Top Music Records

East Central University

(MATH-1223-02)

Nicholas Jacob, PhD

Date

Artist	Album	Released	Genre	Claimed sales in millions
Michael Jackson	Thriller	1982	Pop, post-disco, funk, rock	70
AC/DC	Back in Black	1980	Hard rock	50
Whitney Houston / various artists	The Bodyguard	1992	R&B, soul, pop, soundtrack	45
Pink Floyd	The Dark Side of the Moon	1973	Progressive rock	45
Eagles	Their Greatest Hits (1971â€"1975)	1976	Country rock, soft rock, folk rock	44
Eagles	Hotel California	1976	Soft rock	42
Shania Twain	Come On Over	1997	Country, pop	40
Fleetwood Mac	Rumours	1977	Soft rock	40
Bee Gees / Various artists	Saturday Night Fever	1977	Disco	40
Led Zeppelin	Led Zeppelin IV	1971	Hard rock, heavy metal, folk rock	37
Michael Jackson	Bad	1987	Pop, rhythm and blues, funk and rock	35
Meat Loaf	Bat Out of Hell	1977	Hard rock, glam rock, progressive rock	34
Alanis Morissette	Jagged Little Pill	1995	Alternative rock	33
Various artists	Dirty Dancing	1987	Pop, rock, R&B	32
Celine Dion	Falling into You	1996	Pop, soft rock	32
Adele	21	2011	Pop, soul	31
The Beatles	1	2000	Rock	31
Metallica	Metallica	1991	Heavy metal	31
Celine Dion	Let's Talk About Love	1997	Pop, soft rock	31
Bob Marley & The Wailers	Legend: The Best of Bob Marley & The Wailers	1984	Reggae	30

Guns N' Roses	Appetite for Destruction	1987	Hard rock	30
Bruce Springsteen	Born in the U.S.A.	1984	Heartland rock	30
ABBA	Gold: Greatest Hits	1992	Pop, disco	30
Dire Straits	Brothers in Arms	1985	Roots rock, blues rock, soft rock	30
Santana	Supernatural	1999	Latin rock	30
Madonna	The Immaculate Collection	1990	Pop, dance	30
Pink Floyd	The Wall	1979	Progressive rock	30
The Beatles	Sgt. Pepper's Lonely Hearts Club Band	1967	Rock	30
Michael Jackson	Dangerous	1991	New jack swing, R&B and pop	30
The Beatles	Abbey Road	1969	Rock	30
Norah Jones	Come Away with Me	2002	Jazz	28
Mariah Carey	Music Box	1993	Pop, R&B	28
Various artists	Grease: The Original Soundtrack from the Motion Picture	1978	Rock and roll	28
Eminem	The Eminem Show	2002	Hip hop	27
James Horner	Titanic: Music from the Motion Picture	1997	Film score	27
Britney Spears	...Baby One More Time	1999	Pop	26
Nirvana	Nevermind	1991	Grunge, alternative rock	26
Eric Clapton	Unplugged	1992	Acoustic rock, acoustic blues	26
Queen	Greatest Hits	1981	Rock	25

Phil Collins	No Jacket Required	1985	Pop rock	25
Eminem	The Marshall Mathers LP	2000	Hip hop	25
Linkin Park	Hybrid Theory	2000	Nu metal, rap metal, alternative metal	25
U2	The Joshua Tree	1987	Rock	25
Whitney Houston	Whitney Houston	1985	Pop, R&B	25
Prince and The Revolution	Purple Rain	1984	Pop rock, new wave, R&B	25
Bon Jovi	Slippery When Wet	1986	Hard rock, glam metal	25
Carole King	Tapestry	1971	Pop	25
Madonna	True Blue	1986	Pop, dance	25
Simon & Garfunkel	Bridge over Troubled Water	1970	Folk rock	25
George Michael	Faith	1987	Pop, R&B, funk, soul	25
Elton John	Greatest Hits	1974	Pop	24
Backstreet Boys	Millennium	1999	Pop	24
Spice Girls	Spice	1996	Pop	23
Ace of Base	Happy Nation/The Sign	1993	Pop	23
Adele	25	2015	Soul, pop, R&B	22
Michael Jackson	HIStory: Past, Present and Future, Book I	1995	R&B, pop and hip hop	22
Celine Dion	All the Way... A Decade of Song	1999	Pop	22
Madonna	Like a Virgin	1984	Pop, dance	21
Bon Jovi	Cross Road	1994	Hard rock, glam metal	21

