Role of Statistics in Geography

What Is Geography?

- 1. Attempt to describe, explain and predict spatial patterns and activities
- 2. How and why do things differ from place to place?
- 3. How do spatial patterns change through time?

How Do Geographers Approach Discipline

- 1. Positivism- objectivity of scientific analysis and testing hypotheses to build knowledge and understanding
- 2. Humanistic- people create subjective worlds in their minds- behavior understood only by a methodology that penetrates the subjectivity
- 3. Structuralists- cannot explain observed pattern by examining pattern itself. But rather establish theories to explain development of societal conditions within which people must act

Role of Statistics

- Room in all the above interpretations for quantitative analysis.
- But increasingly both quantitative and qualitative analysis are important
- Qualitative analysis involves?
- Statistics and measurement are used commonly in our lives
  - A. Making home purchase decisions
  - B. Setting up investments
  - C. Weather variations are expressed as probabilities

How Do Geographers Use Statistics?

- 1. Describe and summarize data
- 2. Make generalizations concerning complex spatial patterns
- 3. Estimate likelihoods of outcomes for events at particular location(s)
- 4. Use sample data to make inferences about a larger set of data (a population)
- 5. Learn whether actual pattern matches an expected or theoretical
- 6. Wish to compare or associate (correlate) patterns of distributions

Formulating the Research Process

- 1. Problem Identification
- 2. Develop Questions to Investigate
- 3. Collect and Prepare Data
- 4. Process descriptive data (maps, graphics)>>>>> Reach conclusions
- 5. Formulate Hypothesis >>>>> Collect and Prepare Sample Data
- 6. Test Hypothesis>>Evaluate Hypothesis
- 7. Develop Model, Law, or Theory

What Are Models?

- Abstractions of the real world
- Simplified versions of reality
- Easier to examine scaled down and simplified structures in attempt to understand
- Iconic models- look like what they represent (
- Analogue models- one property used to represent another
- Symbolic models- equations

Basic Terms and Concepts

- Data element- basic element of information which we measure
- Data Set- groups of data (commuting sheds of industries)
- Observations-Cases-Individuals- elements of phenomena under study
- Variable- property or characteristics of each observation that can be measured, classified or counted
- Values may vary among set of observations: rainfall, per capita income, years of schooling

Geographic Data

- 1. What sources of data are available?
- 2. Which methods of data collections should be used?
- 3. What type of data will be collected and then analyzed statistically?

Types of Data

- Primary Data- acquired directly from original source
- 1. Information collected in the field
- 2. Usually very time consuming
- 3. Involves decision about a sample design so representative data may be obtained

Types of Data

- Secondary Data (or Archival Data)
- 1. Usually collected by some organization (United Nations, U S Bureau of Census)
- 2. Often easily accessible- hardcopy or CD rom
- 3. Less time consuming but also more limiting
- 4. Often need to inspect historical records and archives for diaries, oral histories, official reports in order to develop a picture of problem

Characteristics of Data

- 1. Some data are explicitly spatial- locations are directly analyzed
- 2. Other data implicitly spatial- data represents places but locations themselves are not analyzed (population sizes of towns)

Measurement Concepts

- 1.Precision- level of exactness associated with measurement (rain gauge to inches or fractions of inches)
- 2. Accuracy- extent of system wide bias in measurement process
- 3. Validity- if geographical concept is complex expressing "true" or "appropriate" meaning of the concept through measurement may be difficult (levels of poverty, economic well being, environmental quality)
- 4. Reliability- changes in spatial patterns are analyzed over time must ask about consistency and stability of data
-

Types of Statistical Analysis

- Descriptive Statistics- concise numerical or quantitative summaries of the characteristics of a variable or data set (e.g. mean, standard deviation, etc)
- Inferential Statistics- here we wish to make generalizations about a statistical population (total set of information or data under investigation) based on the information from a sample
- Sample- typical or representative or unbiased subset of the broader, larger more complete statistical population

MEASUREMENT SCALES

- The "levels of measurement" is an expression which typically refers to the theory of scale types developed by the psychologist Stanley Smith Stevens.
- Stevens proposed his theory in a 1946 article titled "On the theory of scales of measurement".
- In this article Stevens claimed that all measurement in science was conducted using four different types of numerical scales which he called "nominal", "ordinal", "interval" and "ratio".

