Degrees of freedom in statistics

consider equation of std.

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

↳ why divide by $n-1$!?

Has to do with degrees of freedom.

1. Intuition

- How many spines do sea urchins have?

↳ say we take a sample [213, 180, 175, 242, 287]

we can calculate sample stats from this.

← estimates of population value. ie from long-run samples

	dof
mean	n
std	$n-1$
skew	$n-2$
kurt	$n-3$

2. How does this work in descriptive stats

consider sample of $n=1$

what is mean? its the value you measured

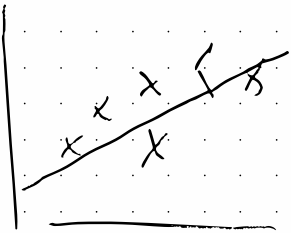
what is std? you can't calculate it !!

Note that for large sample sizes def doesn't matter ↳ (but it can, it just means we are closely approx pop. params.)

3. Application in regression

What's the min n of obs required to make this reg?

2 points.



↳ Problem, we can't make an error term w/ a line through 2 pts.

∴ dof = 2 − 1 − 1 = 0 w/ 3 pts dof = 1 ✓

but what if we have multiple vars?

dof is then $n - k - 1$ ← number of vars.

So, for k dims, we need at least

$dof = 1 = n - k - 1$

… $n = 6$ before we can calculate our error.

4. Application in χ^2 tests

Counts of sea urchins

P	25
E	22
S	13
	60

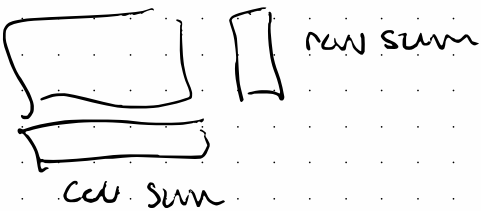
↳ Are all pts needed to know complete picture?

No. we can remove one & still infer it

$13 = 60 - 22 - 25$

$\therefore df = n - 1 = 2$

When calculating the χ^2 stat we need to right multiply the two sum vectors to get a matrix of expected vals



Thus we are left $(n-1)(n-1)$ pieces of independent information

In conclusion:

- dof encodes the redundancy / independence of information within a dataset & even describes the lower limit of information needed to calculate a ~~value~~ statistic.
  - ↳ it is widely used through stats, so its very handy to know this.