Elvis Presley	Elvis' Christmas Album	1957	Christmas, pop, gospel, rock and roll	20
Boston	Boston	1976	Arena rock, hard rock	20
Mariah Carey	Daydream	1995	Pop, R&B	20
Green Day	Dookie	1994	Pop punk, punk rock, alternative rock	20
Shania Twain	The Woman in Me	1995	Country, pop	20
Britney Spears	Oops!... I Did It Again	2000	Pop	20
Whitney Houston	Whitney	1987	Pop, R&B	20
Def Leppard	Hysteria	1987	Hard rock, glam metal	20
Lauryn Hill	The Miseducation of Lauryn Hill	1998	Neo Soul, R&B and Hip Hop	20
Tracy Chapman	Tracy Chapman	1988	Folk rock	20
Lionel Richie	Can't Slow Down	1983	Pop, R&B, soul	20
Michael Jackson	Off the Wall	1979	Disco, pop, funk and R&B	20
Fugees	The Score	1996	Alternative hip hop	20
Oasis	(What's the Story) Morning Glory?	1995	Britpop, rock	20
Celine Dion	The Colour of My Love	1993	Pop	20
Elton John	Goodbye Yellow Brick Road	1973	rock, pop rock, glam rock	20

			Progressive rock, art	
Pink Floyd	Wish You Were Here	1975	rock, experimental rock	20
Various artists	Flashdance: Original Soundtrack from the Motion Picture	1983	Electro, synthpop	20

Introduction to Data

I have, unfortunately, never been too interested in music, and have recently started listening to some in the car and such. I think that studying this data set would be an excellent way to learn more about music and famous creators.

The data presented here is only the top 30 records listed in the datasheet. The reset can be found on GitHub as linked in the work cited section of the paper.

Something that I want to know is how the group/artist and the number of records produced by each affect the number of records that make this list. I'd also like to know if some genres are more popular than others. In order to find the answers to these questions I may have to expand the data sheet to include other variables such as the number of records recorded by each group. The main thing I currently want to learn from this project is if the number of records created by each group has any correlation to the number of top records they have.

The data currently displays the categorical variables: genre, record rank, year, and artist/group. The quantitative variables are records sold. I plan to add the number of times an artist/group is listed as well as the total number of records each has, both of which are quantitative variables.

Records per-year	Frequency	Relative frequency
1950s	1	1.30%
1960s	2	2.60%
1970s	16	20.78%
1980s	22	28.57%
1990s	28	36.36%
2000s	6	7.79%
2010s	2	2.60%

Two Way Table

|	Pop	Electro	Rock	hip hop	Disco	Funk	R&B	Soul	Metal	Country	Xmas	Gospel	Dance	New Wave	Rap	Grunge	Jazz	Filmscore	New jack swing	Sound Track	Reggae
1950s	1		1								1	1									
1960s			2																		
1970s	3		12		2	1	1		1	1											
1980s	12	1	13		1	3	7	2	2				2	1							1
1990s	19		9	3	1		7	2	2	2			1			1		1	1	1	
2000s	1		1	2				1							1			1			
2010s	2						1	2													

Two Way Table

In this table we see a comparison between the release date and genres of the records which ranked in the top 77. The majority of records fell into some subcategory of pop or rock during the 70-90s. There was a strong correlation between R&B music during the 80-90s and it may be a result of events that occurred in that time period. Records were the primary method of

listening to music during the rock and roll era which explains why so many albums that included that genre made the top 77 with a few outliers such as Elvis Presley with his Christmas Album in 1957. The relationship between the two variables is apparent and we can say that the amount of records sold is related to the year. Today we have more music listeners than ever, yet the number of top hits teeters off further into the 2000s.

