Assignment 3&4

Principles of Data Science

Karthikeya Arra

16354793

[Ka7hc@umsystem.edu](mailto:Ka7hc@umsystem.edu)

A screenshot of a computer program

Description automatically generated

Code:

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A green and blue squares with white text

Description automatically generated

Based on the analysis of the picture and the context of the Python code, the conclusions drawn from the two bar graphs can be summed up as follows:   
  
Temperature: The lower to mid-20s degrees Celsius are shown as the temperature in Kansas City at the time the data was retrieved. Based on the time of year and the average temperature in the area, this indicates a comparatively warm or mild temperature.  
Humidity: Kansas City's humidity level is described as being just above 20%, which is low. If this reading is paired with the moderate temperatures shown by the temperature graph, it may indicate comfortable conditions.   
  
All things considered, these graphs provide an overview of the weather in Kansas City, which seems to be

A white paper with black text

Description automatically generated

Code:

A screenshot of a computer

Description automatically generated

A chart with different colored rectangles

Description automatically generated

The mean and maximum glucose levels for a sample and the total population are contrasted in the bar chart you provide.   
• Sample Mean: The sample appears to have a mean glucose level of just over 100 mg/dL.  
• Population Mean: Slightly less than 125 mg/dL is the mean glucose level for the population, which is higher.   
  
• Sample Max: At approximately 175 mg/dL, the sample's maximum glucose level is noticeably greater than the sample mean.   
• Population Max: At over 200 mg/dL, the maximal glucose level in the population is even greater than the sample max.

Results:

• The sample's average glucose level is lower than the mean for the population. This could indicate that the sample is not entirely typical of the population in terms of average glucose levels, or it could be the result of random variance in the sampling procedure.   
• The highest glucose level is rather high in the population as well as the sample, which may mean that some people in both groups have very high glucose readings. The population's highest glucose level, however, is higher than the sample's.When drawing conclusions about a population from a sample, sample size and representativeness are crucial, as demonstrated by the distinctions between sample and population statistics.

Top of Form

Bottom of Form

2B)

A screenshot of a computer

Description automatically generated

A blue and green squares

Description automatically generated

The 98th percentile of BMI values for a sample and the population the sample was obtained from are displayed in a bar chart.   
• Sample 98th Percentile BMI: The sample's 98th percentile BMI is around 42. This shows that the BMI readings of 98% of the sample are less than 42.  
• Population 98th Percentile BMI: The population's BMI is situated just over the 50th percentile. This implies that 98% of the population's BMI values are below this threshold.

Conclusions:

• The population's 98th percentile BMI is much higher than the sample's. This could indicate that either the population has outliers with significantly higher BMI values than the sample as a whole, or the sample may not accurately reflect the higher BMI values found in the population as a whole. • Such a difference can be significant, indicating potential sampling bias or natural variation within the sample compared to the whole population.   
  
The results may indicate that the sampling strategy needs to be reviewed in order to better reflect the distribution of BMI values in the population, particularly at the top end of the range, if this sample was meant to represent the population for BMI.

2C)Top of FormBottom of Form

A screenshot of a computer program

Description automatically generated

A screenshot of a graph

Description automatically generated

The graph and output compare the mean and standard deviation of blood pressure using bootstrap sampling to the actual population data.   
• The bootstrap mean blood pressure is roughly 69 mmHg, with error bars showing variability among samples.  
• The population's mean blood pressure is approximately 69.1 mmHg, which is close to the bootstrap mean. The smaller error bar indicates less fluctuation as this is a population-wide measure.   
• The bootstrap standard deviation for blood pressure is approximately 19.2 mmHg, with broad error bars suggesting variability among samples.

• The population standard deviation for blood pressure is around 19.4 mmHg, with no error bar as it represents a single measurement from the entire population.

Findings:

• The bootstrap mean closely matches the population mean, indicating an excellent representation of central tendency.   
• The bootstrap means have natural variability due to sample-to-sample variation, as evidenced by the length of error bars.The standard deviations of the bootstrap samples are likewise quite near to the population standard deviation, implying that the bootstrap samples adequately capture the distribution of the population data.

• The population 2.5th and 97.5th percentile values are exactly 0 and 96, respectively. This could indicate a data entry error or censoring of the 'BloodPressure' data in medical datasets due to privacy or measurement limitations.   
Overall, the bootstrap sampling technique appears to provide a decent approximation of the population parameters for mean and standard deviation of blood pressure, based on the similarities between the calculated values from the bootstrap samples and the real population.