The Theory Of Scale Types

Stevens (1946, 1951) proposed that measurements can be classified into four different types of scales. These were:

- Nominal
- Ordinal
- Interval
- Ratio

NOMINAL SCALE

- A **categorical variable**, also called a **nominal variable**, is for mutual exclusive, but not ordered, categories.
- Nominal scales are mere codes assigned to objects as labels, they are not measurements.
- Not a measure of quantity. Measures identity and difference. People either belong to a group or they do not.
- Sometimes numbers are used to designate category membership.

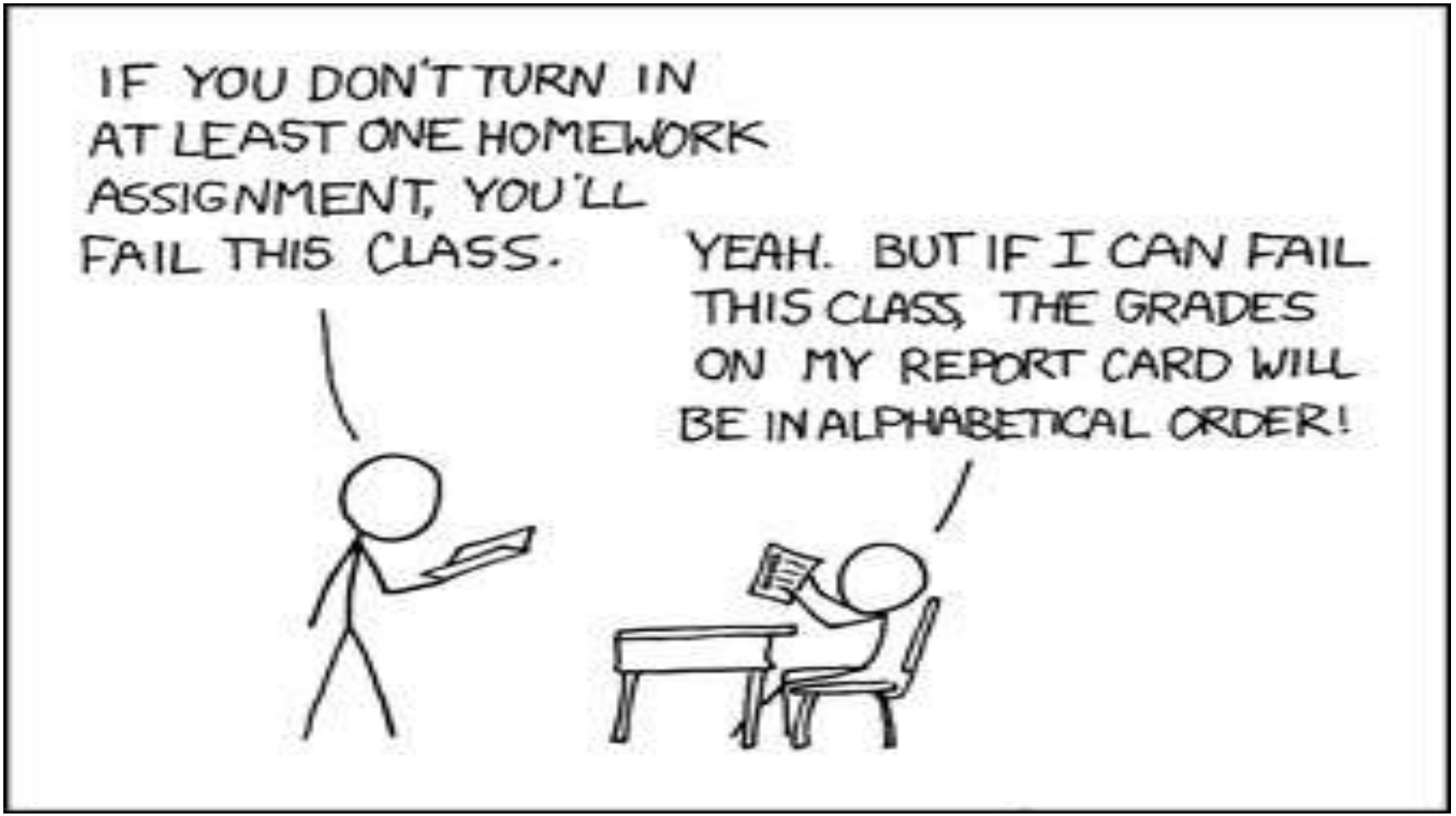
Examples

- Eye color: blue, brown, green, etc.
- Biological sex (male or female)
- Democrat, republican, green, libertarian, etc.
- Married, single, divorced, widowed
- Country of Origin
- 1 = United States
- 2 = Mexico
- 3 = Canada