Claimed Sales histogram
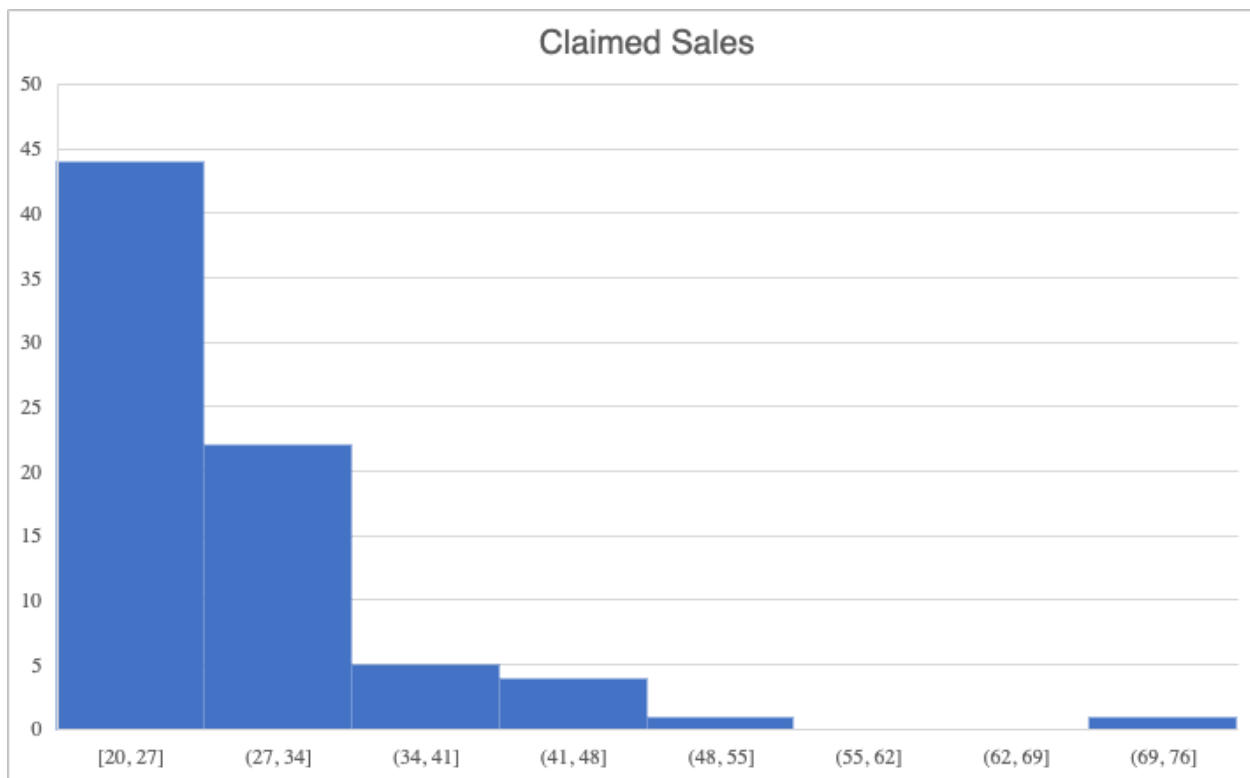
Sale in Millions: Mean- 42.9091 Standard deviation- 23.6916, Q1 20, Median 