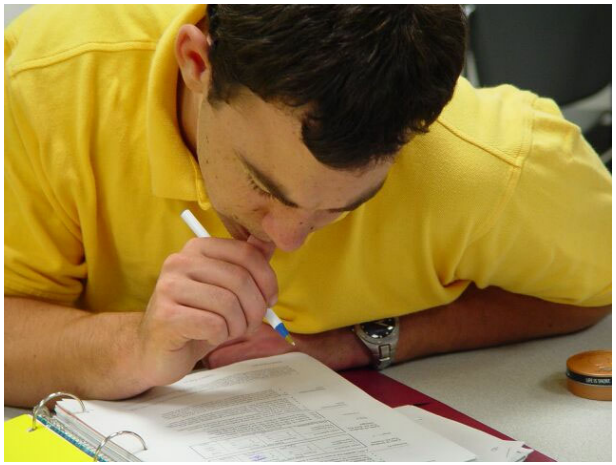
Statistical Literacy©

critical thinking about statistics as evidence in arguments



GRAMMAR Of CHANCE

StatLit-Project-Brochure 3A 9/29/2002 ©

Grammar of Statistics

As a discipline, statistics is traditionally a branch of mathematics. In describing uncertainty, mathematicians use statements of probability to describe events or entities that are indistinguishable except for the outcome of interest.

In describing observed relative frequencies statisticians tend to use the language of probability or chance rather than the language of rates and percentages.

This creates a major problem for those studying statistics:

- It blurs the distinction between the evidence found in the world and the conclusions drawn from that evidence.
- It dulls the critical thinking capability of those studying statistics.
- US women age 20-25 have a 20% chance of having a baby in the next year.

Grammar of Uncertainty

Uncertainty is a prominent element of reality. And for living things, it has life-and death consequences. For humans, a precise description of uncertainty is critical.

Consider two kinds of uncertainty.

- 10% of US women age 20-24 had babies in 1994.
- US women age 20-25 have a 20% chance of having a baby in the next year.

The uncertainty in the first statement is benign. The identity and number of US women age 20-24 in 1994 is known; the identity and number of these having babies in 1994 is also known. The ratio of these two is a fact: a statement about the world.

The uncertainty in the second statement is pernicious. While the identity and number of US women age 20-24 in 1994 is known, the identity and number of those that will have babies in the next 12 months is not known with the same certainty.

Moral:

Statements of rates and percentages are evidence about the world. Statements of chance, risk and probability are conclusions about the world. Notice the difference!

Ratio Grammar
Rates and Percentages

In English, there are four different families of words used to describe ratios.

The most common is the "% of/are" family. Other members include share.
In the US in 1994,

- 20% of women 20-24 gave birth.
- Among women 20-24, 20% gave birth.

The percentage family includes fraction.

- The percentage of women age 20-24 who gave birth was 20%.
- Among women age 20-24, the percentage who gave birth was 20%.
- Among women age 20-24, the percentage of birth-moms was 20%.

The rate family includes incidence and prevalence. "Incidence" describes a rate per unit time. "Prevalence" describe a ratio at a particular moment in time. Immigration is an incidence ratio; unemployment is a prevalence ratio.

Moral:

Users must treat all these contextual words as causally ambiguous even when they are used by statisticians or social scientists!

Likely-Prevalent Grammar

Likely is often used to make comparisons of probabilities and observed relative frequencies. Grammar is a common way English can distinguish association (correlates) from causation (causes, effects).

- Ambiguous: 80% of the lung-cancer deaths among smokers were **due to** smoking. 80,000 smokers died **because of** cigarette smoke. 80,000 smokers died **from** smoking. Smoking was **significantly linked** with 80,000 lung-cancer deaths.
- Pure causation: Smoking **caused** 80,000 lung-cancer deaths.

Ambiguity in Likely Language

Suppose that 10% of US women age 20-24 gave birth last year. If these women are half the population in their age range, then we can say:

- 5% of US adults age 20-24 gave birth last year.

We can transform these percentages into statements of chance:

- US women age 20-24 had a 10% chance of giving birth last year.
- US adults age 20-24 had a 5% chance of giving birth last year.

The next step is to distribute these to each member.

- A US women age 20-24 had a 10% chance of giving birth last year.
- A US adult age 20-24 had a 5% chance of giving birth last year.

Moral:

The truth of A (association) or of B does not guarantee the truth of C (causation).