Null Hypothesis Significance Testing
Gallery of Tests

18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom

General pattern of NHST

You are interested in whether to reject H_0 in favor of H_A.

Design:

- Design experiment to collect data relevant to hypotheses.
- Choose text statistic x with known null distribution $f(x \mid H_0)$.
- Choose the significance level α and find the rejection region.
- For a simple alternative H_A, use $f(x \mid H_A)$ to compute the power.

Alternatively, you can choose both the significance level and the power, and then compute the necessary sample size.

Implementation:

- Run the experiment to collect data.
- Compute the statistic x and the corresponding p-value.
- If $p < \alpha$, reject H_0.

Concept question

We run a two-sample t-test for equal means, with $\alpha = .05$, and obtain a p-value of .04. What are the odds that the two samples are drawn from distributions with the same mean?

(a) $19/1$ (b) $1/19$ (c) $1/20$ (d) $1/24$ (e) unknown

answer: (e) unknown. Frequentist methods only give probabilities for data under an assumed hypothesis. They do not give probabilities or odds for hypotheses. So we don't know the odds for distribution means

Chi-square test for homogeneity

In this setting homogeneity means that the data sets are all drawn from the same distribution.

Three treatments for a disease are compared in a clinical trial, yielding the following data:

	Treatment 1	Treatment 2	Treatment 3
Cured	50	30	12
Not cured	100	80	18

Use a chi-square test to compare the cure rates for the three treatments

Solution

$H_0 =$ all three treatments have the same cure rate.
$H_A =$ the three treatments have different cure rates.

Under H_0 the MLE for the cure rate is

$$(\text{total cured})/(\text{total treated}) = 92/290 = .317 \quad .$$

Given H_0 we get the following table of observed and expected counts. We include the fixed values in the margins

	Treatment 1	Treatment 2	Treatment 3	
Cured	50, 47.6	30, 34.9	12, 9.5	92
Not cured	100, 102.4	80, 75.1	18, 20.5	198
	150	110	30	

Likelihood ratio statistic: $G = 2 \sum O_i \ln(O_i/E_i) = 2.12$

Pearson's chi-square statistic: $X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 2.13$

continued

Solution continued

Because the margins are fixed we can put values in 2 of the cells freely and then all the others are determined: degrees of freedom $= 2$.

```
p = 1 - pchisq(2.12, 2) = .346
```

The data does not support rejecting H_0. We do not conclude that the treatments have differing efficacy.

Board question: Khan's restaurant

Sal is thinking of buying a restaurant and asks about the distribution of lunch customers. The owner provides row 1 below. Sal records the data in row 2 himself one week.

	M	T	W	R	F	S
Owner's distribution	.1	.1	.15	.2	.3	.15
Observed # of cust.	30	14	34	45	57	20

Run a chi-square goodness-of-fit test on the null hypotheses:

H_0: the owner's distribution is correct.

H_A: the owner's distribution is not correct.

Compute both G and X^2

Solution

The total number of observed customers is 200.

The expected counts (under H_0) are 20 20 30 40 60 30

$$G = 2\sum O_i \log(O_i/E_i) = 11.39$$

$$X^2 = \sum \frac{(O_i - E_i)^2|}{E_i} = 11.44$$

$df = 6 - 1 = 5$ (6 cells, compute 1 value –the total count– from the data)

$p =$ `1-pchisq(11.39,5) = .044.`

So, at a significance level of 0.5 we reject the null hypothesis in favor of the alternative the the owner's distribution is wrong.

Board question: genetic linkage

In 1905, William Bateson, Edith Saunders, and Reginald Punnett were examining flower color and pollen shape in sweet pea plants by performing crosses similar to those carried out by Gregor Mendel.

Purple flowers (P) is dominant over red flowers (p).
Long seeds (L) is dominant over round seeds (l).

