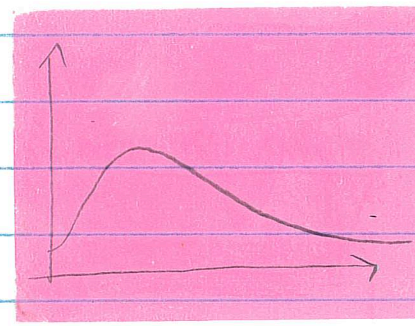
Unit 8 - x^2 homogenity and Independence tests

We use x^2 tests to observe if different variables are Independent or dependent, and homogeneous/heterogeneous

Chi-Squared Dist:
- Almost Always right skewed
- Gets more normal as sample size goes up



♡ Always do CHAD (conditions, hypothesis, alpha, DoF)

Conditions
- random says
- normal $n \cdot p \geq 5$ for all expected values
- Independent $n < .1N$

DoF

1 var: (# columns − 1)

2 vars: (rows − 1) · (columns − 1)

α = usually .05 unless says otherwise

Hypothesises

	homo	Independence
Ho	no differ. dist	no association
Ha	differ dist	association

Differences of Homogeneous and Independent

homogeneous-	Independent-
· Compares 1 variable across 2 populations	· Compares 2 variables for one population
· Asks if difference in the distribution of populations	· Asks if association between 2 variables in a population

Main order of Steps

1. Find expected values, plug in obs data L_1, exp data L_2
2. CHAD, plug lists into formula, Sto → L_3, Stats Calc vars on L_3, Look at Σx. That's the test statistic → 2nd vars, x^2cdf, compare p-value

Calculating X^2

Expected values -

- If %'s, multiply each % by the total sample size.
- If table, do $\left(\frac{\text{row total} \cdot \text{column total}}{\text{total}}\right)$

Ex)

	kayak	bike	raft	
NOC	79 observed			720
WW				
Gorge				
	123			1745

To find the expected # of kayakers at NOC, do: $\left(\frac{720 \cdot 123}{1745}\right)$

- Plug obs into L_1, expected into L_2

$$X^2 = \sum \frac{(obs-exp)^2}{exp} \quad \text{so...} \quad L3 = \frac{(L_1 - L_2)^2}{L_2}$$

Sto that into L3, then Calc L3 and $\sum x$ is test stat.
Convert to p-value: $\sum x \rightarrow$ 2nd vars, #8 x^2cdf.

If $p > \alpha$ not reject null, results not significant, no association between vars or no difference in dist. of populations

If $p < \alpha$ reject null accept alt of association or difference in dist. of populations.

Residuals - part of unit 9

- It is possible to make residual points from $L_2 - L_3$ or (obs-exp) → sto L_4 for res

To get L_3: y = slope(L_1) + y-int. This gives expected values of LSRL. This tests if the LSRL is a good fit. Do (L_1, L_4) as residual points. If no pattern, linear is good fit.