Madison Ward

Intro to Probability and Stats

Dr. Nicholas Jacob

January 16, 2023

**First Hit Songs**

I looked at some top artists and their first hit songs. I wanted to find out if there was a correlation between the time it takes to release a breakthrough song and when they started making music. I found all of my information on Wikipedia by independently searching the artists’ names. My quantitative variables are the artists’ ages, starting year, and the time they took to release their first hit song. The categorical variables are their music genres and first hit songs (Claim to Fame). I chose this topic because I really enjoy listening to music and wanted to explore how these artists came to be known.

| **Name** | **Age** | **Starting Year** | **Claim to Fame** | **Genres** | **Time Taken To Release First Hit Song** |
| --- | --- | --- | --- | --- | --- |
| Taylor Swift | 33 | 2004 | “Tim McGraw” (2006) | Pop | 2 years |
| Billie Eilish | 21 | 2015 | “Ocean Eyes” (2015) | Pop | <1 year |
| Dua Lipa | 27 | 2014 | “New Love” (2015) | Pop | 1 year |
| The Weeknd | 32 | 2009 | “Kiss Land” (2013) | R&B | 4 years |
| Justin Beiber | 28 | 2007 | “One Time” (2009) | Pop | 2 years |
| Cardi B | 30 | 2015 | “Bodak Yellow” (2018) | Hip Hop | 3 years |
| Selena Gomez | 30 | 2002 | “Come and Get It” (2013) | Pop | 11 years |
| Ed Sheeran | 31 | 2004 | “The A Team” (2011) | Pop | 7 years |
| Harry Styles | 28 | 2010 | “Sign of the Times” (2017) | Pop | 7 years |
| Shawn Mendes | 24 | 2013 | “Life of the Party” (2014) | Pop | 1 year |
| Ariana Grande | 29 | 2008 | “The Way” (2013) | Pop | 5 years |
| Rhianna | 34 | 2003 | “Pon de Replay” (2005) | Pop | 2 years |
| Charlie Puth | 31 | 2009 | “Marvin Gaye” (2015) | Pop | 6 years |
| Olivia Rodrigo | 19 | 2015 | “Drivers License” (2020) | Pop | 5 years |
| Katy Perry | 38 | 2001 | “I Kissed a Girl” (2008) | Pop | 7 years |
| Doja Cat | 27 | 2012 | “So High” (2014) | Hip Hop | 2 years |
| Post Malone | 27 | 2011 | “White Iverson” (2015) | Hip Hop | 4 years |
| Lizzo | 34 | 2010 | “Truth Hurts” (2019) | Hip Hop | 9 years |
| Lil Nas X | 23 | 2018 | “Old Town Road” (2019) | Pop Rap | 1 year |
| Adele | 34 | 2006 | “Hometown Glory” (2007) | Pop | 1 year |
| Khalid | 24 | 2016 | “Location” (2017) | R&B | 1 year |
| Jack Harlow | 24 | 2011 | “Whats Poppin” (2020) | Hip Hop | 9 years |
| Conan Gray | 24 | 2013 | “Idle Town” (2017) | Pop | 4 years |
| Drake | 36 | 2001 | “Over” (2010) | Hip Hop | 9 years |
| Lady Gaga | 36 | 2001 | “Just Dance” (2008) | Pop | 7 years |

**Table 1.1**

| **Years to Release Hit Song** | **Frequency** | **Relative Frequency** |
| --- | --- | --- |
| <1 Year | 1 | 0.04 |
| 1-3 Years | 10 | 0.4 |
| 4-6 Years | 6 | 0.24 |
| 7-9 Years | 7 | 0.28 |
| 10+ Years | 1 | 0.04 |
| **Total** | **25** | **1** |

