Crime and Unemployment Tables

# Unemployment

To interpret these statistics, let’s look at each one in the context of understanding the shape and spread of the data distribution.

1. **Mean = 7.39**:
   * The mean (average) of the data is 7.39. This gives an idea of the central tendency, or the "typical" value in the dataset.
2. **Min = 2.87**:
   * The minimum value is 2.87, which is the smallest observation in the data. It shows the lower boundary of your data.
3. **Max = 14.12**:
   * The maximum value is 14.12, the largest observation in the dataset, indicating the upper boundary of your data.
4. **Range = 11.24**:
   * The range, calculated as max - min, is 11.24. This represents the span of the data, showing how spread out the values are from the smallest to the largest.
5. **Skew = 0.35**:
   * Skewness measures the asymmetry of the distribution around the mean.
   * A skewness of **0.35** indicates a slight positive skew, meaning that there are some values in the data that are a bit higher than the mean, pulling the tail of the distribution slightly to the right.
   * Generally, a skew close to 0 (e.g., between -0.5 and 0.5) suggests a fairly symmetrical distribution.
6. **Kurtosis = -0.51**:
   * Kurtosis describes the "tailedness" of the distribution, or the prevalence of extreme values.
   * A kurtosis of **-0.51** suggests the distribution has slightly lighter tails and a flatter peak than a normal distribution (which has a kurtosis of 0).
   * This negative kurtosis indicates a **platykurtic** distribution, where extreme values (outliers) are less common, and the distribution is slightly more flat-topped.

**Overall Interpretation**

This data appears to have a moderate spread (range of 11.24) with values clustered around the mean of 7.39. The distribution is roughly symmetrical with a slight right skew and a somewhat flat peak, indicating few extreme values on either end. The negative kurtosis and small skew value both suggest that the data is close to normally distributed, with only a mild rightward skew and a flatter-than-normal peak.

# Crime

Given the values and constraints of this dataset (where values range between 0 and 1), let's break down the interpretation of each statistic:

1. **Mean = 0.37**:
   * The mean (average) of the data is 0.37. This suggests that, on average, values are closer to the lower end of the range (0) than to the upper end (1). Since the mean is below 0.5, it implies that the distribution is skewed toward the lower half of the possible values.
2. **Min = 0.1** and **Max = 0.8**:
   * The minimum observed value is 0.1, and the maximum is 0.8, indicating that this dataset does not fully span the possible range from 0 to 1.
   * This limited range shows that no observations reach the theoretical maximum (1) or minimum (0), with values clustering more in the middle-to-lower part of the possible range.
3. **Range = 0.7**:
   * The range, calculated as max - min, is 0.7. This confirms that while there is some spread in the data, it only covers 70% of the potential range from 0 to 1, implying a moderate variation among the observed values.
4. **Skew = 0.48**:
   * A skewness of 0.48 suggests a mild positive skew, meaning that there are slightly more values closer to the minimum (0.1) than to the maximum (0.8), pulling the distribution slightly to the right.
   * In practical terms, this skewness indicates that most values are slightly below the mean, with a few higher values extending toward the upper end (0.8).
5. **Kurtosis = -0.22**:
   * The kurtosis of -0.22 indicates a distribution that is somewhat flatter than a normal distribution, though not drastically so.
   * This slight negative kurtosis (platykurtic) means that the distribution has fewer extreme values (outliers) and a broader, less sharply peaked shape.

**Overall Interpretation**

The data is moderately spread between 0.1 and 0.8, with most values concentrated closer to the lower end (mean of 0.37). The mild positive skew suggests a slight tendency for values to be lower rather than higher, but there are some values reaching toward the upper part of the observed range. The slightly negative kurtosis further supports a relatively flat distribution without extreme peaks or outliers. Given the constraints (with possible values between 0 and 1), this distribution shows a modest concentration below the midpoint of 0.5, with limited spread near the maximum possible value of 1.

Some Correlation

To characterize the relationship between the two sets of values, let’s compare them in terms of central tendency, spread, skewness, and kurtosis.

**1. Central Tendency (Mean)**

* **First Set**: Mean = 7.39
* **Second Set**: Mean = 0.37

The first set of values has a much higher mean, which reflects its higher overall scale compared to the second set. This difference in means could indicate that the datasets are on different scales or represent fundamentally different kinds of measurements.

