**Santino Celebre**

**Analysis of Residential Sales in the United States**

**INDEX**

**1.0 Data Set and Background .......... 2**

**1.1 Mean, Median, Mode, Variance, and Standard Deviation ........... 2**

**1.2 Combination and Permutation …....... 2**

**1.3 Binomial Distribution …...... 2-3**

**1.4 Chebyshev's Theorem .......... 3-4**

* 1. **Example 1 and Example 2**

**2.0 Real Life Examples of Probability and Statistics …....... 3-4**

**3.0 Graphs …....... 7-8**

* 1. **Median and Average Sales Prices of New Homes Sold in United States**
  2. **b. Houses Sold by Region**

**3.1 Work Cited …....... 8**

**1. Data Set Background Knowledge**

The data set that I chose was “Houses Sold by Region” this data includes data from 1963-2021. This data was pulled straight from The United States Census Bureau. Pulled from the same reference was “Median and Average Sales Prices of New Homes Sold in United States”. All data was for new, single-family houses only.

**1.1 Mean, Median, Mode, Variance, and Standard Deviation.**

When looking at this data set to ask more probability bases questions, we need to first need to understand some information about the set. Mean can be closely compared to the average of a set and for this set the mean of houses sold in the United States from 1963-2021 would be 699,000. This number was acquired by taking all values, adding them up and dividing by the total this number will later help us be able to find variance and standard deviation. The median or middle number was calculated to be 671,000 and since there were no repeating number of houses sold in the United States from 1963-2021 there is no mode. Standard deviation and variance can also be found in this data set. Standard deviation (Definition 1.3) from the book, “The standard deviation can be used to give a fairly accurate picture of data variation for a single set of measurements” (Mathematical Statistics with Applications 7th Edition). In this case the single set of measurements is the number of houses sold in the United Stated for given years. Variance (Definition 1.2) from the book, “The larger the variance of a set of measurements, the greater will be the amount of variation within the set. The variance is of value in comparing the relative variation of two sets of measurements, but it gives information about the variation in a single set only when interpreted in terms of the standard deviation” (Mathematical Statistics with Applications 7th Edition). That is why we were able to find out that the standard deviation of this specific data set was 205,000.

**1.2 Combination and Permutation**

The number of **combinations** of n objects taken r at a time is the number of subsets, each of size r, that can be formed from the n objects.

Example 1:

**There are 4 separate ways to finance a house, Conventional, FHA, VA, and Cash. If you were to buy four houses in a year how many ways can you finance the house without repeating the same technique?**

= 4\*3\*2\*1 = 24

= 1 + 4 + 12 + 24 + 24

Thus, there are 64 ways to finance a house.

An ordered arrangement of r distinct objects is called a **permutation**. The number of ways of ordering n distinct objects taken r at a time will be designated by the symbol .

**Permutation Example:**

**1.3 Binomial Distribution**

For binomial distribution the properties were the most important thing to understand to determine whether the equation is a binomial equation. Something that I found especially important is how the things we learn like binomial distribution apply to the real world. Well, a lot of things in our daily life can be represented using a binomial distribution. For instance, when you are looking for a loan to put towards a house you must deal with the bank.

“So, what does this have to do with finance? More than you might think. Let us say you are a bank, a lender, who knows within three decimal places the likelihood of a particular borrower defaulting. What are the chances of so many borrowers defaulting that they would render the bank insolvent? Once you use the cumulative binomial distribution function to calculate that number, you have a better idea of how to price insurance, and how much money to loan and how much to keep in reserve.”

Example 1

According to the US Department of Commerce in 2002, 62% of people had a cellphone. If 10 people were selected at random what is the probability that 3 have a cellphone?

n = 10; X = 3; P = 62.0; 1-P = 38.0

Binomial Probability = .032 or 3% chance that out of 10 people three have a cellphone.

**1.4 Chebyshev's Theorem**

Theorem 3.14

Chebyshev's Theorem Let Y be a random variable with mean μ and finite variance σ 2. Then, for any constant k > 0,

Chebyshev's Theorem can be used to determine where data falls within a distribution of values. For instance, if we look at “Median and Average Sales Prices of New Homes Sold in United States” we can pose this question.

Example 1

From 1963-2021 the average house sold for about $172,180 with a standard deviation of $124,459. What was the minimum percentage of houses sold between $100,000 and $400,000?