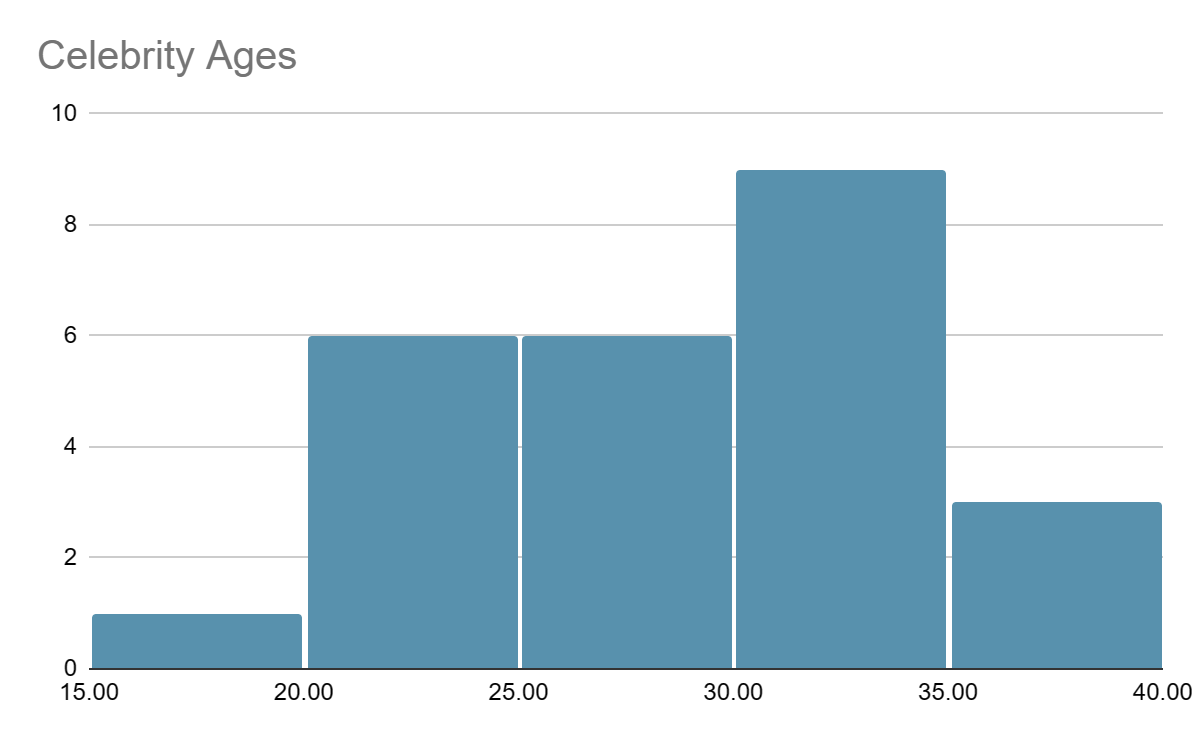
**Table 2.1**

Table 2.1 shows the frequency of the range of years these artists released their first hit songs. Most hit songs fell between 1-3 years, so I’d say that is a common time to become an up-and-running star.

|  | **Male** | **Female** | **Total** |
| --- | --- | --- | --- |
| ≤ 5 Years | 7 | 9 | 16 |
| ≥ 6 Years | 5 | 4 | 9 |
| Total | 12 | 13 | 25 |

