**Descriptive Statistics of Income Levels by Gender: A Comparative Analysis**

**Stefani D.W. Yates**

**RES814-2404C-02**

**Dr. Richard Miller**

**Colorado Technical University**

**Descriptive Statistics of Income Levels by Gender: A Comparative Analysis**

This paper analyzes income levels among males and females using the SPSS Explore command. The analysis computes and compares descriptive statistics such as mean, median, quartiles, range, variance, standard deviation, skewness, and kurtosis for both groups. Additionally, several plots are generated to visualize income distribution across genders. The analysis aims to understand the differences in income distribution between males and females and to interpret the statistical measures concerning the normal distribution.

**Methods**

The dataset used in this analysis is from the General Social Survey (GSS), with the variable *rincdol* representing income and *sex* representing gender. SPSS was utilized to compute descriptive statistics and generate visualizations, including histograms, box plots, and stem-and-leaf plots.

**Results**

**Descriptive Statistics**

The income distribution for males and females was analyzed and the following statistics were obtained:

* **Males:**
  + **Mean Income:** $34,603.75
  + **Median Income:** $32,500.00
  + **Quartiles:**
    - Q1 (25th percentile): $16,250.00
    - Q3 (75th percentile): $45,000.00
  + **Range:** $1,095,000
  + **Variance:** 5,929,565,578.8
  + **Standard Deviation:** $24,350.70
  + **Skewness:** 1.123 (indicating a right skew)
  + **Kurtosis:** 1.276 (indicating a slightly peaked distribution)
* **Females:**
  + **Mean Income:** $27,729.70
  + **Median Income:** $21,250.00
  + **Quartiles:**
    - Q1 (25th percentile): $11,250.00
    - Q3 (75th percentile): $37,500.00
  + **Range:** $1,095,000
  + **Variance:** 5,438,978,874.0
  + **Standard Deviation:** $23,321.62
  + **Skewness:** 1.425 (indicating a more pronounced right skew than males)
  + **Kurtosis:** 2.234 (indicating a more peaked distribution than males)

**Skewness and Kurtosis**

Skewness measures the asymmetry of the distribution, where positive skewness indicates a right-tailed distribution, often caused by outliers in the higher range. On the other hand, Kurtosis measures the "tailedness" of the distribution, with higher values indicating more extreme outliers or a sharper peak than a normal distribution (Seijas-Macías et al., 2021).

Both male and female income distributions exhibit positive skewness, indicating that many individuals earn less than the mean, with a small number of high-income earners pulling the distribution to the right. The skewness is more pronounced in the female income distribution, suggesting greater inequality or a more significant number of low-income earners than males.

Both skewness and kurtosis provide insight into the underlying distribution of data. Positive skewness suggests that most income values are lower, with a few individuals earning significantly more, creating a long tail. Leptokurtic distributions with higher kurtosis indicate a distribution with more frequent extreme values, as seen in income distributions with substantial disparities between lower and higher earners (Jammalamadaka et al., 2020).

The mean income in a positively skewed distribution is higher than the median because the extreme values pull the mean upwards. In such cases, the median often serves as a better measure of central tendency because it is less influenced by outliers (Yu & Yin, 2023).

Both skewness and kurtosis are critical for understanding income distribution across genders, particularly in identifying the impact of extreme values on summary statistics. The analysis supports the conclusion that the median is a more reliable measure of central tendency in skewed distributions, reflecting the typical income level more accurately than the mean.

Both male and female income distributions exhibit positive skewness, indicating right-skewed distributions. This suggests that a more significant proportion of individuals in both groups earn below the mean income, with a tail extending toward higher income values (Cain et al., 2021). The female income distribution shows a more pronounced right skew (1.425) than males (1.123).

The kurtosis values for both distributions are positive, indicating leptokurtic distributions. This means both distributions have heavier tails and sharper peaks than normal distributions (Yadav & Singh, 2021). The female income distribution has a higher kurtosis (2.234) than the male distribution (1.276), suggesting more extreme values in the female group.

