The statistical analysis of film style: tests of proportion

Nick Redfern

Although it has become increasingly common for film scholars to cite statistics of film style, few (if any) employ statistical tests to justify their conclusions. This paper describes three statistical tests of proportions: the one-sample z-test, the two-sample z-test, and the chi-square test of homogeneity of proportions for 3 or more samples. These tests can be used to analysis the frequencies with which different types of shots occur in films. These methods are simple to understand and easy to calculate, and when presented as part of a study of film style allow the reader to evaluate the methodology used and the interpretation of the results.

z-test of one proportion: medium long shots in *La Bête Humaine*

Suppose we have read in an article that medium long shots (MLS) typically account for a quarter of the shots (0.25) in a film directed by Jean Renoir, and we want to see if this is the case in *La Bête Humaine* (1938). To compare the proportion of a shot type in a film with a hypothesized value we use the z-test of one proportion. The data set we will be using in this example is the file submitted to the Cinemetrics database by David van de Geer (2008) for this film.

Step 1: state the null hypothesis, the level of significance, and the tails

The null hypothesis is that 'the proportion of medium long shots in Jean Renoir's *La Bête Humaine* is equal to 0.25,' and the level of significance is 0.10. Therefore, if p>0.10, we will conclude that there is insufficient evidence to reject the null hypothesis. This is a two tailed test of whether the proportion of MLSs in *La Bête Humaine* is significantly less than *or* significantly greater than 0.25. A one-tailed test specifies the direction of the difference being tested, and would require the null hypothesis to be rephrased to reflect this.

Step 2: calculate the proportion of medium long shots in La Bête Humaine

From the data set we see that the number of medium long shots in *La Bête Humaine* is 35 and that the total number of shots in the film is 320. Therefore the proportion of MLSs in this film is P = 35/320 = 0.11.

Step 3: calculate the test statistic and p-value

The formula for the one-sample z-test is

$$z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}},$$

where P is the observed proportion of the variable in which we are interested, π_0 is the hypothesized proportion, and n is the size of the sample. For *La Bête Humaine*, these values are P = 0.11, π_0 = 0.25, and n = 320. Substituting these values into the above equation gives

$$z = \frac{0.11 - 0.25}{\sqrt{\frac{0.25(1-0.25)}{320}}} = \frac{-0.14}{0.0242} = -5.80\,.$$

If $n \geq$ 30, $n\pi \geq 5$, and $n(1-\pi) \geq 5$, the value of z is approximately normally distributed, and the two-tailed *p*-value for the test statistic can be calculated using the Microsoft Excel function =2*(1–NORMSDIST(5.80)). This returns the result *p* = <0.0001, which less is than our significance level of 0.10 and so we can conclude that there is sufficient evidence to reject the null hypothesis. Note that the above function uses the absolute value of z (its numerical value without regard to its sign) for convenience when using Microsoft Excel. If either $n\pi < 5$ or $n(1-\pi) < 5$, and/or the sample size is small, it is necessary to use the binomial distribution to calculate an exact *p*-value.

Step 4: present the results
In presenting the results of the one-sample z-test we need to include the name of the test, the hypothesis being tested, the level of significance, the number of tails, the test statistic, and the *p*-value.

> A two-tailed z-test was employed to test the null hypothesis that the proportion of medium long shots in Jean Renoir's *La Bête Humaine* (1938) is equal to 0.25, with a significance level of 0.10. Medium long shots accounted for 35 out of a total of 320 shots for this film, and the proportion of MLSs is calculated to be *P* = 0.11. The result shows that this is significantly different from the estimated value (z = -5.80, *p* = <0.0001), and we conclude that *La Bête Humaine* has fewer medium long shots than expected.

