Lecture 18 - χ^2 Tests

Sta102 / BME102

April 6, 2016

Colin Rundel

χ^2 test of GOF

Weldon's dice

- Walter Frank Raphael Weldon (1860 - 1906), was an English evolutionary biologist and a founder of biometry. He was the joint founding editor of Biometrika, with Francis Galton and Karl Pearson.
- In 1894, he rolled 12 dice 26,306 times, and recorded the number of 5s or 6s (which he considered to be a success).
    - 5s or 6s occurred more often than expected - Pearson hypothesized that since inexpensive dice have hollowed-out pips the face with 6 pips is lighter than its opposing face, which has only 1 pip.

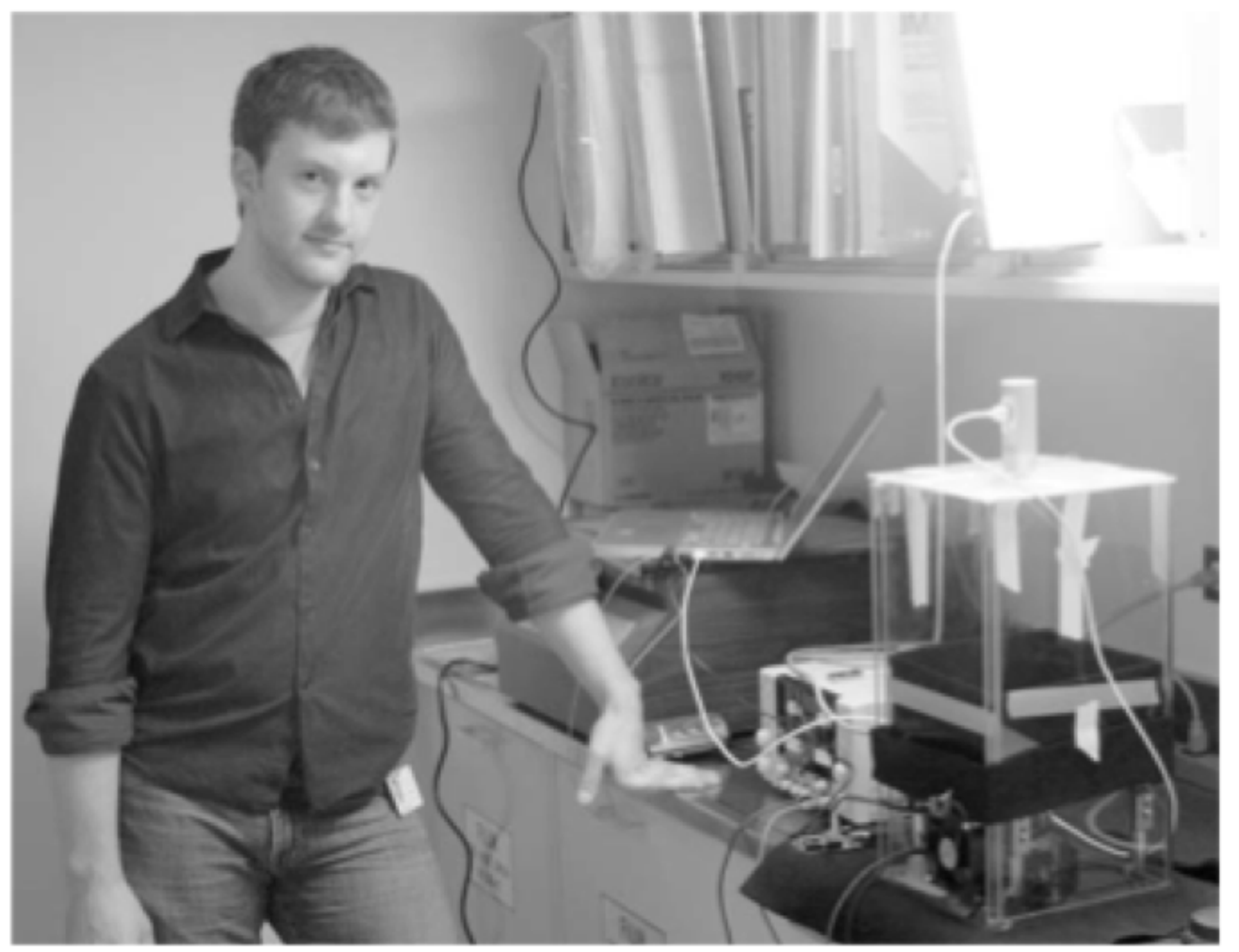
Weldon's Data

Number of Successes	Observed Frequency	Theoretical Frequency, p = 1/3	Deviation
0	185	203	-18
1	1149	1216	-67
2	3265	3345	-80
3	5475	5576	-101
4	6114	6273	-159
5	5194	5018	176
6	3067	2927	140
7	1331	1255	76
8	403	392	11
9	105	87	18
10	14	13	1
11	4	1	3
12	0	0	0
Total	**26,306**	**26,306**	

Labby's dice

In 2009, Zacariah Labby (U of Chicago), repeated Weldon's experiment using a homemade dice-throwing, pip counting machine.

http://www.youtube.com/watch?v=95EErdouO2w

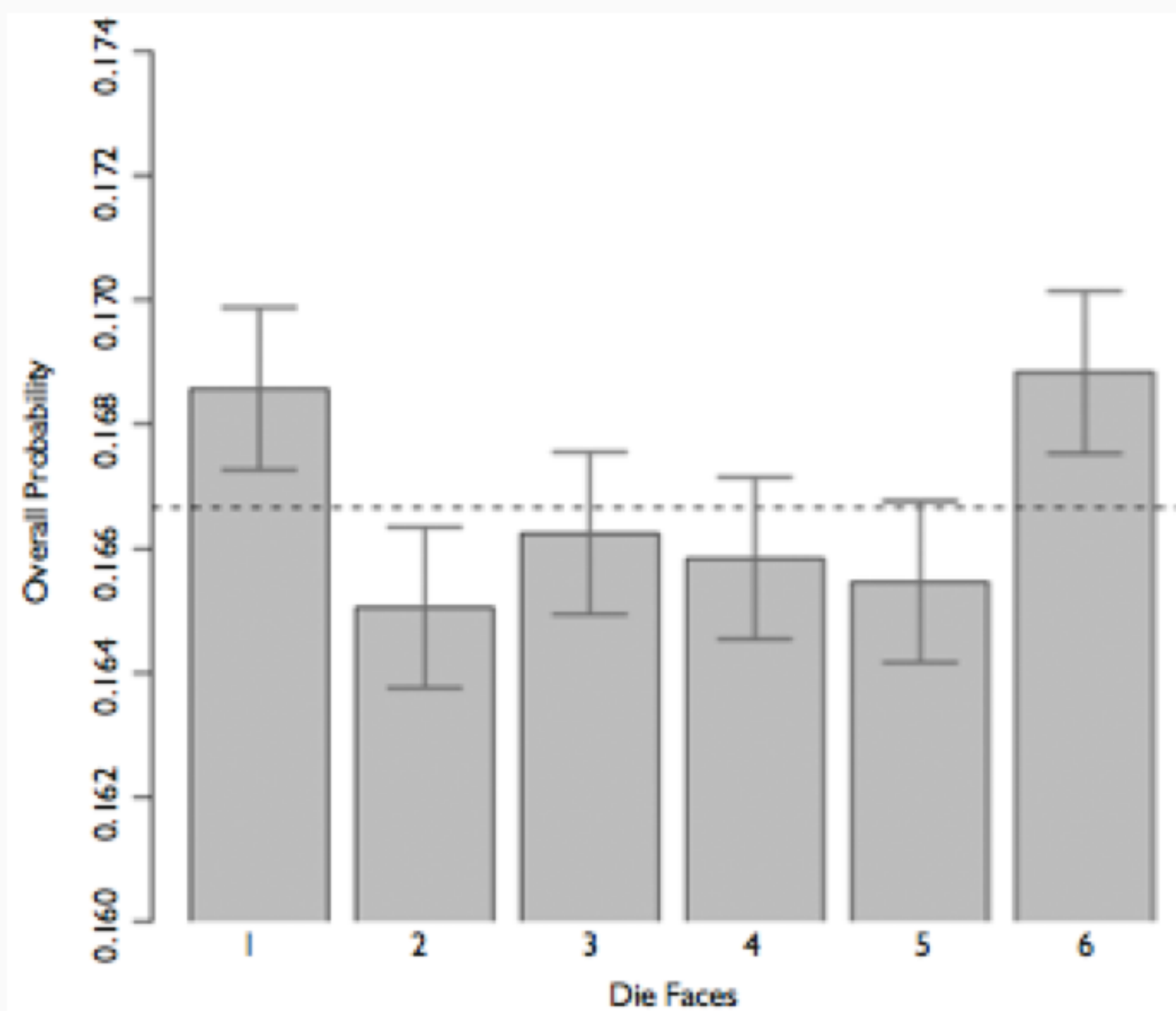


- Able to recreate Weldon's experiment in about six days (with complete data).
- Further reading: *http://news.uchicago.edu/static/newsengine/pdf/labby09dice.pdf*

Labby's dice (cont.)

Labby did not actually observe the same phenomenon that Weldon observed (higher frequency of 5s and 6s).

Automation allowed Labby to collect more data than Weldon did in 1894, instead of recording "successes" and "failures", Labby recorded the individual number of pips on each die.



Summarizing Labby's results

Labby rolled 12 dice 26,306 times. If each side is equally likely to come up, how many 1s, 2s, $\cdots$, 6s would he expect to have observed?

Summarizing Labby's results

Labby rolled 12 dice 26,306 times. If each side is equally likely to come up, how many 1s, 2s, $\cdots$, 6s would he expect to have observed?

$$\frac{12 \times 26,306}{6} = 52612$$

Summarizing Labby's results

Labby rolled 12 dice 26,306 times. If each side is equally likely to come up, how many 1s, 2s, $\cdots$, 6s would he expect to have observed?

$$\frac{12 \times 26,306}{6} = 52612$$

Outcome	Observed	Expected
1	53,222	52,612
2	52,118	52,612
3	52,465	52,612
4	52,338	52,612
5	52,244	52,612
6	53,285	52,612
Total	315,672	315,672

Setting the hypotheses

Do these data provide convincing evidence to suggest an inconsistency between the observed and expected counts?

Setting the hypotheses

Do these data provide convincing evidence to suggest an inconsistency between the observed and expected counts?

H_0: There is no inconsistency between the observed and the expected counts. *The observed counts follow the same distribution as the expected counts.*

Setting the hypotheses

Do these data provide convincing evidence to suggest an inconsistency between the observed and expected counts?

H_0: There is no inconsistency between the observed and the expected counts. *The observed counts follow the same distribution as the expected counts.*

$H_0: p_i = 1/6 \quad \forall \, i = \{1 \dots 6\}$

H_A: There is an inconsistency between the observed and the expected counts. *The observed counts do not follow the same distribution as the expected counts.* (There is a bias in which side comes up on the roll of a die)

Evaluating the hypotheses

To evaluate these hypotheses, we quantify how different the observed counts are from the expected counts.

Evaluating the hypotheses

To evaluate these hypotheses, we quantify how different the observed counts are from the expected counts.

Large deviations from what would be expected based on sampling variation (chance) alone provide strong evidence against the null hypothesis.

Evaluating the hypotheses

To evaluate these hypotheses, we quantify how different the observed counts are from the expected counts.

Large deviations from what would be expected based on sampling variation (chance) alone provide strong evidence against the null hypothesis.

This is called a *goodness of fit* test since we're evaluating how well the observed data fit the expected distribution.

Anatomy of a test statistic

The general form of the test statistics we've seen this far is

$$\frac{\text{point estimate} - \text{null value}}{\text{SE of point estimate}}$$

Anatomy of a test statistic

The general form of the test statistics we've seen this far is

$$\frac{\text{point estimate} - \text{null value}}{\text{SE of point estimate}}$$

This construction is based on

1. measure the effect size - i.e. the difference between the point estimate and the null value, and
2. standardizing that difference using the standard error of the point estimate.

Anatomy of a test statistic

The general form of the test statistics we've seen this far is

$$\frac{\text{point estimate} - \text{null value}}{\text{SE of point estimate}}$$

This construction is based on

1. measure the effect size - i.e. the difference between the point estimate and the null value, and
2. standardizing that difference using the standard error of the point estimate.

These two ideas will help in the construction of an appropriate test statistic for count data.

χ^2 statistic

When dealing with counts and investigating how far the observed counts are from the expected counts, we will use a new test statistic called the *χ^2 (chi-squared) statistic*.

χ^2 statistic

When dealing with counts and investigating how far the observed counts are from the expected counts, we will use a new test statistic called the χ^2 *(chi-squared) statistic*.