The kurtosis values for both distributions are above zero, indicating leptokurtic distributions. This suggests that both distributions have fatter tails and a sharper peak than a normal distribution, meaning that extreme values (outliers) are more common than expected under a normal distribution. The female income distribution, with a higher kurtosis value, indicates more pronounced outliers or extreme income values.

**Stem-and-Leaf Plots**

Stem-and-leaf plots are a sophisticated data visualization method that combines the graphical nature of histograms with the precision of raw data. These plots are particularly compelling for smaller datasets, offering a unique balance between data representation and distribution visualization (Boccuzzo & Salmaso, 2022).

The structure of a stem-and-leaf plot is composed of two main elements: the "stem," which represents the leading digit(s) of each value, and the "leaf," which denotes the final digit of each value. Each row in the plot corresponds to a range of values sharing the same stem. To interpret the plot, one reads from left to right and then top to bottom, with each number formed by combining the stem (left of the vertical line) with a leaf (right of the line).

The utility of stem-and-leaf plots in data analysis is multifaceted. They provide insights into the distribution shape (symmetric, skewed, etc.), while the row length indicates the frequency of values within a specific range. Outliers are readily identifiable at the extremes of the plot.

In the context of income analysis, stem-and-leaf plots offer several advantages. They provide a visual representation of income distribution, complementing numerical statistics. They facilitate easy identification of the most common income ranges for each gender and aid in detecting outliers, which is crucial given this dataset's high skewness and kurtosis values. Furthermore, they enable detailed comparisons between male and female income distributions.

When analyzing stem-and-leaf plots, it's essential to focus on various aspects. The concentration of leaves indicates the most common income ranges. The rate at which the number of leaves decreases moving up the stems suggests right-skew. Isolated leaves far above the main data body may represent potential high-income outliers. Comparative patterns between male and female plots can reveal gender-based differences in income distribution.

The detailed perspective offered by stem-and-leaf plots enriches the analysis by reinforcing and adding nuance to conclusions drawn from other statistical measures and visualizations. This comprehensive approach ensures a thorough understanding of the income distribution patterns across genders (Boccuzzo & Salmaso, 2022).

**Mean vs. Median**

For males, the mean income ($34,603.75) is higher than the median income ($32,500.00). Similarly, for females, the mean income ($27,729.70) is higher than the median income ($21,250.00). This difference between mean and median is consistent with the right-skewed nature of both distributions (Yue et al., 2022).

The mean income is higher than the median for males and females, consistent with the positive skewness observed. In skewed distributions, the mean is more affected by extreme values than the median, making the median a better measure of central tendency for these income distributions.

Given the skewed nature of both distributions, the median is a better measure of central tendency than the mean. The median is less affected by extreme values or outliers, which are present in both distributions, as evidenced by the skewness and kurtosis values (Gómez-Cruz et al., 2023).

