Santino Celebre

Analysis of Residential Sales in the United States

INDEX

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?

$P_4 = 4! = 4*3*2*1 = 24$

$P_0 + P_1 + P_2 + P_3 + P_4 + P_5 = 1 + 4 + 12 + 24 + 24$

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 P_r^n.

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

$P(x) = nCx \cdot p^x \cdot (1-p)^{n-x}$

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,

$$P\left(|Y-\mu| < \kappa\sigma|\right) \geq 1-\frac{1}{k^2} \; or \; (P|Y-\mu| > \kappa\sigma) \leq \frac{1}{k^2}$$

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 $\mu + \kappa\sigma$. Using the ranges given in the question we get:

$$172{,}180 + \kappa(124{,}459) = 400{,}000 \text{ thus } k = 1.8$$

$$\mu = 172{,}180, k = 1.8, \sigma = 124{,}459$$

Now that we have k we can assume $1 - \frac{1}{k^2} = 1 - \frac{1}{(1.8)^2} = .69\ or\ 69\%$

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 $[\mu - \kappa\sigma,\ \mu + \kappa\sigma]$. Using Chebyshev's Theorem and applying it to the question leads to:

$1 - \frac{1}{k^2} = .42$ thus, k = 2.89

$$\mu = 172{,}180, k = \pm 1.3, \sigma = 124{,}459$$

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

$[172{,}180 - 1.3(124{,}459), 172{,}180 + 1.3(124{,}459)]$ = [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 Data Sets

A. Median and Average Sales Prices of New Homes Sold in United States

Annual Data

Period	Median	Average
1963	$18,000	$19,300
1964	$18,900	$20,500
1965	$20,000	$21,500
1966	$21,400	$23,300
1967	$22,700	$24,600
1968	$24,700	$26,600
1969	$25,600	$27,900
1970	$23,400	$26,600
1971	$25,200	$28,300
1972	$27,600	$30,500
1973	$32,500	$35,500
1974	$35,900	$38,900
1975	$39,300	$42,600
1976	$44,200	$48,000
1977	$48,800	$54,200
1978	$55,700	$62,500
1979	$62,900	$71,800
1980	$64,600	$76,400
1981	$68,900	$83,000
1982	$69,300	$83,900
1983	$75,300	$89,800
1984	$79,900	$97,600
1985	$84,300	$100,800
1986	$92,000	$111,900
1987	$104,500	$127,200
1988	$112,500	$138,300
1989	$120,000	$148,800
1990	$122,900	$149,800
1991	$120,000	$147,200
1992	$121,500	$144,100
1993	$126,500	$147,700
1994	$130,000	$154,500
1995	$133,900	$158,700
1996	$140,000	$166,400
1997	$146,000	$176,200
1998	$152,500	$181,900
1999	$161,000	$195,600
2000	$169,000	$207,000
2001	$175,200	$213,200

2002	$187,600	$228,700
2003	$195,000	$246,300
2004	$221,000	$274,500
2005	$240,900	$297,000
2006	$246,500	$305,900
2007	$247,900	$313,600
2008	$232,100	$292,600
2009	$216,700	$270,900
2010	$221,800	$272,900
2011	$227,200	$267,900
2012	$245,200	$292,200
2013	$268,900	$324,500
2014	$288,500	$347,700
2015	$294,200	$352,700
2016	$307,800	$360,900
2017	$323,100	$384,900
2018	$326,400	$385,000
2019	$321,500	$383,900
2020	$336,900	$391,900
2021	$398,800	$462,000

Note: The sales price includes the land.

B. Houses Sold by Region

Annual Data

(Components may not add to total because of rounding. Number of housing units in thousands.)

