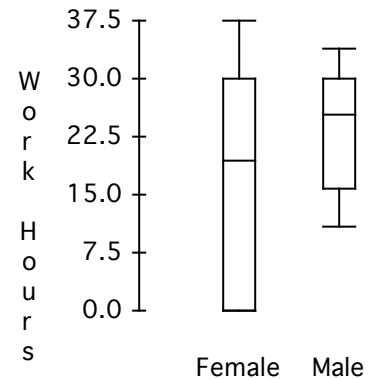


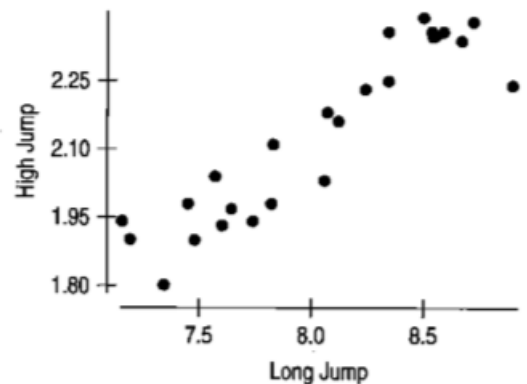
Test 1 practice worksheet

- A distribution of 100 data values is unimodal, with no outliers. Its mean=21 and median=32.
 - Is the distribution symmetric, skewed left, or skewed right? Why?
 - Suppose the largest value is increased by 20 units. Find the new mean and median.
- A study of employment trends among college students recorded each respondent's age, sex and number of hours worked per week. The boxplots show their results. Write a few sentences interpreting what the display tells us.



- Find the standard deviation of the following data values: $-2, 0, 4, 6$.
The intent here is to do it without the use of your calculator's statistics functions, and to show the needed computational steps.
- Two quantitative variables $V1$ and $V2$ are approximately linearly associated, with a correlation of $r = -0.9$. Assume both variables contain only positive values.
 - Sketch a qualitatively reasonable scatter plot showing the association.
 - Sketch another qualitatively reasonable scatter plot showing the association after standardizing $V1$ and $V2$.
- Identify the sampling technique proposed in each of the following situations:
 - To conduct an opinion poll about a proposed referendum on the ballot in Indiana, a random sample of 6 telephone prefixes (first 3 digits of the phone number) were selected and all households in those prefixes were called.
 - Quality control inspectors opened and tested every 40th bottle of soda coming off a production line at a factory.
 - Subscribers to the magazine *Consumers Digest* were assigned unique numbers. Then a sample of 300 subscribers were selected and invited to participate in a survey.
 - Producers of a proposed TV sitcom want to test the appeal of their show for people of different ages. They divide their population of interest into 3 different age groups, and select a random sample of 50 people from each.
- For each of the following variables, indicate the expected shape of the distribution (how many modes; type of skew; etc.):
 - household income in a geographic region

- (b) heights of singers in a co-ed choir
 (c) weights of adult female elephants of a particular species
 (d) number of times each face of a die shows up in 100 tosses
7. A children's research organization studied the television viewing habits of 822 children between the ages of 8 and 14 years. They found the average time these children spent in front of a television per year was about 1054 hours, with standard deviation 450 hours. Identify the following as precisely as possible: the population, the sample, the parameter(s), and the statistic(s).
8. A car dealer wants to model the association between age and price of used Ford Mustang cars. They collect data on a random sample of 50 cars, and find the scatterplot shows a negative and approximately linear association. A regression to predict price (in \$) from age (in years) has an R^2 of 69.2%.
- (a) Identify the explanatory and response variables (with correct units).
 (b) What units would the slope of the regression line have?
 (c) Explain what the value of R^2 means in this application context.
 (d) Find the value of the correlation coefficient r .
 (e) Suppose a certain car's age is 2 standard deviations below the mean age, what can we predict about its price?
9. The scatterplot shows winning long-jump and high-jump distances, in meters, for all Summer Olympic games between 1912 and 2008.
- (a) Describe the association in a couple of complete and grammatically correct sentences. Be sure to address the 3 key things you are supposed to observe in scatterplots.
- (b) The correlation in this scatterplot is 0.92. If we convert the jump lengths to centimeters by multiplying by 100, would the correlation increase or decrease? Reason?
- (c) Suppose we add a new data point to the scatterplot, with coordinates (8.0, 1.80). Would the correlation increase or decrease? Reason?
10. The 5-number summary of midterm scores for an Elementary Statistics class are



Min	Q1	Median	Q3	Max
17	35	43	45	48

- (a) Would you expect the mean score to be higher or lower than the median? Why?
 (b) Would you expect any outliers in this dataset? Justify your answer.

11. The SIPRI 2014 annual report on worldwide arms sales showed that Lockheed Martin (LM) was the world's largest seller of weapons and military equipment. A closer look at the data shows that two important variables affecting annual profits at LM are calendar year, and fluctuations in the number of armed conflicts occurring in the world.

Analysis of data from SIPRI for the last 10 years gave the correlations shown in the table. The variables are: Profit=LM's annual profits; Year=calendar year; and Conflicts=# of armed conflicts in that year.

	Profit	Year	Conflicts
Profit	1.000		
Year	0.623	1.000	
Conflicts	0.697	0.388	1.000

[SIPRI=Stockholm International Peace Research Institute]

- (a) Which variable seems to have the strongest effect on LM's profits? Give reason.
 (b) Which 2 variables are the weakest in their correlation with each other?
 (c) Suppose this year turns out to be particularly peaceful, and the number of worldwide armed conflicts is 2 SD below average. What can we predict about LM's profits?
 (d) State the key assumption made when interpreting correlations in this way.
12. Given the information in the previous question, suppose we want to build a linear regression model to predict Lockheed Martin's annual profits from the number of armed conflicts in the world. The following additional summary statistics are available for the 10-year data set that comprises our sample

	Mean	SD
# of armed conflicts	35.6	2.46
LM's annual profits(million \$)	2855	472.3

- (a) Construct the equation for the regression line.
 (b) Explain the meaning of the slope and intercept in the context of this application.
 (c) Find R^2 and explain what it means.
13. A psychology study collected the data given below that shows the time (in seconds) that it took 20 subjects to complete a paper-and-pencil maze:
 24, 25.8, 30, 31.6, 32, 34, 34.1, 42.2, 42.8, 43, 44, 46.6, 47, 48, 58.1, 66.5, 74, 78
 (a) Make a suitably scaled, accurate, histogram for these data. Show detailed steps.
 (b) Describe the distribution.

14. Students in a statistics class at a university reported the number of credit hours they were taking that semester. Summary statistics are shown in the table.

Suppose the university charges tuition of \$73 per credit hour plus a flat student fee of \$35 per semester. For example, a student taking 12 credit hours would pay $\$35 + \$73(12) = \$911$ for that semester.

\bar{x}	16.65
s	2.96
minimum	5
Q1	15
median	16
Q3	19
maximum	28

- (a) Find the mean, standard deviation, median, and IQR of the fees paid.

- (b) The maximum credit hours shown in the table are 28. Would you consider that unusually high based on the given data? Explain/ justify your claim.
15. Replacement times for a certain line of TV sets have a normal distribution with mean of 8.6 years and standard deviation of 1.2 years.
- (a) Find the percent of TV sets that must be replaced within 5 years.
- (b) A company that sells this line of TV sets wants to offer a warranty so that no more than 2.5% of their sets need to be replaced within the warranty period. Determine the time-period of the warranty they should offer.