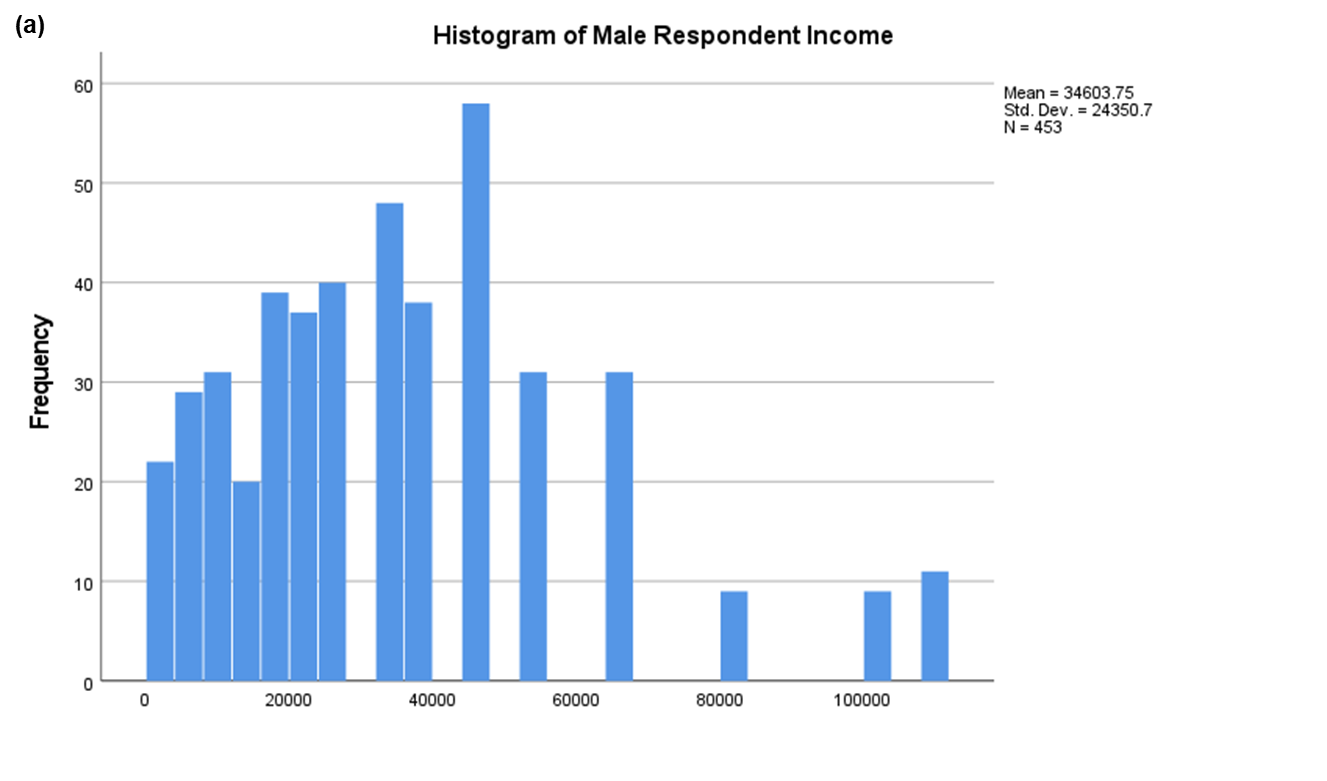
**Impact of Extreme Scores**

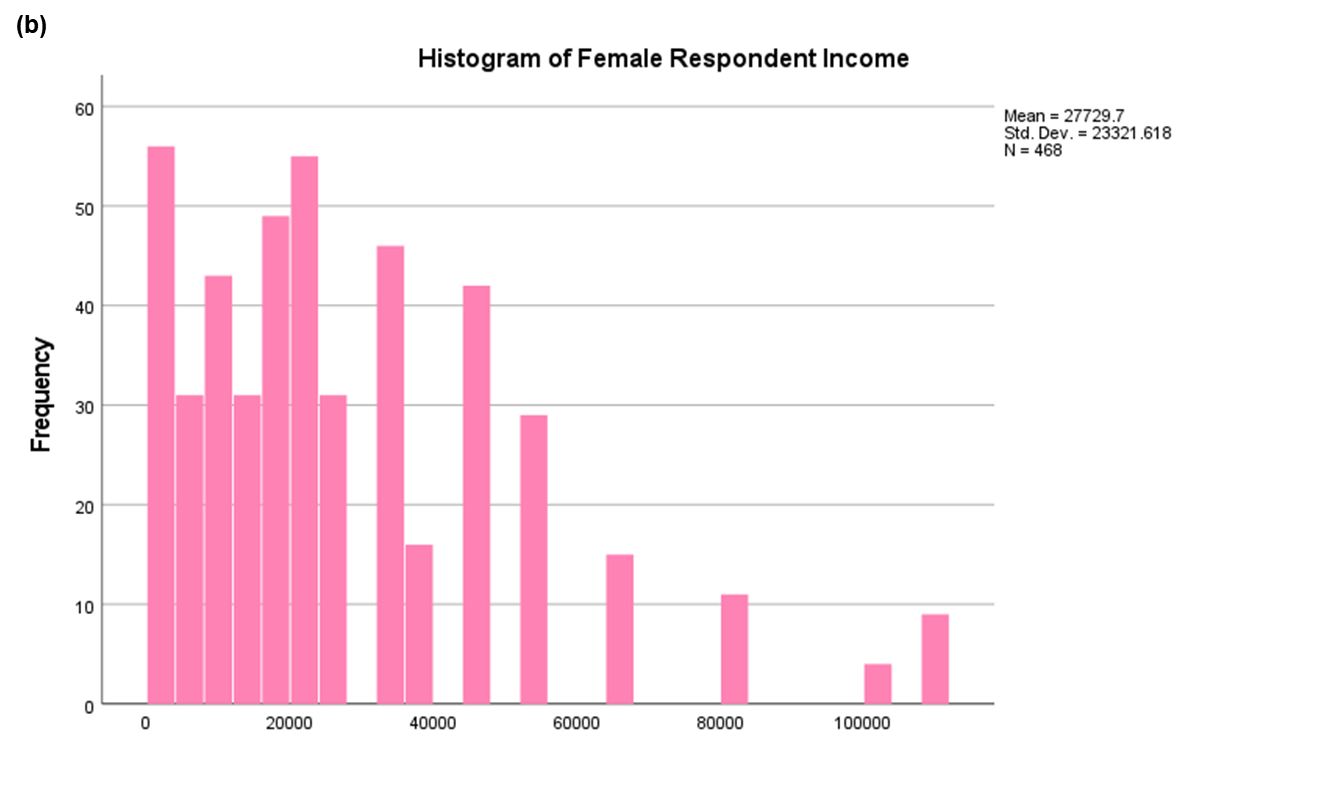
Extreme scores, particularly high-income outliers, have significantly impacted the mean values for both genders. These outliers pull the mean higher than the median, especially in the female distribution, where the difference between the mean and median is more pronounced. This effect is more pronounced among females, as indicated by their higher skewness and kurtosis values (Seo & Kim, 2022). The presence of extreme values suggests that the median is a more robust and representative measure of central tendency for these distributions.

