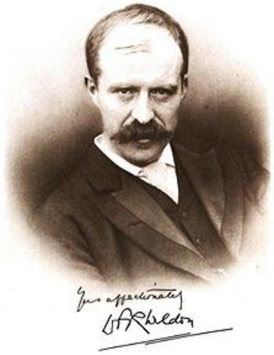
Lecture 17 - χ^2 Tests

Sta102 / BME102

Colin Rundel

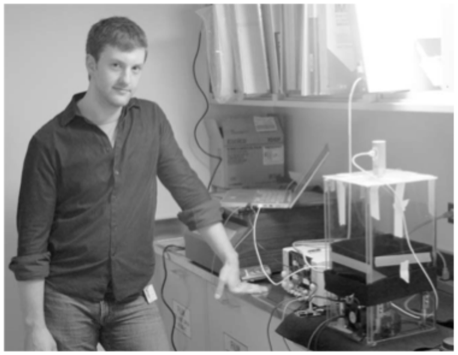
November 2, 2015

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).
- It was observed that 5s or 6s occurred more often than expected, and Pearson hypothesized that this was probably due to the construction of the dice. Most inexpensive dice have hollowed-out pips, and since opposite sides add to 7, the face with 6 pips is lighter than its opposing face, which has only 1 pip.

Labby's dice

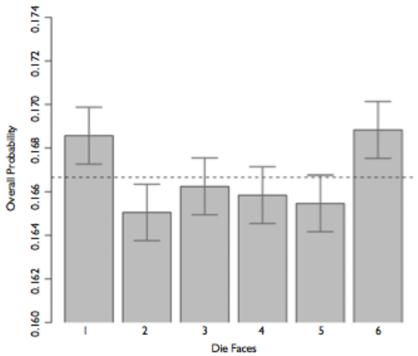


- In 2009, Zacariah Labby (U of Chicago), repeated Weldon's experiment using a homemade dice-throwing, pip counting machine.

  http://www.youtube.com/watch?v=95EErdouO2w
- The rolling-imaging process took about 20 seconds per roll.
- Each day there were ∼150 images to process manually.
- At this rate Weldon's experiment was repeated in about six days.
- Recommended reading: *http://galton.uchicago.edu/about/docs/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, · · · , 6s would he expect to have observed?

The table below shows the observed and expected counts from Labby's experiment.

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?

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_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.
- 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

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

- This construction is based on
  1. identifying the difference between a point estimate and an expected value if the null hypothesis was true, 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 use a new test statistic called the *χ^2 (chi-squared) statistic*.

The χ^2 statistic is defined to be

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad \text{where } k = \text{total number of cells/categories}$$

Calculating the χ^2 statistic

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

- 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?

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 affect the test's error rates.

The χ^2 distribution

- In order to determine if the χ^2 statistic we calculated is considered unusually high or not we need to first describe its distribution.
- The χ^2 distribution has just one parameter called *degrees of freedom (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 two other continuous distributions:
  - Normal distribution - unimodal and symmetric with two parameters: mean (center) and standard deviation (spread)
  - T distribution - unimodal and symmetric with one parameter: degrees of freedom (spead, kurtosis)

The χ^2 distribution (Theory)

Where does the χ^2 distribution come from?

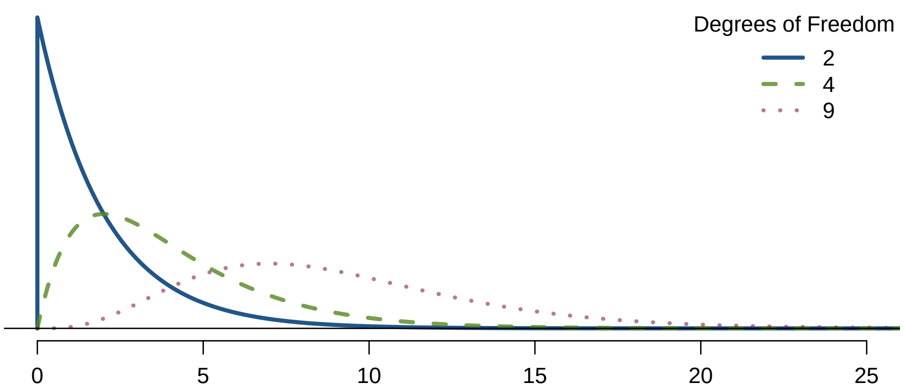
- Like the T and the Z, the χ^2 distribution is an example of a continuous probability distribution (it is actually a special case of another distribution called the Gamma distribution)
- If we define $Z \sim N(0, 1)$ then the quantity,

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

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

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

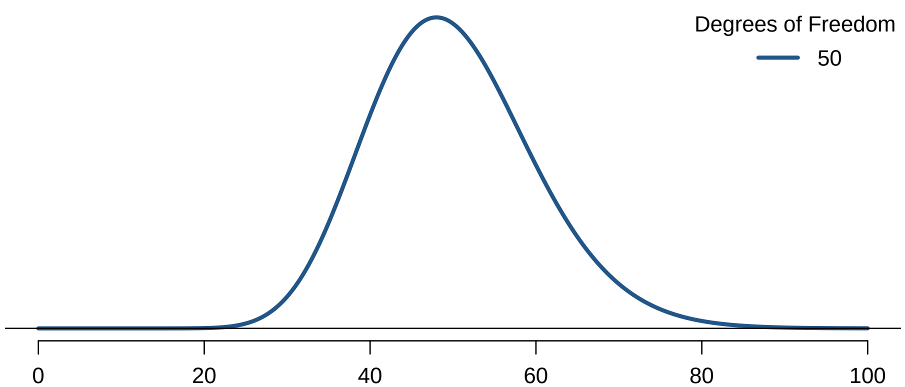
The χ^2 distribution (cont.)