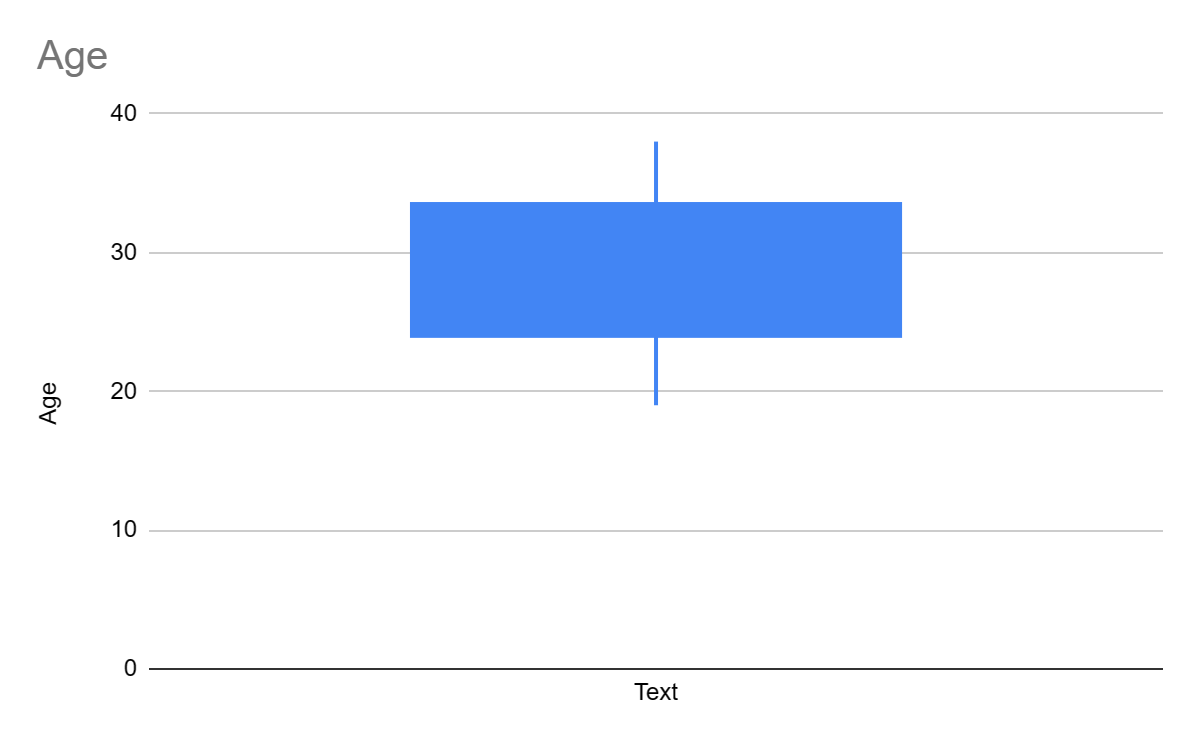
**Table 2.2**

Table 2.2 shows first hit release years sorted by male or female. I chose this because I wondered if there was a relation between these celebrities’ gender and how long their hit songs took. They came out reasonably close! This leads me to believe that an artist’s gender does not correlate with their breakthrough year.



**Table 3.1**

Table 3.1 shows a histogram graph of how many of the celebrities I listed fell into these age groups. This graph appears to be skewed left with no outliers.



**Table 3.2**

Table 3.2 shows the same data from Table 3.1 in a box plot. There are no outliers. The five-number summary is as follows:

**Min= 19**

**Max= 38**

**Q1= 24**

**Median= 29**

**Q3= 33.5**

Project Part 4

For my quantitative variable, I would like to look at the ages of the singers I listed in Table 1.1. I would like to determine the average age of the current celebrities on my list. This continues on the query that someone’s age has to do with their popularity. Do most singers make their break before the age of 30? The average age I found was 28.96 years old.

**H0: μ=30**

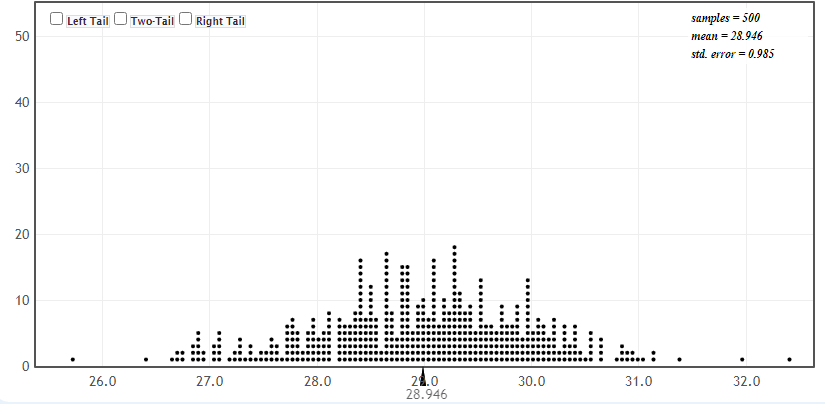
**Ha: μ<30**

For my categorical value, I chose to find the highest occurring genre of music. I think the genre of music has a significant factor in how popular an artist and their songs are. If most people are listening to pop music, singers are more likely to shoot for being pop stars. For this, I think the highest-occupied genre is going to be pop and that more than 50% of those on my list make pop music. I found that the frequency of the genre Pop is 64%.