**Visualizations**

The histograms, box plots, and stem-and-leaf plots created for both genders confirm the statistical findings. The histograms show right-skewed distributions for both male and female income data. The box plots highlight the presence of outliers and the central tendency for both groups, while the stem-and-leaf plots provide a detailed view of the data distribution.

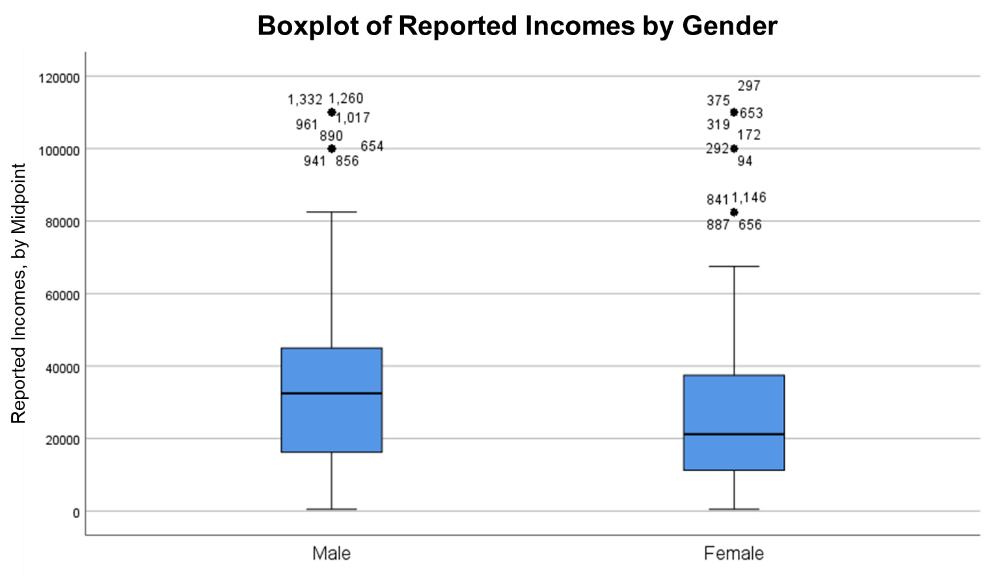
The histograms in Figures 1(a) and 1(b) illustrate that the income distribution for both males and females is right-skewed. This visual representation aligns with the positive skewness values calculated in the descriptive statistics, showing a concentration of incomes at lower values with a long tail extending towards higher incomes for both genders.





