Stat 515: Introduction to Statistics

Chapter 13

The Multinomial Experiment

- The **multinomial experiment** is very similar to the binomial experiment but instead of binary output we have k possible outcomes

- The **binomial** is a special case of the **multinomial** with k=2

The Multinomial Experiment

1. The experiment consists of n identical trials
2. There are k possible outcomes to each trial
3. The probabilities of the k outcomes $p_1, p_2, \ldots p_k$ are the same across trials
4. Trials are independent
5. The random variables $X_1, X_2, \ldots X_k$ are the number of observations that fall into each of the k outcomes

A One Way Table

Outcome	1	2	…	k
Count	n_1	n_2	…	n_k

- For each outcome, i, we have an associated count n_i

A One Way Table

Outcome	1	2	…	k
Count	p_1	p_2	…	p_k

- For each outcome, i, we have an associated population proportion, or probability, p_i

- Here, it is of interest to know whether or not these probabilities are equal or not

Hypothesis Test for Multinomial: Step 1

- It is of interest to know whether or not the k probabilities are equal or not
  - **Null hypothesis**: we assume that the population proportion equals some p_o
    - $H_o: p_1 = p_2 = \cdots = p_k = \frac{1}{k}$
  - **Alternative hypothesis:** What we're interested in
    - $Ha: at\ least\ one\ p_i\ is\ different.$

Hypothesis Test for Multinomial: Step 2

- **Check the assumptions:**

1. The experiment follows the multinomial
2. We have a random sample of the population of interest
3. The sample size n is large enough so that we expect at least 5 observation for each observation

Hypothesis Test for Multinomial: Step 3

- **Calculate Test Statistic, χ^{2*}**
  - The test statistic measures how different the sample probabilities we have are from the null hypothesis
  - We calculate the χ^2-score by assuming that all k probabilities are equal (we use $\mathrm{p_1 = p_2 = \cdots = p_k}$)

$$\chi^{2^*} = \frac{(n_1 - E_1)^2}{E_1} + \frac{(n_2 - E_2)^2}{E_2} + \cdots + \frac{(n_k - E_k)^2}{E_k}$$

  - Where $E_i = \mathrm{np_i} = \mathrm{n} * \left(\frac{1}{\mathrm{k}}\right) = \frac{\mathrm{n}}{\mathrm{k}}$

Hypothesis Test for Multinomial: Step 4

- **Reject when**

$$\chi^{2^*} > \chi^2_{1-\alpha,k-1}$$

Example

Candidate Chosen	Number of Votes
Bachmann	491
Cain	6,338
Gingrich	244,065
Huntsman	1,173
Johnson	211
Paul	78,360
Perry	2,534
Romney	168,123
Santorum	102,475
Total	603,770

Example

Candidate Chosen	Proportion
Bachmann	.0008
Cain	.0150
Gingrich	.4042
Huntsman	.0019
Johnson	.0003
Paul	.1298
Perry	.0042
Romney	.2785
Santorum	.1697
Total	1

Hypothesis Test for Multinomial: Step 1

- It is of interest to know whether or not the 9 probabilities are equal or not
  - **Null hypothesis**: we assume that the population proportion equals some p_o
    - $H_o: p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9 = \frac{1}{9}$
  - **Alternative hypothesis:** What we're interested in
    - $Ha: at\ least\ one\ p_i\ is\ different.$

Hypothesis Test for Multinomial: Step 2

- **Check the assumptions:**

1. The experiment follows the multinomial
2. We have a random sample of the population of interest – this is questionable
3. The sample size n is large enough so that we expect at least 5 observation for each observation

Hypothesis Test for Multinomial: Step 3

- **Calculate Test Statistic, χ^{2*}**
  - The test statistic measures how different the sample probabilities we have are from the null hypothesis
  - We calculate the χ^2-score by assuming that all k probabilities are equal (we use $\mathrm{p_1 = p_2 = \cdots = p_k}$)

$$\chi^{2^*} = \frac{(n_1 - E_1)^2}{E_1} + \frac{(n_2 - E_2)^2}{E_2} + \frac{(n_3 - E_3)^2}{E_3} + \frac{(n_4 - E_4)^2}{E_4} + \frac{(n_5 - E_5)^2}{E_5} + \frac{(n_6 - E_6)^2}{E_6} + \frac{(n_7 - E_7)^2}{E_7} + \frac{(n_8 - E_8)^2}{E_8} + \frac{(n_9 - E_9)^2}{E_9} = 954{,}281.3$$

