Slide Deck E4:

A Goodness-of-Fit Test

The section in which we cover the Chi-square Goodness-of-Fit test. This test is used to compare an observed (categorical) distribution to a hypothesized one. While Pearson developed this test for the stated purpose in 1900, Fisher attempted to extend it as a test of Normality before giving up on it for that purpose.

Today's Objectives

By the end of this slidedeck, you should

1. understand the theory behind, and test hypotheses about:
   - comparing an observed (categorical) distribution to a hypothesized one
2. better understand the p-value and how to test hypotheses
3. understand why confidence intervals are not appropriate for this test

Note that we are moving beyond the general theory of confidence intervals and hypothesis testing. We are looking at how to specifically perform the procedures. Make sure you pay attention to the statistical process we follow.

Goodness of Fit

Parametric Procedure: Chi-Square Goodness-of-Fit Procedure

- Null hypothesis: Data are generated by the hypothesized distribution
- Graphic: Binomial plot
  `binom.plot(x=c(x1,x2,...,xk), n=c(n1,n2,...,nk))`
- Requires: Expected number of successes is at least 5 in each group*
- R function:
  `chisq.test(x=c(x1,x2,...,xk), p=c(p1,p2,...,pk))`

Note: This function is *not* what Hawkes covers. They use something close to this, but this procedure makes adjustments for the fact that the Binomial distribution is discrete and the Normal distribution is not. The "hand" calculations agree with Hawkes, however.

Framing Example

Example

I would like to test if my three-sided die is fair. To do this, I roll it $n = 600$ times and tabulate the observed frequency distribution.

In those 600 rolls, I got $n_1 = 180$ ones, $n_2 = 215$ twos, and $n_3 = 205$ threes.

Framing Example

Example

I would like to test if my three-sided die is fair. To do this, I roll it $n = 600$ times and tabulate the observed frequency distribution.

In those 600 rolls, I got $n_1 = 180$ ones, $n_2 = 215$ twos, and $n_3 = 205$ threes.

That is, we are given the following information:

- Observed counts, $\{180, 215, 205\}$
- Expected counts, $\{200, 200, 200\}$

The Theory

Note what information the above gives to us:

- Observed counts: $\{x_1, x_2, \ldots, x_k\}$
- Expected counts: $\{\mu_1, \mu_2, \ldots, \mu_k\}$

The goal is to create a test statistic that measures how far the observed counts are from the expected counts, while still having a distribution we know (or can determine).

The Theory

It can be shown* that this test statistic approximately follows a Chi-square distribution with $k-1$ degrees of freedom (if the μ_i are large enough):

$$TS = \sum_{i=1}^{k} \frac{(x_i - \mu_i)^2}{\mu_i}$$

It just requires that we can determine the expected value of each count, μ_i.

* This is proven in STAT 225 and STAT 321. While the proof is a three-liner, it does require a couple of definitions. As such, it is beyond the scope of this course.

Framing Example, continued

Recall from earlier that we have

- Observed counts, $\{180, 215, 205\}$
- Expected counts, $\{200, 200, 200\}$

Where did the expected counts come from?

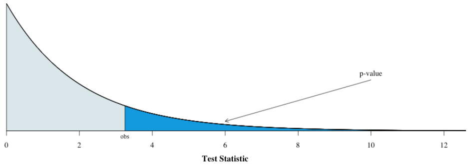
- Recall from the Binomial distribution that $\mu_i = np_i$
- Here,
  - n is the number of rolls
  - p_i is the probability of the i^{th} outcome

Framing Example, continued

$$
\begin{aligned}
TS &= \sum_{i=1}^{k} \frac{(x_i-\mu_i)^2}{\mu_i} \\
&= \frac{(x_1-\mu_1)^2}{\mu_1} + \frac{(x_2-\mu_2)^2}{\mu_2} + \frac{(x_3-\mu_3)^2}{\mu_3} \\
&= \frac{(180-200)^2}{200} + \frac{(215-200)^2}{200} + \frac{(205-200)^2}{200} \\
&= \frac{(-20)^2}{200} + \frac{(15)^2}{200} + \frac{(5)^2}{200} \\
&= \frac{400}{200} + \frac{225}{200} + \frac{25}{200} \\
&= 2.000 + 1.125 + 0.125 \\
&= 3.250
\end{aligned}
$$

Framing Example, continued

How does this test statistic value (3.250) compare to the Chi-square distribution with $k-1=2$ degrees of freedom?



