

Assignment 2: Report

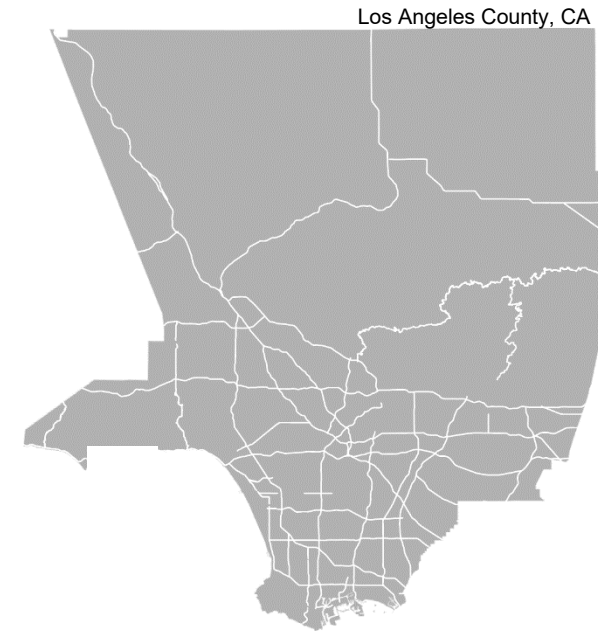
Catherine Saint

9/10/2020

Instructions: For each continuous variable in your dataset, calculate the sample mean, the sample standard deviation, the 95-percent confidence interval for the population mean, and the interquartile range. Create a histogram to illustrate the distribution of each variable and describe the distribution in a sentence or two. For each categorical variable in your dataset, calculate the 95-percent confidence interval for the proportion of the population in each category.

List of variables (from Assignment 1)

1. Median Age (continuous) : med_ageE
2. Part of the total population unemployed in civilian labor force (age 16 years and above) (continuous): pct_unemp_labor
3. Part of the total population who are high school graduates (or equivalent) in labor force (continuous): pct_hs_grad
4. Majority race of the population (categorical) : maj_race_tract
5. Whether the majority of the population is foreign born or not(categorical): maj_foreign



