

AP Stats Summer Packet

Name: _____

1. Fuel Efficient Car

Here is a small part of a data set that describes the fuel economy in miles per gallon of 1998 model motor vehicles

Make and Model	Vehicle Type	Transmission Type	Number of Cylinders	City MPG	Highway MPG
BMW 318I	Sub-compact	automatic	4	22	31
BMW 318I	Sub-compact	manual	2	23	32
Buick Century	Mid-size	automatic	6	20	29
Chevrolet Blazer	4 wheel drive	automatic	6	16	20

- Which are the individuals of this data set?
- For each individual what variables are given? Which of these variables are quantitative and which are categorical?

(Hint: **Categorical variable** places an individual into one of several groups or categories. A **Quantitative Variable** takes numerical values for which arithmetic operations such as adding and averaging make sense.)

2. Medical Study Variables

Data from a medical study contain many variables for each of the subjects of the study. Which of the following variables are categorical, and which are quantitative?

- gender (male or female)
- age (years)
- race (Asian, black, white, other)
- smoker (yes or no)
- systolic blood pressure (mL of mercury)
- level of calcium in blood (micrograms per mL)

3. SSHA Scores

Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

140 152 154 178 101 103 126 126 137 165 165 129 200 148 154 109 137 115

and for 20 first-year men:

108 140 114 91 180 115 126 92 169 146 109 132 75 88 113 151 70 115 187 104

- Make side by side box plots to compare the distribution
- Compute numerical summaries for the two distributions
- Write a paragraph comparing the SSHA scores for men and women

4. Olympic Gold

Athletes like Cathy Free, Rulon Gardner, Ian Thorpe, Marion Jones, and Jenny Thompson captured public attention by winning gold medals in the 2000 Summer Olympic Games in Sydney, Australia. The table displays the total number of gold medals won by several countries in the 2000 Summer Olympics.

Country	Gold Medals	Country	Gold Medals
Sri Lanka	0	Netherlands	12
Qatar	0	India	0
Vietnam	0	Georgia	0
Great Britain	28	Kyrgyzstan	0
Norway	10	Costa Rica	0
Romania	26	Brazil	0
Switzerland	9	Uzbekistan	1
Armenia	0	Thailand	1
Kuwait	0	Denmark	2
Bahamas	1	Latvia	1
Kenya	2	Czech Republic	2
Trinidad and Tobago	0	Hungary	8
Greece	13	Sweden	4
Mozambique	1	Uruguay	0
Kazakhstan	3	United states	39

Make a dot plot or bar graph to display these data. Describe the distribution of number of gold medals won.

5. \bar{x} -bar and s Are Not Enough

The mean \bar{x} -bar and standard deviation s measure center and spread but are not a complete description of a distribution. Data sets with different shapes can have the same mean and standard deviation. To demonstrate this fact, use your calculator to find \bar{x} -bar and s for the following two small data sets. Then make a stem plot of each and comment on the shape of each distribution. (Mean = \bar{x} -bar)

A:	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74
B:	6.58	5.76	7.71	8.84	8.47	7.04	5.25	5.56	7.91	6.89	12.50

(Hint: Input the data values in List1 using Stat-Edit.

Press Stat – Calc and Select 1:1 – Var Stats and then Press Enter Key.)

6. Who Makes More?

A manufacturing company is reviewing the salaries of its full-time employees below the executive level at a large plant. The clerical staff is almost entirely female, while a majority of the production workers and technical staff are male. As a result, the distributions of salaries for male and female employees may be quite different. The following table gives the frequencies and relative frequencies for women and men.

- Make histograms for these data, choosing a vertical scale that is most appropriate for comparing the two distributions.
- Describe the shape of the overall salary distributions and the chief differences between them.
- Explain why the total for women is greater than 100%.

Salary (\$1000)	Women		Men	
	Number	Percent	Number	Percent
10-15	89	11.8	26	1.1
15-20	192	25.4	221	9
20-25	236	31.2	677	27.9
25-30	111	14.7	823	33.6
30-35	86	11.4	365	14.9
35-40	25	3.3	192	7.4
40-45	11	1.5	91	3.7
45-50	3	0.4	33	1.4
50-55	2	0.3	19	0.8
55-60	0	0	11	0.4
60-65	0	0	0	0
65-70	1	0.1	3	0.1
Total	756	100.1	2,451	100

7. Presidential Elections

Here are the percents of the popular vote won by the successful candidate in each of the presidential elections from 1948-2000

Year	1948	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992	1996	2000
Percent	49.6	55.1	57.4	49.7	61.1	43.4	60.7	50.1	50.7	58.8	53.9	43.2	49.2	47.9

- Make a stem plot of the winners' percents. (Round to whole numbers and use split stems.)
- What is the median percent of the votes won by the successful candidate in presidential elections? (Work with the unrounded data)
- Call an election a landslide if the winners' percent falls at or above the 3rd quartile. Which elections were landslides?

8. Better Corn

Corn is an important animal food. Normal corn lacks certain amino acids, which are building blocks for protein. Plant scientists have developed new corn varieties that have more of these amino acids. To test a new corn as an animal food, a group of 20 one-day-old male chicks was fed a ration containing the new corn. A control group of another 20 chicks was fed a ration that was identical except that it contained normal corn. Here are the weight gains (in grams) after 21 days.

Normal Corn					New Corn			
380	321	366	356		361	447	401	375
283	349	402	462		434	403	393	426
356	410	329	399		406	318	467	407
350	384	316	272		427	420	477	392
345	455	360	431		430	339	410	326

- Compute five-number summaries (Min, Q1, Median, Q3, Max) for the weight gains of the two groups of chicks. Then make box plots to compare the two distributions. What does the data show about the effect of the new corn?
- The researchers actually reported means (\bar{x}) and standard deviations (s) for the two groups of chicks. What are they? How much larger is the mean weight gain of chicks fed the new corn?
- The weights are given in grams. There are 28.35 grams in an ounce. Use the result of part b) to compute the means and standard deviations of the weight gains measured in ounces.

