STA221

Neil Montgomery

Last edited: 2017-01-23 15:07

χ^2 test for homogeneity

Do the rows (columns) have the same distributions? The mechanics were:

1. Because the row (column) probabilities are the same under H_0, we ended up with:

$$E_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{n_{\cdot\cdot}}$$

χ^2 test for homogeneity

Do the rows (columns) have the same distributions? The mechanics were:

1. Because the row (column) probabilities are the same under H_0, we ended up with:

$$E_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{n_{\cdot\cdot}}$$

2. Then we compared $\sum_{i,j} (E_{ij} - O_{ij})^2 / E_{ij}$ with a χ^2 distribution with $(r-1)(c-1)$ degrees of freedom.

“homogeneity” versus “independence””

We are going to learn a different test (of “independence”) that is mechanically identical. The difference is subtle, and is in how the data are collected.

“homogeneity” versus “independence”"

We are going to learn a different test (of “independence”) that is mechanically identical. The difference is subtle, and is in how the data are collected.

In a test of homogeneity, we are seeing if two or more sub-populations (the rows or columns) have the same distribution with respect to the levels of a factor variable.

“homogeneity” versus “independence”"

We are going to learn a different test (of “independence”) that is mechanically identical. The difference is subtle, and is in how the data are collected.

In a test of homogeneity, we are seeing if two or more sub-populations (the rows or columns) have the same distribution with respect to the levels of a factor variable.

e.g. NEJM papers from subpopulations: 1978-79, 1989, and 2004-05. Factor variable was “did the paper have statistics in it, yes/no?”

“homogeneity” versus “independence”"

We are going to learn a different test (of “independence”) that is mechanically identical. The difference is subtle, and is in how the data are collected.

In a test of homogeneity, we are seeing if two or more sub-populations (the rows or columns) have the same distribution with respect to the levels of a factor variable.

e.g. NEJM papers from subpopulations: 1978-79, 1989, and 2004-05. Factor variable was “did the paper have statistics in it, yes/no?”

In a test of independence, we have one population and are seeing if categorization into levels of two factor variables is done independently.

“homogeneity” versus “independence””

We are going to learn a different test (of “independence”) that is mechanically identical. The difference is subtle, and is in how the data are collected.

In a test of homogeneity, we are seeing if two or more sub-populations (the rows or columns) have the same distribution with respect to the levels of a factor variable.

e.g. NEJM papers from subpopulations: 1978-79, 1989, and 2004-05. Factor variable was “did the paper have statistics in it, yes/no?”

In a test of independence, we have one population and are seeing if categorization into levels of two factor variables is done independently.

Consider Q13 “Childbirth” from the textbook. Researchers followed up on 1178 births, classifying them as “did mother have epidural” and “was child breastfeeding at six months.”

testing independence

What would it mean for epidural status and breastfeeding status to be independent?

testing independence

What would it mean for epidural status and breastfeeding status to be independent?

It means: if you know a woman had an epidural, it would not change the distribution of the two levels of breastfeeding.

testing independence

What would it mean for epidural status and breastfeeding status to be independent?

It means: if you know a woman had an epidural, it would not change the distribution of the two levels of breastfeeding.

Or vice-versa - independence is symmetric.

testing independence

What would it mean for epidural status and breastfeeding status to be independent?

It means: if you know a woman had an epidural, it would not change the distribution of the two levels of breastfeeding.

Or vice-versa - independence is symmetric.

The null hypothesis is best stated in words that describe the two variables, e.g. "H_0: epidural status is independent of breastfeeding status."

testing independence

What would it mean for epidural status and breastfeeding status to be independent?

It means: if you know a woman had an epidural, it would not change the distribution of the two levels of breastfeeding.

Or vice-versa - independence is symmetric.

The null hypothesis is best stated in words that describe the two variables, e.g. "H_0: epidural status is independent of breastfeeding status."