Distribution of Continuous Variables

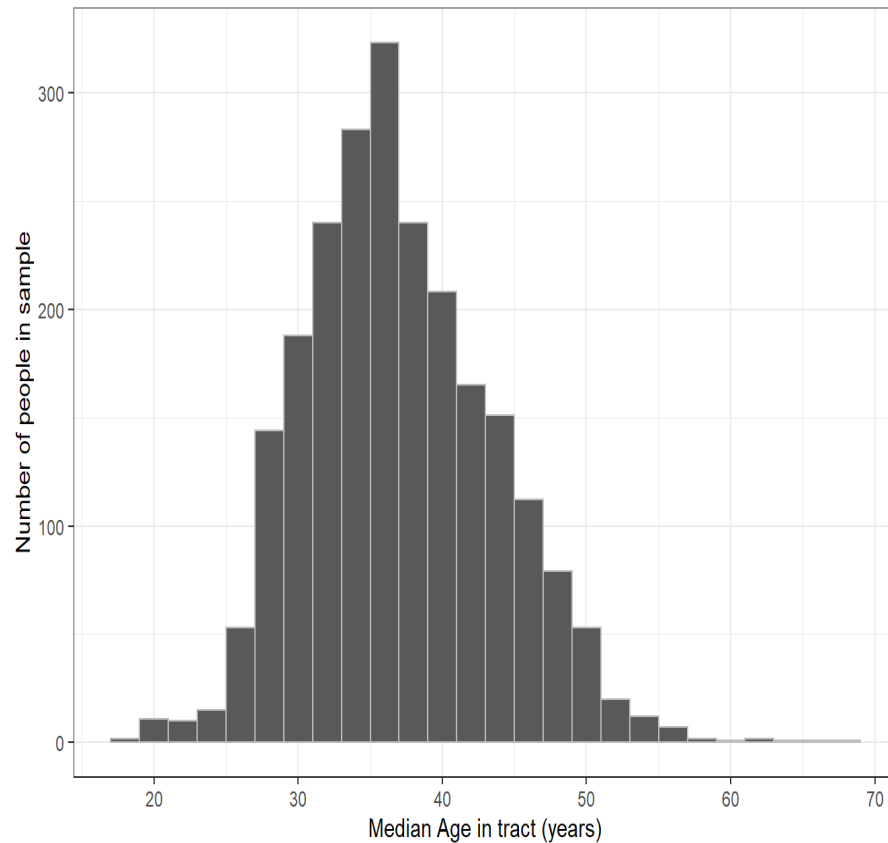
Descriptive Statistics

Characteristics of Census Tracts in Los Angeles County

<i>Statistic</i>	<i>Median Age (years)</i>	<i>Percent unemployed</i>	<i>Percent high-school graduate workers</i>
<i>Sample Mean</i>	37.1	3.55%	8.4%
<i>Median</i>	36.4	3.32%	8.67%
<i>Standard Deviation</i>	6.59	1.71%	4.25%
<i>Interquartile Range</i>	32.4 to 41.3	2.38% to 4.51%	5.69% to 11%
<i>Population Mean</i>	36.8 to 37.3	3.48% to 3.62%	8.23% to 8.58%

Distribution of Continuous Variables (cont.d)

Histograms



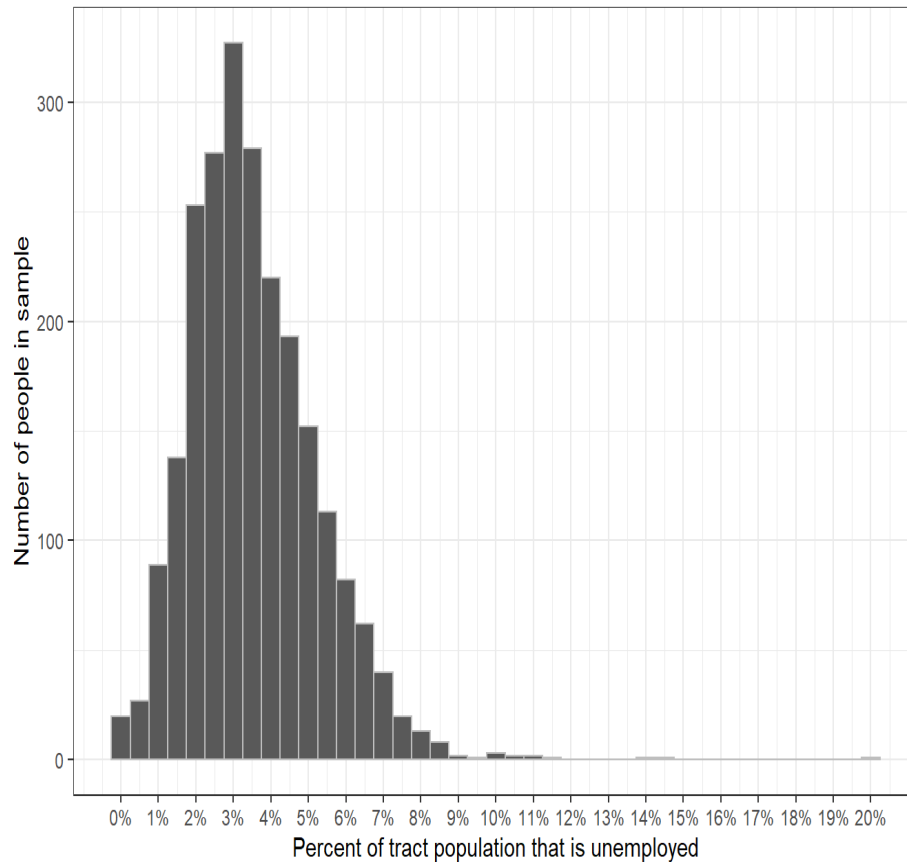
Median Age

The distribution looks relatively normal, with a slight skew to the right. The median value is less than the corresponding mean value, because the outliers are located further out along the x-axis, where age increases.

Statistic	Median Age (years)	Percent unemployed	Percent high-school graduate workers
Sample Mean	37.1	3.55%	8.4%
Median	36.4	3.32%	8.67%
Standard Deviation	6.59	1.71%	4.25%
Interquartile Range	32.4 to 41.3	2.38% to 4.51%	5.69% to 11%
Population Mean	36.8 to 37.3	3.48% to 3.62%	8.23% to 8.58%

Distribution of Continuous Variables (cont.d)

