

The What and Why of Statistics

What is Science?

The development of knowledge, ultimately in the form of laws and theories and based on a systematic examination of facts (research methods and statistics).

What is the purpose of science?

To distinguish what is real from what is illusion.

What is the relationship between science and statistics?

Statistics help to test the accuracy of theories.

Research methods and social statistics are sometimes used to answer a research question (e.g., a practical concern) rather than test a theory.

What would be an example where statistics are used to better understand a "practical" concern?

The Research Process can be used to answer research questions:

1. Ask research question (e.g., are college students open to change more than others?)
2. Formulate hypothesis (e.g., college students are more open to change than their parents)
3. Collect data
4. Analyze data
5. Answer the question and perhaps developing a theory based on the results (e.g., A theory of age and social change)

The Research Process can also be used to test existing theories but the "process" order is changed:

1. State the theory (e.g., A theory of age and social change)
2. Formulate hypotheses (e.g., college students are more open to change than their parents)
3. Collect data
4. Analyze data
5. Evaluate accuracy of hypotheses and subsequently the theory

To test a hypothesis, the concepts must be stated in "measurable" terms or "variables" (e.g., college student, attitude toward change, parent).

What are dependent variables?
(e.g., social change)

Independent variables?
(e.g., age)

What is a "unit of analysis"?

An object for observation
such as people or cities

What is the difference
between a "population"
and a "sample"?

What is a "random"
sample and why would we
want one?

Random samples are
representative of the
population from which
they are drawn.

Inferential Statistics

Statistics typically used with samples to determine the probability that a relationship, found between two variables in our sample, is true for the whole population.

In other words, what is the probability that what we found in our sample is true for the whole population?

(Inferential Statistics)

If we are given data showing that the average age of UNT men students is 23 and the average age of women students is 21:

Does it matter whether these numbers were derived from the whole population of UNT students or whether they were derived from a sample of UNT students?

We use Inferential Statistics to determine the probability that the relationship between variables we find in the sample data is the same relationship that exists in the population.

That is, if the average ages of 23 for men and 21 for women come from a sample, what is the probability that these are the average ages for men and women in the population?

When analyzing sample data, the level of measurement is used to determine which statistic to use.

There are many statistics such as the:

lambda (used with nominal level data), gamma (used with ordinal level data), or correlation coefficient (used with interval/ratio level data)

Levels of Measurement:

1. Nominal: values of a variable are not ordered (e.g., religion, gender)
2. Ordinal: values of a variable are ordered (e.g., level of job satisfaction, desirability of ice cream)
3. Interval/Ratio: values of a variable are ordered and have the same distance between each value (e.g., income, age)

Exercise: Determine the level of measurement of these variables

1. Brands of toothpaste used by people
2. Number of cars per day that pass exit 469 on I-35
3. People's desire for baseball season to start (e.g., high, medium, low)
4. telephone numbers
5. IQ
6. Place of residence (urban, rural, suburban)

Dichotomous Variables are somewhat unique.

Properties

- A nominal variable with two values (usually 0 and 1)
- Often treated as if it were ordinal or interval/ratio

Examples

- Sex (0 = female; 1 = male)
- Marital status (0 = not married; 1 = married)

What is the difference between "statistical significance" and "measures of association"?

Statistical significance is the probability that the sample statistic (e.g., average age of women) is true for the population.

If the probability is high, then we assume that the sample statistic is representative of the population.

For example:

If a difference is found between two groups in the sample,

what is the probability that this difference actually exists in the population (and thus is not due to sampling error).

Measures of association are statistics that show how strongly related two variables are and the direction of their relationship.

(By direction of relationship is meant, for example, as one variable gets larger the other gets smaller)

Can statistics be used "subjectively" and if so, how?

Thank You.