The test is done once again *with row and column totals taken as constant*.

independence—the details

The joint distribution along with the marginals (in a 2 by 3 example):

	1	2	3	Row Marginal
1	p_{11}	p_{12}	p_{13}	$p_{1\cdot}$
2	p_{21}	p_{22}	p_{23}	$p_{2\cdot}$
Column Marginal	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	

independence—the details

The joint distribution along with the marginals (in a 2 by 3 example):

	1	2	3	Row Marginal
1	p_{11}	p_{12}	p_{13}	$p_{1\cdot}$
2	p_{21}	p_{22}	p_{23}	$p_{2\cdot}$
Column Marginal	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	

Independence just means:

$$P(\text{Row Level 1 } \textit{and} \text{ Column Level 1}) = P(\text{Row Level 1})P(\text{Column Level 1})$$

and so on for all row and column levels.

independence—the details

The joint distribution along with the marginals (in a 2 by 3 example):

	1	2	3	Row Marginal
1	p_{11}	p_{12}	p_{13}	$p_{1\cdot}$
2	p_{21}	p_{22}	p_{23}	$p_{2\cdot}$
Column Marginal	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	

Independence just means:

$$P(\text{Row Level 1 } \textit{and} \text{ Column Level 1}) = P(\text{Row Level 1})P(\text{Column Level 1})$$

and so on for all row and column levels.

In short:

$$p_{ij} = p_{i\cdot}p_{\cdot j}$$

independence—the details with fixed row/column totals

	1	2	3	Sum
1	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
2	n_{21}	n_{22}	n_{23}	$n_{2\cdot}$
Sum	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$n_{\cdot\cdot}$

independence—the details with fixed row/column totals

	1	2	3	Sum
1	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
2	n_{21}	n_{22}	n_{23}	$n_{2\cdot}$
Sum	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$n_{\cdot\cdot}$

Translating the last probability statement into a cell count version gives:

$$\frac{n_{ij}}{n_{\cdot\cdot}} = \frac{n_{i\cdot}}{n_{\cdot\cdot}}\frac{n_{\cdot j}}{n_{\cdot\cdot}}$$

independence—the details with fixed row/column totals

	1	2	3	Sum
1	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
2	n_{21}	n_{22}	n_{23}	$n_{2\cdot}$
Sum	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$n_{\cdot\cdot}$

Translating the last probability statement into a cell count version gives:

$$\frac{n_{ij}}{n_{\cdot\cdot}} = \frac{n_{i\cdot}}{n_{\cdot\cdot}} \frac{n_{\cdot j}}{n_{\cdot\cdot}}$$

And from here we get the (same!) expected cell count as before:

$$E_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{n_{\cdot\cdot}}$$

Those are the cell counts one would get under perfect independence.

the χ^2 test for independence

Compare:

$$\chi^2_{obs} = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

with a χ^2 distribution with $(r-1)(c-1)$ degrees of freedom.

the χ^2 test for independence

Compare:

$$\chi^2_{obs} = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

with a χ^2 distribution with $(r-1)(c-1)$ degrees of freedom.

Childbirth example in full:

the χ^2 test for independence

Compare:

$$\chi^2_{obs} = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

with a χ^2 distribution with $(r-1)(c-1)$ degrees of freedom.

Childbirth example in full:

H_0 : epidural status is independent of breastfeeding status.

H_a : (usually omitted) ... is not independent ...

childbirth example

The observed data:

```
##           Breastfeeding
## Epidural   No  Yes  Sum
##       No  284  498  782
##       Yes 190  206  396
##       Sum 474  704 1178
```

childbirth example

The observed data:

```
##          Breastfeeding
## Epidural   No  Yes  Sum
##       No  284  498  782
##       Yes 190  206  396
##       Sum 474  704 1178
```

The expected cell counts:

```
##          Breastfeeding
## Epidural    No   Yes
##       No  314.7 467.3
##       Yes 159.3 236.7
```

childbirth example

The results:

```
##
##  Pearson's Chi-squared test
##
## data:  childbirth
## X-squared = 14.869, df = 1, p-value = 0.0001152
```

Conclusion: there is evidence that epidural status is not independent of breastfeeding status.

example “Twins” (Part VI Review Q15)

A JAMA paper studied the relationship between the quality of prenatal care and the birth circumstances.

example "Twins" (Part VI Review Q15)

A JAMA paper studied the relationship between the quality of prenatal care and the birth circumstances.

H_0: is quality of prenatal care independent of the birth circumstances?

example "Twins" (Part VI Review Q15)

A JAMA paper studied the relationship between the quality of prenatal care and the birth circumstances.

