Fall, 2021 **Name:** _____
APSTA-GE.2001
Midterm Exam

Instructions

1. This exam becomes available Tuesday, October 12, 2021, at 6pm NYC/Eastern time, and it will close on Thursday, October 14, 2021, at 6pm NYC/Eastern time.

2. You will upload your response under the "Assignments" tab, just as you do your weekly assignments.

3. You must complete the exam on your own, but you may use your textbook, course notes, PowerPoint slides, and videos.

4. Most of the questions for this exam are based on the Stata dataset "Days Skipped Lunch.dta" – this dataset can be found in the same location as the test document.

5. For the questions requiring you to draw something, you can either draw it by hand and scan it in to your exam, or you can use the Stata .dta file and create the output or graphic in Stata and paste that in to your exam. If you use Stata, copy and paste your command codes as well (e.g., "regress days_skipped male, beta" if running a regression predicting the number of days skipping a meal by the student's gender and requiring the beta weight option).

6. Save a file with your answers and upload it to the classes site.

In a study on eating habits of college sophomores, 40 students (20 males and 20 females) were asked to complete a survey that included a number of relevant questions.

These data can be found in the "Days Skipped Lunch.dta" file.

1. (51 pts) One of the questions asked students to indicate the number of days each skipped lunch within a given, specified 20-day period during their sophomore year. The data are below. The male students' data points are listed in bold.

6, 4, 13, 17, **7**, 9, **4**, **4**, **13**, **8**, 9, **4**, 12, **11**, 12, 18, **6**, **4**, 12, 10, **4**, **4**, **5**, **12**, 10, **4**, **8**, **8**, **6**, **4**, 15, 4, 20, 16, 6, 11, 15, 11, 11, **7**

(a) (5 pts.) Construct a stem-and-leaf plot with frequency distribution, using the template below. (Notice that each interval is divided into fifths – two leaves per stem -- * is the stem for 0 & 1;**t** is for 2 & 3;**f** is for four & five; **s** is for six & seven; **.** is for 8 & 9)

Stem & Leaf

0f	|
0s	|
0.	|
1*	|
1t	|
1f	|
1s	|
1.	|
2*	|

(b) (5 pts.) What is the mean (rounded to the nearest tenth)?

(c) (3 pts.) What is the median?

(d) (2 pts.) What is the mode?

(e) (3 pts.) What is Q1?

(f) (3 pts.) What is Q3?

(g) (3 pts.) What is the IQR? Interpret what it means.

(h) (5 pts.) Draw a boxplot with the five values (Q1, Q2, Q3, minimum and maximum) labeled on the graph either on the vertical axis or alongside the boxplot itself.

(i) (2 pts.) Describe the shape of the distribution (circle your response below):

reasonably symmetric positively skewed negatively skewed

(j) (6 pts.) If the time period were doubled from 20 days to 40 days, and everyone's eating habits remained the same, what would be the new mean, median, and mode based on this new 40-day time period?
Note: By "eating habits remained the same" over the next 20 days, that means someone who skipped 2 meals in the first 20 days would skip another 2 meals over the next 20 days, for a new total meals skipped for that person of 4 (over the 40 days).

New Mean: _____ New Median: _____ New Mode: _____

(k) (2 pts.) If the standard deviation based on the 20-day time period equals 4.5, what would the new standard deviation be based on the new 40-day time period?

New Standard Deviation: _____

(l) (2 pts.) If you were to draw a boxplot to reflect this new 40-day time period, the shape of the new boxplot relative to the original boxplot would be (circle your response below):

the same more symmetric more positively skewed more negatively skewed

(m) (4 pts.) Briefly explain why you selected the answer you did in question (l) above.

(n) (6 pts.) Given that the mean and standard deviation of the number of days out of 20 days that a student skipped lunch by gender are the following:

FEMALES: MEAN = 11.75; SD = 4.3 ; MALES: MEAN = 6.5; SD = 2.8.

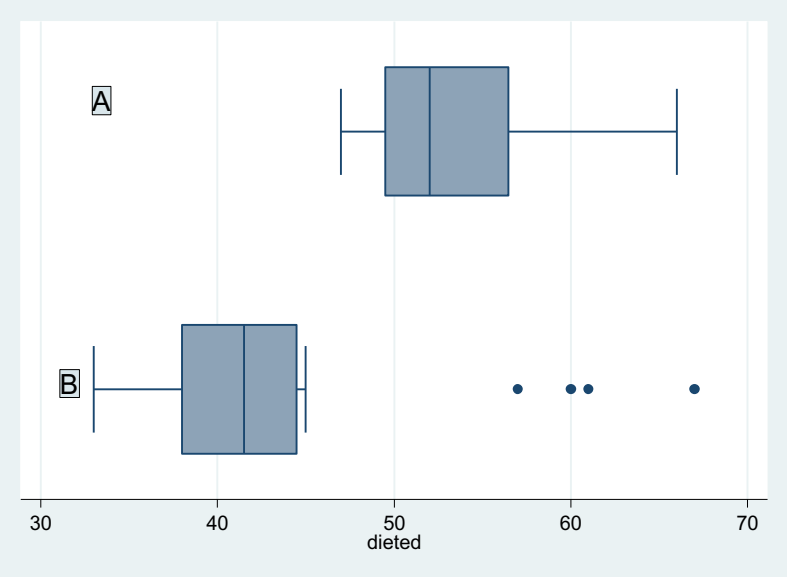
If John and Mary each skipped lunch 9 days, would you say that *relative to their gender cohorts*, John and Mary skipped lunch to the same extent? Why or why not? Be specific and show your work. Interpret your results.