The χ^2 statistic is defined to be

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{(E_i - O_i)^2}{E_i}$$

where k = total number of cells/categories

Calculating the χ^2 statistic

Outcome	Observed	Expected	$\frac{(O-E)^2}{E}$
1	53,222	52,612	$\frac{(53{,}222-52{,}612)^2}{52{,}612} = 7.07$

Calculating the χ^2 statistic

Outcome	Observed	Expected	$\frac{(O-E)^2}{E}$
1	53,222	52,612	$\frac{(53{,}222-52{,}612)^2}{52{,}612} = 7.07$
2	52,118	52,612	$\frac{(52{,}118-52{,}612)^2}{52{,}612} = 4.64$

Calculating the χ^2 statistic

Outcome	Observed	Expected	$\frac{(O-E)^2}{E}$
1	53,222	52,612	$\frac{(53{,}222-52{,}612)^2}{52{,}612} = 7.07$
2	52,118	52,612	$\frac{(52{,}118-52{,}612)^2}{52{,}612} = 4.64$
3	52,465	52,612	$\frac{(52{,}465-52{,}612)^2}{52{,}612} = 0.41$

Calculating the χ^2 statistic

Outcome	Observed	Expected	$\frac{(O-E)^2}{E}$
1	53,222	52,612	$\frac{(53{,}222-52{,}612)^2}{52{,}612} = 7.07$
2	52,118	52,612	$\frac{(52{,}118-52{,}612)^2}{52{,}612} = 4.64$
3	52,465	52,612	$\frac{(52{,}465-52{,}612)^2}{52{,}612} = 0.41$
4	52,338	52,612	$\frac{(52{,}338-52{,}612)^2}{52{,}612} = 1.43$

Calculating the χ^2 statistic

Outcome	Observed	Expected	$\frac{(O-E)^2}{E}$
1	53,222	52,612	$\frac{(53{,}222-52{,}612)^2}{52{,}612} = 7.07$
2	52,118	52,612	$\frac{(52{,}118-52{,}612)^2}{52{,}612} = 4.64$
3	52,465	52,612	$\frac{(52{,}465-52{,}612)^2}{52{,}612} = 0.41$
4	52,338	52,612	$\frac{(52{,}338-52{,}612)^2}{52{,}612} = 1.43$
5	52,244	52,612	$\frac{(52{,}244-52{,}612)^2}{52{,}612} = 2.57$

Calculating the χ^2 statistic

Outcome	Observed	Expected	$\frac{(O-E)^2}{E}$
1	53,222	52,612	$\frac{(53{,}222-52{,}612)^2}{52{,}612} = 7.07$
2	52,118	52,612	$\frac{(52{,}118-52{,}612)^2}{52{,}612} = 4.64$
3	52,465	52,612	$\frac{(52{,}465-52{,}612)^2}{52{,}612} = 0.41$
4	52,338	52,612	$\frac{(52{,}338-52{,}612)^2}{52{,}612} = 1.43$
5	52,244	52,612	$\frac{(52{,}244-52{,}612)^2}{52{,}612} = 2.57$
6	53,285	52,612	$\frac{(53{,}285-52{,}612)^2}{52{,}612} = 8.61$

Calculating the χ^2 statistic

Outcome	Observed	Expected	$\frac{(O-E)^2}{E}$
1	53,222	52,612	$\frac{(53{,}222-52{,}612)^2}{52{,}612} = 7.07$
2	52,118	52,612	$\frac{(52{,}118-52{,}612)^2}{52{,}612} = 4.64$
3	52,465	52,612	$\frac{(52{,}465-52{,}612)^2}{52{,}612} = 0.41$
4	52,338	52,612	$\frac{(52{,}338-52{,}612)^2}{52{,}612} = 1.43$
5	52,244	52,612	$\frac{(52{,}244-52{,}612)^2}{52{,}612} = 2.57$
6	53,285	52,612	$\frac{(53{,}285-52{,}612)^2}{52{,}612} = 8.61$
Total	315,672	315,672	24.73

Why square?

Squaring the difference between the observed and the expected outcome does two things:

Why square?

Squaring the difference between the observed and the expected outcome does two things:

- Any standardized difference that is squared will now be positive.

Why square?

Squaring the difference between the observed and the expected outcome does two things:

- Any standardized difference that is squared will now be positive.
- Differences that already looked unusual will become much larger after being squared.

Why square?

Squaring the difference between the observed and the expected outcome does two things:

- Any standardized difference that is squared will now be positive.
- Differences that already looked unusual will become much larger after being squared.

Where have we seen this before?

Variance ANOVA etc

Conditions for the χ^2 test

1. *Independence:* Each case that contributes a count to the table must be independent of all the other cases in the table.

Conditions for the χ^2 test

1. *Independence:* Each case that contributes a count to the table must be independent of all the other cases in the table.
2. *Sample size:* Each particular scenario (i.e. cell) must have at least 5 *expected* cases.

Conditions for the χ^2 test

1. *Independence:* Each case that contributes a count to the table must be independent of all the other cases in the table.
2. *Sample size:* Each particular scenario (i.e. cell) must have at least 5 *expected* cases.

Failing to check conditions may unintentionally effect the test's error rates.

The χ^2 distribution

The χ^2 distribution has one parameter, *df*, which influences the shape, center, and spread of the distribution.

The χ^2 distribution

The χ^2 distribution has one parameter, *df*, which influences the shape, center, and spread of the distribution.

- For a goodness of fit test the degrees of freedom is the number of categories - 1 ($df = k - 1$).

The χ^2 distribution

The χ^2 distribution has one parameter, *df*, which influences the shape, center, and spread of the distribution.

- For a goodness of fit test the degrees of freedom is the number of categories - 1 ($df = k - 1$).

So far we've seen three other continuous distributions:

- Normal - unimodal and symmetric with two parameters: μ (center) and σ^2 (spread)
- T - unimodal and symmetric with one parameter: *df* (spead, kurtosis)
- T - unimodal and non-symmetric with two parameters: df_1, df_2

The χ^2 distribution (Theory)

Where does the χ^2 distribution come from?

The χ^2 distribution (Theory)

Where does the χ^2 distribution come from?

- Like the T and the Z, the χ^2 distribution is an example of a continuous probability distribution (it is actually a special case of another distribution called the Gamma distribution)

The χ^2 distribution (Theory)

Where does the χ^2 distribution come from?

- Like the T and the Z, the χ^2 distribution is an example of a continuous probability distribution (it is actually a special case of another distribution called the Gamma distribution)
- If we define $Z \sim N(0, 1)$ then the quantity,

$$Z^2 \sim \chi^2_{df=1}$$

The χ^2 distribution (Theory)

Where does the χ^2 distribution come from?

- Like the T and the Z, the χ^2 distribution is an example of a continuous probability distribution (it is actually a special case of another distribution called the Gamma distribution)

$$\frac{(O_i - E_i)^2}{E_i} \sim Z^2$$

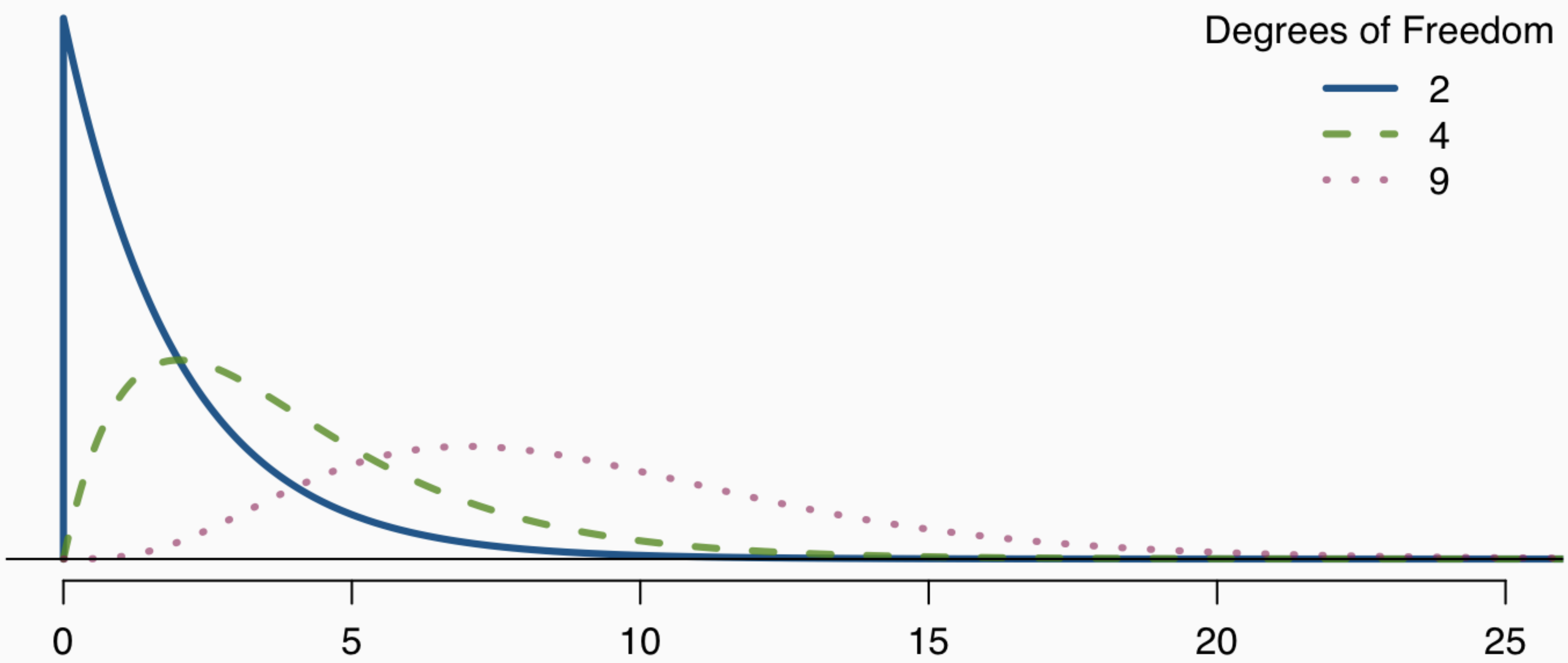
- If we define $Z \sim N(0, 1)$ then the quantity,

$$Z^2 \sim \chi^2_{df=1}$$

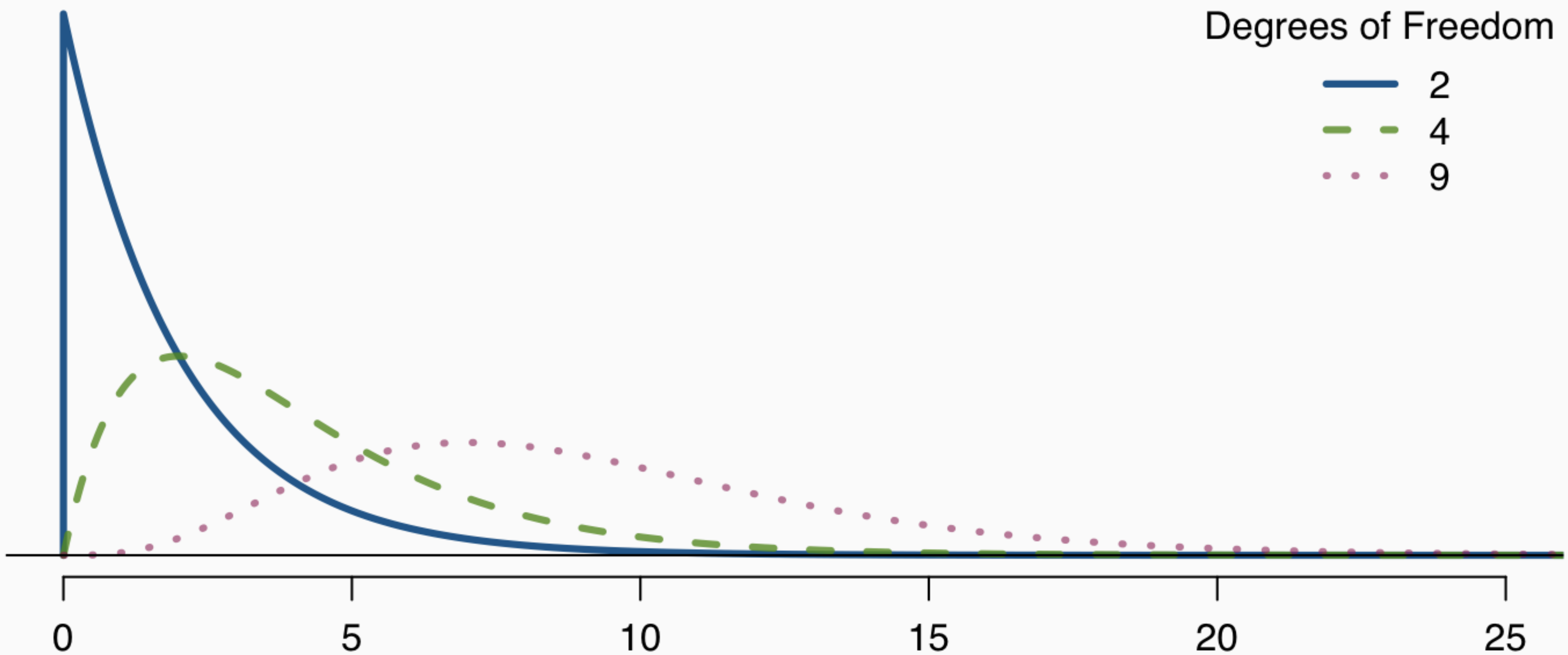
- If we have $Z_1, \ldots, Z_n \sim N(0, 1)$ then the quantity,

$$\underbrace{{Z_1}^2 + {Z_2}^2 + \ldots + {Z_n}^2}_{} \overset{iid}{\sim} \chi^2_{df=n}$$

The χ^2 distribution (cont.)



