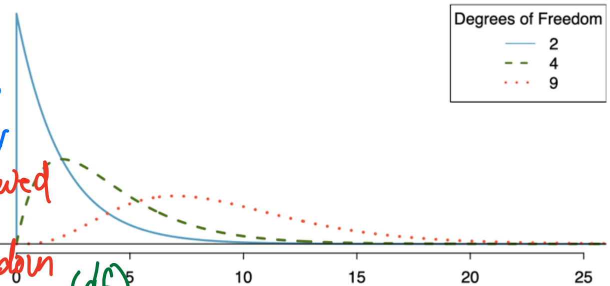
MAT1372, Classwork21, Fall2025

6.3 Testing for Goodness of Fit Using Chi-square

1. The Chi-square Distribution



When to use it? It's used to characterize data sets and statistic that are always positive and typically right skewed

How to describe it?

It has just one parameter called degree of freedom (df)

2. The feature of chi-square distribution.

Distribution's center: The center becomes larger when df is larger.

Mean: The mean of each distribution is equal to distribution's df.

Variability (spread): The variability increases as df increases

Shape: The distribution is very strong skewed for df = 2, and then it becomes more symmetric for the larger df = 4 or 9

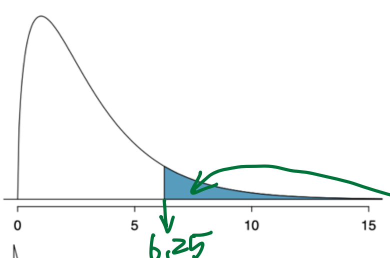
3. Use a chi-square probability table to find tail area. (p. 416-417)

① $P(X^2 > a)$

② Probability

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	2	2.41	**3.22**	**4.61**	5.99	7.82	9.21	10.60	13.82
	3	*3.66*	*4.64*	6.25	*7.81*	*9.84*	*11.34*	*12.84*	*16.27*
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52
	6	7.23	8.56	10.64	12.59	15.03	16.81	18.55	22.46
	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32

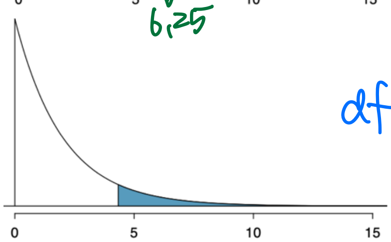
Figure C.5: A section of the chi-square table. A complete table is in Appendix C.3.



(a) df = 3 and an upper shaded tail starting at 6.25. Find the shaded area.

If we use table above,

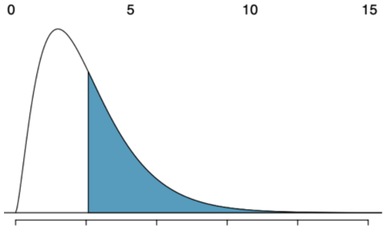
$P(X > 6.25) = 0.1$



(b) df = 2 and an upper shaded tail starting at 4.3. Find the shaded area.

df = 2, $P(X > 4.3) = ?$

$0.1 < P(X > 4.3) < 0.2$



(c) df = 5 and an upper shaded tail starting at 5.1. Find the shaded area.

df = 5. $P(X > 5.1) > 0.3$

4. When to use a chi-square test:

Given a sample of cases that can be classified into several groups, determine if the sample is representative of the general population

5. Example to use a chi-square test for one-way table.

Consider data from a random sample of 275 jurors in a small county. Jurors identified their racial group, as shown in the figure below, and we would like to determine if these jurors are racially representative of the population. If the jury is representative of the population, then the proportions in the sample should roughly reflect the population of eligible jurors, i.e. registered voters.

Race	White	Black	Hispanic	Other	Total
observation Representation in juries	O_W 205	O_B 26	O_H 25	O_O 19	275
Registered voters	0.72	0.07	0.12	0.09	1.00
Expected counts	E_W 198	E_B 19.25	E_H 33	E_O 24.75	

(a) If the individuals in this small county are randomly selected to serve on a jury, about how many of the 275 people would we expect to be White, Black, Hispanic, and other, respectively?

Expected count of White: $275 \times 0.72 = 198$, Expected count of Black $= 275 \cdot 0.07 = 19.25$

Expected count of Hispanic: $275 \times 0.12 = 33$ Expected count of other $= 275 \cdot 0.09 = 24.75$

(b) The sample proportion was not a precise match with the expected count for any ethnic group. We need to test whether the differences are strong enough to provide convincing evidence that the jurors are not a random sample. These ideas can be organized into **hypotheses**:

H_0: The jurors are a random sample, i.e. there is no racial bias, and the observed counts reflect natural sampling fluctuation

H_A: The jurors are not randomly sampled, there is racial bias.

(c) Find a p-value for a **chi-square distribution**.

(1) How many categories (k) were there in the juror example? K=4 (White, Black, Hispanic, other)

(2) What is the degree of freedom df in this case? $df = K-1 = 3$

(3) Find the test statistic X^2 (which follows a chi-square distribution with df (*why? check it later*)):

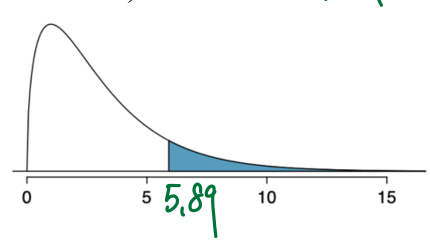
$$X^2 = \frac{(O_w-E_w)^2}{E_w} + \frac{(O_b-E_b)^2}{E_b} + \frac{(O_h-E_h)^2}{E_h} + \frac{(O_o-E_o)^2}{E_o} =$$

$$\frac{(205-198)^2}{198} + \frac{(26-19.25)^2}{19.25} + \frac{(25-33)^2}{33} + \frac{(19-24.75)^2}{24.75} = (0.5)^2 + (1.54)^2 + (-1.39)^2 + (-1.16)^2 = 5.89$$

(4) Find the p-value (the area of an upper tail starting at the test statistic 5.89 with df = 3):

p-value $= P(X^2 > 5.89)$ when $df=3$

which is between 0.1 and 0.2 (0.1171)



(d) Can we reject H_0 based on the p-value we found in (c) 4.?

If $\alpha = 0.05$, $P(X^2 > 5.89)$ is more than α, we do not reject H_0