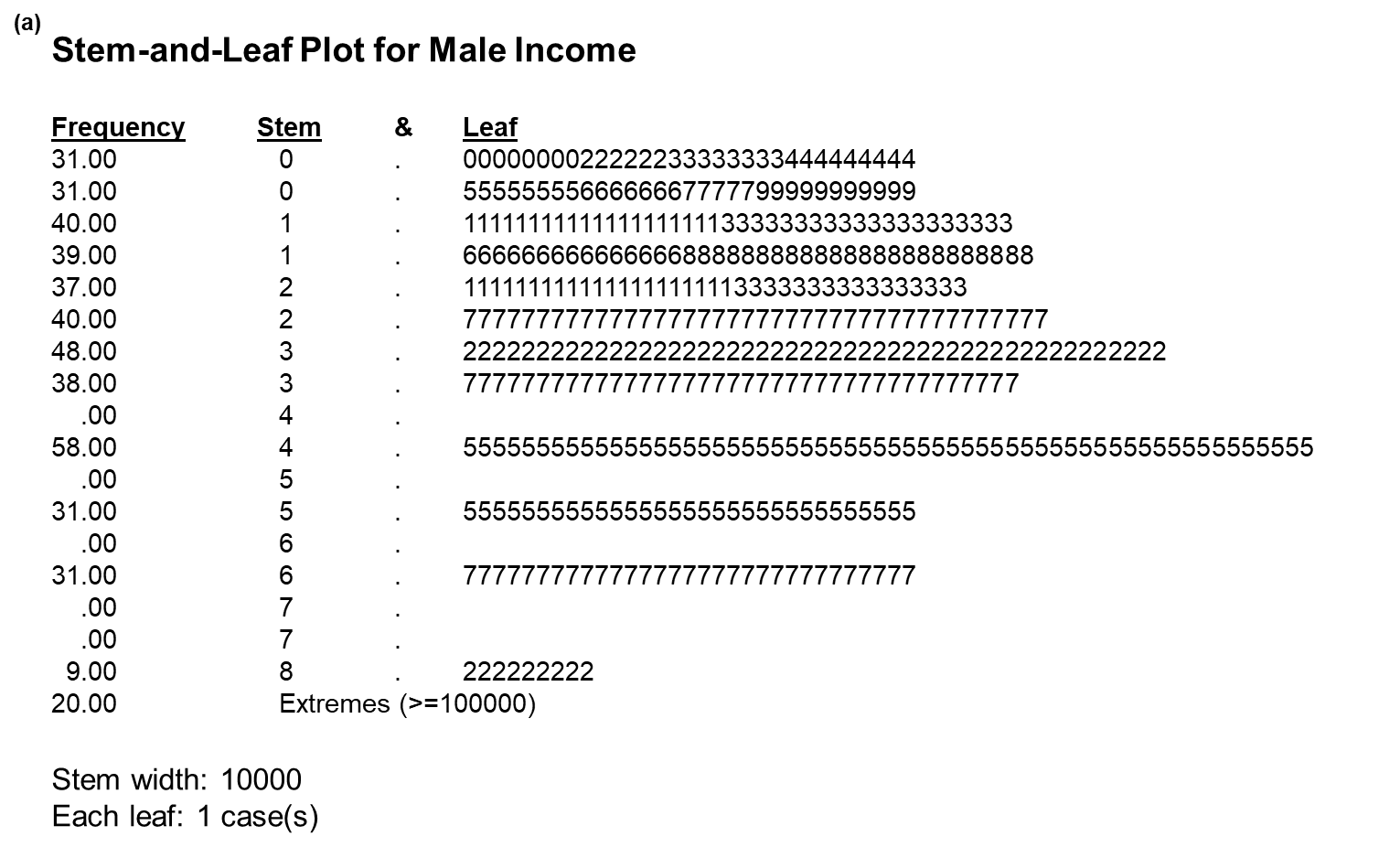
**Figure 1. Histograms of Frequency of Midpoint Income based on Sex. (a).** Male reported income. **(b).** Female reported income. The individual mean, standard deviation, and n (the total number of valid responses for that group) are given in both plots.

