Test for Independence

- two categorical variables
- Are they related?

Test for Homogeneity

- one categorical variable
- Does it act the same for different populations?

Test for Independence

- took a random sample of 300 students from some university
- for each student, recorded their status as an undergraduate student or a graduate student
- also recorded whether the majority of their classes are given in-person, remotely, or in a hybrid manner

Question: Is student status independent of class delivery type?

Test for Independence

	in-person	remote	hybrid
ugrads	121	47	31
grads	33	57	11

Test for Independence

	in-person	remote	hybrid	
ugrads	121	47	31	199
grads	33	57	11	101
	154	104	42	300

Test for Independence

	in-person	remote	hybrid	
ugrads	121	47	31	199
grads	33	57	11	101
	154	104	42	300

- estimated probability person in sample is an undergraduate is

$$\frac{199}{300}$$

Test for Independence

	in-person	remote	hybrid	
ugrads	121	47	31	199
grads	33	57	11	101
	154	104	42	300

- estimated probability person in sample is taking a majority of in-person courses is $\frac{154}{300}$

Test for Independence

	in-person	remote	hybrid	
ugrads	121	47	31	199
grads	33	57	11	101
	154	104	42	300

- If delivery type and student type are independent, the probability we fall in this square should be $\frac{199}{300} \cdot \frac{154}{300}$

Test for Independence

	in-person	remote	hybrid	
ugrads	121	47	31	199
grads	33	57	11	101
	154	104	42	300

- The expected number of people, out of the total of 300 sampled, who fall in that square should be

$$\frac{199}{300} \cdot \frac{154}{300} \cdot 300 = \frac{(199)(154)}{300} \approx \boxed{102.15}$$

Test for Independence

- Under the assumption of independence, the expected counts are

	in-person	remote	hybrid	
ugrads	$\frac{(199)(154)}{300}$	$\frac{(199)(104)}{300}$	$\frac{(199)(42)}{300}$	**199**
grads	$\frac{(101)(154)}{300}$	$\frac{(101)(104)}{300}$	$\frac{(101)(42)}{300}$	**101**
	154	**104**	**42**	**300**

Test for Independence

H_0 : Student status and class delivery type are independent

H_1 : Student status and class delivery type are not independent

Test Statistic:

$$W := \sum_i \frac{(O_i - E_i)^2}{E_i} \sim \chi^2(?)$$

Test for Independence

Degrees of freedom parameter is

(number of rows – 1)(number of cols – 1)

Reject H_0, in favor of H_1 if W is “large”.

Test for Independence

For our example, the test statistic is

$$W \approx 32.193$$

The critical value ($\alpha = 0.10$) is

$$\chi^2_{0.10,2} = 4.60517$$

Reject H_0!

Test for Independence

There is sufficient evidence in the data to conclude that student status and class delivery type are dependent at level 0.10.

Test for Homogeneity

- took a random sample of 199 undergraduate students
- took an independent random sample of 101 graduate students
- For each sample, recorded whether the majority of their classes are given in-person, remotely, or in a hybrid manner.

Test for Homogeneity

Question: Is the distribution of in-person, to remote, to hybrid the same for both groups?

	in-person	remote	hybrid	
ugrads	121	47	31	199
grads	33	57	11	101
	154	104	42	300

Test for Homogeneity

H_0 : The distribution of class type is the same for undergraduate and graduate students.

H_1 : Not H_0.

Test for Homogeneity

	in-person	remote	hybrid	
ugrads	121	47	31	199
grads	33	57	11	101
	154	104	42	300

- Overall "in-person" probability is estimated to be

$$\frac{154}{300}$$

Test for Homogeneity

	in-person	remote	hybrid	
ugrads	121	47	31	199
grads	33	57	11	101
	154	104	42	300

- Since there are 199 undergrads, the expected number in the in-person group under H_0 is $\frac{154}{300} \cdot 199$

Test for Homogeneity

Expected number under the assumption of independence:

$$\frac{199}{300} \cdot \frac{154}{300} \cdot 300$$

Expected number under the assumption of homogeneity:

$$\frac{154}{300} \cdot 199$$

The Same!