- 4 = Other

(Here, the numbers do not have numeric implications; they are simply convenient labels)

What Statistic Can I Apply?

OK to compute....	Nominal
Frequency Distribution and mode	Yes
Median And Percentiles.	No
Add Or Subtract.	No
Mean, Standard Deviation, Standard Error Of The Mean.	No
Ratio, Or Coefficient Of Variation.	No
Chi-square	Yes



Ordinal Scale

Ordinal Scale

- This scale has the ability to rank the individual attributes of to items in same group but unit of measurement is not available in this scale, like student A is taller than student B but their actual heights are not available.
- Designates an ordering: greater than, less than.
- Does not assume that the intervals between numbers are equal.

EXAMPLES

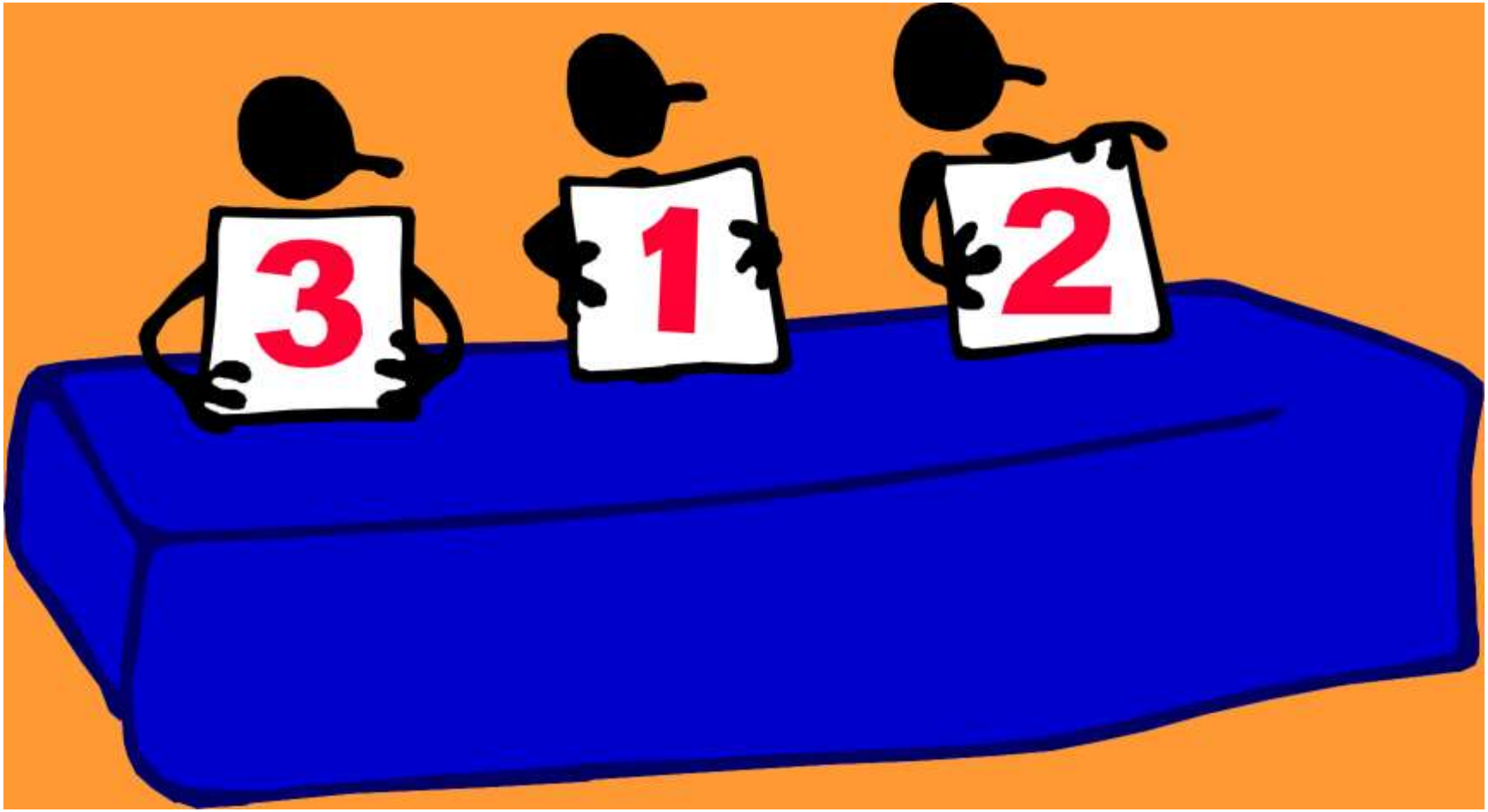
- Rank your food preference where 1 = favorite food and 4 = least favorite:

