

Student name:

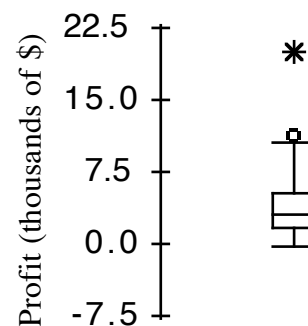
MATH 120: Elementary Statistics  
Spring 2022

Test 1  
March 2, 2022

**Instructions:**

- This is a regular “closed-book” test, and is to be taken without the use of notes, books, or other reference materials. Collaboration or group work is not permitted.
- Cell-phone usage of any kind is prohibited for the entire duration of the test. This also applies to any restroom breaks taken during the test.
- The time limit for taking this test is 50 minutes from the scheduled start time. Please turn in your test promptly when time is called to avoid late penalties.
- This test adds up to 50 points. It contains questions numbered 1 through 7.

1. [6 pts.] The boxplot shows the distribution of monthly profits (in thousands of \$) over a 3-year period at a retail store. Write a few sentences describing this distribution. Be sure to include the application context and units.



**Solution:** The distribution of monthly profits at this store is skewed right (i.e., towards the higher values). The median profits are around \$3000, and the interquartile range is about \$4000. There are two high outliers, one at \$10,000 and another at \$20,000. The shape may be unimodal, but it is not possible to tell from a boxplot.

Grade: 1 pt. each for correctly discussing skew, center, spread, outliers.  
1 pt. for complete and gramatically correct sentences.  
1 pt. for including application context & units.

2. [6 pts.] A school district sets office supply budgets for their high schools based on enrollment of students. The district budgets \$15, plus \$2 per enrolled student. Thus a class with one student gets \$17, and classes with 40 students get  $15+2(40) = \$95$ . One school in the district has the following enrollment statistics for their classes: mean=33.39 students, SD=5.66 students, median=32, IQR=10. Find the median and IQR of the office supply budget for their classes. Show calculation steps.

**Solution:** Office supply budget =  $\$15 + \$2 \times (\# \text{ of enrolled students})$   
The budget data is, essentially, a rescaling of the student enrollment data. Adding \$15 only affects the median, but not the IQR.  
Thus, median office supply budget =  $\$15 + \$2 \times (\text{median of enrolled students})$   
 $= \$15 + \$2 \times 32 = \$79$  (answer)  
IQR of office supply budget =  $\$2 \times (\text{IQR of enrolled students})$   
 $= \$2 \times 10 = \$20$  (answer)

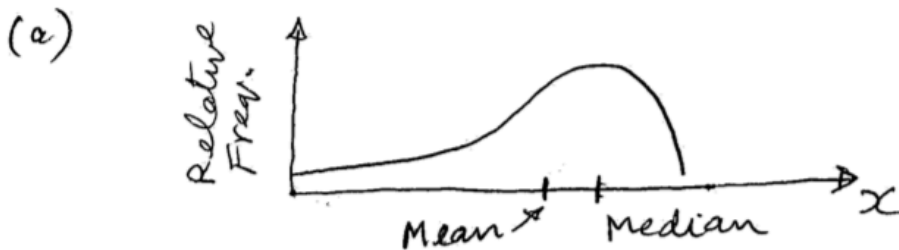
Grade: 2 pt. each for correct answer for median + IQR.  
2 pt. for correct reasons or steps.

3. [6 pts.] Sketch a rough (but qualitatively correct) histogram profile for each of the following types of quantitative variables:

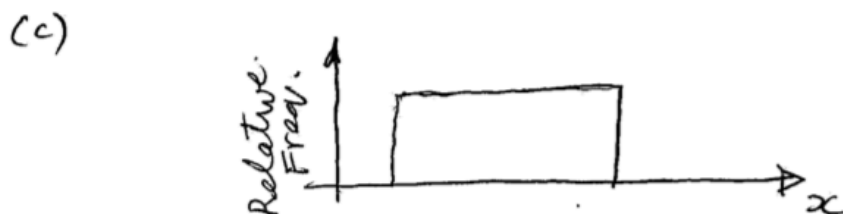
(a) The variable has mean less than its median.

(b) The variable is skewed strongly to the left.

(c) The variable has uniform distribution.



(b) The same sketch shown in (a) would work



Grade: (a)=(b)=(c)=2 pt. each for correct answer.

4. [8 pts.] The following statistics summarize the prices (in dollars) of 10 items in a customer's Ebay shopping cart:

Median	Mean	IQR	SD
25.8	32	8.4	10.9

Suppose the cheapest item, which costs \$12, is replaced by one that costs \$8. Compute, if possible, the new values of the above summary statistics. Show reasoning.

**Solution:**

- The median will not change. Reason: The median is the mid point of ordered data, and decreasing the minimum value will have no effect on it.
- The mean will change, since it depends on the total sum.  
Total before change =  $32 \times 10 = 320$ . Total after =  $320 - 12 + 8 = 316$ .  
New mean =  $316/10 = 31.6$ .
- The IQR will not change. Reason: Decreasing the minimum does not affect the middle 50% of ordered data.
- The SD will change, since the mean and the deviations change. But the new SD cannot be determined without more information.