2. (12 pts) -A different question asked students to report the proportion of days they dieted during sophomore year. The data, by gender, are given in stem-and-leaf form below.

```
-> male = 0 (females)

Stem-and-leaf plot for dieted

  4. | 77999
  5* | 01112244
  5. | 56779
  6* | 2
  6. | 6
```

```
-> male = 1 (males)

Stem-and-leaf plot for dieted

  3* | 33
  3. | 56889
  4* | 01122234
  4. | 5
  5* |
  5. | 7
  6* | 01
  6. | 7
```

(a) (2 pts.) Given these data, which boxplot represents the Males (Underline One)? **A** **B**



(b) (2 pts.) What is the median for Distribution B. _____

(c) (2 pts.) What is the most appropriate estimate of the mean for Distribution B (Select one).

33% 37% 44% 67%

(d) (6 pts.) Draw a histogram of the data for Distribution A (label the axes).

3. (15 pts. – part h is worth 1 pt.; all others are worth 2 pts.) A regression equation was computed to see to what extent the number of days a person skipped lunch was related to his/her gender. If males are coded 1 and females coded 0 for these data, we obtain the following equation and related output.

stats	skipped	male
N	40	40
mean	9.1	0.50
sd	4.49	0.51

```
      Source |       SS           df       MS      Number of obs   =        40
-------------+----------------------------------   F(1, 38)        =     21.15
       Model |       280.9         1       280.9   Prob > F        =    0.0000
    Residual |       504.7        38  13.2815789   R-squared       =    0.3576
-------------+----------------------------------   Adj R-squared   =    0.3407
       Total |       785.6        39  20.1435897   Root MSE        =    3.6444

------------------------------------------------------------------------------
days_skipped |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
        male |       -5.3   1.152457    -4.60   0.000               -.5979641
       _cons |      11.75   .8149104    14.42   0.000                       .
------------------------------------------------------------------------------
```

(a) Explain why the mean of the variable male equals .50? _____

(b) Write out the unstandardized regression equation: _____

(c) Use the equation to predict the number of days a male skips lunch: _____

(d) Use the equation to predict the number of days a female skips lunch. _____

(e) What is the difference between these two predicted values? _____

(f) Based on this result, how would you interpret, more generally, the value of a b-weight in a regression equation with a single dichotomous predictor?

_____

_____

(g) Why is the sign of the gender b-weight negative in this case? _____

_____

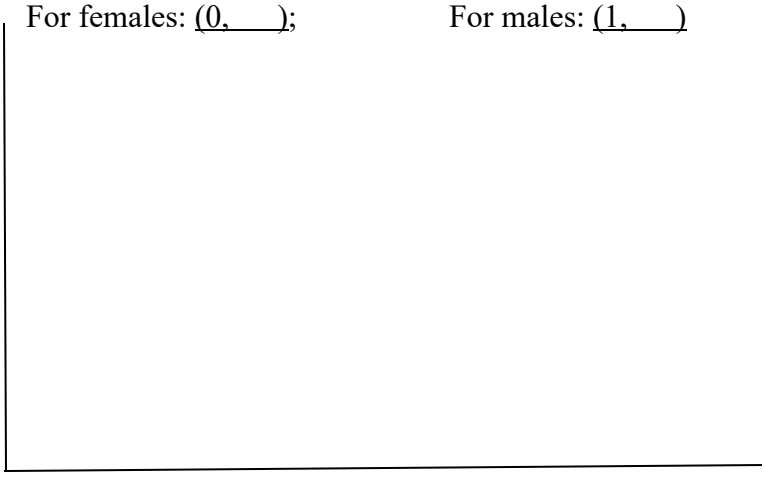
_____

(h) What proportion of the variance in days skipped lunch is associated with gender?

_____

4. (6 pts.) Using the data from the first page of the exam, draw a scatterplot with the variable skipped on the vertical axis and the variable male on the horizontal axis. Be sure to label the axes. Superimpose the regression line on the scatterplot, giving the values of the two points the regression line passes through for females and males, respectively.

For females: (0, ___); For males: (1, ___)



5. (4 pts.) Another question on this survey assesses Positive Body Image. It asks students to indicate their degree of agreement with the statement *I like how I look* by selecting one of the following:

1=strongly agree; 2=agree; 3=neither agree nor disagree; 4=disagree; 5=strongly disagree.

(a) (2 pts.) Write an equation to transform this variable called, I_LOOK, to a form that would make interpretation easier, so that a higher score represents more positive body image, rather than less.

(b) (2 pts.) Is this transformation an example of a linear or non-linear transformation? _____ Explain your answer.

6. (8 pts.) Classify each of the following variables as nominal, ordinal, interval, or ratio.

(a) Ranked preference for Vanilla, Chocolate, or Strawberry ice cream _____

(b) Type of diet followed (Atkins, South Beach, Weight-Watchers, Personally-Designed) _____

(c) Amount of weight lost (or gained) sophomore year, as measured between Sept and May. _____

(d) Energy level on the 1st day of school, as measured on a 100-point standardized scale. _____

7. (4 pts.) A recent NY Times article (published on September 19, 2013) with the headline, "Poverty Rate Up in City, and Income Gap is Wide, Census Data Show," reports the following:

"Citywide, the mean income of the lowest fifth was $8,993, while the highest fifth made $222,871."

Would the difference in income between the lowest and highest fifth be as dramatic if one were to report the medians of these two groups instead of the means? Explain why or why not.