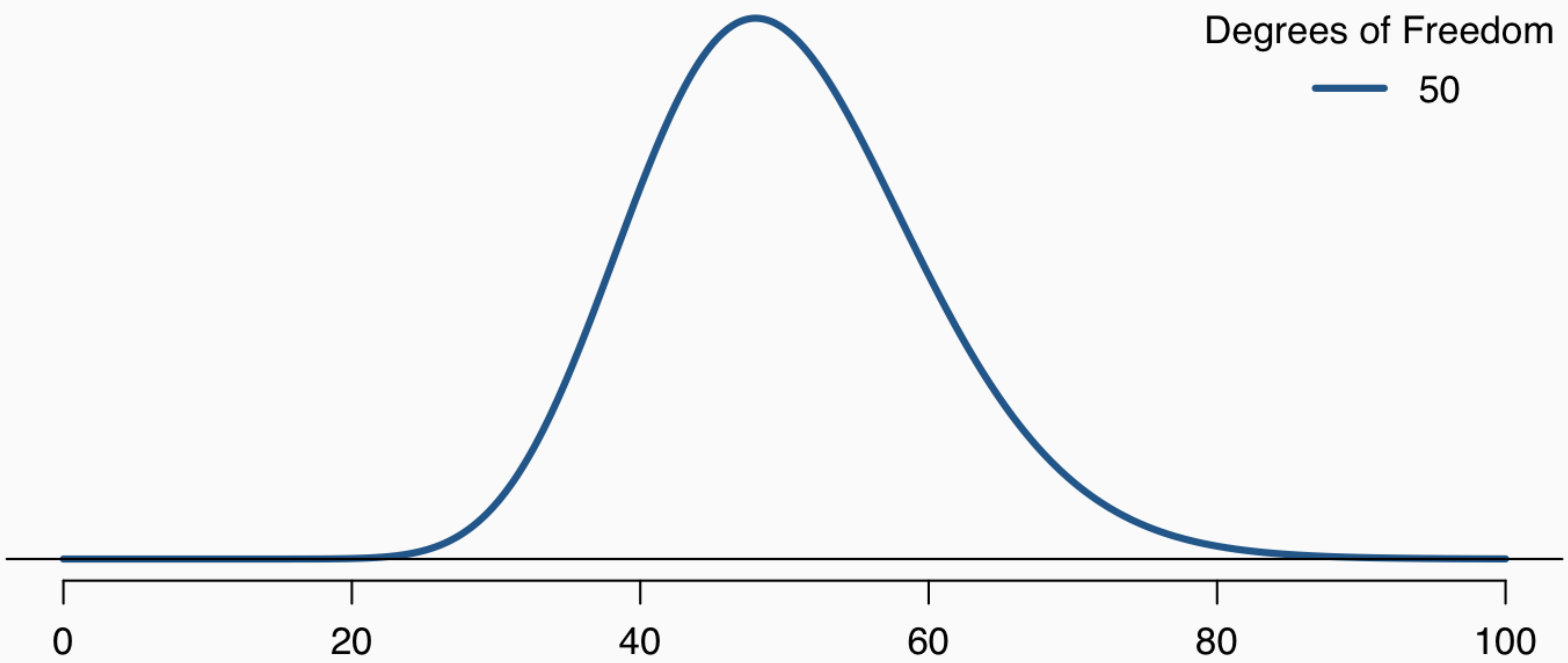
The χ^2 distribution (cont.)



As the df increases:

- the center of the χ^2 distribution increases
- the variability of the χ^2 distribution increases

The χ^2 distribution (cont.)



Also, for large *df* the χ^2 distribution converges to the normal distribution with

$$\mathcal{N}(\mu = df,\ \sigma^2 = 2\,df).$$

Finding areas under the χ^2 curve

- p-value = tail area under the χ^2 distribution (as usual)

Finding areas under the χ^2 curve

- p-value = tail area under the χ^2 distribution (as usual)
- For this we can use technology, or a χ^2 *table*.

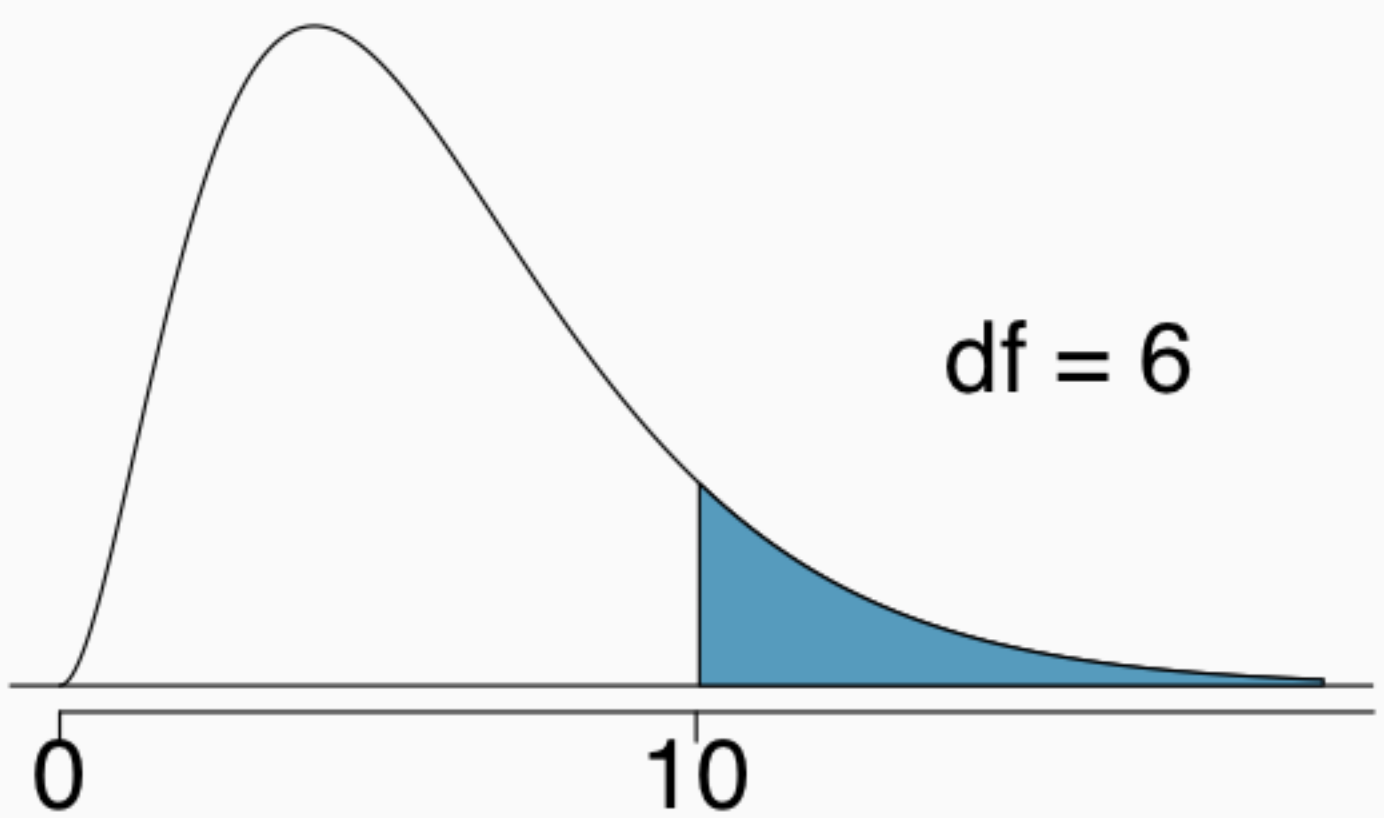
Finding areas under the χ^2 curve

- p-value = tail area under the χ^2 distribution (as usual)
- For this we can use technology, or a χ^2 *table*.
- This table is similar to the *t* table, it provides upper tail probabilities.

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	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
	...								

Finding areas under the χ^2 curve (cont.)

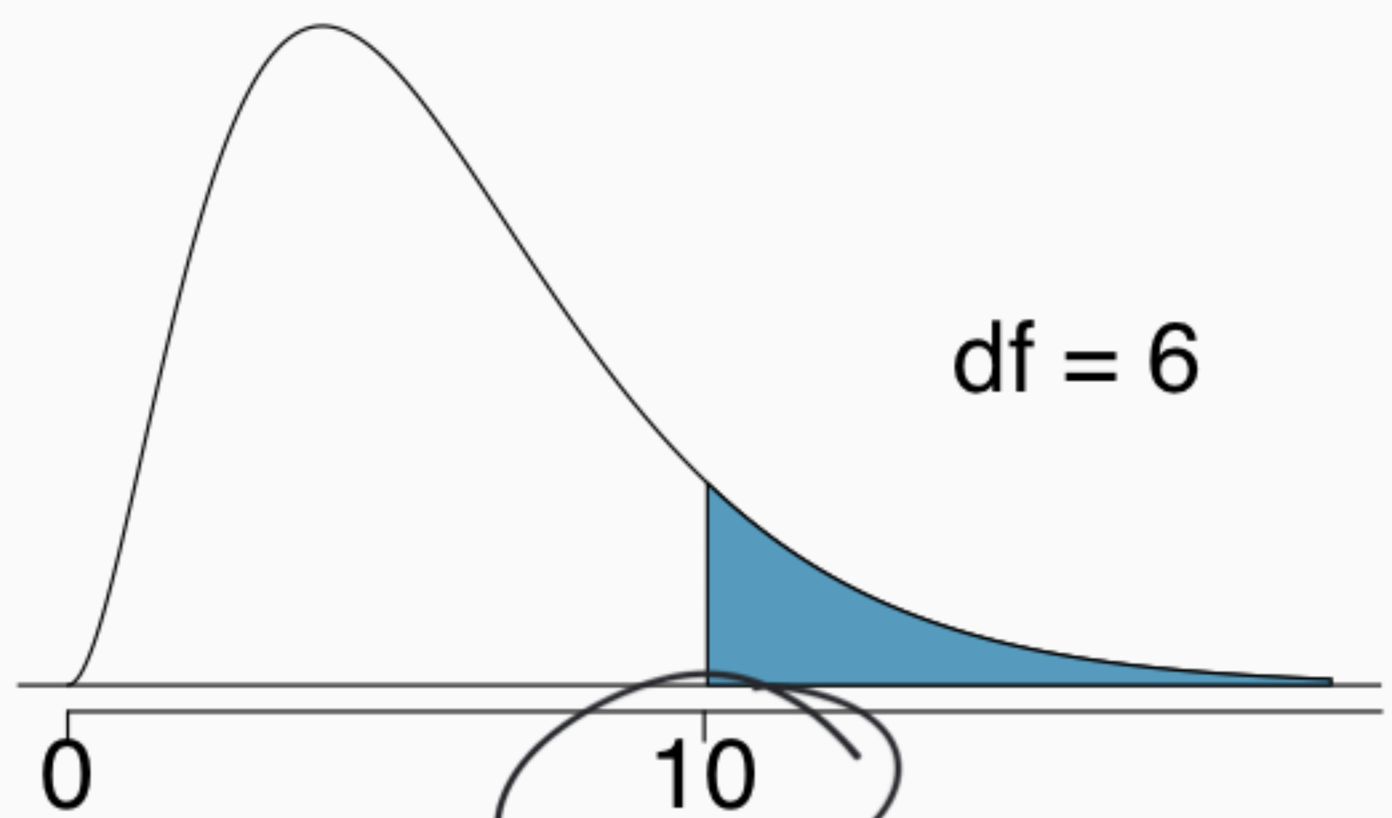
Estimate the shaded area under the χ^2 curve with $df = 6$.



Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	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

Finding areas under the χ^2 curve (cont.)

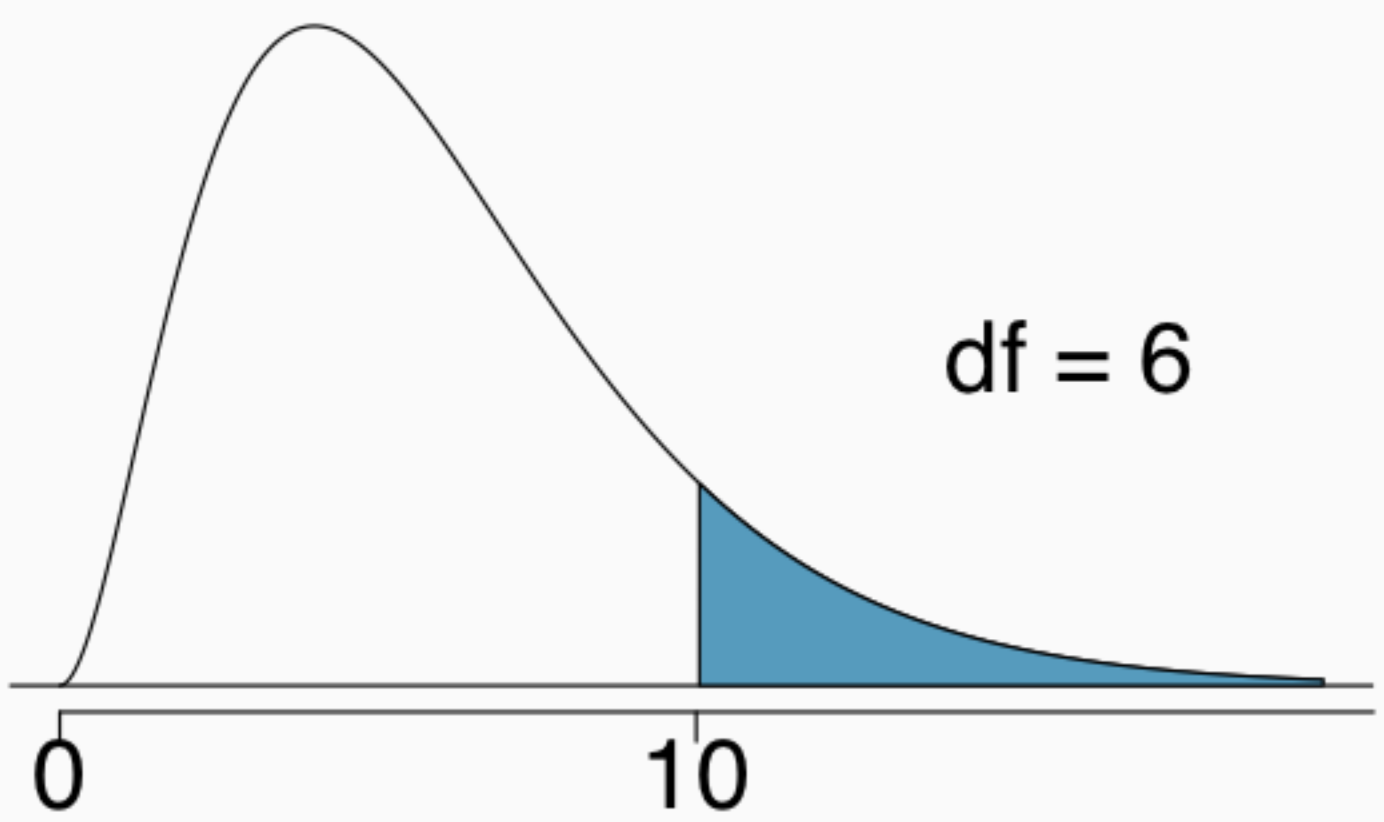
Estimate the shaded area under the χ^2 curve with $df = 6$.



Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df 1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
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

Finding areas under the χ^2 curve (cont.)

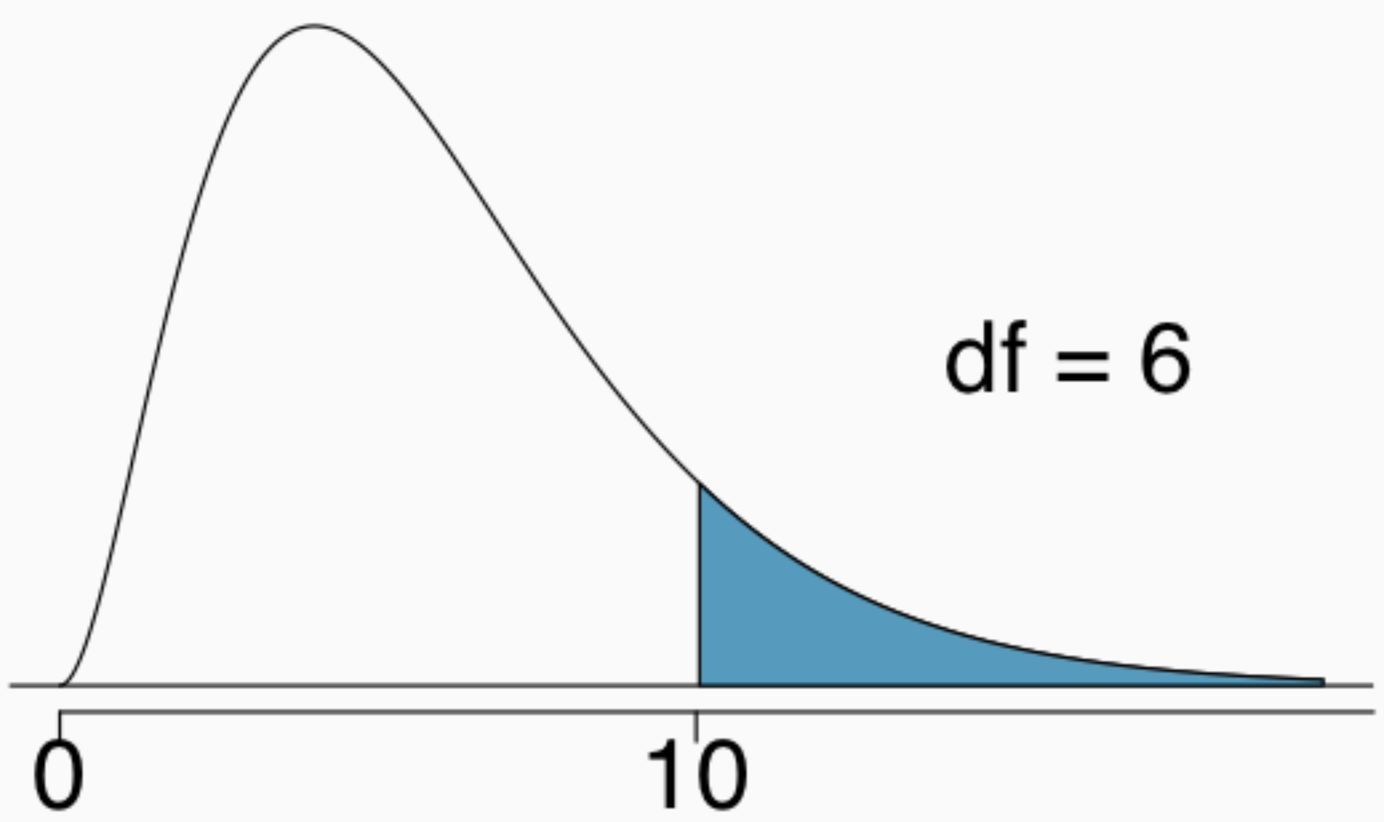
Estimate the shaded area under the χ^2 curve with *df* $= 6$.



Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	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

Finding areas under the χ^2 curve (cont.)

Estimate the shaded area under the χ^2 curve with $df = 6$.

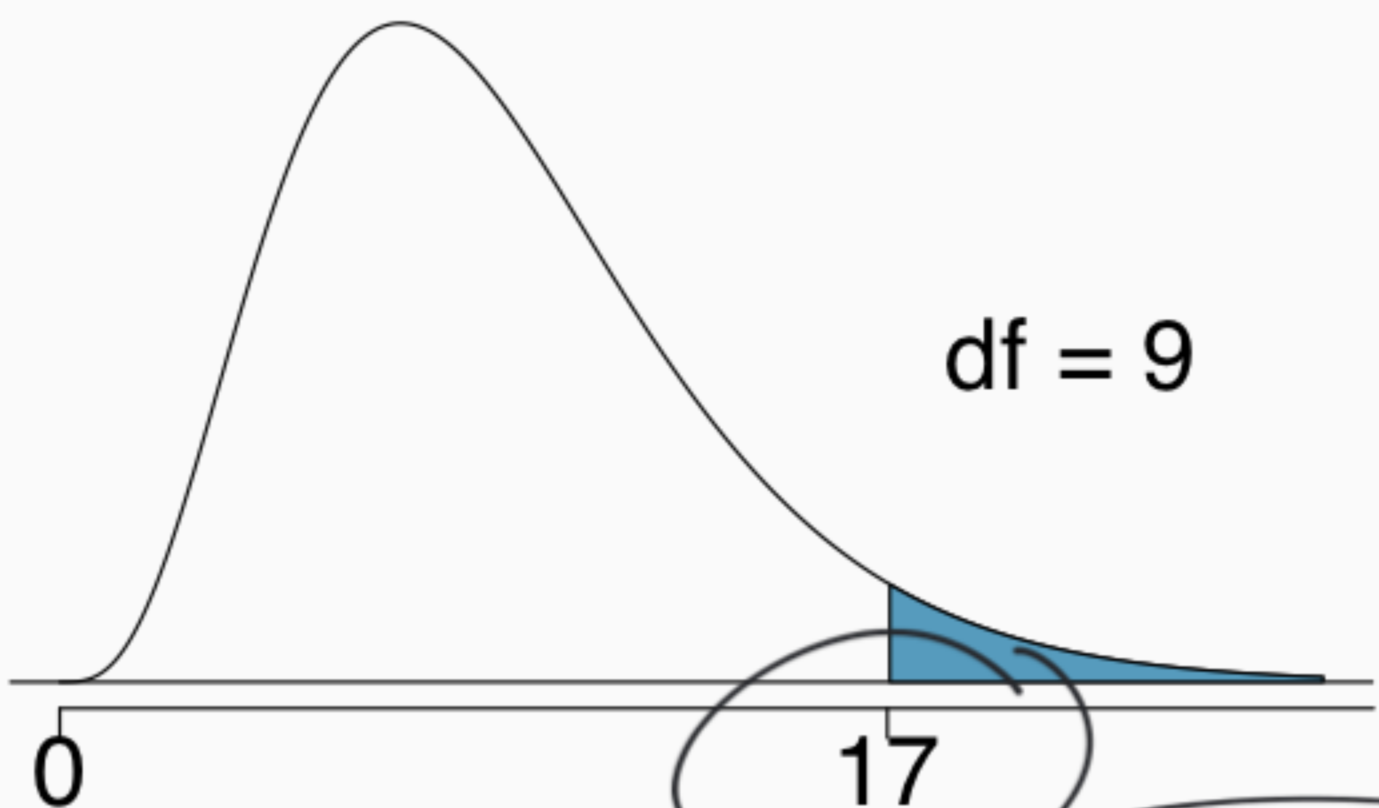


$P(\chi^2_{df=6} > 10)$
is between 0.1 and 0.2

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	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

Finding areas under the χ^2 curve (cont.)

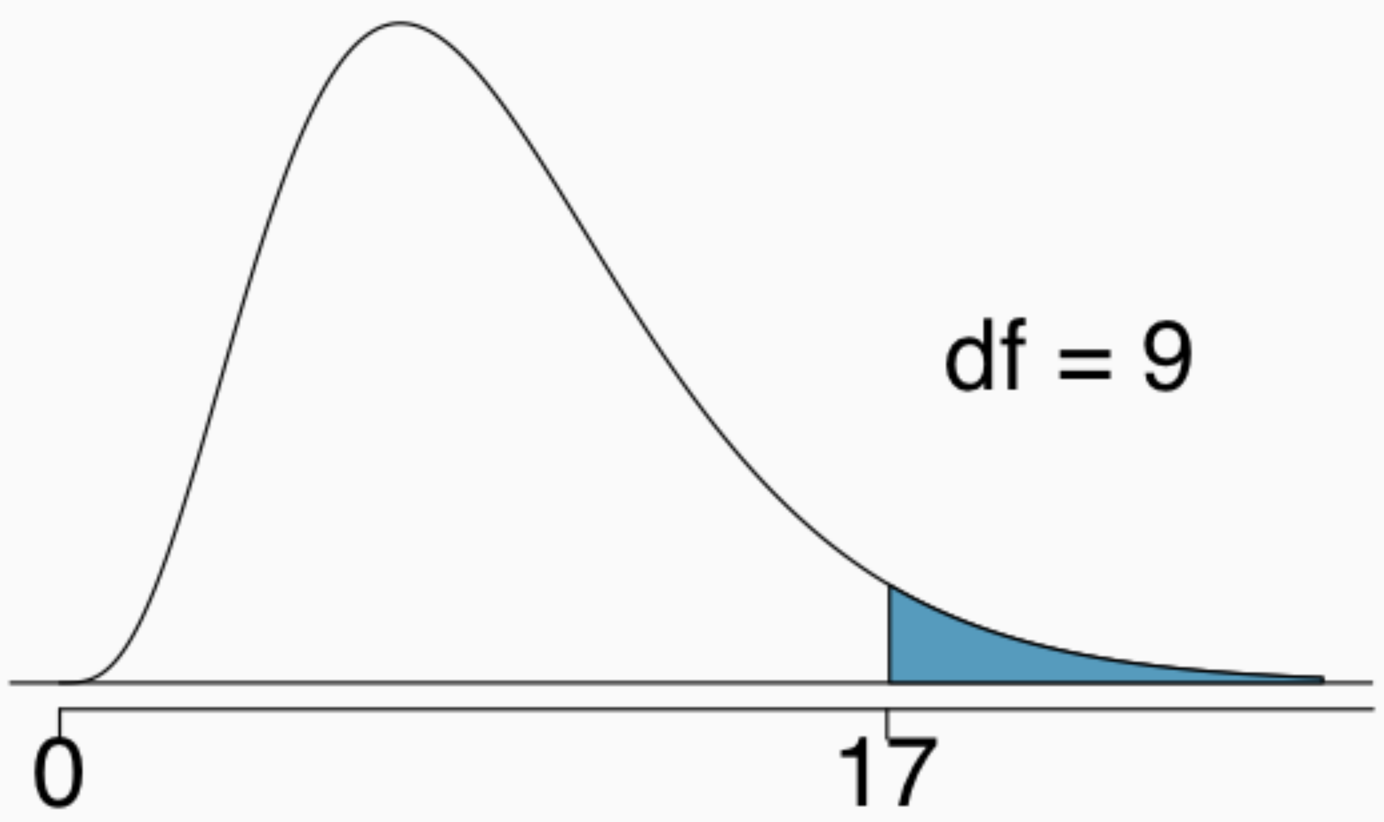
Estimate the shaded area (above 17) under the χ^2 curve with $df = 9$.



Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df 7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59
11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26

Finding areas under the χ^2 curve (cont.)

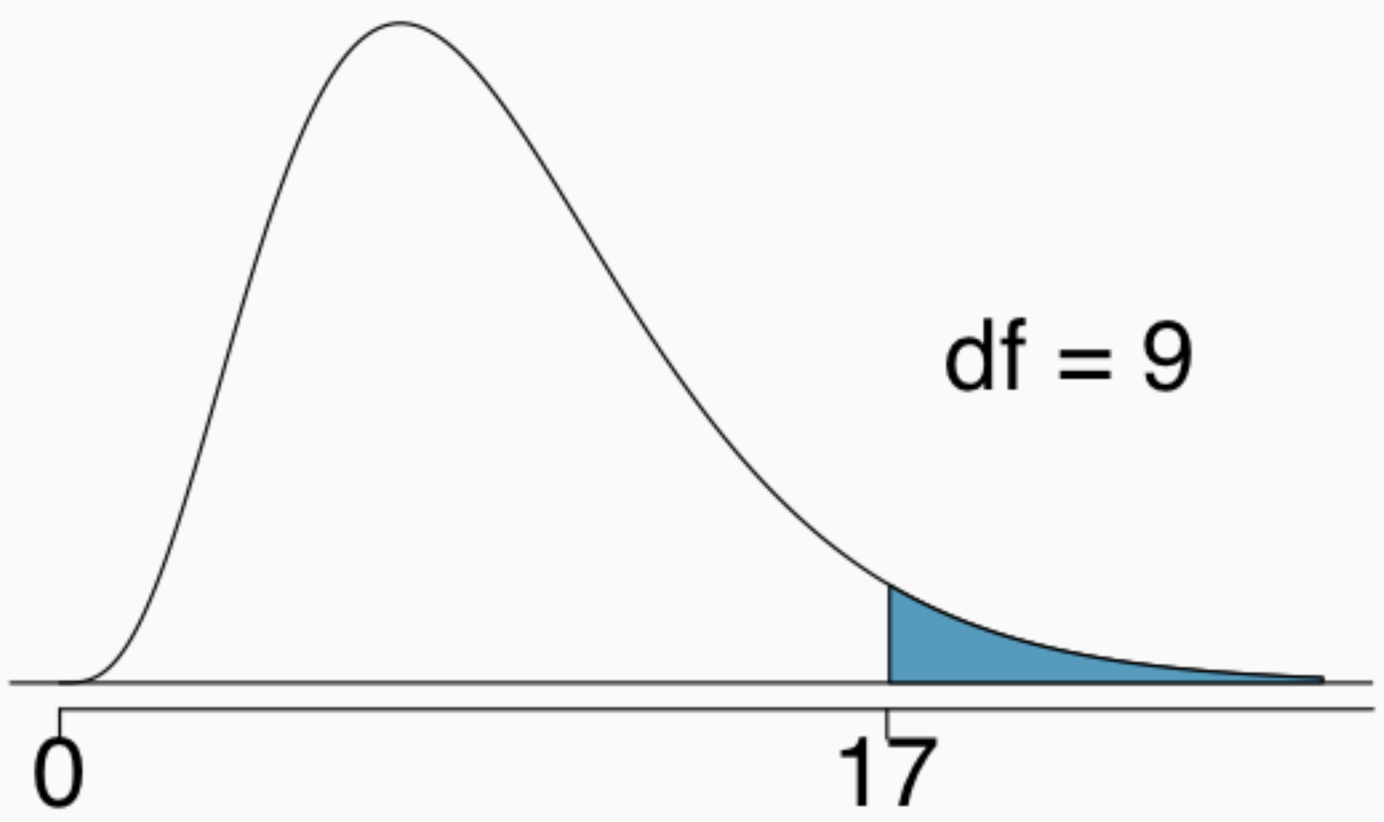
Estimate the shaded area (above 17) under the χ^2 curve with $df = 9$.