Although this paragraph contains only three sentences, it gives us everything we need to know about the test conducted, the outcome of the test, and the researcher's interpretation of the result. Anyone with a basic understanding of statistics can evaluate the methodology and the interpretation of the results, and can (if necessary) repeat the process to verify the result.

z-test of two proportions: medium close-ups in *Blackmail*

In 1929, Alfred Hitchcock directed two versions of *Blackmail* – one silent and one sound. We are interested in discovering what impact – if any – sound technology had on the proportion of medium close-ups (MCUs) in the two versions of this film. If we want to compare the proportions of shot types in two films we can use the z-test for two proportions. The logic behind this test is the same as for the one-sample test above, but the calculations are different because we have two observed proportions. The data used here is from Barry Salt's database of shot scale data available at the Cinemetrics website, where the frequency of each scale in a motion picture is normalized to correspond to the number that would have occurred if the film was comprised of 500 shots.[1]

Step 1: state the null hypothesis, the level of significance, and the tails
The null hypothesis is that 'there is no difference in the proportion of medium close-ups in the two versions of Alfred Hitchcock's *Blackmail*,' and the level of significance is 0.05. Therefore, if p>0.05, we will conclude that there is insufficient evidence to reject the null hypothesis. As before, this is a two-tailed test.

Step 2: calculate the proportion of medium close-ups in the two versions of Blackmail
From the data we see that in the silent and sound versions of *Blackmail* there are 67 and 87 medium close-ups, respectively. Dividing these figures by 500, we have the proportion of MCUs in the silent version (P_1 = 0.134) and the proportions of MCUs in the sound version (P_2= 0.174).

Step 3: calculate the test statistic and p-value
The formula for the two-sample z-test is

$$z = \frac{P_1 - P_2}{\sqrt{\hat{P}\left(1 - \hat{P}\right)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}},$$

where P_1 and n_1 are the proportion and the sample size of the first sample, P_2 and n_2 are the proportion and the sample size of the second sample, and $\hat{P}$ is the pooled proportion of the two samples and is calculated as

$$\hat{P} = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}.$$

Substituting the data for the two versions of *Blackmail*,

$$\hat{P} = \frac{(500 \times 0.134) + (500 \times 0.174)}{500 + 500} = 0.154\,,$$

and

[1] The database can be accessed at http://www.cinemetrics.lv/satltdb.php.

$$z = \frac{0.134 - 0.174}{\sqrt{0.154(1 - 0.154)\left[\frac{1}{500} + \frac{1}{500}\right]}} = -1.752\,.$$

As before, if $n \geq 30$, $nP_i \geq 5$ and $n(1 - P_i) \geq 5$ then the distribution of z is approximately normal and the two-tailed *p*-value can be calculated using Microsoft Excel with the function =2*(1–NORMSDIST(1.752)). For this experiment, $p = 0.08$ and we do not have sufficient evidence to reject the null hypothesis. Note that we cannot say that the null hypothesis is proven or that the *p*-value is the probability that the null hypothesis is true. A statistical hypothesis test is only a test of the plausibility of a model given a set of data, and the lower the *p*-value the more implausible the model for that data. If the sample size is small, and/or either $n\pi < 5$ or $n(1 - \pi) < 5$, it is necessary to use the Fisher Exact Test to calculate an exact *p*-value.

Step 4: calculate the confidence interval for the difference between P_1 and P_2
If the data set we are using is perfectly accurate then we can simply say that the true difference between the two proportions of close-ups is simply –0.04, but as the interpretation of shot scales is subjective and/or different versions of the same film exist (e.g., the difference between the theatrical release and the DVD) we can calculate a confidence interval for the difference to give an estimate of the error of the difference. The size of the confidence interval is 100(1–α)%, where α is the significance level. If α = 0.05, the size of the confidence interval is 95%, and we are 95% confident that this interval includes the true difference in the proportions.

To calculate the confidence interval for the difference between the proportions of close-ups in these films, we need to calculate the point estimate of the difference and the margin of error for the difference:

$$(P_1 - P_2) \pm z_{\alpha/2}\left(\sqrt{\frac{P_1(1 - P_1)}{n_1} + \frac{P_2(1 - P_2)}{n_2}}\right).$$

For the shot scale data used above, the critical value of z when α = 0.05 is 1.96 and we have

$$(0.134 - 0.174) \pm 1.96\left(\sqrt{\frac{0.134(1 - 0.134)}{500} + \frac{0.174(1 - 0.174)}{500}}\right).$$

The margin of error is 0.045, and the 95% confidence interval for the difference in the proportions is -0.085 to 0.005. Note that the confidence interval includes a difference of zero as a possible value.

