**In-Class Exercise: Descriptive Statistics Review**

**Part 1: Median, Average, and Outliers**

Consider the following are the incomes of 22 households from two neighborhoods.

|  |  |  |
| --- | --- | --- |
| Neighborhood A |  | Neighborhood B |
| $22,000  |  | $32,000  |
| $30,000  |  | $33,000  |
| $35,000  |  | $35,000  |
| $38,000  |  | $36,000  |
| $40,000  |  | $40,000  |
| $42,000  |  | $42,000  |
| $55,000  |  | $45,000  |
| $62,000  |  | $60,000  |
| $65,000  |  | $70,000  |
| $250,000  |  | $74,000  |
| $350,000 |  | $75,000 |

1. What is the average income of each neighborhood? Which neighborhood is higher?

**A is higher: $89,909 vs. $49,273**
2. What is the median income of each neighborhood?

**$42,000 for both**
3. Now remove the top two incomes from each group. Excluding those households, which neighborhood has the highest average income?

**Neighborhood B: $43,667 vs. $43,222**
4. If you include all the data, which measure (median or average) better explains the income distribution of the two neighborhoods? Why?

**Median is better because the two outliers in neighborhood A skew the results**

**Part 2: Interpreting a Histogram**

**The following is a histogram of salaries for the 483 players in the NBA along with some summary statistics, all provided by R.
*Source: draftexpress.com*

Mean: 4142000

Std.Dev = 4610687

N=483

1. What is the mean player salary?

**$4,142,000**
2. Do most players make more or less than the mean? Explain.

**Most make less than the mean. You know this because the largest bars are to the left of the peak of the normal curve.**
3. Are player salaries normally distributed? Explain.

**No, they aren’t. The histogram doesn’t follow the shape of the normal curve.**
4. What do you learn about player salaries based on the standard deviation being greater than the mean?

**This means there probably are a lot of outliers (a lot of players making low salaries and some making really high salaries).**

**Part 3: Interpreting Statistical Tests**

The following is the output of statistically testing the average NBA salaries for point guards versus shooting guards *(Source: draftexpress.com)*:

Point Guards: $4,076,414.56 (n=110)
Shooting Guards: $4,158,783.82 (n=115)
*F-value = 0.017, p-value = 0.895*

From this, do you conclude that the two player groups have a statistically significant difference in their salaries? Why?

**No, because the p-value is much larger than 0.05 (or even 0.01).**

Now let’s look at the output of statistically testing the difference in average salary of the 50 highest paid basketball players versus the 50 highest paid baseball players
*(Source: draftexpress.com and newsday.com):*

Baseball: $18,538,001.90 (n=50)
Basketball: $15,458,543.62 (n=50)
*F-value = 17.509, p-value = 0.000*

From this, do you conclude that the two sports have a statistically significant difference in their salaries? Why?

**Yes, because the p-value is much smaller than 0.05 (or even 0.01).**

**Part 3: Probability**

Consider flipping a “fair” coin (50% chance of heads, 50% chance of tails).

1. What’s the probability of getting “tails” two times in a row?

**0.5 \* 0.5 = 0.25 (or 25%)**
2. What’s the probability of getting “tails” three times in a row?

**0.5 \* 0.5 \* 0.5 = 0.125 (or 12.5%)**

Now imagine there is a bag with four red marbles and one green marble.

1. What’s the chance of drawing a red marble?

**4 out of 5 🡪 4/5 = 0.8 (or 80%)**
2. Let’s say you get that red marble. Now what’s the chance of drawing another red marble from the remaining marbles?

**3 out of 4 🡪 3/4 = 0.75 (or 75%)**