Histograms



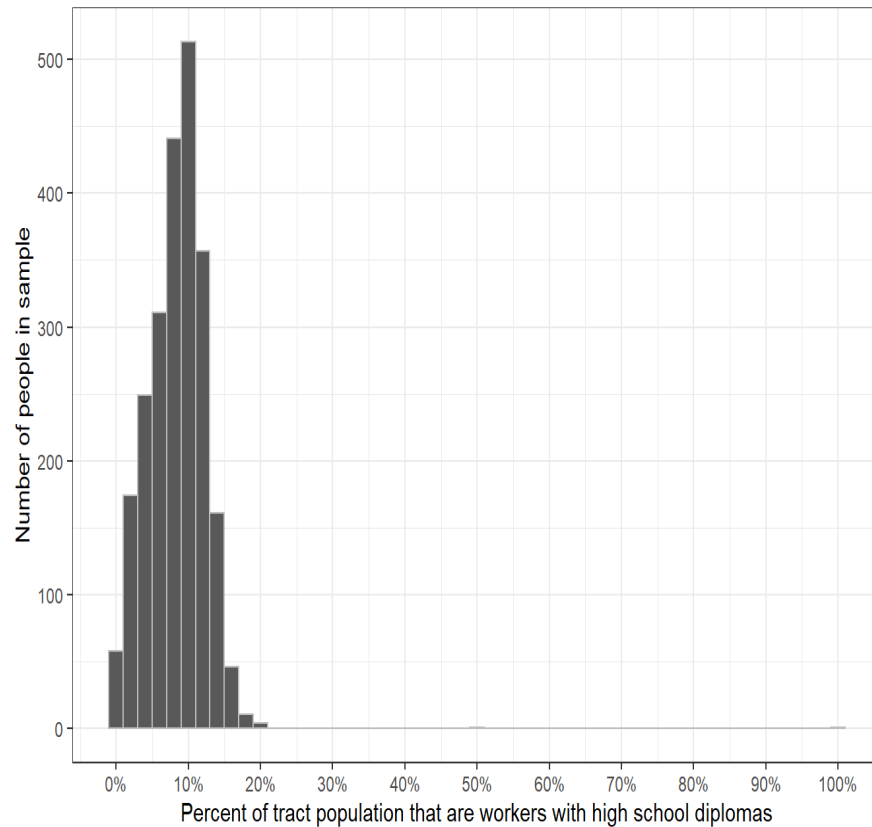
Unemployment

The distribution of unemployment is positively skewed with a long tail. Outliers around 20% bring the average of the distribution up, with most of the observations concentrated to the left.

Statistic	Median Age (years)	Percent unemployed	Percent high-school graduate workers
Sample Mean	37.1	3.55%	8.4%
Median	36.4	3.32%	8.67%
Standard Deviation	6.59	1.71%	4.25%
Interquartile Range	32.4 to 41.3	2.38% to 4.51%	5.69% to 11%
Population Mean	36.8 to 37.3	3.48% to 3.62%	8.23% to 8.58%

Distribution of Continuous Variables (cont.d)

Histograms



Worker Education

The data skews negatively to the left. Although there are outliers located to the right of the mean and median, there still remains a greater number of observations closer to 0%, bringing down the average.

Statistic	Median Age (years)	Percent unemployed	Percent high-school graduate workers
Sample Mean	37.1	3.55%	8.4%
Median	36.4	3.32%	8.67%
Standard Deviation	6.59	1.71%	4.25%
Interquartile Range	32.4 to 41.3	2.38% to 4.51%	5.69% to 11%
Population Mean	36.8 to 37.3	3.48% to 3.62%	8.23% to 8.58%

Distribution of Categorical Variables

Descriptive Statistics

Race/Ethnicity Statistics for Census Tracts in Los Angeles County

<i>Majority Race</i>	<i>Sample Proportion</i>	<i>95-percent confidence interval</i>
<i>Hispanic/Latino</i>	46.2%	44.2% to 48.2%
<i>White (not Hispanic/Latino)</i>	23.2%	21.5% to 24.9%
<i>No Majority</i>	23%	21.3% to 24.7%
<i>Asian (not Hispanic/Latino)</i>	5.29%	4.38% to 6.19%
<i>Black (not Hispanic/Latino)</i>	2.26%	1.66% to 2.86%

Nativity Statistics for Census Tracts in Los Angeles County

<i>Majority Nativity</i>	<i>Sample Proportion</i>	<i>95-percent confidence interval</i>
<i>Native Born</i>	88.3%	87% to 89.6%
<i>Foreign Born</i>	11.7%	10.4% to 13%