Upper tail		0.3	0.2	0.1	*0.05*	*0.02*	0.01	0.005	0.001
df	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
	9	10.66	12.24	14.68	*16.92*	*19.68*	21.67	23.59	27.88
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59
	11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26

Finding areas under the χ^2 curve (cont.)

Estimate the shaded area (above 17) under the χ^2 curve with $df = 9$.

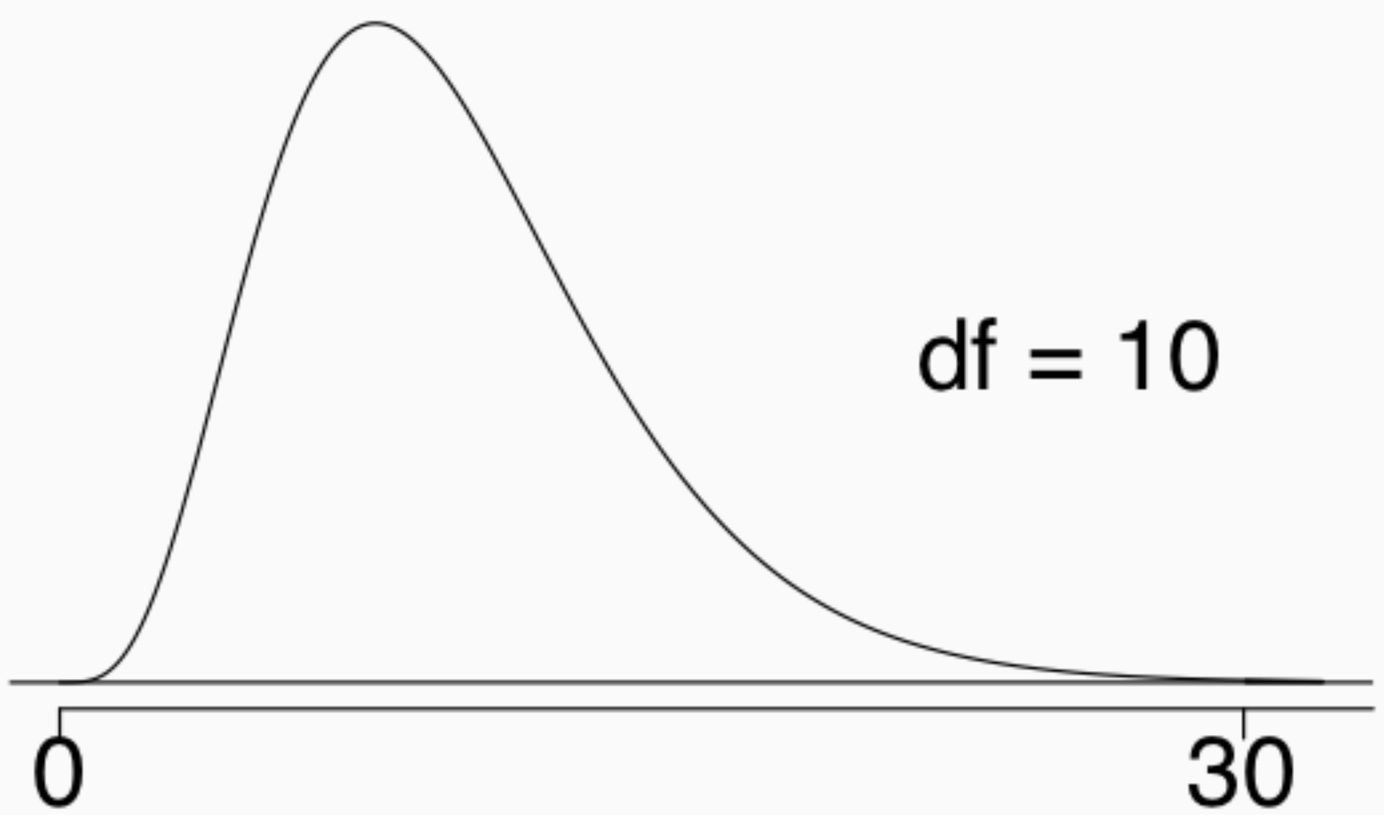


$P(\chi^2_{df=9} > 17)$
is between 0.02 and 0.05

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
	9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59
	11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26

Finding areas under the χ^2 curve (one more)

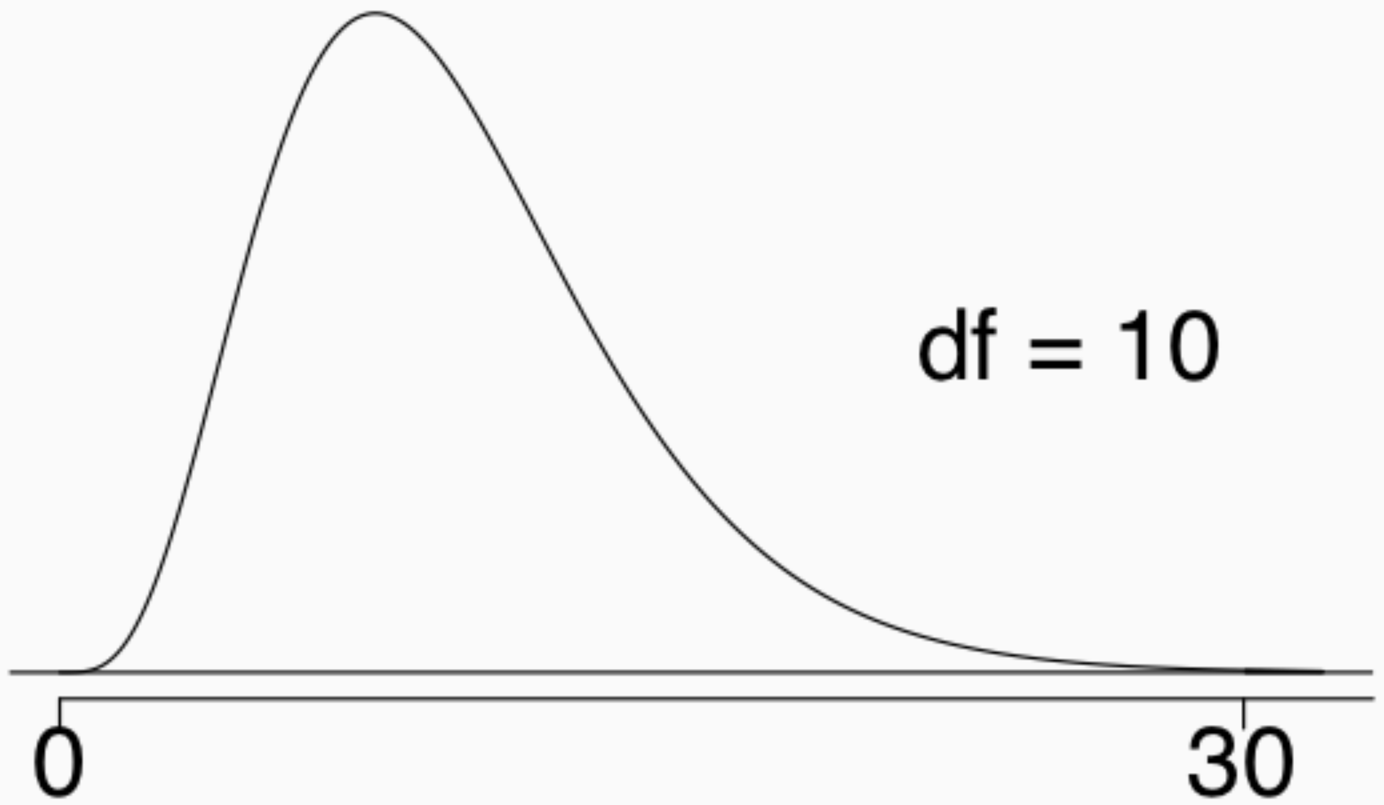
Estimate the shaded area (above 30) under the χ^2 curve with $df = 10$.



Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32
	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12
	9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59
	11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26

Finding areas under the χ^2 curve (one more)

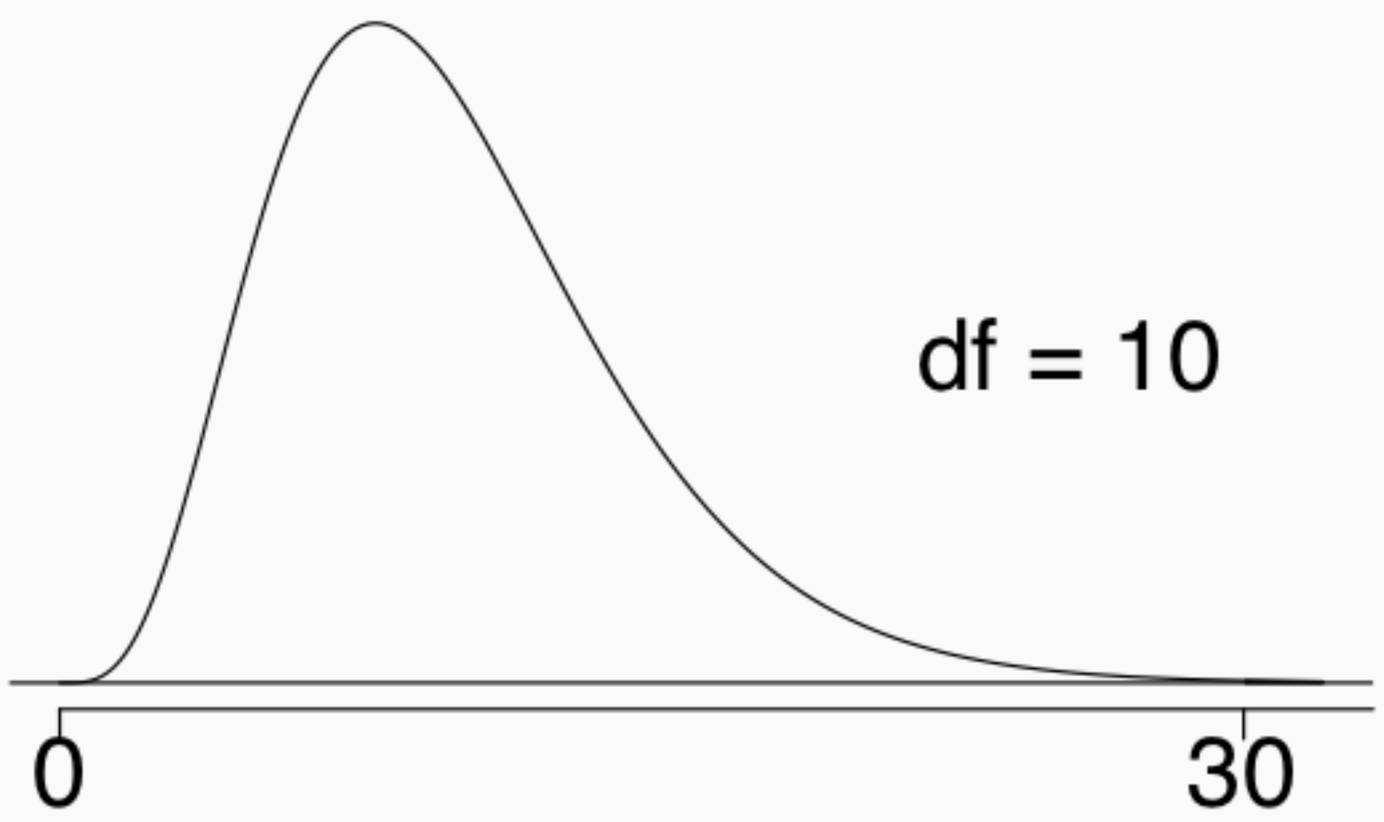
Estimate the shaded area (above 30) under the χ^2 curve with $df = 10$.



Upper tail	0.3	0.2	0.1	0.05	0.02	0.01	0.005	*0.001*	→
df 7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32	
8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12	
9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88	
10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	*29.59*	→
11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26	

Finding areas under the χ^2 curve (one more)

Estimate the shaded area (above 30) under the χ^2 curve with $df = 10$.



