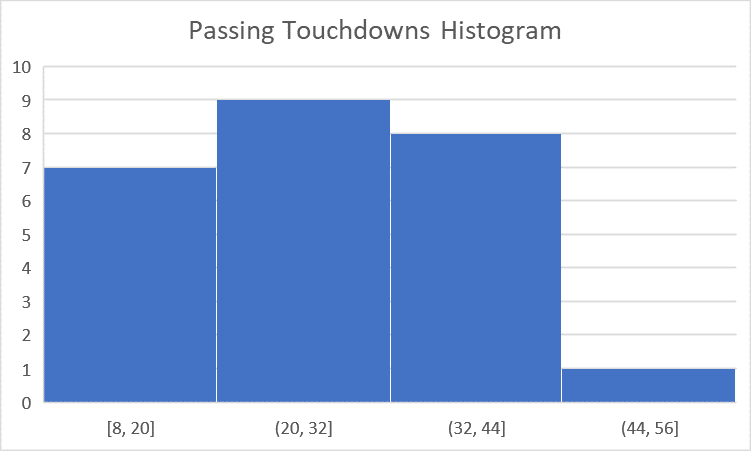
For the second part of the project, I am looking at the frequency and relative frequency of the player’s conference divisions. The table below shows the data I collected. There are a total of 8 different conference divisions in the data I collected. The 8 different conference divisions are listed below. The AFC West, AFC North, AFC South, and NFC South all had the highest percentage of quarterbacks at 16%.

Next, I created a two-way table to compare a player’s conference to their age group.

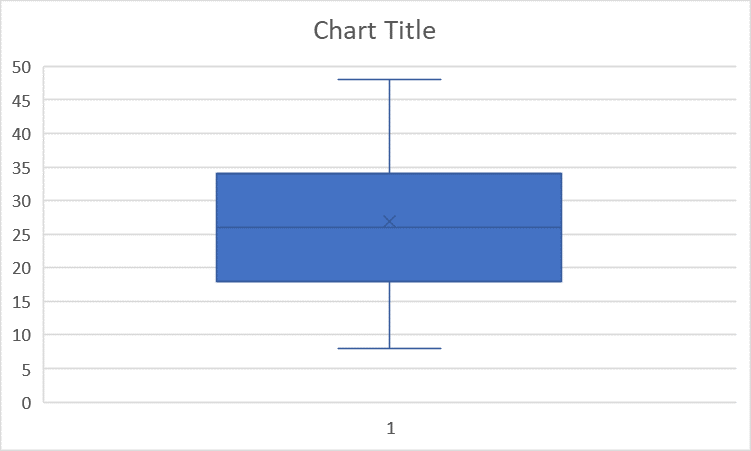
This table showed me that the majority of my chosen players were between the ages of 21-25. You can also see a relationship between the two variables and see that the AFC is really heavy on starting young quarterbacks, and the NFC not so much.

**Part 3**

For the third part of the project, I am looking at passing touchdowns (TD). I will be looking at 25 different quarterbacks that have thrown passing touchdowns in this data set. I will find the mean, standard deviation, five-number summary and display a histogram and box plot that will be listed below. There are no outliers in this data set. You can see that there are none by looking at Table 3.2. The distribution of the data is skewed to the right, because the mean is larger than the median. You can see this by looking at Table 3.3. The five number summary is also listed in Table 3.3 as well.

 **Table 3.1**

 **Table 3.3**

**Table 3.2**

**Part 4**

For part four of this project, I am creating a hypothesis test for a quantitative and categorical variable. I will be examining completion percentage (CMP%) and the conference of the top 25 quarterbacks in the NFL during the 2020 season.

**Quantitative Hypothesis:**

My hypothesis is that the average completion percentage (CMP%) for the top 25 quarterbacks in the NFL is 0.60. My alternative hypothesis is that it does not equal 0.60. Completion percentage is calculated by taking the completed pass attempts and dividing them by the passing attempts.

H0: 𝜇 = 0.60

Ha: 𝜇 ≠ 0.60

**Categorical Hypothesis:**

My hypothesis is that out of the top 25 NFL quarterbacks for the 2020 season, 40% are in the National Football Conference (NFC). My alternative hypothesis is that it is greater than or equal to 40%. I picked the AFC because there are a lot more veteran quarterbacks in that conference than there are in the American Football Conference (AFC). This is why I think that 40% of the top 25 quarterbacks for 2020 are in the NFC.

H0: p = 0.4

Ha: p>0.4

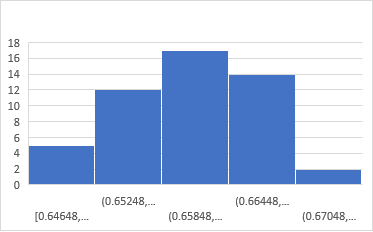
**Part 5**

For the fifth part of this project, I am computing the standard error for my quantitative and categorical hypothesis using bootstrapping. I will then reject or fail my null hypothesis depending on what the results are.