25, Q3 70, Min 20, Max 70 Range- 50
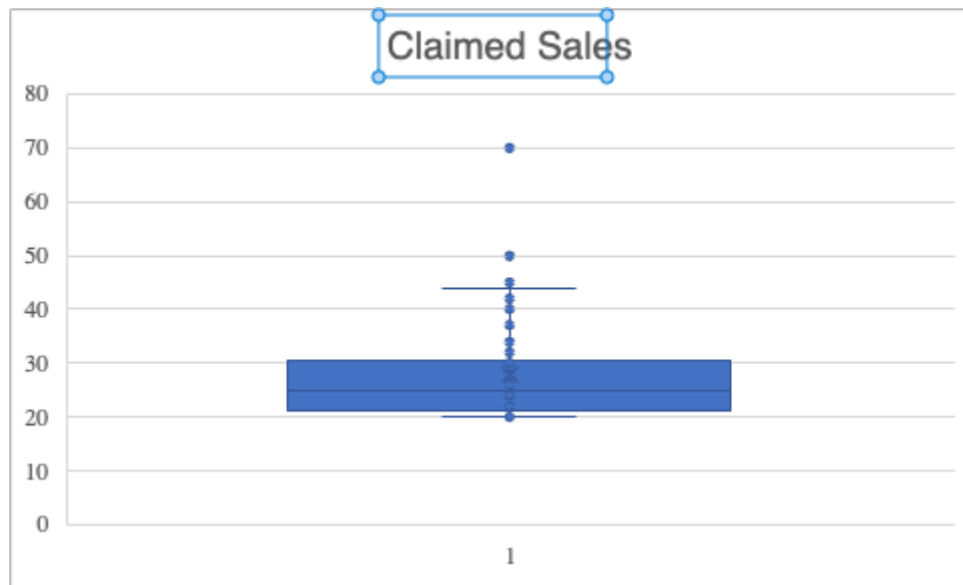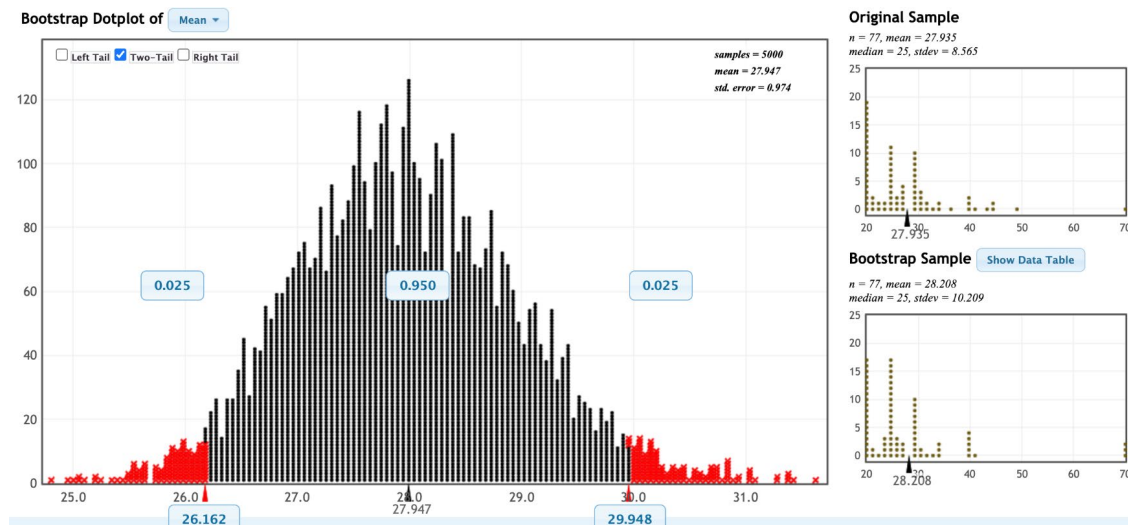
Claimed Sales