_____sushi	_____chocolate
_____hamburger	papaya

- Final position of horses in a thoroughbred race is an ordinal variable. The horses finish first, second, third, fourth, and so on. The difference between first and second is not necessarily equivalent to the difference between second and third, or between third and fourth.

What Statistic Can I Apply?

OK To Compute....	Ordinal
Frequency Distribution.	Yes
Median And Percentiles.	Yes
Add Or Subtract.	No
Mean, Standard Deviation, Standard Error Of The Mean.	No
Ratio, Or Coefficient Of Variation.	No



Interval Scale

Interval Scale

- Classifies data into groups or categories
- Determines the preferences between items
- Zero point on the internal scale is arbitrary zero, it is not the true zero point
- Designates an equal-interval ordering.
- The difference in temperature between 20 degrees f and 25 degrees f is the same as the difference between 76 degrees f and 81 degrees f.

EXAMPLES

- Temperature in Fahrenheit is interval.
- Celsius temperature is an interval variable. It is meaningful to say that 25 degrees Celsius is 3 degrees hotter than 22 degrees Celsius, and that 17 degrees Celsius is the same amount hotter (3 degrees) than 14 degrees Celsius. Notice, however, that 0 degrees Celsius does not have a natural meaning. That is, 0 degrees Celsius does not mean the absence of heat!
- Common IQ tests are *assumed* to use an interval metric.

Examples

Likert scale: How do you feel about Stats?

1 = I’m totally dreading this class!

2 = I’d rather not take this class.

3 = I feel neutral about this class.

4 = I’m interested in this class.

5 = I’m SO excited to take this class!

What Statistic Can I Apply?

OK To Compute....	Interval
Frequency Distribution.	Yes
Median And Percentiles.	Yes
Add Or Subtract.	Yes
Mean, Standard Deviation, Correlation, Regression, Analysis Of Variance	Yes
Ratio, Or Coefficient Of Variation.	No

Ratio Scale

- This is the highest level of measurement and has the properties of an interval scale; coupled with fixed origin or zero point.
- It clearly defines the magnitude or value of difference between two individual items or intervals in same group.