Grade: 2 pt. for each of the 4 items – median + mean + IQR + SD.

For each of them, 1.5 pt for correct answer + 0.5 pt for reasons or steps.

5. [8 pts.] A research study explored the relationship between working memory and several other variables in a sample of subjects. Here are summary statistics for working memory capacity of the subjects, measured in the form of a (unitless) numerical score, and their age in years

	Mean	SD
Age (years)	55.6	6.3
WMC (no units)	9.5	3.2
Correlation coefficient, $r = -0.59$		

The researchers want to develop a model for predicting WMC from age. They have found the association between age and WMC is approximately linear via a scatter plot.

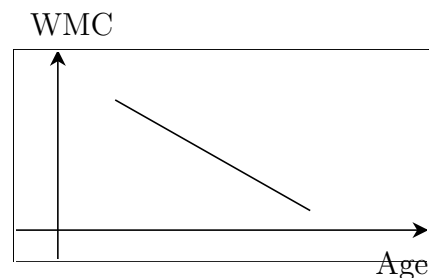
- Identify the explanatory and response variables (with correct units).
- Find the equation for the regression line.
- Explain the meaning of the slope in the context of this application.  
If you don't know how to solve (b), attempt (c) assuming the regression equation is:  $\hat{y} = -12.34 - 56.78x$ . Note, this is NOT the correct answer to (b).
- Find the  $R^2$  value for your model.

**Solution:**

- Explanatory variable = Age in years.  
Response variable = WMC (no units)

- Equation of the regression line:

$$\hat{y} = b_0 + b_1x, \text{ where } y = \text{WMC and } x = \text{Age (years)}$$



Information given in the problem:  $\bar{x} = 55.6, s_x = 6.3$  (years)

$$\bar{y} = 9.5, s_y = 3.2 \text{ (no units)}$$

The correlation is:  $r = -0.59$

The slope formula is:  $b_1 = r \frac{s_y}{s_x} = -0.59 \left( \frac{3.2}{6.3} \right) = -0.2997$  per year.

So, we have:  $\hat{y} = b_0 - 0.2997x$

To find  $b_0$ , plug in  $(\bar{x}, \bar{y})$  and solve:  $9.5 = b_0 - 0.2997 \times 55.6 \Rightarrow b_0 = 26.16$

Therefore, the regression model is:

$$\hat{y} = 26.16 - 0.2997x$$

OR

$$\widehat{\text{WMC}} = 26.16 - 0.2997 \times \text{Age in years}$$

- Meaning of slope: "For each year that a person ages, the working memory capacity is predicted to decrease by about 0.2997 units, on average."
- $R^2 = r^2 \times 100 = (-0.59)^2 \times 100 = 34.8\%$

Grade: (a)=2 pt, (b)=3.5 pt, (c)=1.5 pt, (d)=1 pt.

For (a); 1 pt. each for correct  $x, y$  variable + correct units.

For (b): 1pt=know/show correct process for  $b_0, b_1$  (e.g., formula for  $b_0$  + plugin  $\bar{x}, \bar{y}$ ).

1+1 pt=compute correct  $b_0$  and  $b_1$ .

0.5 pt=state final regression model correctly.

For (c): No partial credit.

For (d): 1 pt=show squaring of  $r$  and get answer. Answer in % or fraction is okay.

6. [8 pts.] A member of the City Council has proposed a resolution opposing the construction of a new state prison there. The council members decide they want to assess public opinion before they vote on the resolution. Below are some of the methods proposed to sample local residents to determine the level of public support for the resolution. Write the letter corresponding to the sampling strategy in the blank next to each plan.

<u>F.</u> i.	Call every 500th person in the phonebook.	A. convenience
<u>G.</u> ii.	Mail postcards to every address in the city, asking people to register their opinion via a toll-free phone number or by visiting a website.	B. stratified random
<u>B.</u> iii.	Randomly select 50 voters from each election district.	C. simple random
<u>E.</u> iv.	Randomly pick five city blocks, then randomly pick 50 residents from each block.	D. cluster
		E. multistage
		F. systematic
		G. voluntary response

Grade: 2 pt. for each of 4 correct answers.

For questions ii. and iv., half credit if the answer is A. and D., respectively.

7. [8 pts.] An orange-juice processing plant will accept a shipment of fruit only after carefully inspecting about 200 oranges selected from various locations in the truck. They estimate the proportion of these oranges showing signs of unsuitability for juice (bruised, rotten, unripe, etc.). If this proportion exceeds a threshold, the whole truckload is rejected. Identify the following as precisely as possible

\* The population:

All oranges in the truck.

\* The sample:

The 200 oranges selected for inspection.

\* The parameter of interest:

The true proportion of oranges in the entire truck unsuitable for juice.

\* The statistic:

The proportion of oranges in the sample of 200 that are unsuitable for juice.

Grade: 2 pt. for each of 4 correct answers.