

Day 11

Outline

1. Fisher's Significance Tests
2. Goodness of fit test

Tl;DR

This model is concerned about the model used rather than actually trying to prove anything. If the initial hypothesis is rejected then we look into why it failed, rather than stopping at that conclusion.

Midterm Exam 1

Lecture

- ~ 60%
- 4 to 5 multi-part problems

Allowed: one-sided formula sheet (can be typed or printed)

Lab

- ~ 40%
- 3 problems, 2 to 3 part problems

Allowed: textbook, notes, software help, anything on Titanium

Fisher's Significance Testing

In Fisher's view, there is only one hypothesis. We see how well sample data "fit" that hypothesis.

In the strictest sense, the hypothesis includes all assumptions about the probability model used to obtain the sampling distribution of the test statistic.

Assumptions are of two kinds:

1. Assumptions about parameters
2. Assumptions about data generation/collection

In practice, we refer to the hypothesis as the **null hypothesis** (H_0).

Using the Null Hypothesis Significance Testing we write H_a : not H_0

Where we see this:

- χ^2 test (Chi-Squared)
- ANOVA

In Fisher's approach, we first specify our model

- Then: Specify sampling distribution of test statistic
- Then: Collect data and compute value of test statistic
- Then: Get P-Value

Recall: P-Value is probability of obtaining our data, or a result with a test statistic signalling equal or greater "distance" from H_0 , if H_0 is true.

Pure Fisher Philosophy: Stop here. P-Value represents "how well" data fit hypothesis.

- Very high P-Value: data fit suspiciously well
- Very low P-Value: data does not fit well at all

In practice: Fisher commends a personal significance level.

- Significance = "Signifying something"

$P\text{-Value} \leq \text{significance level}$: our results are a meaningful difference from the model. We should investigate!

- Reject H_0

$P\text{-Value} > \text{significance level}$: our results are consistent with the model. We did not prove it correct but the model is a reasonable approximation of reality.

- Fail to reject H_0

In practice: we define one main assumption about the parameters to be the H_0 that can get rejected.

Goodness of Fit Testing

Most of the “classic” goodness of fit tests involve genetics.

Example (Theory) : Mendel’s Pea Plants

Dihybrid cross for seed shape & seed color Mendel’s Laws: Should see a 9:3:3:1 ratio

Hypothesis: $\frac{9}{16}$ round/yellow, $\frac{3}{16}$ round/green, $\frac{3}{16}$ wrinkled/yellow, $\frac{1}{16}$ wrinkled/green

In practice:

$$H_0: P_{RY} = \frac{9}{16}, P_{RG} = \frac{3}{16}, P_{WY} = \frac{3}{16}, P_{WG} = \frac{1}{16}$$

We know the observed sample will probably not have those proportions.

We need a measure of “how far off” our sample is from what we expect. [χ^2 test statistic]

We define (Pearson) residuals for the different categories:

$$\text{Residual} = \frac{O-E}{\sqrt{E}}$$

- O : # of observed in sample
- E : # of expected in sample

$$\chi^2 = \sum \text{residual}^2 = \sum \frac{O-E}{\sqrt{E}}$$

Where Σ is all categories

Mendel observed:

- 315 RY
- 108 RG
- 101 WY
- 32 WG

Total: 556 seeds

What do we expect?

- $\frac{9}{16}(556) = 312.75$ RY
- $\frac{3}{16}(556) = 104.25$ RG
- $\frac{3}{16}(556) = 104.25$ WY
- $\frac{1}{16}(556) = 34.75$ WG

Pearson Residuals:

- RY : $\frac{315-312.75}{\sqrt{312.75}} = 0.127$
- RG : $\frac{108-104.25}{\sqrt{104.25}} = 0.367$
- WY : $\frac{101-104.25}{\sqrt{104.25}} = -0.318$
- WG : $\frac{32-34.75}{\sqrt{34.75}} = -0.467$

Goodness of Fit Testing (Continued)

Contribution of a category χ^2 = square of its Pearson residual

In our example:

$$\chi^2 = (0.127)^2 + (0.367)^2 + (-0.318)^2 + (-0.467)^2 = 0.47$$

Our sample data is 0.47 “off” from what we expected.

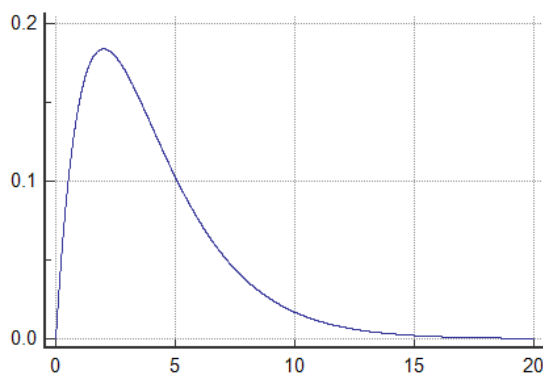
χ is unit-less.

To compute the P-Value, we need a sampling distribution of χ^2

$$\text{P-Value} = P(\chi^2 \geq 0.47 \mid H_0 \text{ is true})$$

Approximate the sampling distribution 1 of 2 ways:

1. Under H_0 , χ^2 has approximately a χ^2 distribution with $(\# \text{ of categories} - 1)$ degrees of freedom



- Strictly non-negative
- “Skewed right”

Software gives us approximate-value $P(\chi^2 \geq 0.47) = 0.9254$

If P-Value is “really small”: reject H_0 means “our proportions are not all correct” - we should investigate to find out which ones & why.

Mendel’s data was probably full crap, not on him but the guy collecting the peas.

2. Simulate very many samples of sizes n , under assumption of H_0 is true, and compute χ^2 for each simulated sample

When we expected ≥ 5 in each category in our sample, both approaches give similar results.

When in any category we expect < 5 counts, we use #2 above.