DEFINITION

Data are collections of observations (such as measurements, genders, survey responses).

Statistics is the science of planning studies and experiments, obtaining data, and then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions based on the data.

A **population** is the complete collection of all individuals (scores, people, measurements, and so on) to be studied.

A **census** is the collection of data from every member of the population.

A **sample** is a subcollection of members selected from a population.

Remember—garbage in, garbage out! Sample data must be collected through a process of random selection. If sample data are not collected in an appropriate way, the data may be completely useless!

1.2 STATISTICAL THINKING

Key Concept…
When conducting a statistical analysis of data we have collected or analyzing a statistical analysis done by someone else, we should not rely on blind acceptance of mathematical calculations. We should consider these factors:

- Context of the data
- Source of the data
- Sampling method
- Conclusions
- Practical implications

650	24249	0
1050	20666	0
967	19413	0
500	21992	0
1700	21399	0
2000	22022	0
1100	25859	0
1300	20390	0
1400	23738	0
2250	23294	0
800	19063	0
3500	30131	0
1200	18698	0
1250	25348	0
2250	25642	1
3000	23074	1
1750	28349	1
1525	24644	1
1500	23245	1
1500	24378	1
1250	23246	1
1200	23695	1
1600	23258	1
425	19325	1
1450	20397	1
900	17256	1
675	19545	1
1450	20780	1

Description: These data for the 1991 season of the National Football League were reported by the Associated Press.

Number of cases: 28

Variable Names:

1. TEAM: Name of team
2. QB: Salary ($thousands) of regular quarterback
3. TOTAL: Total team salaries ($thousands)
4. NFC: National Football Conference (1) or American Football Conference (0)

The Data:

TEAM	QB	TOTAL	NFC
BILLS	650	24249	0
BENGALS	1050	20666	0
BROWNS	967	19413	0
BRONCOS	500	21992	0
OILERS	1700	21399	0
COLTS	2000	22022	0
CHIEFS	1100	25859	0
RAIDERS	1300	20390	0
DOLPHINS	1400	23738	0
PATRIOTS	2250	23294	0
JETS	800	19063	0
STEELERS	3500	30131	0
CHARGERS	1200	18698	0
SEAHAWKS	1250	25348	0
FALCONS	2250	25642	1
BEARS	3000	23074	1
COWBOYS	1750	28349	1
LIONS	1525	24644	1
PACKERS	1500	23245	1
RAMS	1500	24378	1
VIKINGS	1250	23246	1
SAINTS	1200	23695	1
GIANTS	1600	23258	1
EAGLES	425	19325	1
CARDINALS	1450	20397	1
49ERS	900	17256	1
BUCCANEERS	675	19545	1
REDSKINS	1450	20780	1

Example 1: Refer to the data in the table below. The *x*-values are weights (in pounds) of cars; the *y*-values are the corresponding highway fuel consumption amounts (in mi/gal).

Car Weights and Highway Fuel Consumption Amounts

WEIGHT	4035	3315	4115	3650	3565
FUEL CONSUMPTION	26	31	29	29	30

a. Context of the data.

i. Are the *x*-values matched with the corresponding *y*-values? That is, is each *x*-value somehow associated with the corresponding *y*-value in some meaningful way?

yes → the weight and mpg seems to refer to a particular car.

ii. If the *x* and *y* values are matched, does it make sense to use the difference between each *x*-value and the *y*-value that is in the same column? Why or why not?

No → different units.

b. Conclusion. Given the context of the car measurement data, what issue can be addressed by conducting a statistical analysis of the values?

The data can address if weight has an impact on fuel consumption.

c. Source of the data. Comment on the source of the data if you are told the car manufacturers supplied the values. Is there an incentive for car manufacturers to report values that are not accurate?

They may be reporting the mpg under optimal conditions, the data could be generously rounded. The manufacturers want to sell cars, so they might be reporting the best performing models.

d. Conclusion. If we use statistical methods to conclude that there is a correlation between the weights of cars and the amounts of fuel consumption, can we conclude that adding weight to a car causes it to consume more fuel?

No!!! Correlation DOES NOT imply cause and effect!

Example 2: Form a conclusion about statistical significance. Do not make any formal calculations. Either use results provided or make subjective judgements about the results.

One of Gregor Mendel's famous hybridization experiments with peas yielded 580 offspring with 152 of those peas (or 26%) having yellow pods. According to Mendel's theory, 25% of the offspring should have yellow pods. Do the results of the experiment differ from Mendel's claimed rate of 25% by an amount that is statistically significant?

The experimental results seem close to the theoretical results (1% off).

1.3 TYPES OF DATA

DEFINITION

A **parameter** is a numerical measurement describing some characteristic of a population.

A **statistic** is a numerical measurement describing some characteristic of a sample.

Example 3: Determine whether the given value is a statistic or a parameter.

a. 45% of the students in a calculus class failed the first exam.

Parameter

b. 25 calculus students were randomly selected from all the sections of calculus I. 38% of these students failed the first exam.

Statistic

DEFINITION

Quantitative (aka numerical) data consist of numbers

representing counts or measurements.

Categorical (aka qualitative or attribute) data consist of names

or labels that are not numbers representing counts or measurements.

Give 2 examples of

a. Quantitative data

① How many people like a certain type of music.

② Distance you drive to school.

b. Categorical data

① Genre of music

② Type of milk (nonfat, 1%, etc)

DEFINITION

Discrete data result when the number of possible values is either a finite number or a countable number.

Continuous (aka numerical) data result from _____ many possible values that correspond to some _____ scale that covers a _____ of values without gaps, interruptions or jumps.

Give 2 examples of

a. Discrete data

① # of bananas in a bunch

② # of fruit loops in a box of cereal

b. Continuous data