1-2:Data Classification

Statistics Chapter 1

Types of Data

- <u>Qualitative data</u> consist of attributes, labels, or non-numerical entries.

- <u>Quantitative data</u> consist of numerical measurements or counts.

Ex 1:

- The base prices of several vehicles are shown in the table. Which data are qualitative data and which are quantitative data? Explain your reasoning.

Model	Base Price
Fusion 14 S	$17,795
F-150 XL	$18,710
Five Hundred SEL	$23,785
Escape XLT Sport	$24,575
2007 Explorer Sport Trac	$26,775
Freestar SEL	$27,500
Crown Victoria LX	$28,830
Expedition	$35,480

Levels of Measurement

- <u>Nominal Level of measurement</u> are qualitative only. Data at this level are categorized using names, labels, or qualities. No mathematical computations can be made at this level.

- <u>Ordinal Level of measurement</u> are qualitative or quantitative. Data at this level can be arranged in order, or ranked, but differences between data entries are not meaningful.

Ex 2:

- Two data sets are shown. Which data set consists of data at the nominal level? Which data set consists of data at the ordinal level? Explain your reasoning.

Top Five TV Programs (from 2/12/07)
1.   American Idol-Tuesday
2.   American Idol-Wednesday
3. Grey's Anatomy
4.   House
5. CSI

Network Affiliates in Pittsburgh, PA
WTAE (ABC)
WPXI (NBC)
KDKA (CBS)
WPGH (FOX)

Answer:

- The first data set lists the rank of five TV programs. The data consist of the ranks 1,2,3,4, and 5. Because the rankings can be listed in order, these data are at the ordinal level. Note that the difference between a rank of 1 and 5 has no mathematical meaning. The second data set consists of the call letters of each network affiliate in Pittsburg. The call letters are simply the names of network affiliates, so these data are at the nominal level.

Ex 2a:

- Consider the following data sets. For each data set, decide whether the data are at the nominal level or at the ordinal level.

- 1. The final standings for the Pacific Division of the National Basketball Association

- Answer: Ordinal, because the data can be put in order (rank)

- 2. A collection of phone numbers

- Answer: Nominal, because you cannot make calculations on the data (labels)

Ex 2b:

- At the beginning of 2007, Forbes Magazine chose the 25 best U.S. cities to get a job. To gather the rankings, Forbes used five data points: unemployment rate, job growth, income growth, median household income, and cost of living. Growth data was then measured in the largest 100 metropolitan areas from 2003 to 2006.
- In this list what is the level of measurement?

Forbes Top Five U.S. Cities for Jobs
1. Raleigh-Cary, NC
2. Phoenix-Mesa-Scottsdale, AZ
3. Jacksonville, FL
4. Orlando-Kissimmee, FL
5. Washington-Arlington-Alexandria, DC-VA-MD-WV

Levels of Measurement (con't)

- Data at the <u>interval level of measurement</u> can be ordered, and you can calculate meaningful differences between data entries. At the interval level, a zero entry simply represents a position on a scale: the entry is **NOT** an inherent zero. (inherent zero is a zero that implies "none")

- Data at the <u>ratio level of measurement</u> are similar to data at the interval level, with the added property that a zero entry **IS** an inherent zero. A ratio of two data values can be formed so that one data value can be meaningfully expressed as a multiple of another.

Inherent Zero example:

- An inherent zero is a zero that implies "none". For instance, the amount of money you have in a savings account could be zero dollars. In this case, the zero represents no money; it is an inherent zero. On the other hand, a temperature of 0 degrees C does not represent a condition in which no heat is present. The 0 degree C temperature is simply a position on the Celsius scale; it is not an inherent zero.

- To distinguish between data at the interval level and at the ratio level, determine whether the expression "twice as much" has any meaning in the context of the data. For instance, $2 is twice as much as $1, so these data are at the ratio level. On the other hand, 2 degrees C is not twice as warm as 1 degree C so these data are at the interval level.

Ex: 3

- Two data sets are shown below. Which data set consists of data at the interval level? Which data set consists of data at the ratio level? Explain your reasoning.

New	York	Yankees	World Series	Victories (year)
1923	1927	1928	1932	1936
1937	1938	1939	1941	1943
1947	1949	1950	1951	1952
1953	1956	1958	1961	1962
1977	1978	1996	1998	2000

2006 AL Home Run Totals	By Team
Balt 164	Bost 192
Chicago 236	Clev 196
Det 203	Kan City 124
LA 159	Minn 143
NY 210	Oak 175
Sea 172	TB 190
Tex 183	Tor 199

Answer:

- Both of these data sets contain quantitative data. Consider the dates of the Yankees World Series victories. It makes sense to find differences between specific dates. For instance, the time between the Yankees first and last World Series victories is

- 2000 – 1923 =77 years

- But it does not make sense to say that one year is a multiple of another. So, these data are at the interval level. Using the home run totals, you can find differences and write ratios. From the data, you can see that Detroit hit 31 more home runs than Seattle hit and that Chicago hit about twice as many home runs as Kansas City hit. So, these data are at the ratio level.

Ex 3a: Decide whether the data are at the interval level or at the ratio level.

- 1. The body temperatures (in degrees Fahrenheit) of an athlete during an exercise session.

- Answer: Interval, because the data can be ordered and meaningful differences can be calculated, but it does not make sense to write a ratio using the temperatures.

- 2. The heart rates (in beats per minute) of an athlete during an exercise session.

- Answer: Ratio, because the data can be ordered, meaningful differences can be calculated, the data can be written as a ratio and the data set contains an inherent zero.

Summary of The Four Levels of Measurement

Level of measurement	Put data in categories	Arrange data in order	Subtract data values	Determine if one data value is a multiple of another and inherent zero
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	yes

	Example of a Data Set	Meaningful Calculations
Nominal Level (qualitative Data)	Types of Music Played by a Radio Station: Pop Contemporary jazz Modern Rock Hip Hop	Put in a category. For instance, a song played by the radio station could be put into one of the four categories shown.
Ordinal Level (Qualitative or quantitative data)	Motion Picture Association of America Ratings Descriptions G general Audiences PG Parental Guidance Suggested PG-13 Parents Strongly Cautioned R Restricted NC-17 No One Under 17 Admitted	Put in a category and put in order. For instance, a PG rating has a stronger restriction than a G rating.

	Example of a Data Set	**Meaningful Calculations**
Interval Level (Quantitative data)	*Average monthly Temperature (in degrees Fahrenheit) for Sacramento, Ca:* 46, 75, 51, 74, 54, 71, 59, 64, 65, 53, 71, 45	Put in a category, put in order, and *find differences between values.* *For instance, 71.5 – 65.5 =6* *So June is 6 degrees warmer than May*
Ratio Level (Quantitative Data)	*Average Monthly Precipitation (in inches) for Sacramento, CA* 4, .1, 3.5, .1, 2.8, .4, 1, 1, .5, 2.2, .2, 2.5	Put in a category, put in order, find differences between values, and *find ratios of values.* For Instance 1/.5 = 2. So, there is twice as much rain in April as in May.

Assignment:

Page 15 #2-24 even