Framing Example, continued

Conclusion:
We would like to test if the three-sided die is fair. To test this, we rolled the die 100 times. In those rolls, ones came us 180 times; twos, 215 times; and threes, 205 times. To test if there is significant evidence that the die is unfair, we use the Chi-squared goodness-of-fit test.

Because the test's p-value of 0.1969 is greater than our selected value of $\alpha = 0.05$, we fail to reject the null hypothesis. There is not enough evidence to conclude that the die is unfair.

As an aside:
We are also unable to conclude that the die is fair. In fact, this experiment gave us no additional information about the outcome distribution of my favorite three-sided die.

Framing Example, continued

R **Code Options** ("by hand"):

```
obs = c(180, 215, 205)
exp = c(200, 200, 200)
TS = sum( (obs-exp)^2/exp )
TS
1-pchisq(TS, df=2)
```

R output:

```
> TS
[1] 3.25
> 1-pchisq(TS, df=2)
[1] 0.1969117
```

Note: This is the process you will have to use if you are doing the **Hawkes homework**.

Framing Example, continued

…or

```
chisq.test(x=c(180,215,205), p=c(1/3, 1/3, 1/3))
```

R output:

```
        Chi-squared test for given probabilities

data:  c(180, 215, 205)
X-squared = 3.25, df = 2, p-value = 0.1969
```

Note: This is the process you should use if you are performing **genuine statistical analyses**.

Example 1: Car Origin

Example

My friend claims that the proportion of cars on the Knox campus that are American is the same as the proportion that are European and the proportion that are Asian.

To test this, I went to the parking lot across Berrien from SMC and counted the cars and their origins. In that sample, there were 19 American, 23 Asian, and 2 European cars.

From this, the observed and expected values are

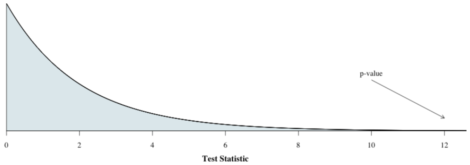
$$\text{Observed} = \{19, 23, 2\}$$
$$\text{Expected} = \{44/3, 44/3, 44/3\}$$

Example 1: Car Origin

$$
\begin{aligned}
TS &= \sum_{i=1}^{k} \frac{(x_i - \mu_i)^2}{\mu_i} \\
&= \frac{(x_1 - \mu_1)^2}{\mu_1} + \frac{(x_2 - \mu_2)^2}{\mu_2} + \frac{(x_3 - \mu_3)^2}{\mu_3} \\
&= \frac{(19 - 44/3)^2}{44/3} + \frac{(23 - 44/3)^2}{44/3} + \frac{(2 - 44/3)^2}{44/3} \\
&= \frac{(4.333)^2}{14.667} + \frac{(8.333)^2}{14.667} + \frac{(-12.667)^2}{14.667} \\
&= \frac{18.778}{14.667} + \frac{69.444}{14.667} + \frac{160.444}{14.667} \\
&= 1.280 + 4.735 + 10.939 \\
&= 16.954
\end{aligned}
$$

Example 1: Car Origin

How does this test statistic value (16.954) compare to the Chi-square distribution with $k - 1 = 2$ degrees of freedom?



Example 1: Car Origin

Conclusion:
We would like to test if the proportions of American, Japanese, and European cars are equal at Knox College. To test this, we examined the country-of-origin of cars in the parking lot on Berrien and Academy. In this lot, the number of American, European, and Japanese cars is 19, 23, and two. To test if there is significant evidence that there is a difference in the origin proportions, we use the Chi-squared goodness-of-fit test.