H_0: is quality of prenatal care independent of the birth circumstances?

Observed:

```
##                Twin Births
## Level of Care Preterm complex Preterm simple Term / postterm Sum
##    Intensive               18             15              28  61
##    Adequate                46             43              65 154
##    Inadequate              12             13              38  63
##    Sum                     76             71             131 278
```

"Twins"

Expected:

```
##                Twin Births
## Level of Care Preterm complex Preterm simple Term / postterm Sum
##    Intensive             16.7           15.6            28.7  61
##    Adequate              42.1           39.3            72.6 154
##    Inadequate            17.2           16.1            29.7  63
##    Sum                   76.0           71.0           131.0 278
```

“Twins”

Expected:

```
##                 Twin Births
## Level of Care Preterm complex Preterm simple Term / postterm Sum
##    Intensive             16.7           15.6            28.7  61
##    Adequate              42.1           39.3            72.6 154
##    Inadequate            17.2           16.1            29.7  63
##    Sum                   76.0           71.0           131.0 278
```

Results:

```
##
##  Pearson's Chi-squared test
##
## data:  twins
## X-squared = 6.1437, df = 4, p-value = 0.1887
```

χ^2 tests—when are the p-values accurate?

The test statistic has an approximate χ^2_ν distribution:

- when the observations in the dataset are independent.

χ^2 tests—when are the p-values accurate?

The test statistic has an approximate χ^2_ν distribution:

- when the observations in the dataset are independent.
  - Not related to the "test of independence"

χ^2 tests—when are the p-values accurate?

The test statistic has an approximate χ^2_ν distribution:

- when the observations in the dataset are independent.
  - Not related to the "test of independence"
  - Cannot usually be verified—only assumed based on the way the data were collected.

χ^2 tests—when are the p-values accurate?

The test statistic has an approximate χ^2_ν distribution:

- when the observations in the dataset are independent.
  - Not related to the "test of independence"
  - Cannot usually be verified—only assumed based on the way the data were collected.
- **and when expected cell counts are all over 5.**

χ^2 tests—when are the p-values accurate?

The test statistic has an approximate χ^2_ν distribution:

- when the observations in the dataset are independent.
  - Not related to the "test of independence"
  - Cannot usually be verified—only assumed based on the way the data were collected.
- **and when expected cell counts are all over 5.**
  - Other suggestions are common—all are fine.

χ^2 tests—when are the p-values accurate?

The test statistic has an approximate χ^2_ν distribution:

- when the observations in the dataset are independent.
  - Not related to the "test of independence"
  - Cannot usually be verified—only assumed based on the way the data were collected.
- **and when expected cell counts are all over 5.**
  - Other suggestions are common—all are fine.
  - Similar in spirit to the $np \geq 5$ suggestion in the "normal approximation to the binomial"

χ^2 tests—when are the p-values accurate?

The test statistic has an approximate χ^2_ν distribution:

- when the observations in the dataset are independent.
  - Not related to the "test of independence"
  - Cannot usually be verified—only assumed based on the way the data were collected.
- **and when expected cell counts are all over 5.**
  - Other suggestions are common—all are fine.
  - Similar in spirit to the $np \geq 5$ suggestion in the "normal approximation to the binomial"

χ^2 tests—when are the p-values accurate?

The test statistic has an approximate χ^2_ν distribution:

- when the observations in the dataset are independent.
  - Not related to the "test of independence"
  - Cannot usually be verified—only assumed based on the way the data were collected.
- **and when expected cell counts are all over 5.**
  - Other suggestions are common—all are fine.
  - Similar in spirit to the $np \geq 5$ suggestion in the "normal approximation to the binomial"

The second condition is the main thing to check.

χ^2 tests - other matters to ponder (or not)

The textbook mentions a few other things to "check" when it comes to goodness-of-fit tests. Here is some commentary:

- "counted data condition," which even the textbook admits is silly.

χ^2 tests - other matters to ponder (or not)

The textbook mentions a few other things to "check" when it comes to goodness-of-fit tests. Here is some commentary:

- "counted data condition," which even the textbook admits is silly.
  - the summary tables with counts are not datasets - they are summaries.