Step 5: present the results
In reporting the results of the two-sample z-test we need to provide the same information as for the one-sample test along with the difference between the two proportions and the 95% confidence interval for the difference:

> A two-tailed z-test was performed to compare the proportion of medium close-ups in the silent and sound versions of Alfred Hitchcock's *Blackmail*, with a significance level of 0.05. The normalized number of close-ups in the silent version is 67 and for the sound versions is 87, and the proportions of MCUs are 0.134 and 0.174, respectively. The difference between the two proportions was estimated to be –0.04 (95% CI: –0.085, 0.005), and this difference is not significant ($z = -1.752$, $p = 0.08$). There is insufficient evidence to conclude that the proportions of close-ups in the two versions of *Blackmail* are different.

As before, a great deal of information can be presented to the reader in a clear format that will allow them to evaluate and interpret the results for themselves. In addition to this information we would also want to include the source from which we obtained the data (i.e. Barry Salt's database) and the fact that this data has been normalized.

Chi-square test for homogeneity of proportions: medium shots in the Hollywood films of Max Ophüls

We are interested to know if the proportions of medium shots in Max Ophüls's four Hollywood films of the 1940s are homogenous. To compare the proportions of shot types in three or more films we can use a chi-square test.[2] Again, this example uses data from Barry Salt's database.

Step 1: state the null hypotheses, the level of significance
The null hypothesis for this test is that 'the proportions of medium shots in the Hollywood films of Max Ophüls are homogenous,' and the significance level is 0.05. The null hypothesis is non-directional and two-tailed. The chi-squared distribution is one-tailed but allows us to test a two-tailed null hypothesis, and the modification for a directional null hypothesis requires using the same test statistic and halving the *p*-value of a non-directional test. Note that the direction of the null hypothesis should be made clear before the test is conducted – you cannot divide the *p*-value of a non-significant non-directional chi-square test by two and then claim to have found a significant result in one direction or the other.

Step 2: arrange the data in a contingency table
A contingency table is a representation of data in which the cells of the table contain the *observed* count belonging to a particular category (i.e. medium shots in *The Exile*). The categories used in a contingency table should be *collectively exhaustive* and *mutually exclusive*. Contingency tables are sorted into

[2] The chi-square test for comparing the proportions of a shot type in two films is equivalent to the two sample z-test because $z^2 = \chi^2$.

rows (r) and columns (c), and the size of the table is r × c. The contingency table for the Max Ophüls is presented in Table 1, where the category of 'other' includes all those shots that are not medium shots. It may be easier to represent this information visually as a bar chart, and visual inspection of the raw data. The bar chart for this data is presented in Figure 1.

Table 1 Contingency table for the normalized number of medium shots in the American films of Max Ophüls

	Medium shots	Other	Total
Exile, The	109	391	500
Letter from an Unknown Woman	116	384	500
Caught	156	344	500
Reckless Moment, The	118	382	500
Total	499	1501	2000