$P(\chi^2_{df=10} > 30)$
is less than 0.001

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001	→
df	7	8.38	9.80	12.02	14.07	16.62	18.48	20.28	24.32	
	8	9.52	11.03	13.36	15.51	18.17	20.09	21.95	26.12	
	9	10.66	12.24	14.68	16.92	19.68	21.67	23.59	27.88	
	10	11.78	13.44	15.99	18.31	21.16	23.21	25.19	29.59	→
	11	12.90	14.63	17.28	19.68	22.62	24.72	26.76	31.26	

Finding the tail areas using computation

As with previous distributions, we can also use a computer to calculate the probabilities exactly:

Finding the tail areas using computation

As with previous distributions, we can also use a computer to calculate the probabilities exactly:

- Using R:

```
pchisq(q = 30, df = 10, lower.tail = FALSE)

## [1] 0.0008566412
```

Finding the tail areas using computation

As with previous distributions, we can also use a computer to calculate the probabilities exactly:

- Using R:

```
pchisq(q = 30, df = 10, lower.tail = FALSE)

## [1] 0.0008566412
```

- Using a web applet - `https://gallery.shinyapps.io/dist_calc/`

Back to Labby's dice

Our original question was: Does Labby's data provide convincing evidence to suggest an inconsistency between the observed and expected counts (assuming the dice are fair)?

Back to Labby's dice

Our original question was: Does Labby's data provide convincing evidence to suggest an inconsistency between the observed and expected counts (assuming the dice are fair)?

The hypotheses were:

H_0: All six die faces are equally likely.

H_A: All six die faces are not equally likely.

Our test statistic was $\chi^2 = 24.67$.

Back to Labby's dice

Our original question was: Does Labby's data provide convincing evidence to suggest an inconsistency between the observed and expected counts (assuming the dice are fair)?

The hypotheses were:

H_0: All six die faces are equally likely.

H_A: All six die faces are not equally likely.

Our test statistic was $\chi^2 = 24.67$.

All we need is the *df* and we can calculate the tail area (the p-value) and make a decision on the hypotheses.

Degrees of freedom for a goodness of fit test

When conducting a goodness of fit test to evaluate how well the observed data follow an expected distribution, the degrees of freedom are calculated as the number of cells (k) minus 1.

$$df = k - 1$$

Degrees of freedom for a goodness of fit test

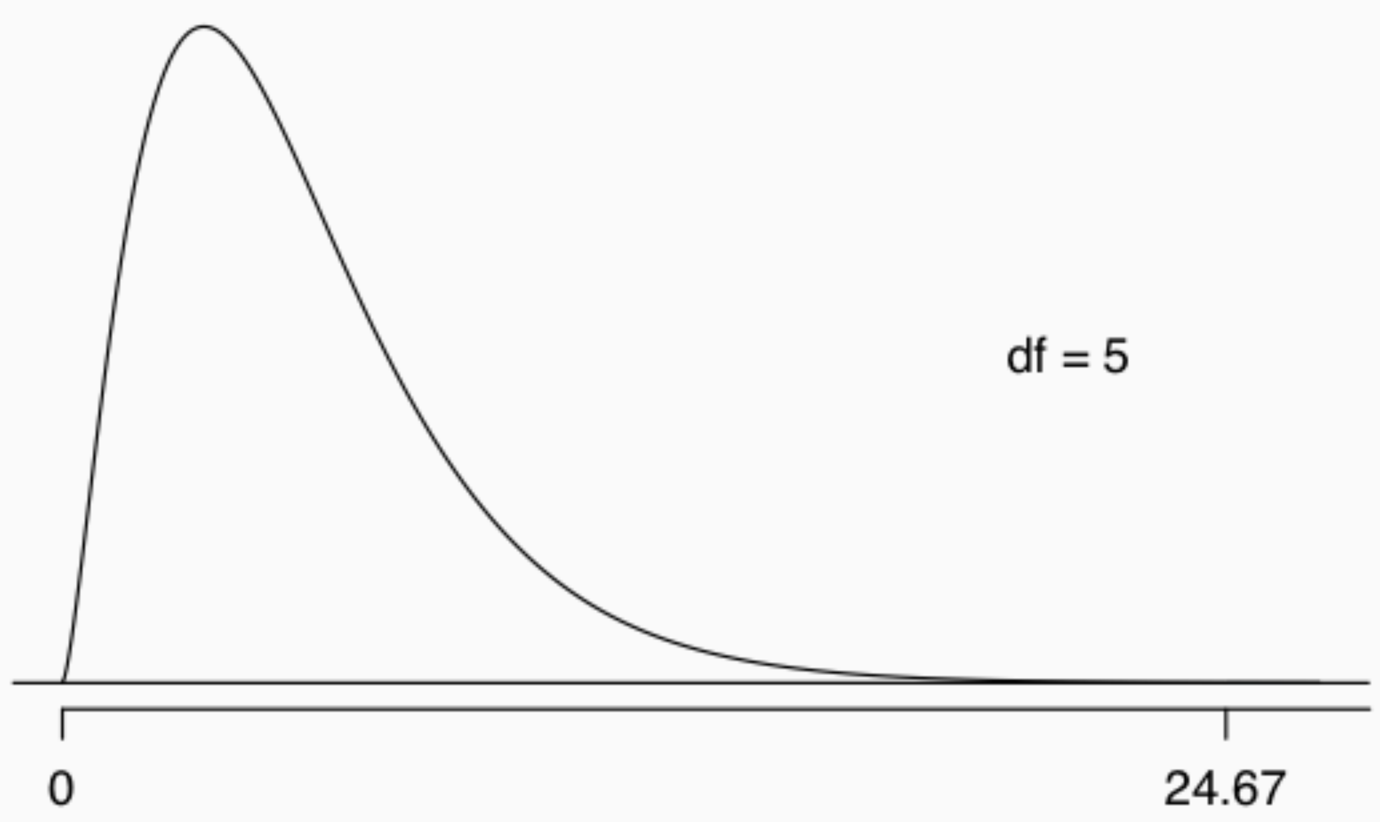
When conducting a goodness of fit test to evaluate how well the observed data follow an expected distribution, the degrees of freedom are calculated as the number of cells (k) minus 1.

$$df = k - 1$$

For the dice, $k = 6$ therefore

$$df = 6 - 1 = 5$$

Finding a p-value for a χ^2 test



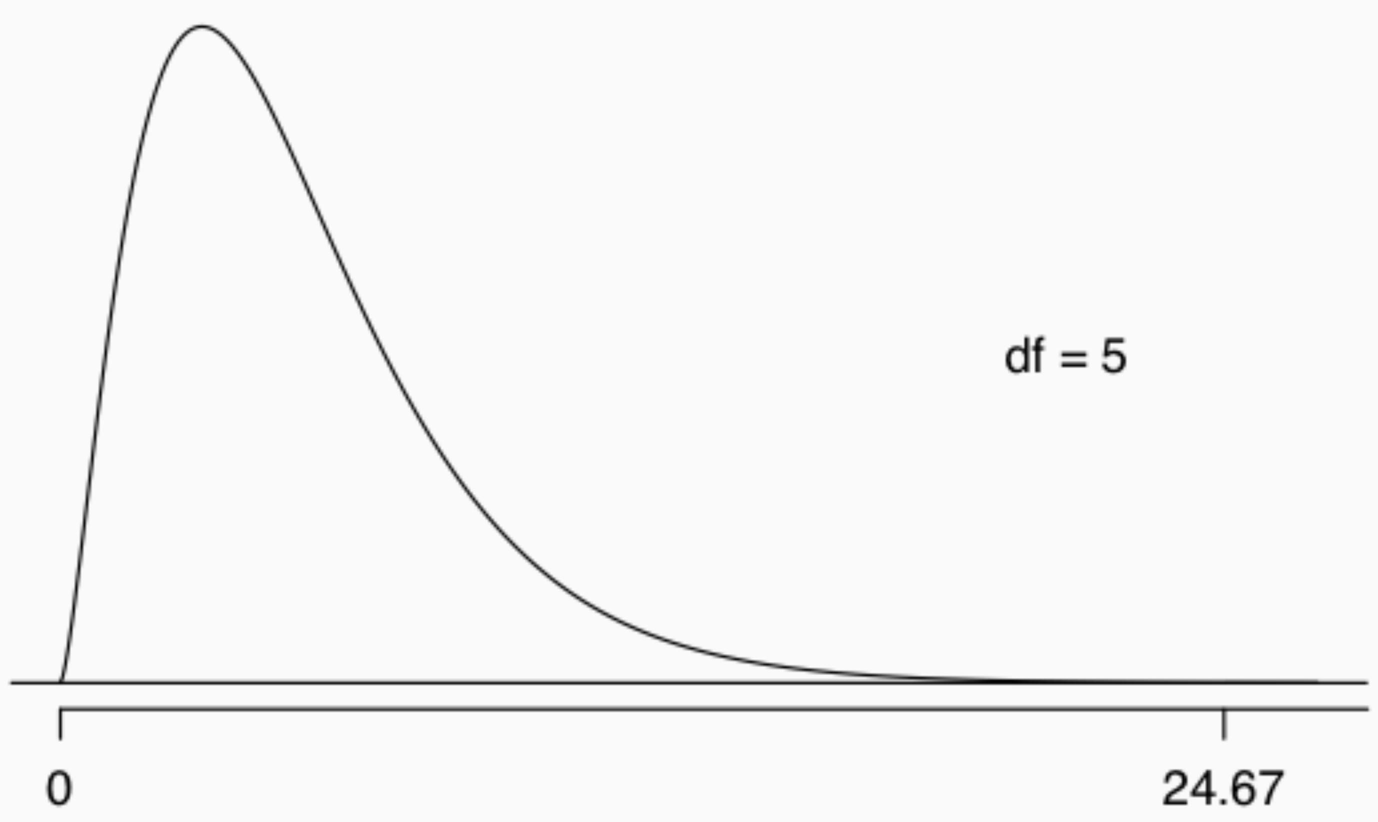
p-value = $P(\chi^2_{df=5} > 24.67)$

< 0.001

$\Rightarrow$ Reject H_0

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	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

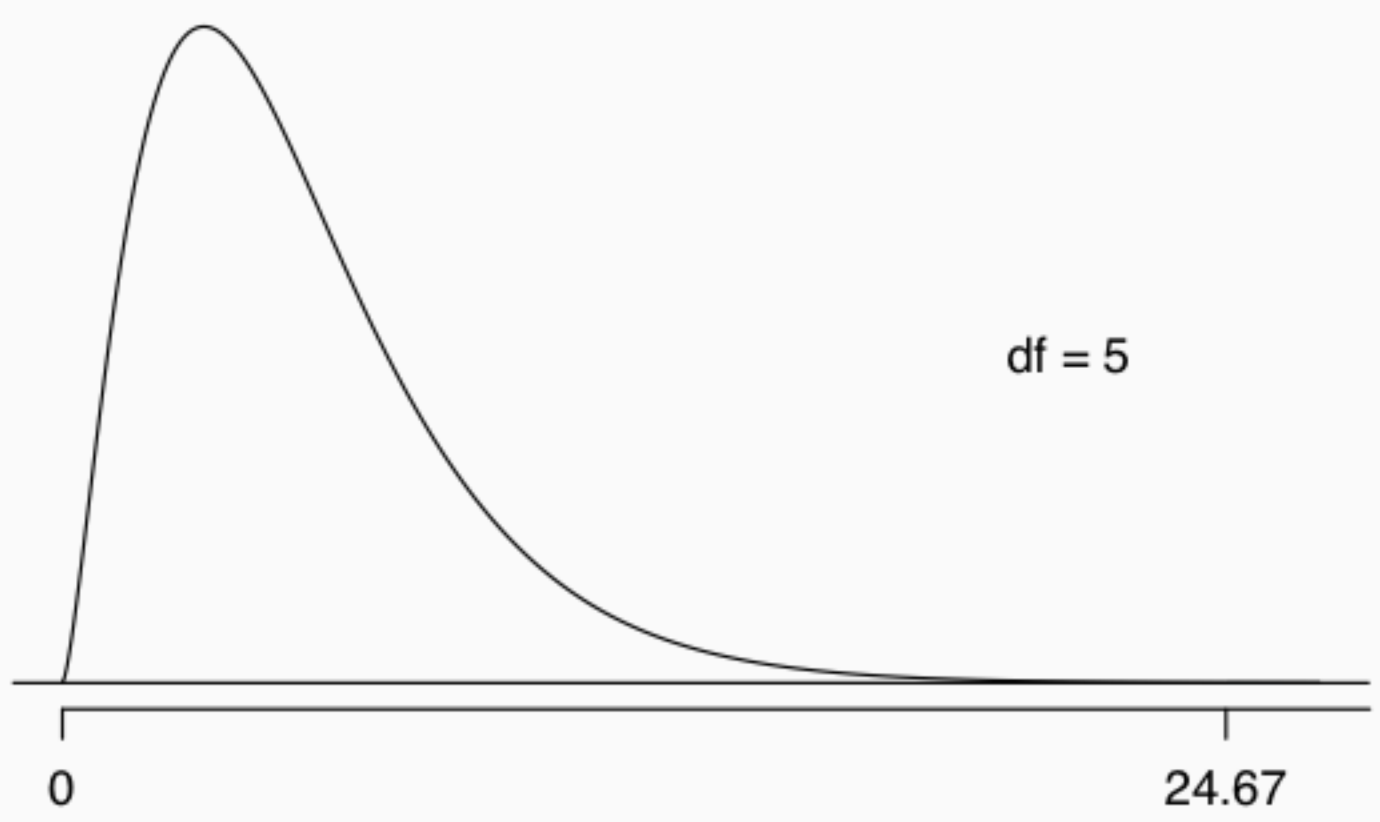
Finding a p-value for a χ^2 test



$p\text{-value} = P(\chi^2_{df=5} > 24.67)$

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001	→
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83	
	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	→

Finding a p-value for a χ^2 test



$p\text{-value} = P(\chi^2_{df=5} > 24.67) < 0.001$

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001	→
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83	
	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	→

Conclusion of the hypothesis test

We calculated a p-value less than 0.001. At the 5% significance level, what is the conclusion of the hypothesis test?

Conclusion of the hypothesis test

We calculated a p-value less than 0.001. At the 5% significance level, what is the conclusion of the hypothesis test?

Reject H_0, the data provide convincing evidence that the dice are biased.

