AMERICAN HOUSING SURVEY: HOUSING AFFORDABILITY DATA SYSTEM

Download: Excel Workbook, ReadMe

Abstract: is a set of files derived from the 1985 and later national American Housing Survey (AHS) and the 2002 and later Metro AHS. This system categorizes housing units by affordability and households by income, with respect to the Adjusted Median Income, Fair Market Rent (FMR), and poverty income. It also includes housing cost burden for owner and renter households. These files have been the basis for the worst case needs tables since 2001. [1]

Data Set Characteristics:	Multivariate	**Number of Instances:**	30193	**Area:**	Housing
Attribute Characteristics:	N/A	**Number of Attributes:**	16	**Missing Values?**	N/A -6

Source:

https://www.huduser.gov/portal/datasets/hads/hads.html

Updated Codebook:

Original Codebook credited to https://www.huduser.gov/portal/datasets/hads/hads.html

Updates made by Kim Kirk "Business Statistics and Analysis" for Rice University.

[1] https://www.huduser.gov/portal/datasets/hads/hads.html

Data Set Information:

The *Housing Affordability Data System (HADS)* is a set of files derived from the 1985 and later national American Housing Survey (AHS) and the 2002 and later Metro AHS. This system categorizes housing units by affordability and households by income, with respect to the Adjusted Median Income, Fair Market Rent (FMR), and poverty income. It also includes housing cost burden for owner and renter households. These files have been the basis for the worst case needs tables since 2001. The data files are available for public use, since they were derived from AHS public use files and the published income limits and FMRs. We are providing these files give the community of housing analysts the opportunity to use a consistent set of affordability measures.[2]

Data variables include:

Name	Data Type	Explanation
AGE	Numerical	Age of head of household
METRO3	Character '1','2', '3', '4' or '5'	Metropolitan status: '1' : Central City '2', '3', '4', '5' :Others
REGION	Character '1','2', '3' or '4'	The four census regions— Northeast, Midwest, South, and West.
LMED	Numerical ($)	Area Median Income
FMR	Numerical ($)	Fair Market Monthly Rent
VALUE	Numerical ($)	Current market value of unit

[2] https://www.huduser.gov/portal/datasets/hads/hads.html

BEDRMS	Numerical	Number of Bedrooms in the unit
BUILT	Numerical	Year the unit was built
ROOMS	Numerical	Number of rooms in the unit
PER	Numerical	Number of persons in Household
ZINC2	Numerical	Annual Household income
ZADEQ	Character	Adequacy of unit '1' Adequate '2' Moderately Inadequate '3' Severely Inadequate '-6' Vacant - No information
UTILITY	Numerical ($)	Monthly utilities cost (gas, oil, electricity, other fuel, trash collection, and water)
OTHERCOST	Numerical ($)	Sum of 'other monthly costs' such as Home owners' or renters' insurance, Land rent (where distinct from unit rent), Condominium fees (where applicable), Other mobile home fees (where applicable).

Data Summaries Calculated:

- Mean
- Standard Error
- Median
- Mode
- Standard Deviation
- Sample Variance
- Kurtosis
- Skewness
- Range
- Minimum
- Maximum
- Sum
- Count

Units for Data Summaries and Variables:

Name	Unit Measurement
AGE	Age in Years
LMED	Dollars ($)
FMR	Dollars ($)
VALUE	Dollars ($)
BEDRMS	Housing Unit
BUILT	Years
ROOMS	Number of Rooms
PER	Number of Persons
ZINC2	Dollars ($)
UTILITY	Dollars ($)
OTHERCOST	Dollars ($)

Data Transformations:

1. Dataset was downloaded from host website: https://www.huduser.gov/portal/datasets/hads/hads.html
2. Data was cleaned to include only single-family houses, flats, apartments with Fair Market Value of at Least $1000.00 owned in 2013 as this was the subset that stakeholders wanted to focus on.
3. VALUE was determined to be the outcome variable.
4. BEDRMS, LMED, FMR, BUILT, ROOMS, UTILITY, REGION, METRO3, OTHERCOST, ZINC2, AGE, ZADEQ, PER, VALUE were determined to be the predictor variables.
5. VALUE, OTHERCOST, ZINC2, LMED, FMR had exponential distributions and a natural logarithmic transformation was applied for a normal distribution.
6. New variables created include:
    - LN_VALUE
    - LN_OTHERCOST
    - LN_ZINC2
    - LN_LMED
    - LN_FMR
7. Worksheets were created to hold four categories for the analysis:
    - Summary Report identifies and answers the business question/what was measured
    - Descriptive Statistics identifies and calculates descriptive statistics
    - Graphs displays histograms, scatterplots, bar graphs
    - Statistical Tests holds results for statistical tests