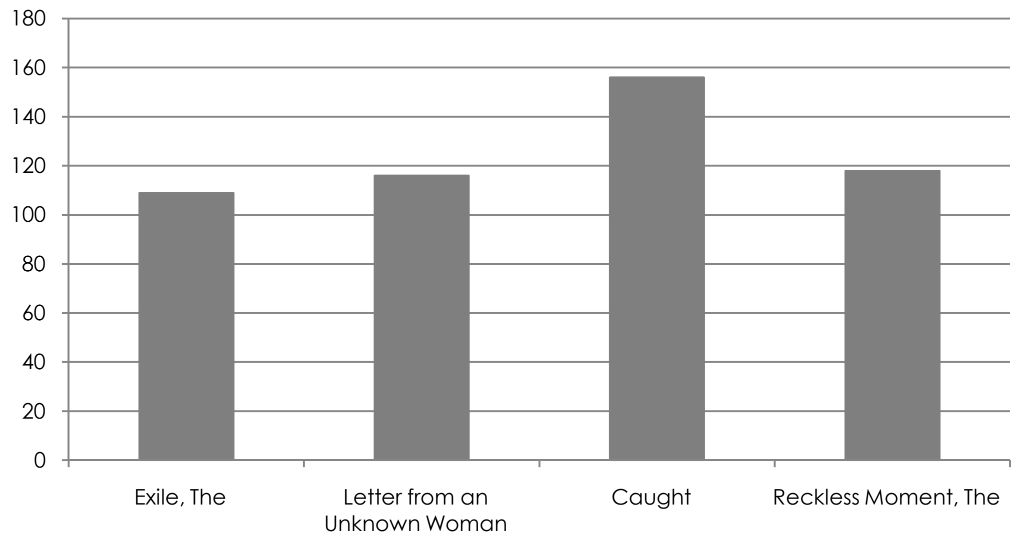

Figure 1 The normalized frequency of medium shots in the Hollywood films of Max Ophüls.

It is immediately apparent from Table 1 and Figure 1, that the proportion of medium shots in *The Exile*, *Letter from an Unknown Woman*, and *The Reckless Moment* are similar both to each other and the expected frequency, while *Caught* stands out as being different. By performing the chi-square test, we are seeking to discover if this is a real difference or if it may be due to random variation alone.

Step 3: calculate the expected frequency of medium shots for each film
The expected frequency of medium shots for each film is calculated from the above contingency table as

$$E = \frac{row\ total \times column\ total}{grand\ total}$$

The expected frequency of medium shots in *The Exile* is

$$E = \frac{500 \times 499}{2000} = 124.75\,,$$

and because this data has already been standardized this value is the expected frequency for all the films in our sample. We also need to calculate the expected frequency of the other shots in these films, and to do this for *The Exile* we simply substitute the appropriate column total into the above equation:

$$E = \frac{500 \times 1501}{2000} = 375.25\,,$$

Again, because we are using standardized data this is the expected frequency for the 'other shots' in all the films. The expected frequencies for this test are presented in Table 2. Note that the sum of the rows and columns in Tables 1 and 2 are the same.

Table 2 Expected frequency of medium shots in the American films of Max Ophüls

	Medium shots	Other	Total
Exile, The	124.75	375.25	500
Letter from an Unknown Woman	124.75	375.25	500
Caught	124.75	375.25	500
Reckless Moment, The	124.75	375.25	500
Total	499	1501	2000

The chi-square test can only be used when all of the expected counts of each cell are greater than or equal to 5. If this is not the case, then it is often appropriate to combine categories to obtain the required expected frequency.

For example, if the expected frequency of very long shots in a film is less than five, then we might combine this scale with long shots to remove this problem. This should then be noted when we describe the methodology used.

Step 4: calculate the test statistic, the degrees of freedom, the p-value, and the effect size
The test statistic for the chi-square test (χ^2) is calculated as the sum of the ratios of the squared difference between the observed and expected frequencies to the expected frequency for each cell in the contingency table:

$$\chi^2 = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

where O_{ij} and E_{ij} are the observed and expected frequencies in the *i*th row and the *j*th columns of Tables 1 and 2, respectively. For medium shots in *The Exile*, imputing the values from the cell where the first row and the first column meet in Tables 1 and 2 gives

$$\frac{(109 - 124.75)^2}{124.75} = 1.9885 .$$

Completing this process for all the cells in the contingency table produces the results in Table 3, and summing them together we get $\chi^2 = 14.3845$.

Table 3 Calculated cell values for the chi-square statistic

	Medium shots	Other	Total
Exile, The	1.9885	0.6611	2.6495
Letter from an Unknown Woman	0.6137	0.2040	0.8178
Caught	7.8282	2.6024	10.4306
Reckless Moment, The	0.3652	0.1214	0.4866
Total	10.7956	3.5889	14.3845

The degrees of freedom of a chi-square test is determined by multiplying the number of rows in the contingency table minus one by the number of columns minus one: for the above contingency table, df = (r–1)×(c–1) = (4–1)×(2–1) = 3.