F0: PPLL x ppll (initial cross)
F1: PpLl x PpLl (all second generation plants were PpLl)
F2: 2132 plants (third generation)

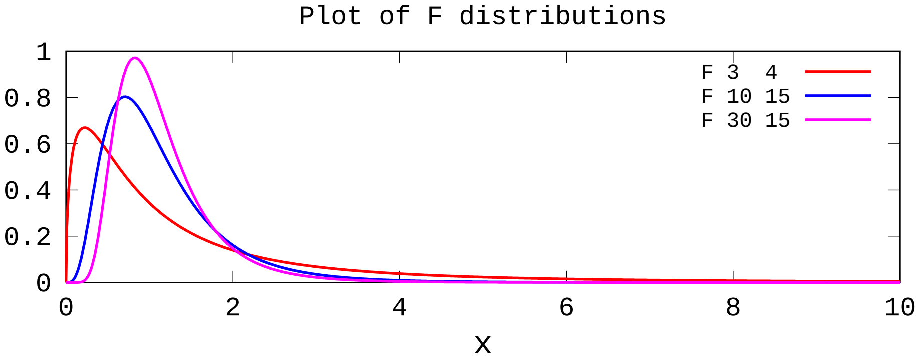
$H_0 =$ independent assortment.

	purple, long	purple, round	red, long	red, round
Expected	?	?	?	?
Observed	1528	106	117	381

Determine the expected counts for F_2 under H_0 and find the p-value for a Pearson Chi-squared test. Explain your findings biologically.

F-distribution

- Notation: $F_{a,b}$, a and b degrees of freedom
- Derived from normal data
- Range: $[0, \infty)$



F-test = one-way ANOVA

Like t-test but for n groups of data with m data points each.

$$y_{i,j} \sim N(\mu_i, \sigma^2), \qquad y_{i,j} = j^{\text{th}} \text{ point in i}^{\text{th}} \text{ group}$$

Null-hypothesis is that means are all equal: $\mu_1 = \cdots = \mu_n$

Test statistic is $\frac{\mathrm{MS}_B}{\mathrm{MS}_W}$ where:

MS_B = between group variance $= \dfrac{m}{n-1} \quad (\bar{y}_i - \bar{y})^2$

MS_W = within group variance = sample mean of $s_1^2, \ldots, s_n^2$

Idea: If μ_i are equal, this ratio should be near 1.

Null distribution is F-statistic with $n-1$ and $n(m-1)$ d.o.f.:

$$\frac{\mathrm{MS}_B}{\mathrm{MS}_W} \sim F_{n-1,\, n(m-1)}$$

Note: Formulas easily generalizes to unequal group sizes: http://en.wikipedia.org/wiki/F-test

Board question

The table shows recovery time in days for three medical treatments.

1. Set up and run an F-test.

2. Based on the test, what might you conclude about the treatments?

T_1	T_2	T_3
6	8	13
8	12	9
4	9	11
5	11	8
3	6	7
4	8	12

For $\alpha = .05$, the critical value of $F_{2,15}$ is 3.68.

Board question: chi-square for independence

(From Rice, *Mathematical Statistics and Data Analysis*, 2nd ed. p.489)

Consider the following contingency table of counts

Education	Married once	Married multiple times	Total
College	550	61	611
No college	681	144	825
Total	1231	205	1436

Use a chi-square test with significance level 0.01 to test the hypothesis that the number of marriages and education level are independent.

Solution

The null hypothesis is that the cell probabilities are the product of the marginal probabilities. Assuming the null hypothesis we estimate the marginal probabilities in red and multiply them to get the cell probabilities in blue.

Education	Married once	Married multiple times	Total
College	.365	.061	611/1436
No college	.492	.082	825/1436
Total	1231/1436	205/1436	1

We then get expected counts by multiplying the cell probabilities by the total number of women surveyed (1436). The table shows the observed, expected counts:

Education	Married once	Married multiple times
College	550, 523.8	61, 87.2
No college	681, 707.2	144, 117.8

Solution continued

We then have

$$G = 16.55 \qquad \text{and} \qquad X^2 = 16.01$$

The number of degrees of freedom is 1. This is because we specified the marginal probabilities and now any one of the cell probabilites determines all the rest. We get

$$p = \texttt{1-pchisq(16.55,1) = } .000047$$

Therefore we reject the null hypothesis in favor of the alternate hypothesis that number of marriages and education level are not independent

Concept question: multiple-testing

1. Suppose we use two-sample t-tests at $\alpha = .05$ level to determine whether 6 treatments all have the same recovery time. How many t-tests might we need to run?

1) 1 2) 2 3) 6 4) 15 5) 30

2. In the situation above, assuming all 6 means are the same, what is the probability that we reject at least one of the 15 null hypotheses?

1) Less than .05 2) .05 3) .10 4) Greater than .50

Discussion: What is an advantage of using the F-test rather than two-sample t-tests?