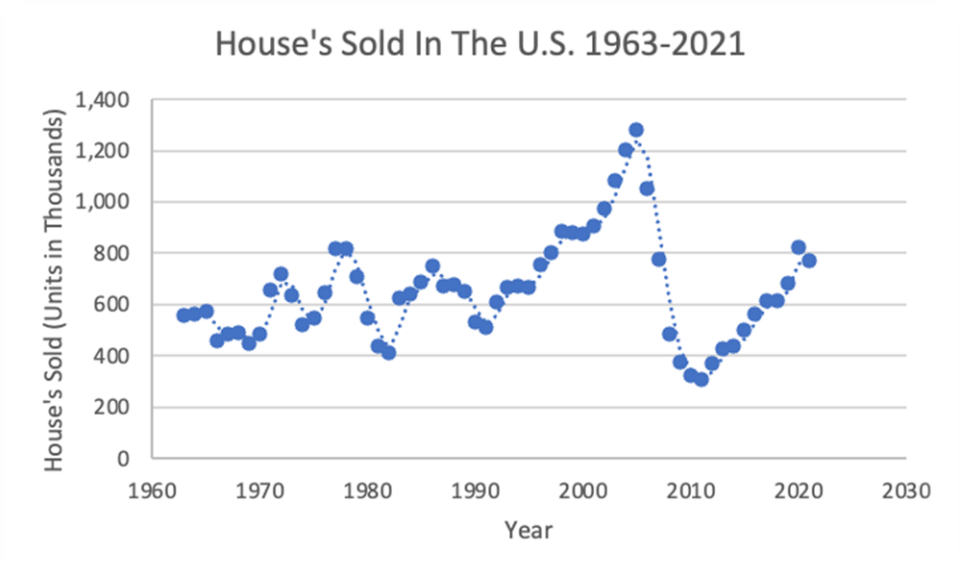
Period	United States	North-east	Mid-West	South	West
1963	560	87	134	199	141
1964	565	90	146	200	129
1965	575	94	142	210	129
1966	461	84	113	166	99
1967	487	77	112	179	119
1968	490	73	119	177	121
1969	448	62	97	175	114
1970	485	61	100	203	121
1971	656	82	127	270	176
1972	718	96	130	305	187
1973	634	95	120	257	161
1974	519	69	103	207	139

1975	549	71	106	222	150
1976	646	72	128	247	199
1977	819	86	162	317	255
1978	817	78	145	331	262
1979	709	67	112	304	225
1980	545	50	81	267	145
1981	436	46	60	219	112
1982	412	47	48	219	99
1983	623	76	71	323	152
1984	639	94	76	309	160
1985	688	112	82	323	171
1986	750	136	96	322	196
1987	671	117	97	271	186
1988	676	101	97	276	202
1989	650	86	102	260	202
1990	534	71	89	225	149
1991	509	57	93	215	144
1992	610	65	116	259	170
1993	666	60	123	295	188
1994	670	61	123	295	191
1995	667	55	125	300	187
1996	757	74	137	337	209
1997	804	78	140	363	223
1998	886	81	164	398	243
1999	880	76	168	395	242
2000	877	71	155	406	244
2001	908	66	164	439	239
2002	973	65	185	450	273
2003	1,086	79	189	511	307
2004	1,203	83	210	562	348
2005	1,283	81	205	638	358
2006	1,051	63	161	559	267
2007	776	65	118	411	181
2008	485	35	70	266	114
2009	375	31	54	202	87
2010	323	31	45	173	74
2011	306	21	45	168	72
2012	368	29	47	195	97
2013	429	31	61	233	105
2014	437	28	59	243	108
2015	501	24	61	286	130
2016	561	32	69	318	142
2017	613	40	72	339	163
2018	617	32	76	348	160
2019	683	30	72	399	182

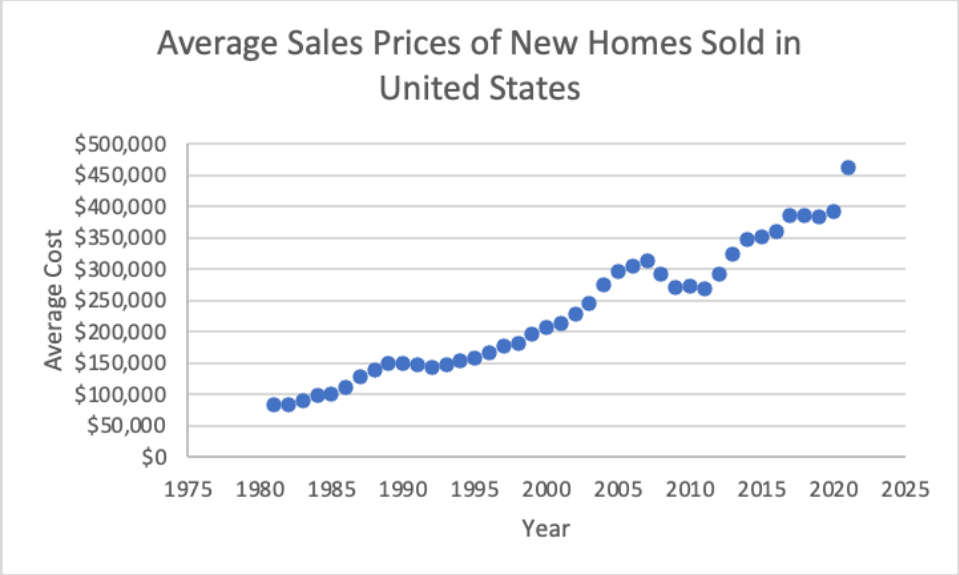
2020	822	37	93	474	218
2021	770	36	86	452	196

Note: Estimates prior to 1999 include an upward adjustment of 3.3 percent made to account for
houses sold in permit-issuing areas that will never have a
permit authorization.

3.1 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.

3.2 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.