Chi-squared Goodness of Fit Test

This is a hypothesis test

1) We're going to assume something about the pop^n distribution
2) Given our assumption we know what we should see from our data.
3) We will compare our data to our assumptions and ask "If H_0 is true, how unlikely is this data?" p-value $< \alpha = .05$
4) If results are unlikely we will reject H_0 and conclude that we have statistically significant results favoring the alternative H_a.
   ↳ If results are not unlikely, we fail to reject H_0.

What if I have a discrete distribution of potential values that I believe a variable follows in a population

Ex: What day of the week do robberies occur on? (Data based FBI statistics)
↳ not weekends

	M	T	W	Th	F
P(Robbery) H_0	.2	.2	.2	.2	.2

Sum = 1
This is a prob dist.

(... each day is equal

P(Robbery) .2 .2 .2 .2 .2 —

H_0

H_0: The prop of crimes each day is equal

$$p_M = p_T = p_W = p_{Th} = p_F = .2$$

H_a: Not this: at least one of these days sees more or less crime than others.

Gather data

M	T	W	Th	F
86	92	84	89	104

$n = 455$

If H_0 is true, how many do we expect each day?

M	T	W	Th	F
91	91	91	91	91

We're going to keep n constant

$.2 \times 455 = 91$

We're going to calculate the Chi-squared test statistic

$$\chi^2 = \sum_{\text{each cell}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Observed: The observed count

Expected: The expected count

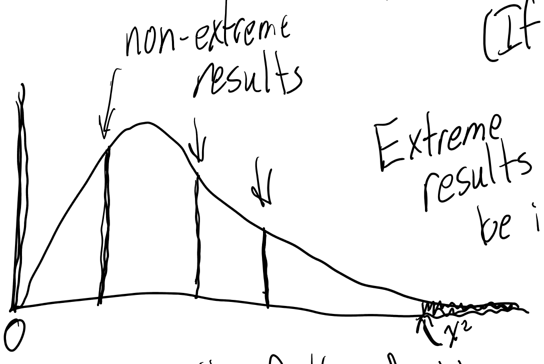
	M	T	W	Th	F
O-E	86-91 -5	92-91 1	84-91 -7	89-91 -2	104-91 13
$(O-E)^2$	25	1	49	4	169
$\frac{(O-E)^2}{E}$					

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{25}{91} + \frac{1}{91} + \frac{49}{91} + \frac{4}{91} + \frac{169}{91}$$

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{25}{91} + \frac{1}{91} + \frac{49}{91} + \frac{4}{91} + \frac{169}{91}$$

$$\approx 2.7253$$

↳ This comes from a χ^2 distribution (If H_0 is true)



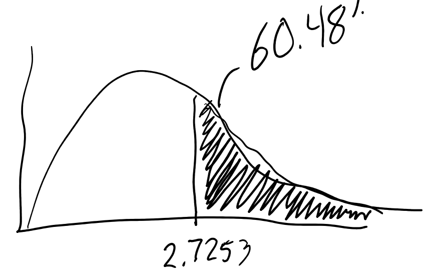
Extreme results will be in the right tail.

If data exactly followed H_0. $\chi^2 = 0$, else $\chi^2 > 0$

This distribution is defined by the degrees of freedom (df)

df = # of days − 1 (# of cells − 1) For Goodness of Fit test.

$df = 5 - 1 = 4$ $P(\chi^2_{df=4} \geq 2.7253)$



p-value 60.48% > $\alpha = 5\%$

Based on this hypothesis test, we do not find significant evidence that the rate of robberies differs by weekday. We fail to reject H_0.