STA 111: Probability & Statistical Inference

Lecture Seventeen – Tests of Independence and Goodness-of-Fit
D.S. Sections 10.1, 10.2 & 10.3

Instructor: Olanrewaju Michael Akande

Department of Statistical Science, Duke University

Outline

- Questions from Last Lecture
- Tests of Independence
- Goodness-of-Fit Tests
- Recap

Introduction

– Today we will continue with our discussions on hypothesis testing.

– We will talk about testing for independence between two "categorical" variables.

– Lastly, we will talk about one way of testing the null hypothesis that data comes from some distribution we have in mind. This is known as goodness-of-fit.

Contingency Tables

Often we observe data where each unit/individual can be categorized according to two different criteria (two categorical variables). For example:

- Each person gets a drug or a placebo, and each person is cured or not;
- Letter grade in a statistics course, and major.

A **contingency table** shows counts for two categorical variables.

Example 1: For example, suppose we observe 90 students at Duke and are interested in the relationship between major and gender, then we can display the data by gender and major as below:

		Major	
	Math	English	History
Male	10	20	15
Female	20	10	15

so that 10 males study math, 20 females study math, and so on.

Contingency Tables (Cont'd)

Example 2: Suppose one took U.S. states and classified them as to whether they supported Romney or Obama, and how many executions each state had in the last five years (e.g., 0, 1-5, more than 5). You might get a **contingency table** that looks like this:

	Obama	Romney	
0	20	1	21
1-5	5	8	13
>5	2	14	16
	27	23	50

Here there are 20 states that supported Obama and had no executions, 1 state that supported Romney and had no executions, and so forth.

Contingency Tables (Cont'd)

Usually, we are only interested in testing if there is some relationship or dependence between the two categorical variables.

Thus we can set up the general null and alternative hypotheses as:

H_0: The two variables are independent.
H_A: Some dependence exists between them.

For a given situation, it is always better to be clear and specific to the context of the problem. For the previous example, the hypotheses are:

H_0: Voting preference has nothing to do with execution rates.
H_A: There is a relationship between voting choice and executions.

Unlike before, there is only one choice for the null and alternative hypothesis. But as with all of our hypothesis tests, there are three parts. We now we need to get a test statistic and a critical value/p-value.

Contingency Tables (Cont'd)

The test statistic is

$$ts = \sum_{\textbf{all cells}} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}.$$

The O_{ij} is the observed count for the cell in row i, column j.

The E_{ij} uses the following formula:

$$E_{ij} = \frac{(i\textbf{th row sum}) * (j\textbf{th column sum})}{\textbf{total}}$$

Contingency Tables (Cont'd)

For example 2, we find:

$$\begin{aligned} E_{11} &= 21 * 27/50 = 11.34 \\ E_{12} &= 21 * 23/50 = 9.66 \\ E_{21} &= 27 * 13/50 = 7.02 \\ E_{22} &= 23 * 13/50 = 5.98 \\ E_{31} &= 27 * 16/50 = 8.64 \\ E_{32} &= 23 * 16/50 = 7.36 \end{aligned}$$

Then the test statistic is:

$$\begin{aligned} ts &= \frac{(20 - 11.34)^2}{11.34} + \frac{(1 - 9.66)^2}{9.66} + \ldots + \frac{(14 - 7.36)^2}{7.36} \\ &= 26.734. \end{aligned}$$

Contingency Tables (Cont'd)

We compare the test statistic to the value from a chi-squared distribution with degrees of freedom equal to

$$k = \textbf{(number of rows - 1)} * \textbf{(number of columns - 1)}.$$

For our example, $k = (3-1)*(2-1) = 2$.

The p-value is the chance of getting a chi-squared random variable greater than or equal to the observed test statistic, or

$$P\textbf{-value} = \mathbf{P}[W \geq ts]$$

where W has the chi-squared distribution with k degrees of freedom.

For a chi-squared random variable with 2 degrees of freedom, the table shows that the chance of getting a value bigger than 26.734 is less than 0.01.

Contingency Tables (Cont'd)

So the p-value is much less than 0.01. We strongly reject the null hypothesis at even $\alpha = 0.01$. There is major evidence that political preference and execution rates are somehow connected.

But the connection can be very subtle. We cannot infer causation, and the apparent relationship may not be at all what we expect. For example, one might argue that voting preferences reflect economic hardship, and states with economic hardship experience more violent crime and thus use the death penalty more often.

Sometimes there are hidden confounders that are more interesting than the relationship between the two classification criteria. It can even happen that the hidden confounder can reverse the apparent relationship in the data. When this happens, it is called **Simpson's Paradox**. This is something we already talked about briefly.