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.)

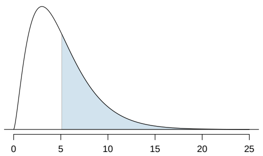


Also, for large df the χ^2 distribution converges to the normal distribution with

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

Finding areas under the χ^2 curve

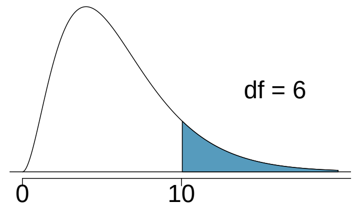
- p-value = tail area under the χ^2 distribution (as usual)
- For this we can use technology, or a *χ^2 probability table*.
- This table works a lot like the t table, but only provides upper tail probabilities.



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.)

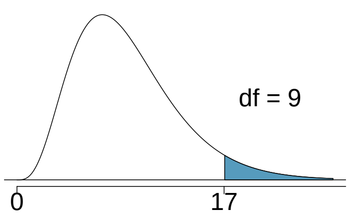
Estimate the shaded area under the χ^2 curve with $df = 6$.



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.)

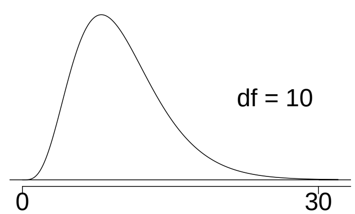
Estimate the shaded area (above 17) under the χ^2 curve with $df = 9$.



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 (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 the tail areas using computation

- While probability tables are very helpful in understanding how probability distributions work, and provide quick reference when computational resources are not available, they are somewhat archaic.
- Using R:

```
pchisq(q = 30, df = 10, lower.tail = FALSE)
## [1] 0.0008566412
```

- Using a web applet - bit.ly/dist_calc

Back to Labby's dice

- The research question was: Does Labby's data provide convincing evidence to suggest an inconsistency between the observed and expected counts?
- The hypotheses were:
  - 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_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)
- We had calculated a test statistic of $\chi^2 = 24.67$.
- 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.

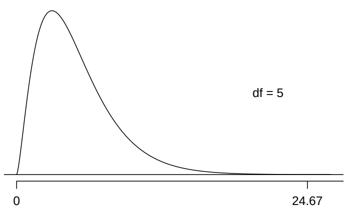
$$df = k - 1$$

- For dice outcomes, $k = 6$, therefore

$$df = 6 - 1 = 5$$

Finding a p-value for a χ^2 test

The *p-value* for a χ^2 test is defined as the *tail area above the calculated test statistic*.



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

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

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?

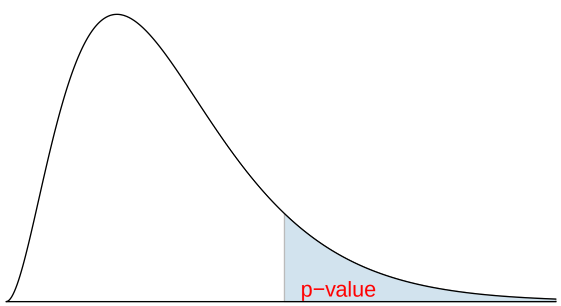
So what does this mean?

Turns out...

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

Recap: p-value for a χ^2 test

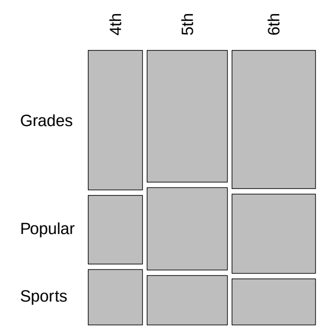
- The p-value for a χ^2 test is defined as the tail area *above* the calculated test statistic.
- This is because the test statistic is always positive, and a higher test statistic means a higher deviation from the null hypothesis.



Popular kids

In the dataset `popular`, 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

- The 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 particular scenario (i.e. cell) must have at least 5 *expected* counts.
- The test statistic is calculated using

$$\chi^2_{df} = \sum_{i=1}^{k} \frac{(O-E)^2}{E} \quad \text{where} \quad df = (R-1) \times (C-1),$$

  where k is the number of cells, R is the number of rows, and C is the number of columns.
- The p-value is the area under the χ^2_{df} curve, above the calculated test statistic.

χ^2 test of independence (cont.)

Expected counts in two-way tables:

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

	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

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

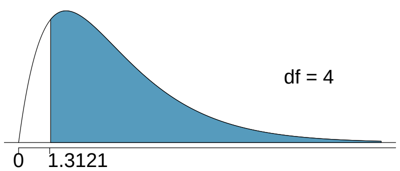
$$\begin{aligned} \chi^2 &= \sum \frac{(63-61)^2}{61} + \frac{(31-35)^2}{35} + \cdots + \frac{(32-34)^2}{34} = 1.3121 \\ df &= (R-1) \times (C-1) = (3-1) \times (3-1) = 2 \times 2 = 4 \end{aligned}$$

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-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.

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:

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

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