Because the p-value is much less than our selected value of $\alpha = 0.05$, we reject the null hypothesis. We are able to conclude that the proportion of cars from the three regions is not the same.

Question: What does this conclusion assume?

Example 1: Car Origin

R **Code Options** (“by hand”):

```
obs = c(19, 23, 2)
exp = c(44/3, 44/3, 44/3)
TS = sum( (obs-exp)^2/exp )
TS
1-pchisq(TS, df=2)
```

R output:

```
> TS
[1] 16.95455
> 1-pchisq(TS, df=2)
[1] 0.0002081456
```

Example 1: Car Origin

. . . or

```
chisq.test( x=c(19,23,2) )
```

R output:

```
        Chi-squared test for given probabilities

data:  c(19, 23, 2)
X-squared = 16.955, df = 2, p-value = 0.0002081
```


Example 2: *Quo vadis?*

Example

The Department of Mathematics claims that the proportion of its graduates who went to graduate school is twice the proportion of any other post-baccalaureate path.

To test this, the Department sent out a questionnaire to all of the alums for whom they had current addresses. Here is a table of our results from those who responded:

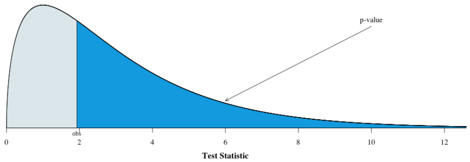
Category	Grad School	Business	Education	Unemployed
Count	13	7	10	5
Expected	14	7	7	7

Example 2: *Quo vadis?*

$$
\begin{aligned}
TS &= \sum_{i=1}^{k} \frac{(x_i - \mu_i)^2}{\mu_i} \\
&= \frac{(x_1 - \mu_1)^2}{\mu_1} + \frac{(x_2 - \mu_2)^2}{\mu_2} + \frac{(x_3 - \mu_3)^2}{\mu_3} + \frac{(x_4 - \mu_4)^2}{\mu_4} \\
&= \frac{(13 - 14)^2}{14} + \frac{(7 - 7)^2}{7} + \frac{(10 - 7)^2}{7} + \frac{(5 - 7)^2}{7} \\
&= \frac{(-1)^2}{14} + \frac{(0)^2}{7} + \frac{(3)^2}{7} + \frac{(-2)^2}{7} \\
&= \frac{1}{14} + \frac{0}{7} + \frac{9}{7} + \frac{4}{7} \\
&= 0.0714 + 0.0000 + 1.2857 + 0.5714 \\
&= 1.9286
\end{aligned}
$$

Example 2: *Quo vadis?*

How does this test statistic value (1.9286) compare to the Chi-square distribution with $k - 1 = 3$ degrees of freedom?



Example 2: *Quo vadis?*

Conclusion:
The Department of Mathematics at Knox College would like to determine if the proportion of its graduates who went to graduate school is twice the proportion of any other post-baccalaureate path. To test this, the department contacted its graduates. Of the 35 who responded, 13 attended graduate school, 7 went into business, 10 went into education, and 5 were unemployed. We used the Chi-square goodness-of-fit test to determine if the claim by the Department of Mathematics is reasonable.

Because the p-value of 0.5874 is greater than our selected value of $\alpha = 0.05$, we cannot reject the null hypothesis. The claim made by the Department of Mathematics that twice as many of its graduates go to graduate school than any other category is reasonable, given the data.

Question: What does this conclusion assume?

Example 2: *Quo vadis?*

R **Code Options** (“by hand”):

```
obs = c(13, 7, 10, 5)
exp = c(14, 7, 7, 7)
TS  = sum( (obs-exp)^2/exp )
TS
1-pchisq(TS, df=3)
```

R output:

```
> TS
[1] 1.928571
> 1-pchisq(TS, df=3)
[1] 0.5873635
```

Example 2: *Quo vadis?*

. . . or

```
chisq.test(x=obs, p=c(14,7,7,7)/35)
```

R output:

```
        Chi-squared test for given probabilities

data:  obs
X-squared = 1.9286, df = 3, p-value = 0.5874
```


Example 3: Knox College

Example

One initiative of Knox College is to become more representative of the US population. This raises a question of whether we have succeeded in terms of numbers.