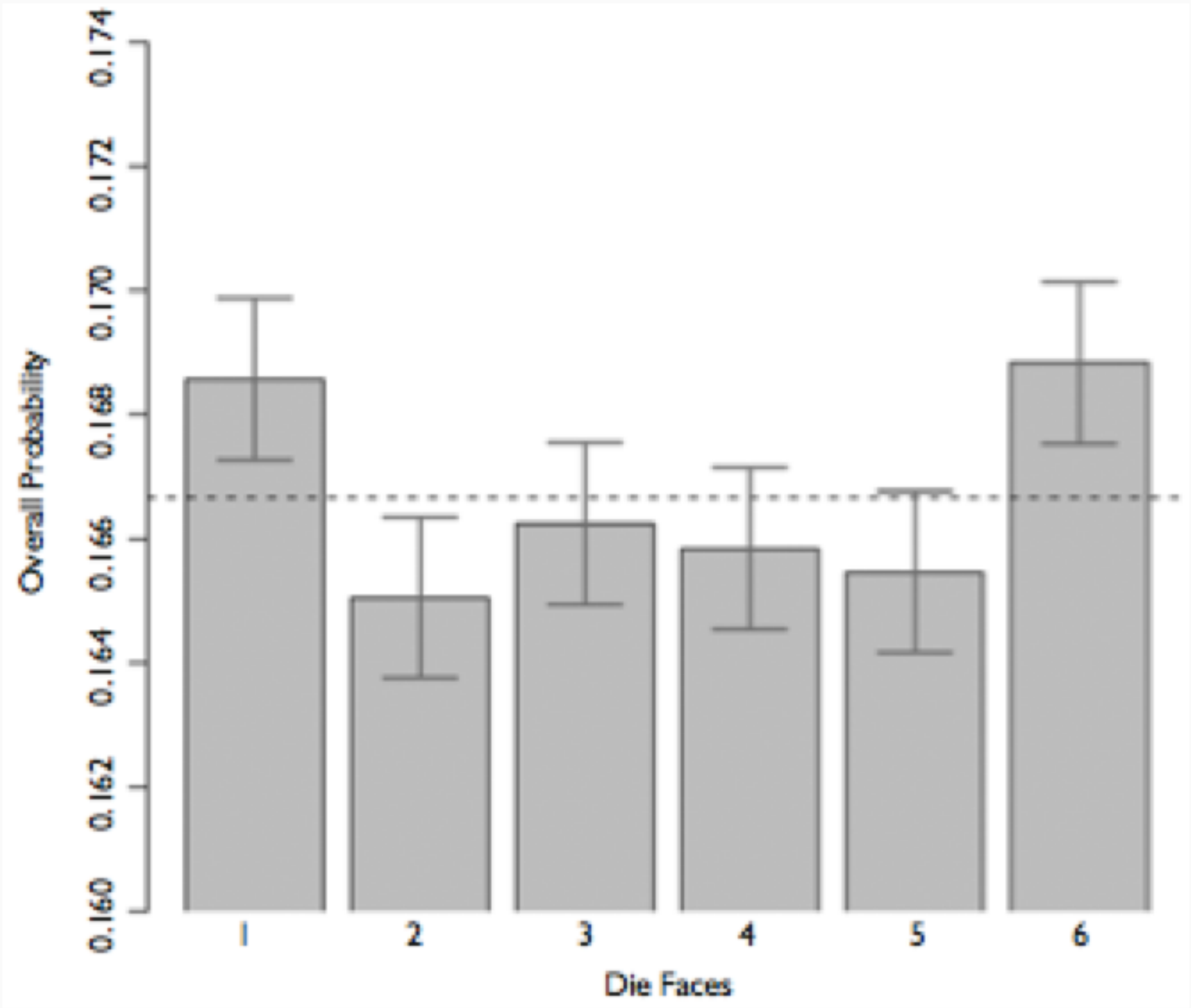
So what does this mean?

Conclusion of the hypothesis test

We calculated a p-value less than 0.001. At the 5% significance level, what is the conclusion of the hypothesis test?

Reject H_0, the data provide convincing evidence that the dice are biased.

So what does this mean?



Turns out...

Pearson's claim that 5s and 6s appear more often due to the carved-out pips is not supported by these data.

Dice used in casinos have flush faces, where the pips are filled in with a plastic of the same density as the surrounding material and are precisely balanced.

Turns out...

Pearson's claim that 5s and 6s appear more often due to the carved-out pips is not supported by these data.

Dice used in casinos have flush faces, where the pips are filled in with a plastic of the same density as the surrounding material and are precisely balanced.

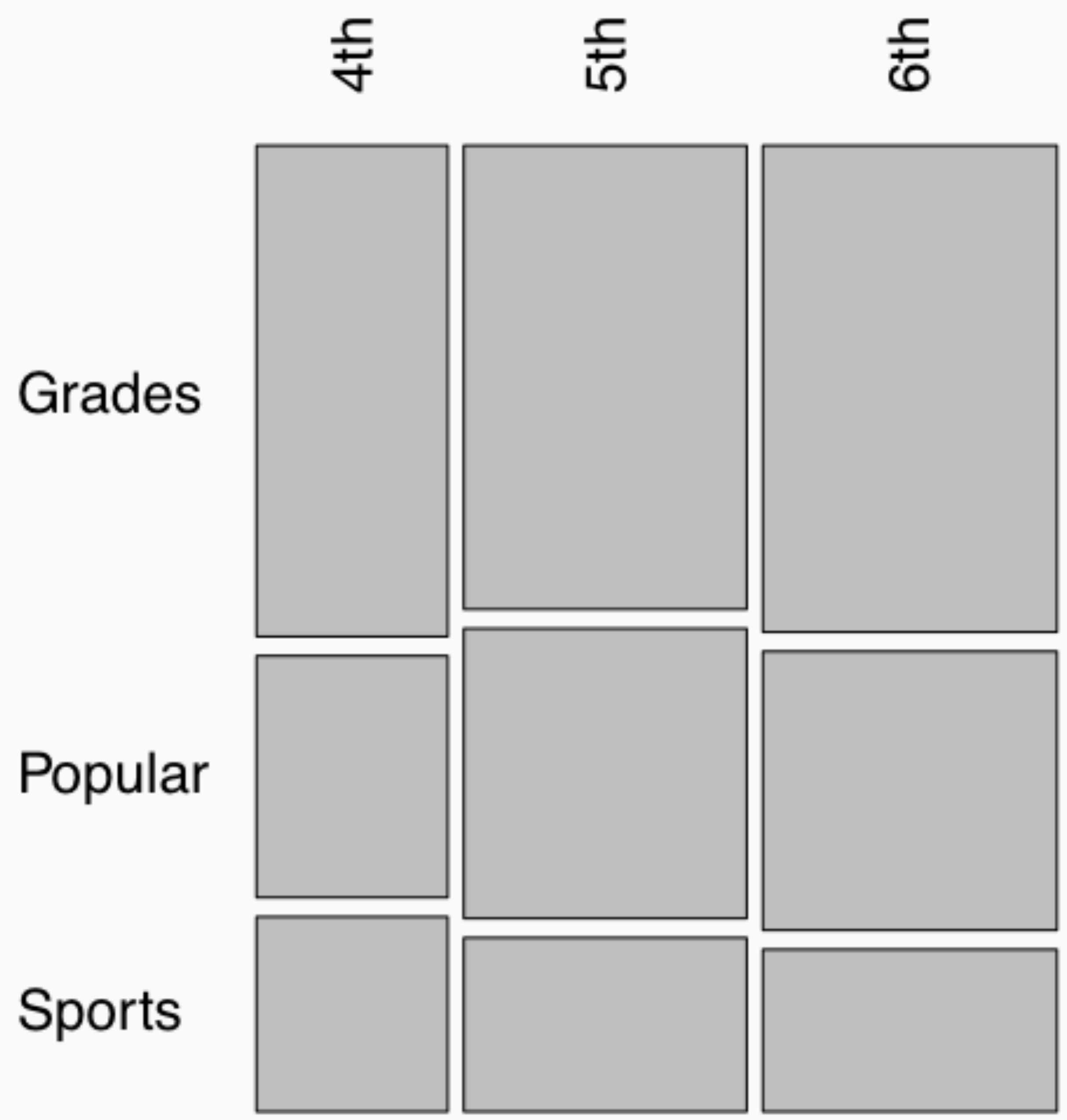
Labby found that the 1-6 axis is consistently shorter than the other two (2-5 and 3-4), the faces with one and six pips are larger than the other faces.

χ^2 test of independence

Popular kids

Students in grades 4-6 were asked whether good grades, athletic ability, or popularity was most important to them. A two-way table separating the students by grade and by choice of most important factor is shown below. Do these data provide evidence to suggest that goals vary by grade?

	Grades	Popular	Sports
4^{th}	63	31	25
5^{th}	88	55	33
6^{th}	96	55	32



χ^2 test of independence

Our hypotheses are:

H_0: Grade and goals are independent. Goals do not vary by grade.

H_A: Grade and goals are dependent. Goals vary by grade.

χ^2 test of independence

Our hypotheses are:

H_0: Grade and goals are independent. Goals do not vary by grade.

H_A: Grade and goals are dependent. Goals vary by grade.

Conditions for the χ^2 test of independence

χ^2 test of independence

Our hypotheses are:

H_0: Grade and goals are independent. Goals do not vary by grade.

H_A: Grade and goals are dependent. Goals vary by grade.

Conditions for the χ^2 test of independence

- *Independence:* Each case that contributes a count to the table must be independent of all the other cases in the table.

χ^2 test of independence

Our hypotheses are:

H_0: Grade and goals are independent. Goals do not vary by grade.

H_A: Grade and goals are dependent. Goals vary by grade.

Conditions for the χ^2 test of independence

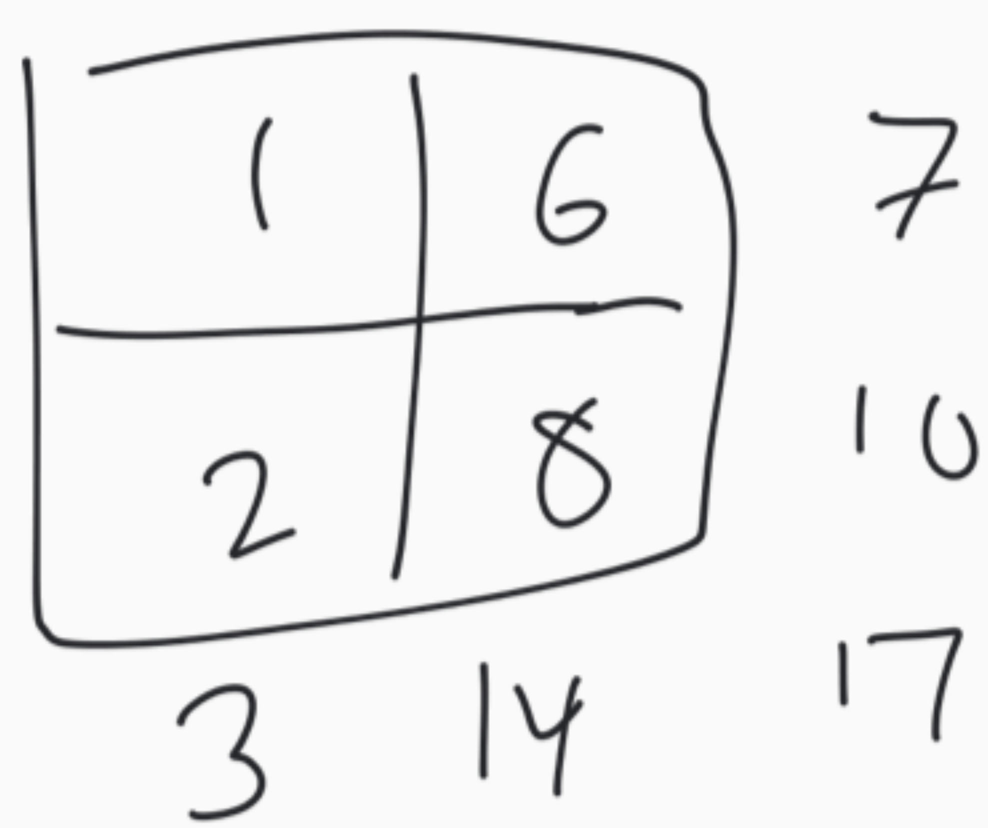
- *Independence:* Each case that contributes a count to the table must be independent of all the other cases in the table.
- *Sample size:* Each cell must have at least 5 *expected* counts.

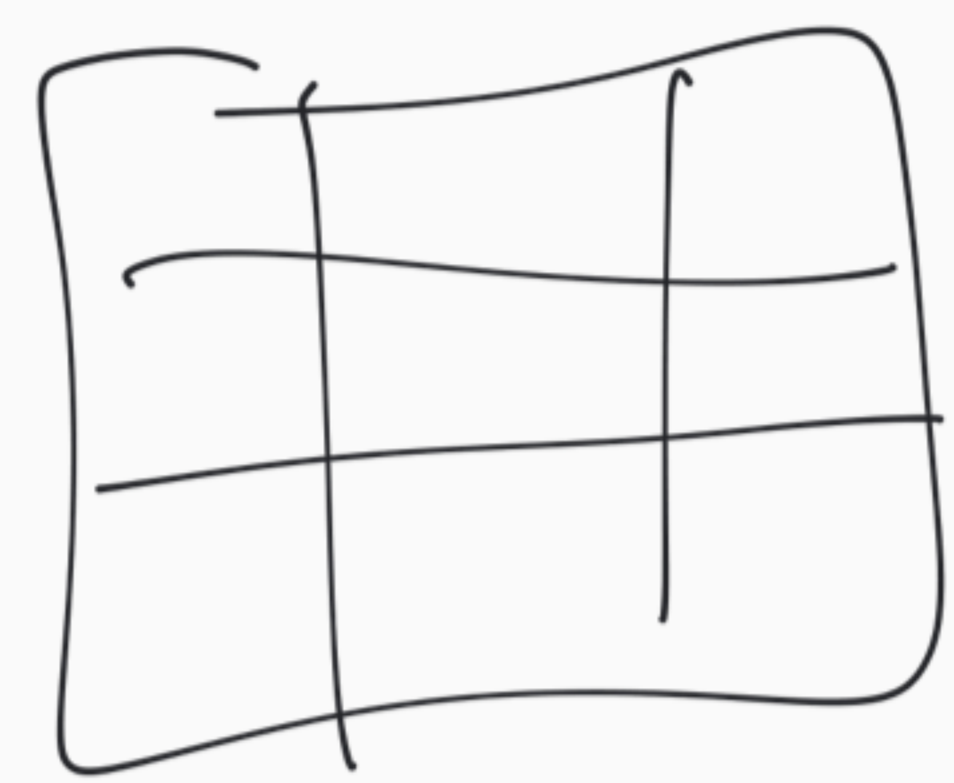
χ^2 test of independence

The test statistic is calculated using

$$\chi^2_{df} = \sum_{i=1}^{k} \frac{(O - E)^2}{E}$$

where k is the number of cells and $df = (R - 1) \times (C - 1)$.





