RenR 480/711

Tests for proportions and binominal data

Binomial data

Two classes expressed in binary format (0 and 1)

For example:

Tossing a coin: Head/tail

Survival: yes/no

Presence/Absence

Susceptible/Resistant

Contingency tables

→Tests for proportions are based on contingency tables

→We have to convert the binomial data into proportions (p) by summing up the responses for each treatment/group

ID	Species	Survival
S1	SpecA	Y
S2	SpecA	N
S3	SpecA	N
S4	SpecB	Y
S5	SpecB	N
S6	SpecB	N
…	…	…
S40	SpecA	Y

Groups/	Survival			
Treatments	Yes	No	n	P Yes
Species A	8	12	20	0.4
Species B	16	4	20	0.8

$$Proportion\ Yes = \frac{16}{20} = 0.8$$

Z-Test for proportions

Binomial equivalent to a T-Test

One sample, one tailed test

What is the probability that the population proportion falls above/below a threshold?

$$H_o: p < threshold$$
$$H_o: p > threshold$$

Question:

Does species B have a survival rate greater than 50%?

Arbitrary threshold

t=threshold

$$z = \frac{signal}{noise} = \frac{p-t}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.8-0.5}{\sqrt{\frac{0.8*(1-0.8)}{20}}}$$

$$z = \underline{3.35}$$

Groups/ Treatments	Survival Yes	No	n	p
Species A	8	12	20	0.4
Species B	16	4	20	0.8

pnorm(3.35) = 0.999
right tail, but we need the left tail

p-value = 1-pnorm(3.35) = 0.0004

Answer:

Species B has a survival rate significantly greater than 50% (z=3.35, n=20, p=0.0004).

Z-Test for proportions

Two sample, two tailed test

Do species A & B have different survival rates?

$$H_o: p_1 = p_2$$

$$z = \frac{signal}{noise} = \frac{difference}{pooled\ SE}$$

$$z = \frac{(p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})} \times \sqrt{\frac{n_1+n_2}{n_1 \times n_2}}}$$

$$\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$\hat{p} = \frac{20 * 0.4 + 20 * 0.8}{20 + 20}$$

$$\hat{p} = 0.6$$

$$z = \frac{(0.8 - 0.4)}{\sqrt{0.6(1-0.6)} \times \sqrt{\frac{20+20}{20 \times 20}}}$$

$z = 2.67$

Groups/ Treatments	Survival Yes	No	n	p
Species A	8	12	20	0.4
Species B	16	4	20	0.8

pnorm(2.67) = 0.999
right tail, but we need the left tail

p-value = 1-pnorm(2.67) = 0.004

Answer:

Species A and B have significantly different survival rates (z=2.67, n=40, p=0.004).

Chi-square test for proportions (χ^2)

Binomial equivalent to an Anova

Comparison of two or more groups

Is there a difference in survival rates among species?

$$H_o: p_A = p_B = p_C$$

$$\chi^2 = \frac{signal}{noise} = \sum_i^n \frac{(observed - expected)^2}{expected}$$

Steps:

1. Calculate totals
2. Calculate expected proportions
3. Calculate expected outcome for each treatment
4. Calculate χ^2

Groups/ Treatments	Survival Yes	No	n	p
Species A	8	12	20	0.4
Species B	16	4	20	0.8
Species C	24	16	40	0.6
Totals:	48	32	80	
Expected p:	0.6	0.4		

Groups/ Treatments	Survival Yes	No
Species A	12	8
Species B	12	8
Species C	24	16

Expected yes/no

$$\chi^2 = \frac{(8-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(24-24)^2}{24} + \frac{(12-8)^2}{8} + \frac{(4-8)^2}{8} + \frac{(16-16)^2}{16}$$

$$\chi^2 = 6.667$$

p-value = 1-pchisq(6.667,2) = 0.0356

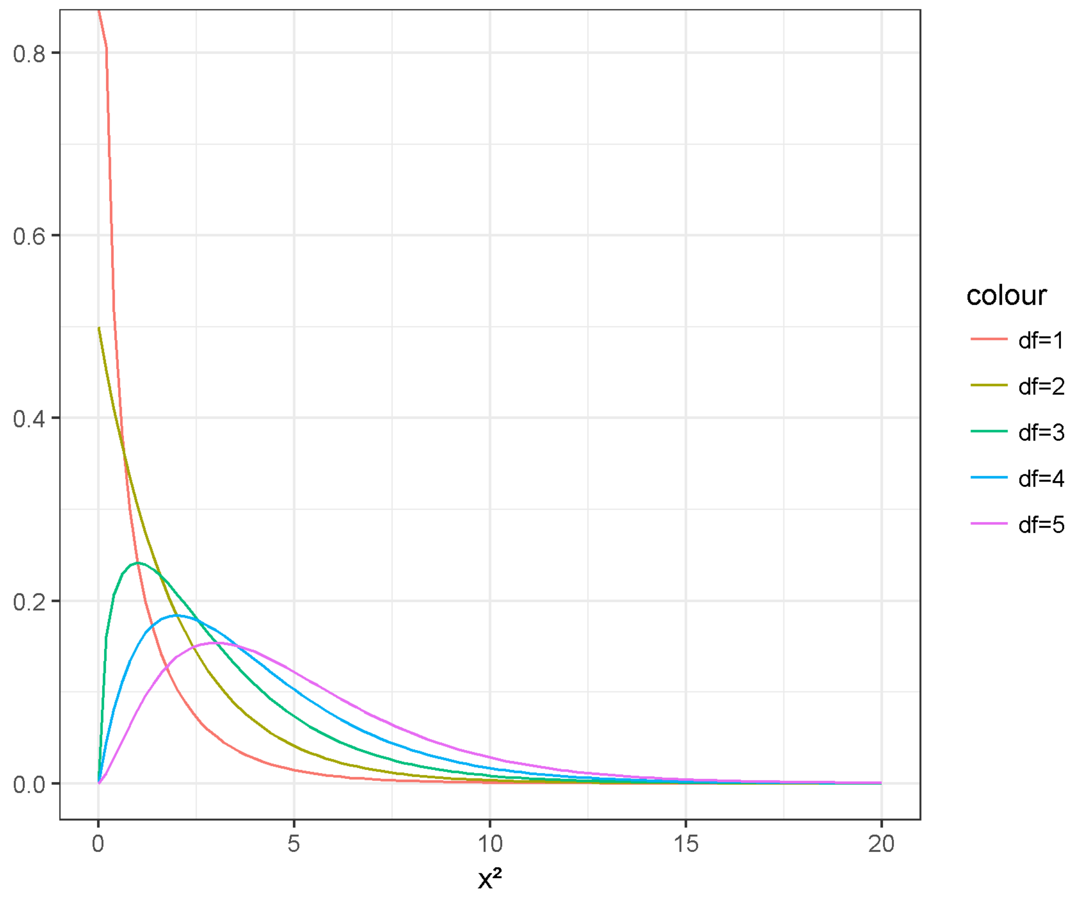
Left tail!

df=n of treatment-1

Answer:

There is a significant difference in survival rates among species ($\chi^2(2,80)$=6.667, p=0.0356).

Chi-square distribution



The more groups, we higher χ^2 we need to reach the threshold of 0.05

qchisq(0.95,1) = 3.84

qchisq(0.95,2) = 5.99

qchisq(0.95,3) = 7.81