To start with a question like this we need to first set up the formula . Using the ranges given in the question we get:

thus

Now that we have k we can assume

The minimum percentage of houses sold between $100,000 and $400,000 was about 69%

In the real world these numbers do matter. When buying a home knowing your price range is very important and if you find that a low percentage of houses were in your price range then you might have a harder time finding real estate. The equation we went through above makes it easy to interchange ranges and find out numbers that work for you.

Example 2

For 1963-2021 the US Census reported the mean price of houses sold was $172,180 and the standard deviation was $124,459. If nothing is known about the shape of the distribution, give an interval of costs that will contain 42% of houses.

To do find the intervals the correct formula would be . Using Chebyshev’s Theorem and applying it to the question leads to:

thus, k = 2.89

Now that we know the value of k, we can substitute into the formula above to find the intervals.

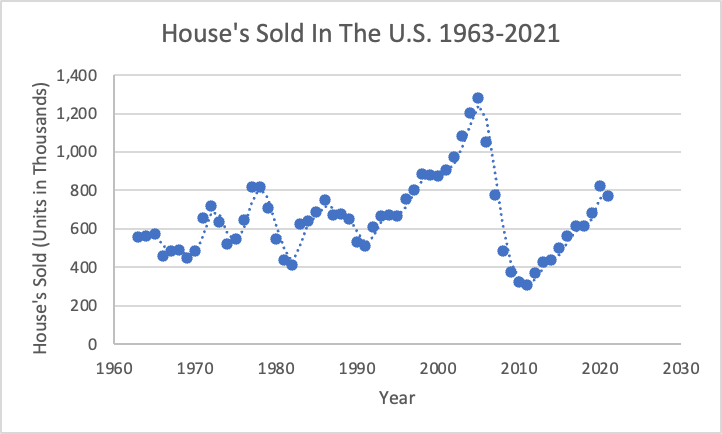
= [10,383, 396,206]

The interval of costs that will contain 42% of houses would be [$10,383, $396,206]

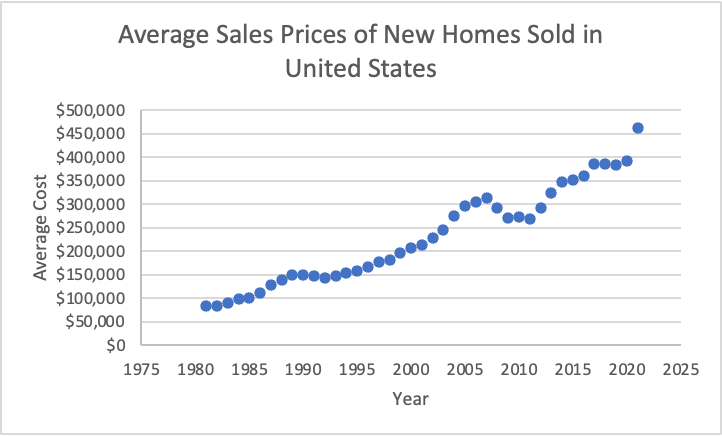
**2.0 Real Life Examples of Probability and Statistics**

When forecasting the future of the housing market, probability and statistics are something that these professionals look at. Projected home growth price rate has skyrocketed compared to following years and therefore sales forecasting methods are such an important tool for real estate agents, and it is made possible by collecting and analyzing data.

**3.0 Graphs**



This graph is interesting because of the huge rise of houses sold from around 2003-2006. This can pose the question of what drove this increase. If you look at what happened during covid something very similar occurred. “Inflows of money into housing markets, loose lending conditions, and government policy to promote homeownership” (Investopedia).



This graph follows a very linear growth. This makes complete sense when you factor in supply and demand, interest rates, and the overall economy.

**3.1 Work Cited**

Miley, Michelle. “What Causes Housing Prices to Rise in the United States?” *Home Guides | SF Gate*, 17 Nov. 2020, <https://homeguides.sfgate.com/causes-housing-prices-rise-united-states-56413.html>.

Owen, William J., and Dennis D. Wackerly. *Student Solutions Manual: Mathematical Statistics with Applications: Seventh Edition*. Brooks/Cole Cengage Learning, 2008.

Team, The Investopedia. “What Is a Housing Bubble?” *Investopedia*, Investopedia, 8 Feb. 2022, <https://www.investopedia.com/terms/h/housing_bubble.asp>.

US Census Bureau (MCD): Cheryl Cornish, Stephen Cooper. “New Residential Sales.” *United States Census Bureau*, 23 Aug. 2011, <https://www.census.gov/construction/nrs/historical_data/index.html>.