**2. Range and Spread**

* **First Set**: Range = 11.24 (from 2.87 to 14.12)
* **Second Set**: Range = 0.7 (from 0.1 to 0.8)

The first set has a much larger range, showing a broader spread in values, while the second set is more constrained. This suggests that the first dataset has greater variability or includes a wider array of values relative to its minimum and maximum.

**3. Skewness**

* **First Set**: Skew = 0.35
* **Second Set**: Skew = 0.48

Both datasets are positively skewed, though only slightly. This means that in both sets, there are more values slightly below the mean, with fewer high-end values extending to the right. However, the second set has a slightly stronger rightward skew, indicating more values clustered toward the lower end of its range compared to the first set.

**4. Kurtosis**

* **First Set**: Kurtosis = -0.51
* **Second Set**: Kurtosis = -0.22

Both sets have negative kurtosis values, meaning they are both platykurtic, or flatter than a normal distribution. This indicates fewer extreme values and a broader, less peaked distribution in both datasets. The first set has a slightly more negative kurtosis, suggesting it may be even flatter, with fewer high or low outliers than the second set.

**5. Scale Differences**

* The first dataset has values that range much higher (2.87 to 14.12) compared to the second (0.1 to 0.8). This scale difference suggests they may represent different measurements or units, and any relationship between them would need to account for this disparity.

**Summary of Relationship**

These two datasets show similar distribution shapes (both are slightly right-skewed and platykurtic), indicating a comparable general pattern of data distribution. However, they differ significantly in scale, with the first set showing much higher values and greater spread. If you’re looking for a correlation between them, it would be necessary to standardize or scale the values first to bring them onto comparable terms.

If the data sets represent measurements taken together or related contexts, it would also be helpful to plot them together or compute their correlation to see if they share any pattern on a comparable scale.

# Correlation

A correlation of **0.20908** between the crime rate and unemployment rate suggests a **weak positive relationship** between the two variables. Here's how to interpret this result:

**1. Strength of the Correlation:**

* The correlation coefficient ranges from **-1 to 1**:
  + A value of **0** indicates no relationship between the variables.
  + Positive values indicate a positive relationship, where an increase in one variable is associated with an increase in the other.
  + Negative values indicate a negative relationship, where an increase in one variable is associated with a decrease in the other.
* A correlation of **0.20908** is **relatively low**, meaning that the crime rate and unemployment rate are weakly related. There is a slight positive trend, but the relationship is not strong enough to be considered significant or predictive on its own.

**2. Interpretation of the Positive Correlation:**

* The positive value indicates that **as the unemployment rate increases, the crime rate tends to increase slightly**, or vice versa. However, this is a **weak** correlation, meaning the relationship is not strong enough to confidently say that one causes the other.
* A weak positive correlation like this means that there may be other factors influencing both variables, and the relationship between the two is not very pronounced.

**3. Implications:**

* The correlation of **0.20908** tells you that there might be a slight trend, but the connection between unemployment and crime is not strong. It's likely that other factors (social, economic, environmental) are influencing these variables, and the relationship between them is not strong enough to draw significant conclusions.
* If you're using this correlation to make predictions or inform policies, you'd want to be cautious about relying too heavily on this relationship, as the data suggests it's weak.

**4. Practical Context:**

* In real-world situations, the correlation between crime and unemployment might be influenced by a variety of other factors such as:
  + Economic conditions (beyond unemployment alone)
  + Social services or community support
  + Crime prevention programs
  + Education levels
  + Family dynamics, etc.
* Therefore, while there is a slight positive correlation, the weak strength suggests that other factors may play a much larger role in determining crime rates than just unemployment.

**Conclusion:**

The correlation of **0.20908** tells you that there is a weak positive relationship between the crime rate and the unemployment rate. However, the correlation is not strong enough to suggest a meaningful or reliable connection, and it's important to consider other potential factors that could be influencing both variables.

# Other Notes

Each state varies from year to year. Each varies more in the area of unemployment. There is less variation in Crime rate. States that lead in unemployment one year will not lead in other years. The same is true with crime. Note authors say there is more property crime caused by unemployment than there is violent crime. Each study seems to vary by their interpretation of data, since that data is from different years.