(Hint: Input the data values in List1 using Stat-Edit.

Press Stat – Calc and Select 1:1 – Var Stats and then Press Enter Key.)

9. Pennies Spinning

Hold a penny upright on its edge under your forefinger on a hard surface, and then snap it with your other forefinger so that it spins for some time before falling. Based on 50 spins, estimate the probability of heads.

10. Matching Probabilities

Probability is a measure of how likely an event is to occur. Match one of the probabilities that follow with each statement about an event (The probability is usually a much more exact measure of likelihood than is a verbal statement.)

0, 0.01, 0.3, 0.6, 0.99, 1

- This event will never occur
- This event is certain. It will occur on every trial of the random phenomenon.
- This event is very unlikely, but it will occur once in awhile in a long sequence of trials
- This event will occur more often than not

11. Wal-Mart Stock

The rate of return on a stock is its change in price plus any dividends paid. Rate of return is usually measured in percent of the starting value. We have data on the monthly rates of return for the stock of Wal-Mart stores for the years 1973-1991, the first 19 years Wal-Mart was listed in the New York Stock Exchange. There are 228 observations.

- Give the five-number summary (Min, Q1, Median, Q3, Max) for monthly returns on Wal-Mart stock.
- Describe in words the main features of the distribution
- Find the interquartile range ($IQR = Q3 - Q1$) for the Wal-Mart data. Are there any outliers according to the $1.5 \times IQR$ criterion? Does it appear that the software uses this criterion in choosing which observations to report separately as outliers?

Figure 1 displays output from statistical software that describes the distribution of these data. The stems in the stem plot are the tens digits of the percent of returns. The leaves are ones digits. The stem plot uses split stems to give a better display. The software gives high and low outliers separately from the stem plot rather than spreading out the stem plot to include them.

Figure 1

Mean = 3.064

Standard Deviation = 11.49

N = 228 Median = 3.4691

Quartiles = -2.950258, 8.4511

Decimal point is 1 place to the right of the colon

Low: -34.04255 -31.25000 -27.06271 -26.61290

-1 : 982

-1 : 444443322222110000

-0 : 99998877766666665555

-0 : 44444444333333322222222222111111100

0 : 0000011111111111222222333333344444444

0 : 55555555555555555556666666666777777888888888899999

1 : 000000001111111122233334444

1 : 55566667889

2 : 011334

High: 32.01923 41.80531 42.05607 57.89474 58.67769

12. Population of the States

Briefly describe the shape, center, and spread of the distribution of population. Explain why the shape of the distribution is not surprising. Are there any states that you consider outliers?

State	Population (1000)	State	Population (1000)
AL	4447	MO	5595
AK	627	MT	902
AZ	5131	NE	1711
AR	2673	NH	1236
CA	33871	NJ	8414
CO	4301	NM	1819
CT	3540	NY	18976
DE	784	NC	8049
DC	572	ND	642
FL	15982	OH	11353
GA	8186	OK	3451
HI	1294	OR	3421
ID	12419	PA	12281
IL	12419	RI	1048
IN	6080	SC	4012
IA	2926	SD	755
KS	2688	TN	5689
KY	4042	TX	20852
LA	4469	UT	2233
ME	1275	VT	609
MD	5296	VA	7079
MA	6349	WA	5894
MI	9938	WV	1808
MN	4919	WI	5364
MS	2845	WY	494
NV	1998		

13. How Many Tosses To Get A Head?

When we toss a penny, experience shows that the probability (long-term proportion) of a head is close to $\frac{1}{2}$. Suppose now that we toss the penny repeatedly until we get a head. What is the probability that the first head comes up in an odd number of tosses (1, 3, 5, etc)? To find out repeat this experiment 50 times, and keep a record of the number of tosses needed to get a head on each of your 50 trials.

14. Three Of Kind

You read in a book on poker that the probability of being dealt three of a kind in a 5 card poker hand $\frac{1}{50}$. Explain in simple language what this means.

15. Winning A Baseball Game

A study of the home-field advantage in baseball found that over the period from 1969 to 1989 the league champions won 63% of their home games. The two league champions meet in the baseball World Series. Would you use the study results to assign a probability 0.63 to the event the home team wins in a World Series game? Explain your answer.

16. Simulating An Opinion Poll

A recent opinion poll showed that about 73% of married women agree that their husbands do at least their fair share of household chores. Suppose that this is exactly true. Choosing a married woman at random then has probability 0.73 of getting one who agrees that her husband does his share. Use software or your calculator to simulate choosing many women independently. (In most software, the key phrase to look for is “Bernoulli trials.” This is the technical term for independent trials with Yes/No outcomes. Our outcomes here are “Agree” or not)

- a) Simulate drawing 20 women, then 80 women, then 320 women. What proportion agrees in this case? We expect (but because of chance variation we can't be sure) that the proportion will be closer to 0.73 in the longer runs of trials
- b) Simulate drawing 20 women 10 times and record the percents in each trial who agree. Then simulate drawing 320 women 10 times and again record the 10 percents. Which set of 10 results is less variable? We expect the results of 320 trials to be more predictable (less variable) than the results of 20 trials. That is “long-run regularity” showing itself

17. Heart Disease And Cancer

Government data assign a single cause for each death that occurs in the United States. The data show that the probability is 0.45 that a randomly chosen death was due to cardiovascular (mainly heart) disease, and 0.22 that it was due to cancer. What is the probability that a death was either due to cancer or cardiovascular disease? What is the probability that the death was due to some other cause?