Statistics by Hand
An Introductory Course for Psychologists

Variability and variance

Version 3.0

Lovely, lovely maths (reprise)

- BEDMAS
  - Order of operations
    - **B**rackets first
    - **E**xponents next (e.g. 2^6)
    - Then **D**ivision and **M**ultiplication
    - Then **A**ddition and **S**ubtraction
  - $3 \times 2 + 6 = 12$, but $3 \times (2 + 6) = 24$
- Sigma (Σ)
  - $X = [\ 1\ 2\ 3]$
  - $\Sigma X = 6$

Revision: Sample and population

- Population:The entire set of measurements in which the investigator is interested.

- Sample: The sub-set of measurements actually collected by the investigator.

Descriptive statistics

- Measures of *central tendency*
  - Mean
  - Mode
  - Median
- Measures of *variability* around the central tendency
  - Variance
  - Standard deviation

Variability, and variance

Tutorials	65	96	84	30	27
Lectures	64	60	47	76	55

- Effects of teaching programme on exam performance.
- Is there a difference?

Variance

Tutorials	65	96	84	30	27	60.4
Lectures	64	60	47	76	55	60.4

- Means are identical, but there's an important difference: distance from the mean.
- Variance is (basically) average distance from the mean.

Variance

- By definition, variance is the average (squared) difference from the mean.

$$\sigma^2 = \frac{\sum\left(X - \overline{X}\right)^2}{N}$$

Biased estimates

- Generally, we have a *sample* but we want to conclude about a *population*.
- Unfortunately the variance of sample tends to under-estimate the population variance.
- This can be demonstrated by the following example:

Biased estimates

- First, take a population of 3 numbers:
  - 1, 2, 3
  - The mean of these numbers = 2
  - Their variance = 2/3 (Population variance)
- Now, let's try to estimate the population variance from all possible samples of size N = 2.

Biased estimates

		Mean	*Variance*
1	1	1.00	0.00
1	2	1.50	0.25
1	3	2.00	1.00
2	1	1.50	0.25
2	2	2.00	0.00
2	3	2.50	0.25
3	1	2.00	1.00
3	2	2.50	0.25
3	3	3.00	0.00
	Average	2.00	0.33

- Variance, on average, is under-estimated.
- We know the population variance is 0.67, but our estimates average to just 0.33.

Sample variance

- The “fix” is to use *N-1* rather than *N*.

$$s^2 = \frac{\sum (X - \overline{X})^2}{N - 1}$$

- Note the name change from σ to *s*.

Unbiased variance

		Mean	*Variance*	*Variance (N-1)*
1	1	1.00	0.00	0.00
1	2	1.50	0.25	0.50
1	3	2.00	1.00	2.00
2	1	1.50	0.25	0.50
2	2	2.00	0.00	0.00
2	3	2.50	0.25	0.50
3	1	2.00	1.00	2.00
3	2	2.50	0.25	0.50
3	3	3.00	0.00	0.00
	Average	2.00	0.33	0.67

- Now it's unbiased.

Calculating variance

						Mean
Lectures	64	60	47	76	55	60.4
$(X-\bar{X})$	3.6	-0.4	-13.4	15.6	-5.4	
$(X-\bar{X})^2$	12.96	0.16	179.56	243.36	29.16	

$\Sigma (X-\bar{X})^2 = 465.2$

$s^2 = 465.2 / 4 = 116.3$

Inferential statistics

- Group differences in *central tendency*
  - Wilcoxon rank-sum test
- Group differences in *variability*
  - Variance test

Homogeneity of variance

						Mean	Var.
Tutorials	65	96	84	30	27	60.4	971.3
Lectures	64	60	47	76	55	60.4	116.3

- Do the groups differ *significantly* in variance?
- Variance test:
  - Divide larger variance by smaller
  - F = 971.3 / 116.3 = 8.35
  - If F exceeds appropriate value in F-table, difference is significant.

Degrees of freedom

- “Degrees of freedom” is the number of numbers free to vary given what we know about them.
  - When we calculate variance, we have to know the mean.
  - If we know the mean, only N-1 numbers can freely vary
    - e.g. 1,2,X with a mean of 2.
    - X has to be 3.
  - Hence, each variance has d.f. of N-1.

Practice session

- Attempt problems V-1, V-2