**H0: μ=50%**

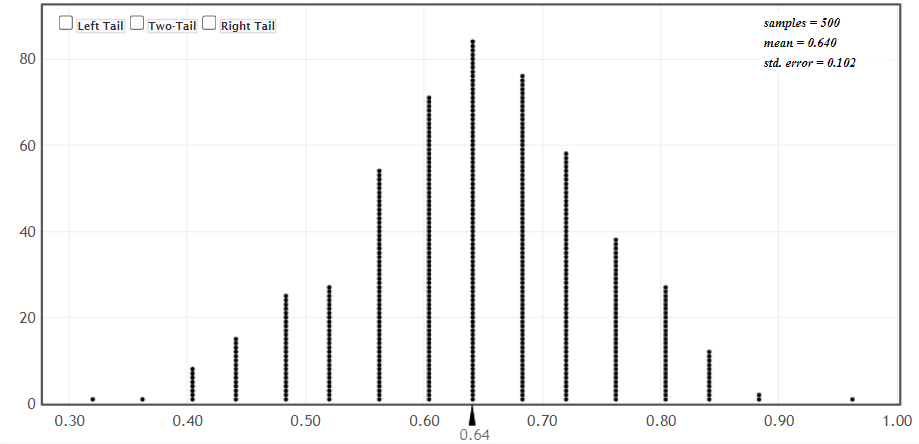
**Ha: μ>50%**

**Project Part 5: Bootstrapping**



**Table 5.1**

I performed a bootstrap for my quantitative variable of the average age of celebrities when they made their break. This histogram table shows an average from a random sampling of my data from part 4.



**Table 5.2**

This table (5.2) shows my categorical value of the most reoccurring genre of music from Table 1.1. There were 4 genres: Pop, Hip Hop, R&B, and Pop Rap. The highest occurring genre was pop.

**Part 6: Categorical Inference with Formulas**

For this part I will be using my categorical hypothesis to find my categorical inference:

**H0: μ=50%**

**Ha: μ>50%**

I failed to reject my null hypothesis (50%), as I found through testing that 64% of the top songs were in the genre Pop. I will start by finding my standard error(SE), then my z value (z), followed by the 95% confidence interval. These are my starting values:

p=0.5 p-hat= 0.64 n=25

SE= SE= = 0.1

Z value= Z= = 1.4

CI(95%)= p-hat (z\*)(SE) CI= 0.64+ (2)(0.1) = 0.84

CI= 0.64 - (2)(0.1) = 0.44

(z\*=2)

Based on this data, I have failed to reject my null hypothesis.

**Project Part 7: Quantitative Inference With Formulas**

For this part I will be using my quantitative hypothesis from part four of the project:

**H0: μ=30**

**Ha: μ<30**

I rejected my null hypothesis as I found through testing that the average age of my list of celebrities is less than 30 (28.96 years old). To get my critical T score I used a 1-tailed test with a 0.05 significance level with 24 degrees of freedom and reached the value of 1.7114. Now I need to calculate the T score to compare this data.

{Sample Mean = x̄ Assumed Mean = μ Sample Size = n Standard Deviation = s}

μ = 30 x̄ = 28.96 n = 25 s=5.2 24 Degrees of Freedom

s = = 5.2

T value= t= = -1

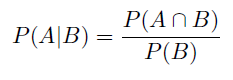
CI(95%) = 28.96 ±1.925 (±6.65%)

Since my T value is less than the critical T value, I have failed to reject my null hypothesis.

**Project Part 8: Conditional Probability**

For this part of the project I will be calculating conditional probability from my previous two-way table:

|  | **Male** | **Female** | **Total** |
| --- | --- | --- | --- |
| ≤ 5 Years | 7 | 9 | 16 |
| ≥ 6 Years | 5 | 4 | 9 |
| Total | 12 | 13 | 25 |

This is the formula used to calculate conditional probability.

1. My first test will be the probability of achieving success within the first 5 years of a celebrity’s career compared to 6 or more years after.

**P(A)= 0.64 P(B)= 0.36 P(A ⋂ B)= 0.2304**

P(A)= Total in ≤ 5 Years / Sample Total P(A)= 16 / 25 = 0.64

P(B)= Total in ≥ 6 Years / Sample Total P(B)= 9 / 25 = 0.36

To find P(A ⋂ B) you use the formula: P(A ⋂ B) = P(A) P(B)

P(A) \* P(B) (0.64)(0.36)= 0.2304

P(A|B) = P(A ⋂ B) / P(B) 0.2304 / 0.36 = 0.64

**P(A|B) = 0.64**

1. My second test will be the probability of males to females achieving popularity.

**P(A)= 0.48 P(B) = 0.52 P(A ⋂ B) = 0.2496**

P(A)= Total Number of Males / Sample Total P(A)= 12 / 25 = 0.48

P(B)= Total Number of Females / Sample Total P(B)= 13 / 25 = 0.52

P(A ⋂ B) = P(A) P(B) P(A ⋂ B)= (0.48) \* (0.52) = 0.2496

P(A|B) = P(A ⋂ B) / P(B) 0.2496 / 0.52 = 0.48

**P(A|B) = 0.48**

|  |  |
| --- | --- |