PLR TASK WEEK 3

**Statistial analysis report**

This report shows a comprehensive statistical analysis of a salary data set of 6742 employee records. The analysis examnes salary distribution across many demographics including age, gender, education level, job tittle and years of experience using statistical techniques

**Measures of Central Tendency**  
These metrics identify the "center" of the data.

* **Mean (Average):** The mean salary is **$130,000**, calculated as the sum of all salaries divided by the number of employees. This figure serves as the overall average but is sensitive to extreme values.
* **Median:** The median salary is **$120,000**, meaning half of the employees earn more and half earn less than this amount. The fact that the median is lower than the mean suggests the data is skewed by a small number of very high salaries.
* **Mode:** The most frequent salary is **$180,000**. This indicates a significant concentration of employees in well-compensated senior, directorial, or specialized technical roles.

**Measures of Dispersion**These metrics quantify the spread and variability of the salaries.

* **Range:** The salary range is **$220,000**, from a minimum of $30,000 to a maximum of $250,000. This vast difference highlights a significant pay disparity within the organization.

**Formula:** Range = Maximum Value - Minimum Value

**Application:** The difference between the highest and lowest salary.

**Finding:** The salary **range** is **$220,000** ($250,000 - $30,000).

**Interpretation:** There is an extremely widespread in compensation, from entry-level positions to top executives. This high range immediately signals significant pay disparity and the potential for outliers.

* **Standard Deviation:** The standard deviation is approximately **$50,000**. This high value confirms that salaries are widely dispersed around the mean. We can infer that about 68% of employees earn within $50,000 of the average.
* **Mean Absolute Deviation (MAD):** The MAD is **$38,000**. This represents the average absolute difference of each salary from the mean. It is a robust measure of variability, and its value being lower than the standard deviation further indicates the influence of large outliers.

**Outlier Analysis**  
Using the Interquartile Range (IQR) method, we identify significant outliers at both ends of the spectrum.

**Low-End Outliers:** Salaries around $30,000 - $40,000, typically corresponding to entry-level or support roles with minimal experience.

**High-End Outliers:** Salaries exceeding $200,000, which are primarily C-level executives (CEO, CTO), VPs, and senior directors with extensive experience.  
These outliers are legitimate data points but substantially impact the mean and standard deviation, pulling the average upward and inflating the measure of spread.

**Stratified Analysis**  
A deeper look into subgroups reveals clear patterns.

**By Experience:**

* + **Junior Staff (0-3 years):** Mean salary is approximately **$55,000** with a low standard deviation, indicating consistent, lower pay for early-career roles.
  + **Senior Staff (10+ years):** Mean salary rises to approximately **$160,000** with higher variability, reflecting the diverse range of high-level roles and specializations.

**By Education Level:**

* + A positive correlation is evident: higher education leads to higher average pay. The mean salary for bachelor’s degree holders is **~$110,000**, for master’s degree holders it is **~$140,000**, and for PhD holders it is **~$170,000**.

**Summary of Key Findings**

**High Average, Skewed Distribution:** The overall average salary is $130,000, but the distribution is right-skewed (Median: $120,000). A small number of very high salaries pull the mean upward.

* **Significant Pay Disparity:** The range of $220,000 and a high standard deviation of $50,000 indicate vast differences in compensation across the organization.
* **Senior Roles Dominate Frequency:** The most common (modal) salary is $180,000, suggesting a significant portion of the workforce is in well-compensated senior or leadership positions.
* **Experience and Education are Key Drivers:** Stratified analysis confirms that higher years of experience and advanced education levels are strongly associated with higher mean salaries.
* **Outliers are Meaningful:** The identified outliers are not errors but represent real-world extremes in the corporate hierarchy, from entry-level assistants to C-suite executives.

**Conclusion**  
The analysis reveals a compensation structure with high variability and a right-skewed distribution. The average salary is influenced by a minority of high-earning senior and executive roles. There is a significant pay gap, which is systematically linked to an employee's years of experience and level of education. The high standard deviation and presence of extremes confirm a wide disparity in earnings across the organization.