χ^2 test of independence

The test statistic is calculated using

$$\chi^2_{df} = \sum_{i=1}^{k} \frac{(O - E)^2}{E}$$

where k is the number of cells and $df = (R - 1) \times (C - 1)$.

The p-value is the area under the χ^2 distribution curve to the right of the calculated test statistic,

$$\text{p-value} = P\left(\chi^2_{df} > \sum_{i=1}^{k} \frac{(O - E)^2}{E}\right).$$

χ^2 test of independence (cont.)

Expected counts in two-way tables:

$$\text{Expected Counts} = \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}$$

χ^2 test of independence (cont.)

Expected counts in two-way tables:

$$\text{Expected Counts} = \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}$$

	Grades	Popular	Sports	Total
4^{th}	*63*	*31*	25	119
5^{th}	88	55	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

χ^2 test of independence (cont.)

Expected counts in two-way tables:

$$\text{Expected Counts} = \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}$$

	Grades	Popular	Sports	Total
4^{th}	63	31	25	119
5^{th}	88	55	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

$$E_{row\ 1, col\ 1} = \frac{119 \times 247}{478} = 61. \ldots$$

don't round

χ^2 test of independence (cont.)

Expected counts in two-way tables:

$$\text{Expected Counts} = \frac{(\text{row total}) \times (\text{column total})}{\text{table total}}$$

	Grades	Popular	Sports	Total
4^{th}	63	31	25	119
5^{th}	88	55	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

$$E_{row\ 1, col\ 1} = \frac{119 \times 247}{478} = 61 \qquad E_{row\ 1, col\ 2} = \frac{119 \times 141}{478} = 35$$

Expected counts in two-way tables

What is the expected count for the highlighted cell?

	Grades	Popular	Sports	Total
4^{th}	63	31	25	119
5^{th}	88	55	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

Expected counts in two-way tables

What is the expected count for the highlighted cell?

	Grades	Popular	Sports	Total
4^{th}	63	31	25	119
5^{th}	88	55	33	176
6^{th}	96	55	32	183
Total	247	141	90	478

$$E_{5,Pop} = \frac{176 \times 141}{478} = 52$$

more 5th graders than expected have a goal of being popular

Calculating the test statistic in two-way tables

Expected counts are shown in (blue) next to the observed counts.

	Grades	Popular	Sports	Total
4^{th}	63 (61)	31 (35)	25 (23)	119
5^{th}	88 (91)	55 (52)	33 (33)	176
6^{th}	96 (95)	55 (54)	32 (34)	183
Total	247	141	90	478

Calculating the test statistic in two-way tables

Expected counts are shown in (blue) next to the observed counts.

	Grades	Popular	Sports	Total
4^{th}	63 (61)	31 (35)	25 (23)	119
5^{th}	88 (91)	55 (52)	33 (33)	176
6^{th}	96 (95)	55 (54)	32 (34)	183
Total	247	141	90	478

$$\chi^2 = \sum \frac{(63-61)^2}{61} + \frac{(31-35)^2}{35} + \cdots + \frac{(32-34)^2}{34} = 1.3121$$

Calculating the test statistic in two-way tables

Expected counts are shown in (blue) next to the observed counts.

	Grades	Popular	Sports	Total
4^{th}	63 (61)	31 (35)	25 (23)	119
5^{th}	88 (91)	55 (52)	33 (33)	176
6^{th}	96 (95)	55 (54)	32 (34)	183
Total	247	141	90	478

$$\chi^2 = \sum \frac{(63-61)^2}{61} + \frac{(31-35)^2}{35} + \cdots + \frac{(32-34)^2}{34} = 1.3121$$

$$df = (R-1) \times (C-1) = (3-1) \times (3-1) = 2 \times 2 = 4$$

Calculating the p-value

What is the correct p-value for this hypothesis test

$$\chi^2 = 1.3121 \qquad df = 4$$

Calculating the p-value

What is the correct p-value for this hypothesis test

$$\chi^2 = 1.3121 \qquad df = 4$$

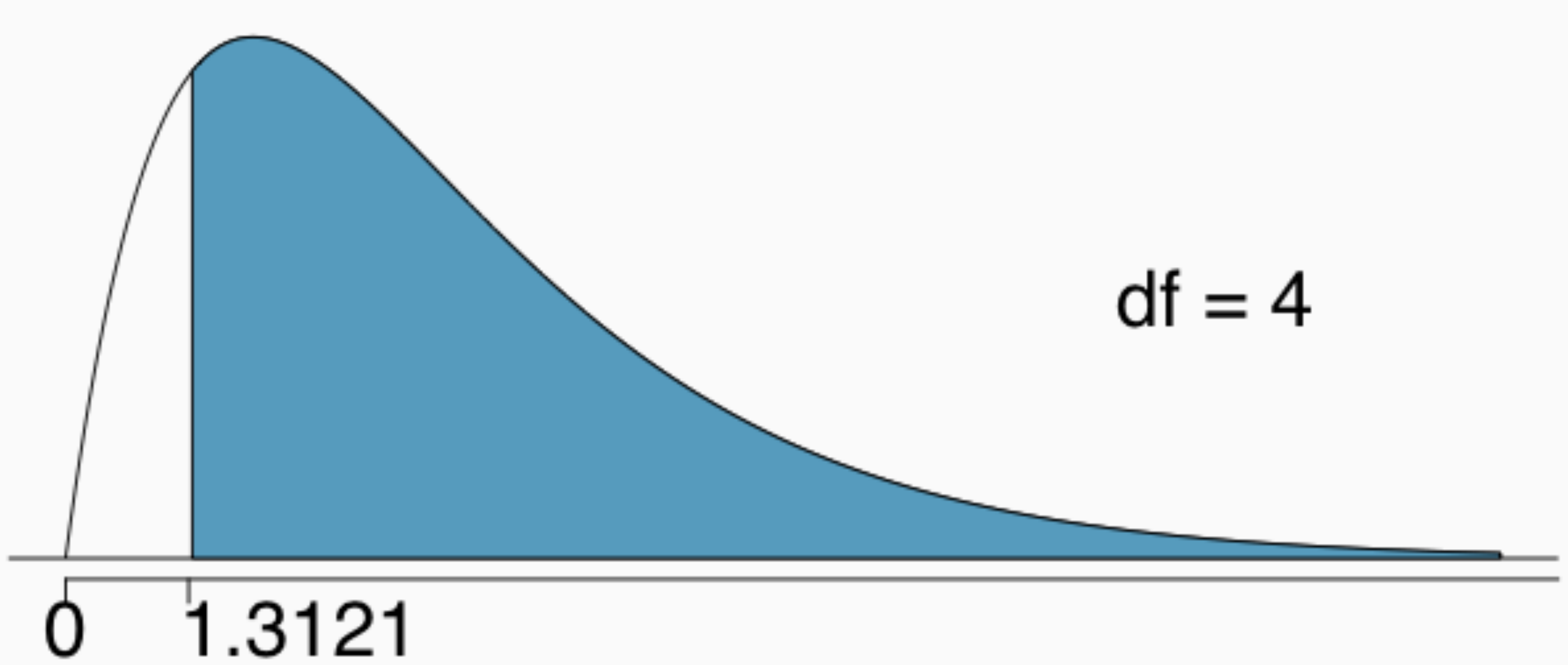
Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	2	2.41	3.22	4.61	5.99	7.82	9.21	10.60	13.82
	3	3.66	4.64	6.25	7.81	9.84	11.34	12.84	16.27
	4	4.88	5.99	7.78	9.49	11.67	13.28	14.86	18.47
	5	6.06	7.29	9.24	11.07	13.39	15.09	16.75	20.52

Calculating the p-value

What is the correct p-value for this hypothesis test

$$\chi^2 = 1.3121 \qquad df = 4$$

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	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

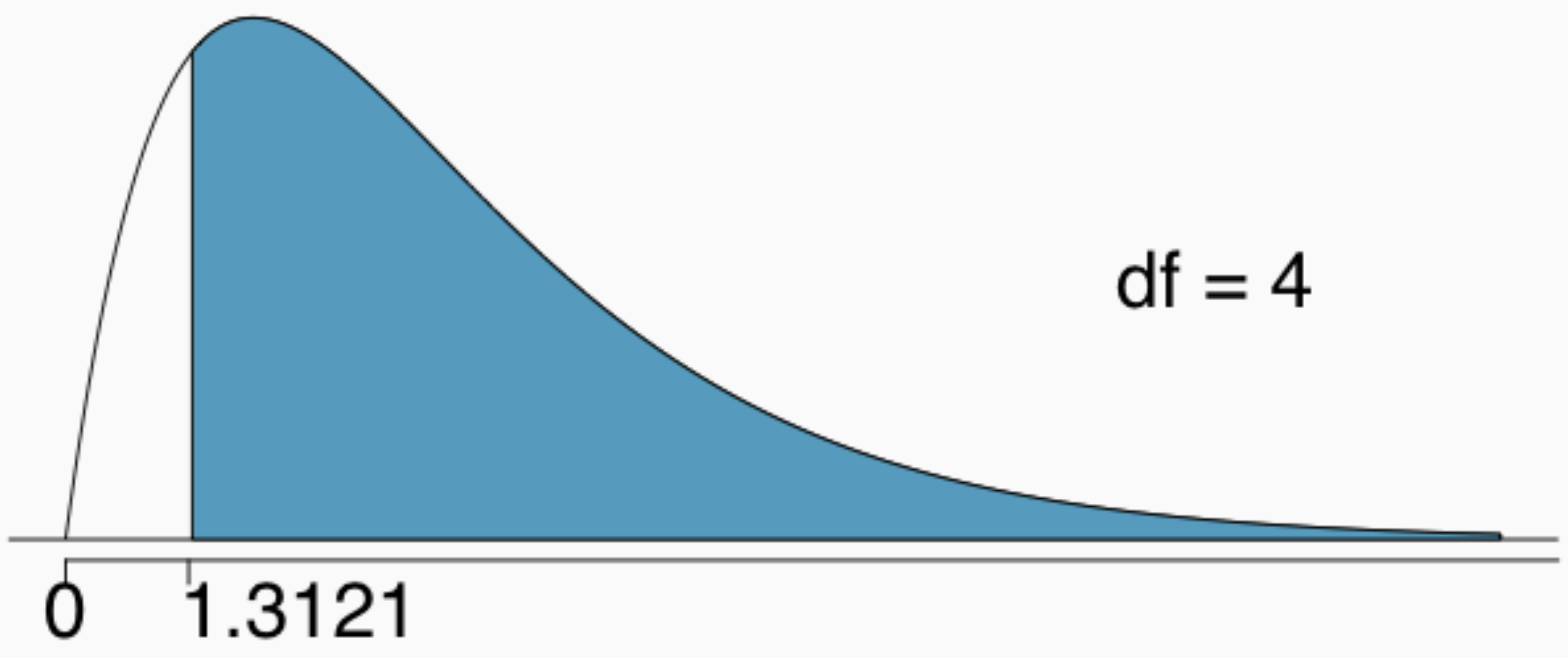


Calculating the p-value

What is the correct p-value for this hypothesis test

$$\chi^2 = 1.3121 \qquad df = 4$$

Upper tail		0.3	0.2	0.1	0.05	0.02	0.01	0.005	0.001
df	1	1.07	1.64	2.71	3.84	5.41	6.63	7.88	10.83
	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



$P(\chi^2_{df=4} > 1.3121)$ is more than 0.3

$P\text{-value} > 0.3$

Conclusion

Do these data provide evidence to suggest that goals vary by grade?

H_0: Grade and goals are independent. Goals do not vary by grade.

H_A: Grade and goals are dependent. Goals vary by grade.

Conclusion

Do these data provide evidence to suggest that goals vary by grade?

H_0: Grade and goals are independent. Goals do not vary by grade.

H_A: Grade and goals are dependent. Goals vary by grade.

Since p-value is large, we fail to reject H_0. The data do not provide convincing evidence that grade and goals are dependent. It doesn't appear that goals vary by grade.

Summary

Summary - χ^2 test of goodness of fit

- Data:

  y - categorical variable w/ 3 or more levels,

  x - none.

- Hypotheses:

  H_0 - data follow the given distribution,

  H_A - data do not follow the given distribution.

- Conditions:

  Independent observations, all $E_i \geq 5$.

- Test statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \quad df = k - 1$$

Summary - χ^2 test of independence

- Data:

  y - categorical variable w/ 2 or more levels,

  x - categorical variable w/ 2 or more levels.

- Hypotheses:

  H_0 - x and y are independent,

  H_A - x and y are dependent.

- Conditions:

  Independent observations, all $E_{i,j} \geq 5$.

- Test statistic:

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}, \quad df = (m-1)(n-1)$$

$$E_{i,j} = \frac{N_{i,\cdot} \times N_{\cdot,j}}{N_{\cdot,\cdot}}$$