  - Where $E_i = \mathrm{np_i} = \mathrm{n} * \left(\frac{1}{\mathrm{k}}\right) = \frac{603770}{9} = 67085.56$

Hypothesis Test for Multinomial: Step 4

- **Reject when**

$$\chi^{2^*} > \chi^2_{1-\alpha,k-1} = qchisq(1-\alpha, k-1)$$

$954{,}281.3 > 15.50731$ so we reject the null hypothesis – there is evidence that the probability the population equally prefers each candidate

Contingency Table

- We use a contingency, or two-way, table to look at two qualitative variables at this same time

- Similar to our analysis of scatterplots we want to see if there's an association or relationship between the two variables

Contingency Table

Outcome 1 Outcome 2	1	2	…	k	Column Total
1	n_{11}	n_{12}	…	n_{1k}	$n_{1.}$
2	n_{21}	n_{22}	…	n_{2k}	$n_{2.}$
…	…	…	…	…	…
m	n_{m1}	n_{m2}	…	n_{mk}	$n_{4.}$
Row Total	$n_{.1}$	$n_{.2}$	…	$n_{.k}$	n

- For each outcome couple, (i,j), we have a count

Contingency Table

Outcome 1 Outcome 2	1	2	…	k	Column Total
1	p_{11}	p_{12}	…	p_{1k}	$p_{1.}$
2	p_{21}	p_{22}	…	p_{2k}	$p_{2.}$
…	…	…	…	…	…
m	p_{m1}	p_{m2}	…	p_{mk}	$p_{4.}$
Row Total	$p_{.1}$	$p_{.2}$	…	$p_{.k}$	1

- For each outcome couple, (i,j), we have an associated population proportion, or probability, p_{ij}
- Here, it is of interest if outcome 2 is associated with outcome 1

Contingency Table

Outcome 1 Outcome 2	1	2	…	k	Column Total
1	p_{11}	p_{12}	…	p_{1k}	$p_{1.}$
2	p_{21}	p_{22}	…	p_{2k}	$p_{2.}$
…	…	…	…	…	…
m	p_{m1}	p_{m2}	…	p_{mk}	$p_{4.}$
Row Total	$p_{.1}$	$p_{.2}$	…	$p_{.k}$	1

- Recall, when we learned probability we used these two-way tables and calculated the table as a percentage and conditional percentage table
- Here, it is of interest if outcome 2 is associated with outcome 1 – if outcome 1 and outcome 2 are dependent

Hypothesis Test for Independence: Step 1

- It is of interest whether or not outcome 2 is associated with outcome 1 – if outcome 1 and outcome 2 are dependent
  - **Null hypothesis**: Outcome 1 and Outcome 2 are independent
  - **Alternative hypothesis:** Outcome 1 and Outcome 2 are independent

Hypothesis Test for Independence: Step 2

- **Check the assumptions:**

1. We have a random sample from the population of interest
2. We can consider this to be a multinomial with $m * k$ outcomes – represented by each cell in the table
3. The sample size n is large enough so that we expect at least 5 observation for each observation

Hypothesis Test for Independence: Step 3

- **Calculate Test Statistic, χ^{2*}**
  - The test statistic measures how different the sample probabilities we have are from the null hypothesis

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

$$= \frac{(n_{11} - E_{11})^2}{E_{11}} + \frac{(n_{12} - E_{12})^2}{E_{12}} + \cdots + \frac{(n_{mk} - E_{mk})^2}{E_{mk}}$$

  - Where $E_{ij} = \frac{n_{i.} * n_{.j}}{n}$

Hypothesis Test for Independence: Step 4

- **Reject when**

$$\chi^{2^*} > \chi^2_{1-\alpha,(m-1)(k-1)}$$

Contingency Table - Example

- Two Categorical Variables
  - Would you keep or turn in a $100 if you found it on the library floor?
  - Do you recycle?