Using Microsoft Excel, we can calculate the *p*-value from χ^2 and the degrees of freedom using the function =CHIDIST(14.3845, 3), which returns $p = 0.0024$ and the null hypothesis is rejected.

Finally, we want to know if this is an important result and so we need to calculate the effect size, Cramer's V, using the formula

$$V = \sqrt{\frac{\chi^2}{n \times (k-1)}},$$

where n is the grand total of the contingency table and k is the smaller of the number of rows or columns in the contingency table.[3] The value of V ranges from 0 to 1, and the greater the value the larger the effect. For the Ophüls data this gives

$$V = \sqrt{\frac{14.3845}{2000 \times (2-1)}} = 0.0848\,.$$

There are no precise guidelines on interpreting the effect size of a chi-square test, but Cohen (1988) recommends that when the chi-square test has 3 degrees of freedom a value for Cramer's V of between 0.06 and 0.17 indicates a small effect, a value between 0.17 and 0.29 is a medium effect, and a value greater than 0.29 is a large effect. We can therefore conclude that the proportions of medium shots in the American films of Max Ophüls are not homogenous, but that there is only a small difference between the films. Note that the *p*-value is *not* an effect size, and that even though observing this data given the model of homogeneity is highly implausible the effect size is small.

Step 5: conduct the post-hoc tests

The chi-square test is an omnibus test and tells us only if there is a difference in the sample of films. It does not tell us where this difference lies.[4] In order to discover which film(s) in our sample contributed to the significant chi-square result we compare the adjusted standardized residuals to a critical z-value. To calculate the adjusted standardized residual we apply the formula

---

[3] For 2×2 contingency tables, the effect size is called *phi* (φ) and is the square root of χ^2/n.

[4] An alternative approach for a small sample such as this is to conduct post-hoc pairwise comparisons between the films using 2×2 contingency tables with a corrected significance level, where *k* is the number of tests to be performed. *k* is calculated as (G(G-1))/2, where G is the number of samples being compared. In this experiment we have four films, and so *k* = (4 × 3)/2 = 6. The Sidak corrected two-tailed *p*-value for the pairwise comparisons would be 0.0085. However, this is impractical for groups of films greater than 5 that require a large number of individual tests.

$$z_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij} \times \left(1 - \frac{r_i}{n}\right) \times \left(1 - \frac{c_j}{n}\right)}}$$

where r_i and c_j are the totals of the *i*th row and *j*th column, respectively. Substituting, we calculate the adjusted standardized residual for medium shots in *The Exile* to be

$$z_{ij} = \frac{109 - 124.75}{\sqrt{124.75 \times \left(1 - \frac{499}{2000}\right) \times \left(1 - \frac{500}{2000}\right)}} = -1.8796 \,,$$

and we can calculate the adjusted standardized residuals of the other cells in the same way. If the adjusted standardized residual is negative, then there are fewer medium shots than expected in that film and if it is positive there is a greater number than expected.

To interpret the adjusted standardized residuals we need to compare them to a test statistic at a specified level of significance, but first we need to correct the significance level to take into account the multiple tests involved in the post-hoc analysis. To adjust the significance level for multiple comparisons we apply the Sidak correction formula

$$\alpha_{adj} = 1 - (1 - \alpha)^{1/k} \,,$$

where α is the desired overall significance level, α_{adj} is the adjusted significance level for the multiple tests, and *k* is the number of cells in the comparison (see Beasley & Schumacker 1995). With four rows and two columns, Table 1 has eight cells and so our adjusted *p*-value for an overall significance level of 0.05 is $\alpha_{adj} = 1 - (1 - 0.05)^{1/8} = 0.0064$.[5] As the null hypothesis was non-directional we are performing a two-tailed test, and it is necessary to divide the corrected *p*-value by two in order to determine the critical value of z. The critical z-value can be determined using Microsoft Excel with the function =NORMSINV(1–(0.0064/2)), and is 2.7266. If an adjusted standardized residual is greater than 2.7266 or less than –2.7266, then we interpret this as statistically significant. It is sufficient that we calculate the adjusted standardized residuals for either of the 2-column categories, as the result will be the same in either case, but for this example we

[5] If we have a large number of films in our sample then the adjusted p-value may become too small to be useful at a significance level of 0.05, and we may increase the overall α for the multiple comparisons to (for example) 0.10. This does not affect the significance level of the omnibus test, which remains 0.05.

will calculate the results for both columns. From the adjusted standardized residuals in Table 4 we can see that the values for *Caught* are greater than the critical value. We can conclude that the frequency of medium shots in *Caught* is greater than expected, while this is not the case for the other films.