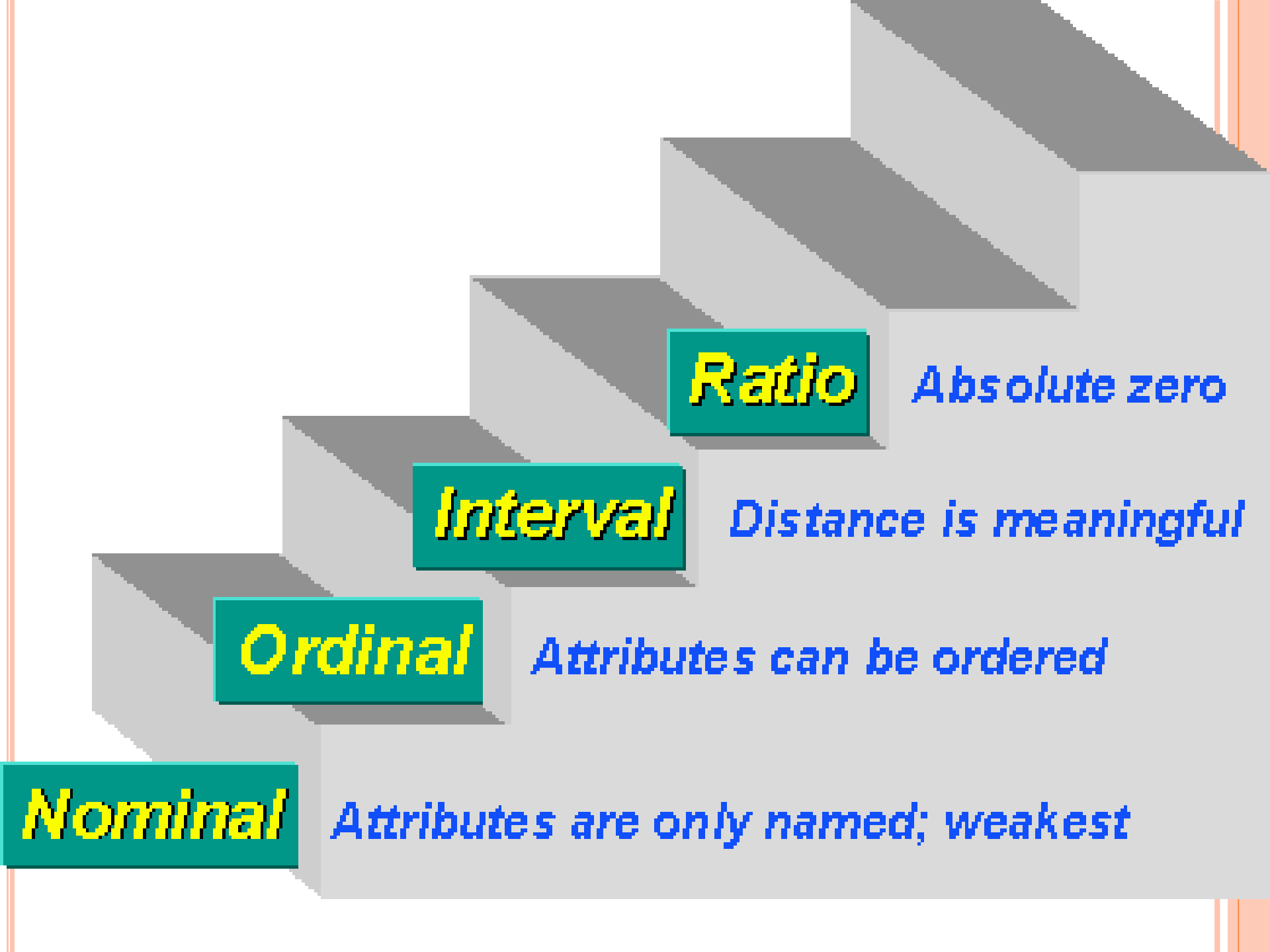
EXAMPLES

- Temperature in Kelvin (zero is the absence of heat. Can’t get colder).
- Measurements of heights of students in this class (zero means complete lack of height).
- Someone 6 ft tall is twice as tall as someone 3 feet tall.
- Heart beats per minute has a very natural zero point. Zero means no heart beats.

What Statistic Can I Apply?

