

BE.104 Spring

Concepts in Variance or "Oh, What a Fuzzy World We Live In"

J. L. Sherley

I. Making Variance

Many know mathematical formalism. Few understand the origin of variance. Generate some & study it.

1. The length of desks versus the length of forearms
2. Pitching pennies
3. Random sampling

Frequency plots, histograms, and distributions

- A) How to think about the data?
- B) How to organize it?
- C) The Freq. Plot = distribution of the data
- D) Vary bins, grouping intervals.
Looks different, but does not alter the distribution's mathematical parameters
- E) What can we say about the variance in the measurements?
Why aren't they single lines?

- 1) **Quantitative error (instrument, user)**
- 2) **Investigator error (mistakes, recording errors, judgement)**
- 3) **Biological Variation and Variability**
- 4) **Seasonal Variation (also circadian variation)**
- 5) **Statistical Variation (Chance variation)**

How do we use distributions?

- A. Define "average" features
- B. Evaluate whether populations differ for a feature on average
- C. How does variance factor into such analyses?

Given the variance, how likely is the average feature to be representative?
Given the variance, does a meaningful difference in a feature exist?
Or do they **appear** to occur as a result of chance and error?
How **confident** are we in our **hypothesis**?

II. Pure Sampling ("All statistics is based on the concept of probabilistic sampling")

"Drawing beads"

(Theoretically, with ideal mixing this is not a conditional probability)

Even in pure sampling other factors also can affect outcome

- A. Quantitative error (miscounts)
- B. Investigator error (poor mixing)
- C. Biological variation (bead size, weight)
- D. Seasonal variation (humidity versus bead color)

But **statistical variation** dominates in this case

How are the shapes of the sampling distributions different than earlier ones?

Why? Low frequency events being measured?

What would the distribution look like if frequency of beads were 50%?

The shape of the distribution is related to the physical process that generates it!

III. Clusters

What is the probability of finding X number of pennies in a given area after a "random toss"?

How do we estimate the expected frequency of such events by statistical variation?

How many pennies in a cluster would we need to see to be confident that we are not seeing them by chance?

Application of Poisson statistics to follow later.

IV. Final Words

In our universe,

Variance (V) is the qualifier of ALL conclusions in science; especially those based on measurement.

We must identify V, quantify V, & when possible control V.

Limit: Physical

Biological

Seasonal

Quantitative

Investigator

But **chance variation** cannot be controlled

So, we must factor it into all of our conclusions

Finally:

What would the mean occurrence per drawing of beads be, if I allowed you to draw bead samples consecutively until the entire sample was used?

Do you ever get to do this in real life? **HARDLY EVER!**

If we could, we wouldn't need statistics. Statistics address the problem of statistical variation in sampling and other sources of variability in measurement.