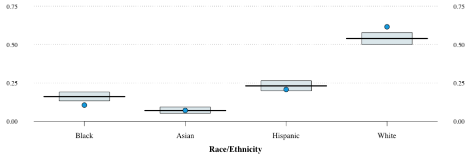
According to Fall 2019 domestic numbers, the data are

	Black	Asian	Hispanic	White
Observed	109	71	194	635
Population	0.1606	0.0709	0.2312	0.5389

Note: Here, we only focus on these four groups. While there are others, their numbers are low.

Example 3: Knox College

Here is a graphic showing where we were in Fall 2019... a snapshot in time. The dots represent the observed proportion. How should we interpret this graphic?



Start of Lecture Material
Goodness-of-Fit
A Few Examples
End of Section Material
Example 1: Car Origin
Example 2: *Quo vadis?*
Example 3: Knox College

Example 3: Knox College

In terms of the calculations, we have

```
chisq.test( x=c(109, 73, 214, 635), p=c(0.1606, 0.0709, 0.2312, 0.5389) )
```

which gives us

```
Error in chisq.test(x = c(109, 73, 214, 635),
    p = c(0.1606, 0.0709, 0.2312,  :
    probabilities must sum to 1.
```

Oops: What happened? Why did it happen? How do we fix it?

Example 3: Knox College

The fix the problem of probabilities not summing to one, include the flag

```
rescale.p = TRUE
```

like this:

```
chisq.test( x=c(109, 73, 214, 635), p=c(0.1606, 0.0709, 0.2312, 0.5389),
    rescale.p = TRUE )
```

The fixed code gives us the following results:

```
        Chi-squared test for given probabilities

data:  c(109, 73, 214, 635)
X-squared = 33.22, df = 3, p-value = 2.895e-07
```


Example 3: Knox College

Conclusion:

We would like to test if the proportions of selected racial-ethnic groupings for domestic students are the same between Knox College and the population of the United States. To test this, we examined the racial-ethnic distribution of students at Knox College in 2019 and compared these counts to the expected counts using the Chi-square goodness-of-fit test.

Because the p-value of less than 0.0001 is much less than our selected value of $\alpha = 0.05$, we reject the null hypothesis. The distribution of domestic students at Knox College does not currently match the demographic distribution of people in the United States.

Today's Objectives

Now that we have concluded this lecture, you should be able to

1. understand the theory behind, and test hypotheses about:
   - comparing an observed (categorical) distribution to a hypothesized one
2. better understand the p-value and how to test hypotheses
3. understand why confidence intervals are not appropriate for this test

Today's R Functions

Here is what we used the following R functions:

- `chisq.test(x, p)`
  This function performs the Chi-square Goodness-of-Fit test. This lecture also covers an important flag used in this function.

Supplemental Activities

The following activities are currently available from the STAT 200 website to give you some practice in performing hypothesis tests concerning the Chi-square Goodness-of-Fit test.

- SCA 32

Source: `https://www.kvasaheim.com/courses/stat200/sca/`

In addition to the SCAs, there are **Laboratory Activity E** (confidence intervals) and **Laboratory Activity F** (hypothesis testing).

Source: `https://www.kvasaheim.com/courses/stat200/labs/`

Supplemental Readings

The following are some readings that may be of interest to you in terms of understanding the theory of hypothesis testing:

• Hawkes Learning:	Section 10.6
• Intro to Modern Statistics:	Section 18.1
• R for Starters:	None
• Wikipedia:	Hypothesis Testing Pearson's chi-squared test

Please do not forget to be familiar with the `allProcedures` document that provides all of the statistical procedures we will use in R.