	Keep It	Turn It In	Total
No	17	8	25
Yes	30	34	64
Total	47	42	89

Contingency Table - Example

Counts

	Keep It	Turn It In	Total
No	17	8	25
Yes	30	34	64
Total	47	42	**89**

Percent:
(Divide each box by the overall total)

	Keep It	Turn It In	Total
No	17/**89**	8/**89**	25/**89**
Yes	30/**89**	34/**89**	64/**89**
Total	47/**89**	42/**89**	89/**89**

	Keep It	Turn It In	Total
No	19.1%	8.989%	28.09%
Yes	33.71%	38.2%	71.91%
Total	52.81%	47.19%	100%

Contingency Table - Example

Counts

	Keep It	Turn It In	Total
No	17	8	**25**
Yes	30	34	**64**
Total	47	42	**89**

Conditional Percent
(Divide each interior box by the row total)

	Keep It	Turn It In	Total
No	17/**25** = .68	8/**25** =.32	25/**25** = 1
Yes	30/**64** = .4688	34/**64** = .5313	64/**64** = 1
Total	47/**89** = .5281	42/**89** = .4719	89/**89** = 1

	Keep It	Turn It In	Total
No	68%	32%	100%
Yes	46.88%	53.13%	100%
Total	52.81%	47.19%	100%

Contingency Table - Example

Counts

	Keep It	Turn It In	Total
No	17	8	25
Yes	30	34	64
Total	47	42	89

Percent

	Keep It	Turn It In	Total
No	19.1%	8.989%	28.09%
Yes	33.71%	38.2%	71.91%
Total	52.81%	47.19%	100%

Conditional Percent

	Keep It	Turn It In	Total
No	68%	32%	100%
Yes	46.88%	53.13%	100%
Total	52.81%	47.19%	100%

Contingency Table Example

Recycle\Money	Keep It	Turn It In	Total
No	68%	32%	100%
Yes	46.88%	53.13%	100%
Total	52.81%	47.19%	100%

– **Explanatory Variable**(rows): Recycling Status
– **Response Variable**(columns): Keep or Return Money

Contingency Table Example

Recycle\Money	Keep It	Turn It In	Total
No	68%	**32%**	100%
Yes	46.88%	**53.13%**	100%
Total	52.81%	47.19%	100%

- Does there appear to be an association between recycling and turning in money found on the floor?
  - Yes – by looking at the conditional percent contingency table it appears that a larger percent of people that recycle turn it in compared to those that keep it
    - For those who recycle more than half would turn it in as where only 32% of those who do not recycle would

Hypothesis Test for Independence: Step 1

- It is of interest whether or not whether or not someone recycles is associated with whether or not they'd turn in money
  - **Null hypothesis**: Someone recycles is not associated with whether or not they'd turn in money
  - **Alternative hypothesis:** Someone recycles is associated with whether or not they'd turn in money

Hypothesis Test for Independence: Step 2

- **Check the assumptions:**

1. We have a random sample from the population of interest
2. We can consider this to be a multinomial with $2*2=4$ outcomes – represented by each cell in the table
3. 89 is large enough for

Hypothesis Test for Independence: Step 3

- **Calculate Test Statistic, χ^{2*}**
  - The test statistic measures how different the sample probabilities we have are from the null hypothesis

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

$$= \frac{\left(17 - \frac{25 * 47}{89}\right)^2}{\frac{25 * 47}{89}} + \frac{\left(8 - \frac{25 * 42}{89}\right)^2}{\frac{25 * 42}{89}} + \frac{\left(30 - \frac{64 * 47}{89}\right)^2}{\frac{64 * 47}{89}}$$

$$+ \frac{\left(34 - \frac{64 * 42}{89}\right)^2}{\frac{64 * 42}{89}} = 3.219262$$

Hypothesis Test for Independence: Step 4

- **Reject when**

$$\chi^{2^*} > \chi^2_{1-\alpha,(m-1)(k-1)} = qchisq(.95,1)$$

$3.219262 > 3.841459$ so we fail to reject the null hypothesis – there isn’t enough evidence to show that someone who recycles is associated with whether or not they’d turn in money