Stat 101
Summer 2023 Session 1
Dr. Lily Yen

Activity2-1

Show all your work

Name: _____
Number: _____
Signature: _____
Score: ___/8

Excel spreadsheet functions allowed

Problem 1: From the spreadsheet acs12, a sample of 2012 US census, how many column headings are there?

13

Problem 2: What is the sample size?

2000

Problem 3: How many of the subjects are female?

969

Problem 4: List all variables in the data set.

Variables: income, employment, hrs work, race, age, gender, citizen, time to work, lang, married, edu, disability, birth grtr.
Categorical variables: employment, race, gender, citizen, lang, edu, disability, birth grtr.
Numerical variables: income, hrs work, age, time to work.

Score: /4

Circle all categorical variables from your list.

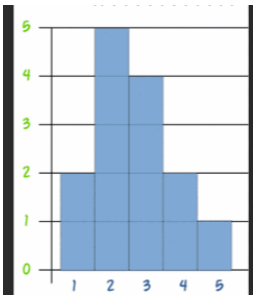
Problem 5: Take the income column on the spreadsheet and find the average income of all those making more than 100 annually.

42940.16

Use $=\text{COUNTIF}(A2:A2001, ">100")$ to find the number of cells under *income* which contains entries greater than 100. The result is 892.
Next, we need to sum the cells under the condition >100, namely,
$=\text{SUMIF}(A2:A2001, ">100")$ to get 38302620. Finally, the average income is $38302620/892 \approx 42940.16$

Score: /1

Problem 6: For the given histogram, answer the following questions: a) Sample size n, b) mean, c) median, d) mode, e) minimum, f) maximum, g) standard deviation. Draw a box plot with clearly labelled axis, and all five values.



Sample size n is the sum of the frequency: $2+5+4+2+1=14=n$. Mean $\overline{x}=\frac{2\times1+5\times2+4\times3+2\times4+1\times5}{14}\approx 2.64$. Median $(2+3)/2=2.5$, the average of the seventh and the eighth entries. Mode is the most frequent entry: 2. Minimum: 1. Maximum: 5. Standard deviation: take the square root of variance to get 7.257054 rounded to 7.3.
The boxplot needs five numbers: minimum, first quartile, median, third quartile, and maximum. We have three of them now. The first quartile is the median of the first seven values: 2. The third quartile is the median of the last seven values: 3. Draw a horizontal axis with marked values.

Score: /3