The spread for Claimed Sales is skewed right with the majority being in the 20-30 million range.

Hypothesis Test 1

Because Michele Jackson is the king of pop Thriller should have more sales than the average.

HO: µ=70   Ha: µ<70

std. error = 0.974

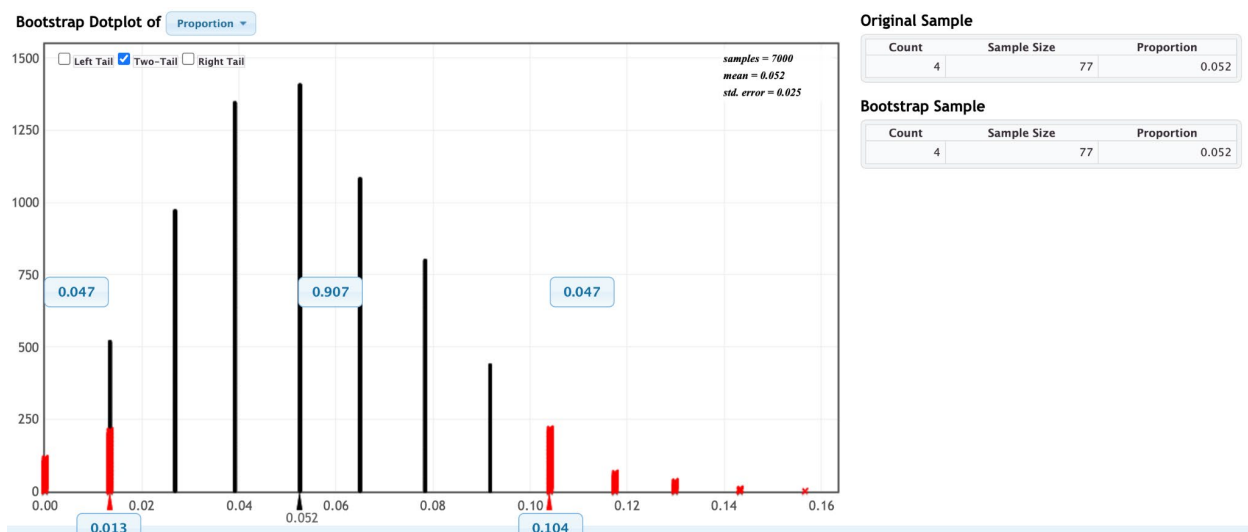*95%Confidence interval= 26.162-29.948*

We can reject the null hypothesis since Thriller, with 70 million sales, is much greater than the

mean (28.208 million) and does not fall within the confidence interval (26.162-29.948).

Hypothesis Test 2

Because there are 21 genres funk should be the genre of 1/21 records

H0:   p = 1/21   Ha: p ≠ 1/21

Bootstrap Dotplot of Proportion ▾

std. error = 0.025

We can reject the alternative hypothesis since the count of funk albums match what we would expect it to be if the genres were distributed equally. Both proportions fall into the confidence interval (0.013<0.052<0.104).

Hypothesis Test 2 Formulas

Working the categorical data using the formulas instead of bootstrapping gives very similar numbers for the standard error and 95% confidence interval for the proportion of top-selling records under the funk genre.

$$z = \frac{phat - p}{\sqrt{\frac{p(1-p)}{n}}} \qquad CI = phat \pm 2 * SE \qquad SE = \sqrt{\frac{p(1-p)}{n}}$$

	Bootstrap	Formula
SE	.025	.024269
95% confidence interval	0.013-0.104	.00341-.100486

Z Score	N/A	.178377

Hypothesis Test 1 Formulas

Working the Quantitative data using the formulas instead of bootstrapping gives very similar almost exact numbers for the standard error and 95% confidence interval for the difference between Thriller and the other top-selling records in the list.

$$z = \frac{\bar{x} - X}{SE} \qquad CI = \bar{x} \pm 2 * SE \qquad SE = s/\sqrt{n}$$

	Bootstrap	Formula
SE	.974	.976

95% confidence interval	26.162-29.948	26.256-30.16
Z Score	N/A	-42.8186721

Probability 1980s & Electro

	Pop	Electro	Rock	hip hop	Disco	Funk	R&B	Soul	Metal	Country	Xmas	Gospel	Dance	New Wave	Rap	Grunge	Jazz	Filmscore	New jack swing	Sound Track	Reggae
1950s	1		1								1	1									
1960s			2																		
1970s	3		12		2	1	1		1	1											
1980s	12	1	13		1	3	7	2	2				2	1							1

1990s	19		9	3	1		7	2	2	2			1		1		1		1	1	
2000s	1		1	2				1							1		1					
2010s	2						1	2														

$$P(A \mid B) = P(A \cap B) / P(B)$$

If 1980s is A and Electro is B then we are claculatign for the probability of a top record classified as Electro being from the 1980s

$P(A \mid B) = P(A \cap B) / P(B)$ is 1/1 or 100% because all records under Electro that made the top records list occurred in the 1980s.

If we switch the 1980s and Electro to find the probability of one of the top records from 1980 being Electro then $P(A \mid B) = P(A \cap B) / P(B)$ would be 1/45 or 2.222···% because of the 45 top records in the 1980s only one fell under Electro

References

https://github.com/nurfnick/Data_Sets_For_Stats/blob/2827ce58f62265ca8f908389c1613fe5d22a7815/CuratedDataSets/topRecords.csv

https://www.lock5stat.com/StatKey/index.html