Contingency Tables (Cont'd)

Example 1 again (to be done in class): Let's revisit example 1.

	Major		
	Math	English	History
Male	10	20	15
Female	20	10	15

Can we test for dependence between major and gender?

Contingency Tables (Cont'd)

Example 3 (D.S. Example 10.3.3 – to be done in class):
The Bureau of Business and Economic Research at the University of Montana conducted a poll of opinions of Montana residents in May 1992. Among other things, respondents were asked whether their personal financial status was worse, the same, or better than one year ago. The participants were also asked about their income range. The table below gives a cross-tabulation of the answers to both questions.

	Personal Financial Status		
Income Range	Worse	Same	Better
Under $20,000	20	15	12
$20,000 – $35,000	24	27	32
Over $35,000	14	22	23

Can we test the null hypothesis that income is independent of opinion on personal financial status?

Goodness-of-Fit Tests

Goodness-of-fit tests are used to decide whether data accord well with a particular theory.

Example 4: For example, recall that Gregor Mendel was an Augustinian monk in charge of the monastery's truck garden. He noted that several traits in pea plants, e.g.:

- color
- height
- wrinkled pods

seemed to be inherited from the parent plants in a predictable way.

To study color Mendel got inbred strains, whose progeny were always yellow or always green. Then he did experiments in which those inbred strains were crossed, and he observed the colors of the offspring.

Goodness-of-Fit Tests (Cont'd)

Based on his experiments, he came up with expected ratios of red, white and pink peonies from crossing peonies.

For this example, we want to know whether the counts of red, white and pink peonies for observed data agree closely with the 1/4, 1/4, 1/2 ratios that are predicted.

In this type of test, the null and alternative are always of the same form just like we had for testing independence. They are:

H_0: The model holds vs. H_A: The model fails.

For our example,

H_0: The ratios of red, white and pink are 1/4: 1/4/: 1/2
H_A: The ratios differ from 1/4: 1/4/: 1/2.

Goodness-of-Fit Tests (Cont'd)

Our test statistic is similar to that for contingency tables (because testing for independence is testing for a specific kind of model). Here, the test statistic is:

$$ts = \sum_{\textbf{all } i} \frac{(O_i - E_i)^2}{E_i}$$

where the sum is taken over all categories (i.e., red, white, and pink). The O_i is the observed count in category i, and E_i is the count predicted in that category by the model.

To be concrete, suppose Darwin had made 100 crosses of pink with pink and had gotten 22 red, 29 white, and 49 pink. So $O_1 = 22$, $O_2 = 29$, and $O_3 = 49$.

The expected counts are those predicted by the model. Thus $E_1 = 25$, $E_2 = 25$, and $E_3 = 50$.

Goodness-of-Fit Tests (Cont'd)

The numerical value of the test statistic is

$$\begin{aligned} ts &= \frac{(22-25)^2}{25} + \frac{(29-25)^2}{25} + \frac{(49-50)^2}{50} \\ &= 1.02. \end{aligned}$$

The significance probability comes from a chi-squared table. Let W be a chi-squared random variable with

$$k = \#\textbf{categories} - 1$$

degrees of freedom. In this example, $k = 3 - 1 = 2$.

The significance probability is:

$$P - \textbf{value} = \mathbf{P}[W \geq ts] = \mathbf{P}[W \geq 1.02].$$

From the table, this is greater than 0.25. So the null is not rejected. The data support Mendel.

Goodness-of-Fit Tests (Cont'd)

Example 5 (D. S. Example 10.1.3 – to be done in class): From a study of blood types among a sample of 6004 Californians, suppose that the actual counts of people with the four blood types are given in the table below:

A	B	AB	O
2162	738	228	2876

Now suppose that the theoretical probabilities of blood types for white Californians is:

A	B	AB	O
$1/3$	$1/8$	$1/24$	$1/2$

Can we test the null hypothesis that the theoretical probabilities hold?

Goodness-of-Fit Tests (Cont'd)

Example 3 again (to be done in class):

	Personal Financial Status		
Income Range	Worse	Same	Better
Under \$20,000	20	15	12
\$20,000 – \$35,000	24	27	32
Over \$35,000	14	22	23

Suppose we collapse the data by responses, can we test the null hypothesis that the probabilities of the three responses are all equal to $1/3$?

Recap

Today we learned about tests of independence and goodness-of-fit tests.

You should be able to set-up such testing problems, calculate p-values, and make the appropriate conclusions.

In the next lecture, we will talk about linear regression, which is one way to examine the relationship between two variables.