OK To Compute....	Ratio
Frequency Distribution.	Yes
Median And Percentiles.	Yes
Add Or Subtract.	Yes
Mean, Standard Deviation, Correlation, Regression, Analysis Of Variance	Yes
Ratio, Or Coefficient Of Variation.	Yes

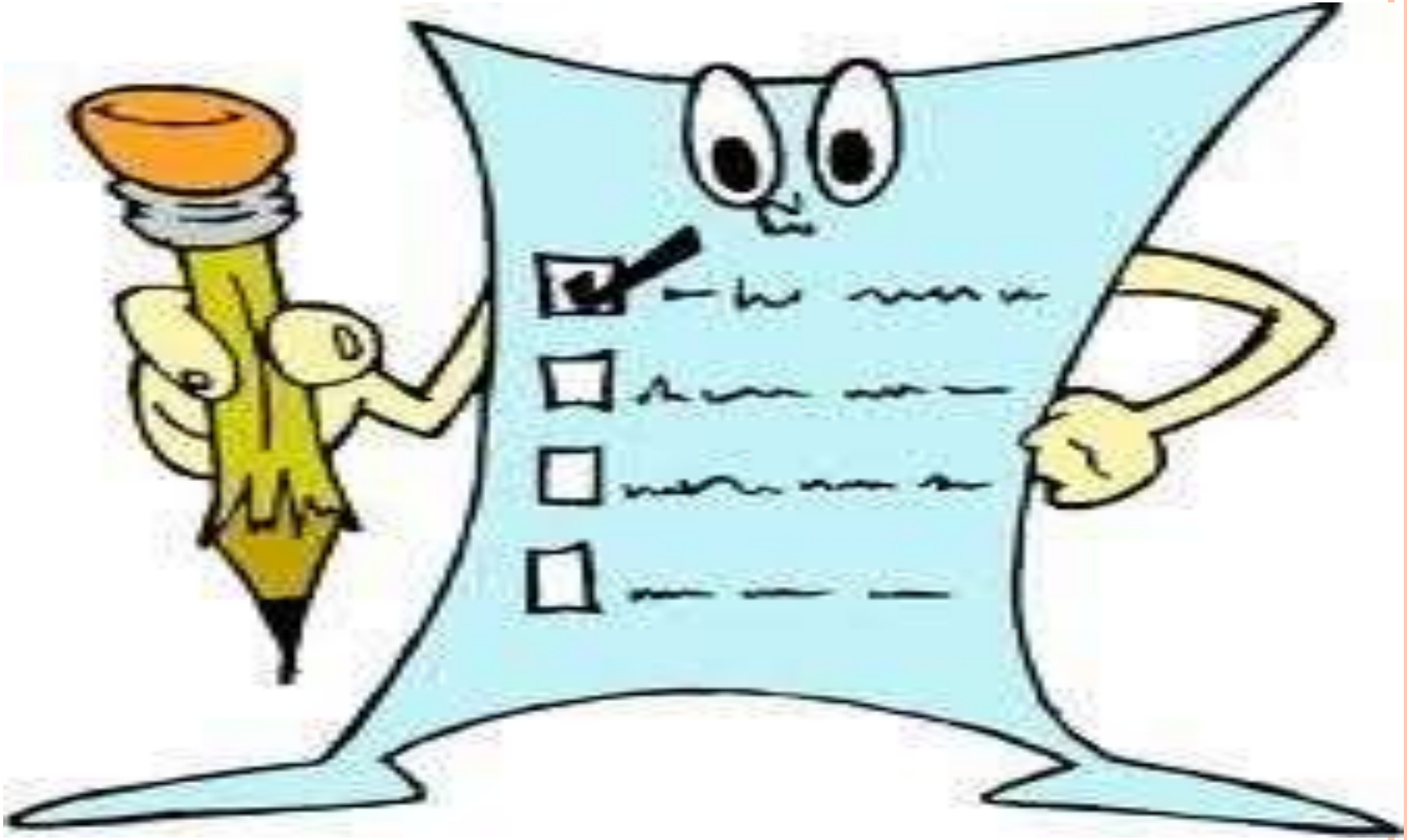
Putting It Together


Ratio
Absolute zero
Interval
Distance is meaningful
Ordinal
Attributes can be ordered
Nominal
Attributes are only named; weakest

Summary of 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
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes

Test Your Knowledge



A professor is interested in the relationship between the number of times students are absent from class and the letter grade that students receive on the final exam. He records the number of absences for each student, as well as the letter grade (A,B,C,D,F) each student earns on the final exam. In this example, what is the measurement scale for number of absences?

a) Nominal **b) Ordinal** **c) Interval** **d) Ratio**