**Quantitative Variable:**

The standard error for my quantitative variable is 0.000629. The 95% confidence interval had a lower bound of 0.64831 and an upper bound of 0.67346. After reviewing the data, I reject the null hypothesis because my initial hypothesis of 0.60 did not fall between the two bounds.

**Quantitative Histogram**

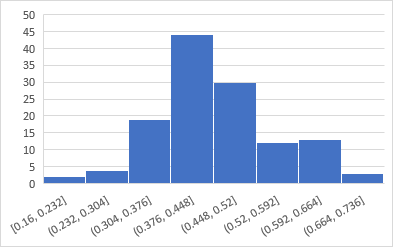




**Categorical Hypothesis:**

The standard error for my categorical variable is 0.10409. The 95% confidence interval had a lower bound of 0.24915 and an upper bound of s. After reviewing the data, I fail to reject the null hypothesis. I do not have enough evidence to suggest the alternative. My initial hypothesis of 0.4 falls between the two bounds.

**Categorical Histogram**



**Part 6**

For this part of the assignment, I am repeating the hypothesis test on the categorical variable utilizing the appropriate formulas. I will also compare my 95% confidence interval from my bootstrapping in part 5 to this one in part 6 using formulas.

|  |  |  |
| --- | --- | --- |
| Statistic | Formula | Result |
| N | Sample Size | 25 |
| X | Number of times (NFC) appears | 11 |
| Proportion | (in hypothesis) | 0.40 |
| P Value | =NORM.S.DIST(Z,TRUE) | 0.792891911 |
| P Hat | =X/N | 0.32 |
| Standard Error | =SQRT(p\*(1-p)/n | 0.09798 |
| Z Statistic | =(P-hat-p)/SE | -0.8165 |
| Z\* | =NORM.S.INV(0.975) | 1.959964 |
| CI low | =p-hat-(2\*SE) | 0.124041 |
| CI high | =p-hat+(2\*SE) | 0.515959 |

The 95% confidence interval had a lower bound of 0.24915 and an upper bound of 0.6655. Using bootstrapping and using the formulas the lower bound was 0.124041 and the upper bound was 0.515959. I fail to reject the null hypothesis because 30% falls within the 95% confidence interval.

**Part 7**

For this part of the assignment, I am repeating the hypothesis test on the quantitative variable utilizing the appropriate formulas. I will also compare my 95% confidence interval from my bootstrapping in part 5 to this one in part 7 using formulas.

|  |  |  |
| --- | --- | --- |
| Statistic | Formula | Result |
| MU | (in hypothesis) | 0.60 |
| X-bar | (sum of the numbers/n) | 0.66 |
| Sigma | (standard deviation) | 0.03408 |
| N | Sample size | 25 |
| Standard Error | Sigma/SQRT(n) | 0.00682 |
| T Statistic | X-bar-mu/SE | 8.797654 |
| Alpha | For 95% CI | 0.05 |
| T\* | T.INV (alpha, n-1) | 1.710882 |
| CI Low | X-bar-T-Critical Value\*SE | -0.05584 |
| CI High | X-bar+T-Critical Value\*SE | 0.06416 |

The 95% confidence interval had a lower bound of 0.64831 and an upper bound of 0.67346. Using bootstrapping and using the formulas the lower bound was -0.05584 and the upper bound was 0.06416. I reject the null hypothesis because my initial hypothesis of 0.60 did not fall between the two bounds.

**Part 8**

For the eighth part of this project, I am going to use the two-way table from part 2 to create 2 conditional probabilities. I will interpret their meaning and explain how they were computed.

**P(A|B) = P(A ∩ B)/P(B)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Two Way Table** | Ages 21-25 | Ages 26-30 | Ages 31-35 | Ages 36-40 | Ages 41-45 | **Total** |
| AFC | 7 | 3 | 2 | 2 | 0 | 14 |
| NFC | 2 | 2 | 3 | 2 | 2 | 11 |
| **Total** | 9 | 5 | 5 | 4 | 2 | 25 |

I am going to be determining the probability of AFC players given they are between the ages 36-40.

**P=2/4≈50%**

I am going to be determining the probability of NFC players given they are between the ages of 26-30.

**P=2/5≈40%**

My first conditional probability looks at the probability of AFC players that are between the ages of 36-40 from the data set I used throughout the project. The percentage was approximately 50% meaning 2 out of 4 AFC players were between the ages of 36-40. This statistic shows the advantage that AFC players between the ages of 36-40 are at being in the top 25 of NFL passing stats for 2020.

My second conditional probability looks at the probability of NFC players that are between the ages of 26-30 from the data set I used throughout the project. The percentage was approximately 40% meaning 2 out of 5 NFC players were between the ages of 26-30. This statistic shows the disadvantage that NFC players between the ages of 26-30 are at being in the top 25 of NFL passing stats for 2020.