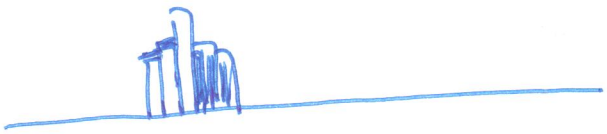
STATISTICAL INFERENCE

LIKELY?

WHY?

- variance is low.
- seem very separate.

- maybe more.



not likely

How much variance is there?

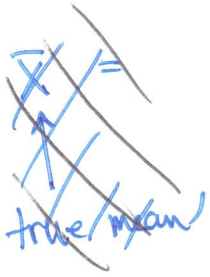
Where is the centre of the data?

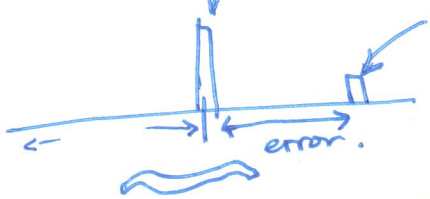
VARIATION.

Tell a class: measure this worm's length.

student	cm.
1	2.1
2	2.2
3	2.0.
4	2.1
5	1.9.
6	2.2.
7	2.0.
8.	2.1

actual length: probably about $\bar{x}$ ↑ hope, but not necessarily estimate.





* assumption: error is "normal"
→ evenly distributed on either side, & follows the usual distribution.

x_{kenny} =  + ε_{kenny}.

x_{bobby} =  + ε_{bobby}

$\bar{X}$ Estimated by $\bar{x}$ + $\bar{x}$

↑ true, mean population

standard error of the mean = $\frac{s}{\sqrt{n}}$

basic idea: more samples mean closer estimate.

Simple Experiment

Compare the perf using text prediction or not.

H_0: perf is same.

Between subjects

	text prediction	no text predict
P_1	231	
P_2	245	
P_3	236.	
P_4	241	
P_5		261
P_6		253
P_7		255
P_8		241

within subjects.
-balance ordering!

	text p.	no text p.
P_1	231	261
P_2	245	255
P_3	236	241
P_4	241	241

~~INSERT P5~~

THOUGHT EXPERIMENT:
Where does variance come from?

D TO 4.5

$$x_i = \bar{X} + (\text{effect of the condition}) + \varepsilon$$

$\bar{X}$: grand mean.

(effect of the condition): deviation

ε: measurement or individual variation

~~$x_i = \bar{x} + \text{text prediction} + \varepsilon_i$~~

~~$x_i = \bar{x} + \text{no text prediction} + \varepsilon_i$~~

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1 X_2} \cdot \sqrt{\frac{2}{n}}}$$

$S_{X_1 X_2}$ ↑ pooled standard deviation.

$$\sigma = \sqrt{\frac{\Sigma(\bar{X} - x_i)^2}{n}}$$

$$S_{X_1 X_2} = \sqrt{\frac{S_{X_1}^2 + S_{X_2}^2}{2}}$$

If t is large, the relative size of group differences is bigger than measurement (variation) ~~within~~ error.

WORKING IT:

$$t = \frac{\bar{X}_{\text{textprediction}} - \bar{X}_{\text{no textpredict}}}{S_{X_1 X_2} \cdot \sqrt{\frac{2}{n}}}$$

t value →

compare this to a t_{critical} ← "based on degrees of freedom" # & α value.

practically, this is related to # of participants

in this case, it is $n-2$.

~~WHERE DOES THIS VARIANCE COME FROM?~~ 4/ ~~45~~

SIDEBAR:

differences among different groups of objects

differences among subjects in the same groups.

If H_0 is true,

this reduces to $\frac{\text{experimental error}}{\text{experimental error}} = 1$ over time.

If H_0 is false:

Something else going on in the numerator!

$$\frac{(\text{treatment effect}) + (\text{experimental error})}{\text{experimental error}}$$

= over time > 1.

* on a given experiment, it is possible to obtain this effect!

4.5 / WHERE DOES VARIATION COME FROM

Deviation

6/.

(deviations) = (deviation due to the condition) + (deviation due to individual) + ε.

practically, we simply roll these terms together

total deviation = between groups deviation + within group deviation

If there is no deviation due to the group condition, then (total deviation = within group variation)

Another way to see this

$$\frac{\text{total deviation}}{\text{within groups deviation}} = \frac{\text{between groups deviation} + \text{within groups deviation}}{\text{within groups deviation}}.$$

If H_0 is true (i.e. between groups deviation)

$\Rightarrow$ ratio = 1.

otherwise: $\frac{bgd + wgd}{wgd} > \frac{\text{treatment effect} + \text{experimental variation}}{\text{experimental variation}}$.

the bigger this is, the bigger the overall value

~~Deviations are related to variance~~
~~Sum of Sum of squares.~~

~~Deviations~~

Sums of squares

Sum of the squared deviations from the mean.

deviation = $(\bar{X} - X_i)$

this variance is captured better w squared deviations: $(\bar{X} - X_i)^2$.

-1	-3
0	0
1	3
-1	-5
1	4
0	5
0	0
1	-4
-1	5
less variance	more variance.

▷ INSERT 7.5

$SS_{TOTAL} = SS_{BETWEEN} + SS_{WITHIN}$.

$SS_{TOTAL} = SS_{TREATMENT} + SS_{ERROR}$.

$SS_{TOTAL} = \sum(x_i - \text{grand mean})^2$.

$SS_{TREATMENT} = n(\text{mean of condition}_1 - \text{grand mean})^2 + n(\text{mean of condition}_2 - \text{grand mean})^2 + \dots$

$SS_{ERROR} = \sum(x_i - \text{mean of condition}_1)^2 + \sum(x_i - \text{mean of condition}_2)^2 + \dots$

F-test (one-way ANOVA)

Source	SS	df	MS	F
between	$SS_{BETWEEN}$	$a-1$	$\frac{SS_{between}}{df_{between}}$	$\frac{MS_{between}}{MS_{within}}$
within	SS_{WITHIN}	$a(n-1)$	$\frac{SS_{within}}{df_{within}}$	
total	SS_{TOTAL}	$an-1$		

F-value, compare to F_{crit}

R will output a p-value too

7.5/ Speed of text entry using: normal typing, iPhone text prediction, Android text predic

H_0 = all the means are equal

IV = text-typing condition

DV = speed (s).

normal	iPhone	Android
10.0	8.8	9.3
9.7	8.6	9.2
8.5	7.0	12.2
10.6	12.0	9.0
10.2	9.1	8.1
9.9	9.2	9.9

avg.

grand avg = /3