χ^2 tests - other matters to ponder (or not)

The textbook mentions a few other things to "check" when it comes to goodness-of-fit tests. Here is some commentary:

- "counted data condition," which even the textbook admits is silly.
  - the summary tables with counts are not datasets - they are summaries.
  - apparently some people try to apply the χ^2 procedures to other kinds of summary tables, which is why the book emphasizes this point as a warning.

χ^2 tests - other matters to ponder (or not)

The textbook mentions a few other things to “check” when it comes to goodness-of-fit tests. Here is some commentary:

- “counted data condition,” which even the textbook admits is silly.
  - the summary tables with counts are not datasets - they are summaries.
  - apparently some people try to apply the χ^2 procedures to other kinds of summary tables, which is why the book emphasizes this point as a warning.
- “randomization condition,” which has more to do with the possibility of *inferring something about a larger population, or not* than anything to do with χ^2 tests per se.

examples of verifying the assumptions and conditions

- In the Lotto 6/49 example:

examples of verifying the assumptions and conditions

- In the Lotto 6/49 example:
  - there was actually a very mild lack of independence among the rows in the dataset, but it wouldn't have effected the results.

examples of verifying the assumptions and conditions

- In the Lotto 6/49 example:
  - there was actually a very mild lack of independence among the rows in the dataset, but it wouldn't have effected the results.
  - **the expected cell counts were all much larger than 5.**

examples of verifying the assumptions and conditions

- In the Lotto 6/49 example:
  - there was actually a very mild lack of independence among the rows in the dataset, but it wouldn't have effected the results.
  - **the expected cell counts were all much larger than 5.**
  - I did indeed analyse counts.

examples of verifying the assumptions and conditions

- In the Lotto 6/49 example:
  - there was actually a very mild lack of independence among the rows in the dataset, but it wouldn't have effected the results.
  - **the expected cell counts were all much larger than 5.**
  - I did indeed analyse counts.
  - The analysis was not on any sample at all - I used *all* numbers ever drawn!

examples of verifying the assumptions and conditions

- In the Lotto 6/49 example:
  - there was actually a very mild lack of independence among the rows in the dataset, but it wouldn't have effected the results.
  - **the expected cell counts were all much larger than 5.**
  - I did indeed analyse counts.
  - The analysis was not on any sample at all - I used *all* numbers ever drawn!
- All other examples satisfied the $E_{ij} \geq 5$ condition, which is the main thing that should always be verified and commented on.

post-hoc investigations of χ^2 tests using residuals

The *chi*^{2} tests are based on the following standardized deviation of *observed* from *expected*:

$$\frac{O_i - E_i}{\sqrt{E_i}}$$

(similar with *ij* subscripts.)

post-hoc investigations of χ^2 tests using residuals

The *chi*^{2} tests are based on the following standardized deviation of *observed* from *expected*:

$$\frac{O_i - E_i}{\sqrt{E_i}}$$

(similar with *ij* subscripts.)

These can be called *standardized residuals*, where the $O_i - E_i$ are just the *residuals.*

post-hoc investigations of χ^2 tests using residuals

The *chi*^{2} tests are based on the following standardized deviation of *observed* from *expected*:

$$\frac{O_i - E_i}{\sqrt{E_i}}$$

(similar with *ij* subscripts.)

These can be called *standardized residuals*, where the $O_i - E_i$ are just the *residuals.*

These are approximately $N(0, 1)$, so one can glance at the cell-by-cell residuals to get information about which cells had the largest deviation from expected.

standardized residuals example - I (pipeline)

```
##       Pressure
## Leak  High  Low  Med  Sum
##   No   277  278  247  802
##   Yes   71   66   61  198
##   Sum  348  344  308 1000

##
##  Pearson's Chi-squared test
##
## data:  table(Leak, Pressure)
## X-squared = 0.16116, df = 2, p-value = 0.9226

##       Pressure
## Leak     High    Low    Med
##   No   -0.125  0.127 -0.001
##   Yes   0.253 -0.256  0.002
```

standardized residuals example - II (births)

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  childbirth
## X-squared = 14.388, df = 1, p-value = 0.0001487
```

```
##          Breastfeeding
## Epidural     No    Yes
##      No  -1.728  1.418
##      Yes  2.429 -1.993
```