It is also useful to know the relative and absolute contributions of the cells in Table 3 to the chi-square test statistic, as this will gives us a greater sense of how the films in the sample differ from the expected value and from one another. To calculate the relative contribution of each cell to the chi-square test statistic, we simply divide the value of a cell by χ^2 and multiply by 100 to express its contribution as a percentage: the relative contribution of MSs in *The Exile* is (1.9885/14.3845) × 100 = 13.82%. Calculating the relative contribution of each cell in the contingency table we see that in *Caught* accounts for more than half χ^2, while the contribution of the other cells is much smaller.

The absolute contribution (r^2) of each cell is determined by dividing the cell values in Table 3 by n: for *The Exile*, the absolute contribution of the MSs to χ^2 is 1.9885/2000 = 0.0010. As before, we calculate the contributions of the other cells in exactly the same way and again we see that it is *Caught* that makes the largest contribution to r^2. Note that the sum of the absolute contributions is equal to the square of the effect size: $r^2 = V^2 = 0.0848^2 = 0.0072$.

Table 4 Adjusted standardized residuals, relative and absolute contributions for the Hollywood films of Max Ophüls

	z_{ij}	Relative contribution (%)	r^2
Exile, The – MS	-1.8796	13.82	0.0010
Letter from an Unknown Woman– MS	-1.0442	4.27	0.0003
Caught – MS	3.7293	54.42	0.0039
Reckless Moment, The – MS	-0.8055	2.54	0.0002
Exile, The – Other	1.8796	4.60	0.0003
Letter from an Unknown Woman – Other	1.0442	1.42	0.0001
Caught – Other	-3.7293	18.09	0.0013
Reckless Moment, The – Other	0.8055	0.84	0.0001
Total		100.00	0.0072

MS = medium shots

The post-hoc tests are only performed if the omnibus test is significant. If the result of the chi-square test is not significant, then there is no need to carry out the post-hoc tests.

Step 6: present the results
As before, we need to describe the analysis performed, the results, and our interpretation.

> A chi-square test was used to analyse the homogeneity of the proportions of medium shots in the four American films of Max Ophüls produced in 1947 and 1948. The results show that there is a statistically significant difference between these films but that this effect is small (χ^2 (3, n = 2000) = 14.3848, p = 0.0024, V = 0.0848). Post-hoc analysis of the adjusted standardized residuals using a critical z-value of 2.7266 revealed that *Caught* has a greater proportion of medium shots than expected, and also made the largest relative contribution to the chi-square statistic and the largest absolute contribution to r^2. The observed frequency of medium shots in the other Hollywood films of Ophüls did not significantly deviate from the expected value.

In addition to the above paragraph we would want to present the raw data along with the adjusted standardized residuals, and the relative and absolute contributions of the cells, and so we would also include Tables 1 and 4. In doing so we do not need to repeat all of the information in these tables, which makes our reporting of the results easier to read and understand. We might also want to include a bar chart such as Figure 1 in our paper, especially if the number of categories in the contingency table is large, but as our sample only includes four films it is redundant to include both Table 1 and Figure 1.

References

Beasley TM & Schumacker RE 1995 Multiple regression approach to analysing contingency tables: post-hoc and planned comparison procedures, *Journal of Experimental Education* 64 (1): 79-93.

Cohen J 1988 *Statistical Power Analysis for the Behavioural Sciences*, second edition. Hillsdale, NJ: Lawrence Earlbaum Associates.

van de Geer D 2008 *La Bête Humaine* (submitted 4 March 2008), Cinemetrics Database, http://www.cinemetrics.lv/movie.php?movie_ID=1257, accessed 21 September 2010.