Stat 135 Lab 9

GSI: Yutong Wang

Apr 3, 2020

Questions?

Amandari Ubud, Bali, Indonesia. Luxury Hotel Review by TravelPlusStyle. Photo © Aman Resorts
https://www.travelplusstyle.com/hotels/amandari-ubud-bali-luxury-hotel-review-travelplusstyle

To-do Today

1. Review: Goodness of Fit Chi-squared test
2. Review: Test of Homogeneity
3. Review: Test of Independence
4. Review: Mann-Whitney Test
5. Practice Problems of 2-sample tests

Berkeley
UNIVERSITY OF CALIFORNIA

Review: Goodness of Fit Chi-squared Test

Does the population have some particular multinomial distribution?

We have a categorical random variable with m outcomes having probabilities $p_1, ..., p_m$. The chance a sample of size n for our box has a certain composition is given by the multinomial formula

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = \binom{n}{x_1, ..., x_m} p_1^{x_1} ... p_m^{x_m}$$

, where $\binom{n}{x_1, ..., x_m} = \frac{n!}{x_1! x_2! ... x_m!}$.

The goal is to test whether a model for the population distribution $(p_1, ..., p_m)$ fits our data. We draw n times with replacement from our box and get **observed counts** $x_1, ..., x_m$ with $\sum_{i=1}^{m} x_i = n$. If the probability of tickets in the box is $p_1(\theta), ..., p_m(\theta)$, we get **expected counts** $np_1(\theta), ..., np_m(\theta)$.

Berkeley
UNIVERSITY OF CALIFORNIA

Review: Goodness of Fit Chi-squared Test

Does the population have some particular multinomial distribution?

We do a **goodness of fit test**.

$$\sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{m-1-K}$$

, where K is the dimension of θ. A goodness of fit test explores how good your probability model fits your data. If the p-value of our Pearson χ^2 test statistic is smaller than α, we reject the null hypothesis.

Assumptions:

1. we have one categorical variable.
2. we have independent observations (draw with replacement).
3. the outcomes are mutually exclusive.
4. we require large n and no more than 20% expected counts are smaller than 5.

Berkeley
UNIVERSITY OF CALIFORNIA

Review: Test of Homogeneity

Do subgroups of population have the same multinomial distribution?

Suppose that we have independent observations from J multinomial distributions, each of which has I cells, and that we want to test whether the cell probabilities of the multinomials are equal—that is, to test the homogeneity of the multinomial distributions.

Observed counts:

n_{11}	n_{12}	n_{13}	$n_{1.}$
n_{21}	n_{22}	n_{23}	$n_{2.}$
$n_{.1}$	$n_{.2}$	$n_{.3}$	n

Expected counts:

$n_{1.}n_{.1}/n$	$n_{1.}n_{.2}/n$	$n_{1.}n_{.3}/n$
$n_{2.}n_{.1}/n$	$n_{2.}n_{.2}/n$	$n_{2.}n_{.3}/n$

The test statistic is $\chi^2_{\text{df}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i,j} \frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}}$

Degree of freedom is $\text{df} = \dim \Omega - \dim \omega_0 = J(I-1) - (I-1) = (J-1)(I-1)$

Review: Test of Independence

Are two categorical variables independent?

In a contigency table with I rows and J columns, the joint distribution of the counts n_{ij} is multinomial with cell probabilities denoted as π_{ij}. Let $\pi_{i.} = \sum_{j=1}^{J} \pi_{ij}$, and $\pi_{.j} = \sum_{i=1}^{I} \pi_{ij}$ denote the marginal probabilities that an observation will fall in the ith row and jth column, respectively.
If the row and column classifications are independent of each other, $\pi_{ij} = \pi_{i.}\pi_{.j}$. We thus consider testing the following null hypothesis: $H_0 : \pi_{ij} = \pi_{i.}\pi_{.j}, \forall i \in \{1, ..., I\}, j \in \{1, ..., J\}$, versus the alternative that the π_{ij} are free.

The test statistic is $\chi^2_{\text{df}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i,j} \frac{(n_{ij} - \frac{n_{i.}n_{.j}}{n})^2}{\frac{n_{i.}n_{.j}}{n}}$

Degree of freedom is $\text{df} = \dim \Omega - \dim \omega_0 = J(I-1) - (I-1) = (J-1)(I-1)$

Review: Mann-Whitney Test (nonparametric)

Suppose that we have m + n experimental units to assign to a treatment group and a control group. The assignment is made at random: n units are randomly chosen and assigned to the control, and the remaining m units are assigned to the treatment. We are interested in testing the **null hypothesis that the treatment has no effect.**

First, we group all m + n observations together and rank them in order of increasing size (we will assume for simplicity that there are no ties, although the argument holds even in the presence of ties). We next **calculate the sum of the ranks of those observations that came from the control group**. If this sum is too small or too large, we will reject the null hypothesis.

Review: Mann-Whitney Test (nonparametric)

Let T_Y denote the sum of the ranks of $Y_1, Y_2, \ldots, Y_m$. Using results from Chapter 7, we can easily find $E(T_Y)$ and $\mathrm{Var}(T_Y)$ under the null hypothesis $F = G$.

THEOREM A

If $F = G$,

$$E(T_Y) = \frac{m(m+n+1)}{2}$$

$$\mathrm{Var}(T_Y) = \frac{mn(m+n+1)}{12}$$

Q1: Unpaired 2-sample test in R (Rice: 11.6.21)

A study was done to compare the performances of engine bearings made of different compounds (McCool 1979). Ten bearings of each type were tested. The following table gives the times until failure (in units of millions of cycles):

Type 1	3.03	5.53	5.60	9.3	9.92	12.51	12.95	15.21	16.04	16.84
Type 2	3.19	4.26	4.47	4.53	4.67	4.69	12.78	6.79	9.37	12.75

Q1: Unpaired 2-sample test in R (Rice: 11.6.21)

1. Use normal theory to test the hypothesis that there is no difference between the two types of bearings.
2. Test the same hypothesis using a nonparametric method. (We did this in class)
3. Which of the methods that of part (a) or that of part (b) do you think is better in this case?
4. Estimate π, the probability that a type I bearing will outlast a type II bearing.
5. Use the bootstrap to estimate the sampling distribution of $\hat{\pi}$ and its standard error.
6. Use the bootstrap to find an approximate 90% confidence interval for π.

Q2: Paired 2-sample test (Rice:11.6.39)

An experiment was done to test a method for reducing faults on telephone lines (Welch 1987). Fourteen matched pairs of areas were used. The following table shows the fault rates for the control areas and for the test areas:

Test	676	206	230	256	280	433	337	466	497	512	794	428	452	512
Control	88	570	605	617	653	2913	924	286	1098	982	2346	321	615	519

Q2: Paired 2-sample test (Rice:11.6.39)

1. Plot the differences versus the control rate and summarize what you see.
2. Calculate the mean difference, its standard deviation, and a confidence interval.
3. Calculate the median difference and a confidence interval and compare to the previous result.
4. Do you think it is more appropriate to use a t test or a nonparametric method to test whether the apparent difference between test and control could be due to chance?

   Why? Carry out both tests and compare.