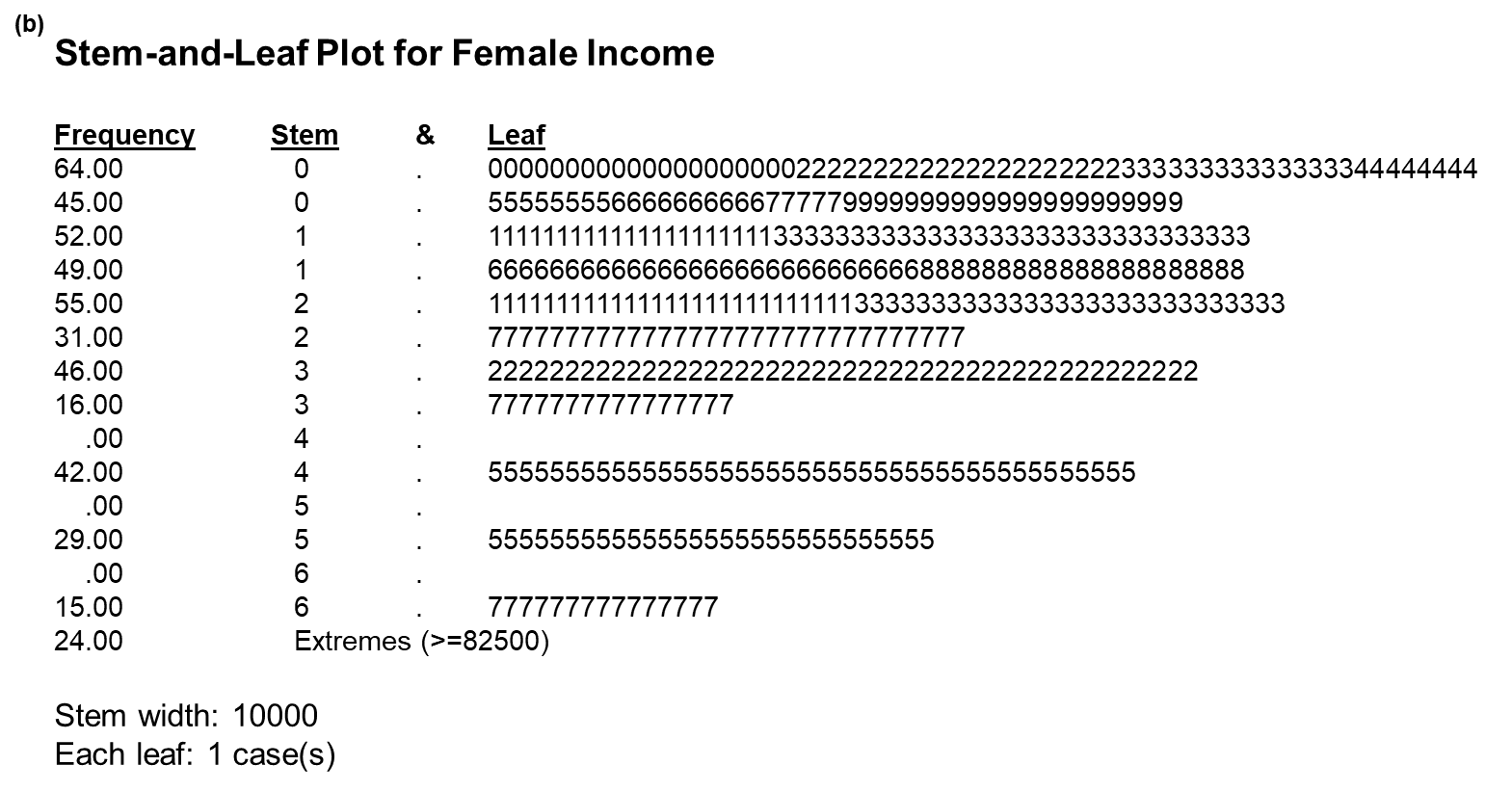
The box plot, shown in Figure 2, effectively displays the median, quartiles, and outliers for each gender. These plots visually confirm the higher median income for males compared to females and the presence of numerous high-income outliers in both groups. The longer upper whisker and the outliers' position illustrate the distributions' right-skewed nature.



**Figure 2. Box-plot of income reported by each gender.** Male reported income. This plot shows that males have a higher reported income overall, but females had more outliers for their group.

Figures 3(a) and 3(b), the stem-and-leaf plots, offer a more granular view of the income data. They allow for a detailed examination of the distribution, showing the frequency of incomes at various levels and making it easier to identify clusters and gaps in the income ranges for both genders.





**Figure 3. Stem-and-leaf plots of income for each gender. (a)** Male reported income, and **(b)** Female reported income. This plot shows that males have a higher reported income overall, but females had more outliers for their group.

Together, these visualizations provide a comprehensive picture of the income distributions, reinforcing the statistical findings and offering intuitive representations of the gender-based income disparities and distribution characteristics.

**Discussion**

The analysis reveals that the income distribution for males and females is positively skewed and leptokurtic, indicating a small number of high-income earners and more extreme values than expected under a normal distribution. The median income, less affected by extreme values, better represents the central tendency in these distributions than the mean.

These findings suggest income inequality within both genders, with females showing more significant skewness and kurtosis, pointing to a more considerable disparity in income distribution. The visualizations complement the statistical analysis, providing clear insights into the data distribution and supporting conclusions.

**Conclusion**

The comparison of income distributions between males and females shows significant differences in skewness and kurtosis, highlighting income inequality and the impact of extreme values on the mean. Both distributions are right-skewed and leptokurtic, with females showing more pronounced skewness and kurtosis. The median income for males is higher than for females, indicating a potential gender pay gap. The presence of high-income outliers in both groups, but more so in the female group, suggests that while some individuals of both genders achieve high incomes, this is less common and more extreme for females.

Given the skewed nature of the distributions, the median serves as a more reliable measure of central tendency for both groups. Future research could investigate the factors contributing to these gender-based income disparities and explore potential interventions to promote income equality.

**References**

Boccuzzo, G., & Salmaso, L. (2022). A contribution to stem-and-leaf plots for univariate and bivariate data. Statistical Methods & Applications, 31(2), 431-453. <https://doi.org/10.1007/s10260-021-00601-6>

Cain, M. K., Zhang, Z., & Yuan, K. H. (2021). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. Behavior Research Methods, 53(1), 450-469. <https://doi.org/10.3758/s13428-020-01452-x>

Gómez-Cruz, N. A., García-Molina, M., & Lozano, S. (2023). Analyzing economic inequality through generalized entropy measures. Journal of Economic Inequality, 21(1), 93-116. <https://doi.org/10.1007/s10888-022-09539-5>

Jammalamadaka, S. R., Taufer, E., & Terdik, G. (2020). On multivariate skewness and kurtosis. *Sankhya A, 83*(4), 607-644. <https://doi.org/10.1007/s13171-020-00211-6>

Seijas-Macías, A., Oliveira, A., & Oliveira, T. A. (2021). The skewness and kurtosis of the product of two normally distributed random variables. *Communications in Statistics - Theory and Methods, 52*(1), 80-93. <https://doi.org/10.1080/03610926.2021.1909734>

Seo, S., & Kim, J. (2022). The effect of outliers on the performance of normality tests. Communications in Statistics - Simulation and Computation, 51(6), 3358-3373. <https://doi.org/10.1080/03610918.2019.1699114>

Yadav, R., & Singh, S. K. (2021). Assessment of univariate normality: A critical review. Journal of Statistics and Management Systems, 24(2), 377-403. <https://doi.org/10.1080/09720510.2020.1833139>

Yu, X., & Yin, C. (2023). Some results on multivariate measures of elliptical and skew-elliptical distributions: Higher-order moments, skewness, and kurtosis. *AIMS Mathematics*. <https://doi.org/10.3934/math.2023370>

Yue, S., Yang, S. Y., & Zeng, Y. (2022). Nonnormality and asymmetry of hydrometeorological data: Causes, impacts, and remedies. Journal of Hydrometeorology, 23(3), 495-517. <https://doi.org/10.1175/JHM-D-21-0088.1>