Sta 111 - Summer II 2017
Probability and Statistical Inference

17. Small sample inference

Lu Wang

Duke University, Department of Statistical Science

July 26, 2017

Outline

1. Small sample inference for a proportion

2. Small sample inference for difference between two proportions

3. Small sample inference for one-way or two-way tables

There is a saying “know something like the back of your hand”. Describe an experiment to test if people really do know the backs of their hands.

In the MythBusters episode, 11 out of 12 people guesses the backs of their hands correctly.

What are the hypotheses for evaluating if people are capable of recognizing the back of their hand at a rate that is better than random guessing. Remember, in the MythBusters experiment, there were 10 pictures to choose from, and only 1 was correct.

H_0 : $p = 0.10$ (random guessing)

H_A : $p > 0.10$ (better than random guessing)

Conditions

1. *Independence:* We can assume that each person's guessing is independent of another.

2. *Sample size:* The number of expected successes is *smaller than 10*.

$$12 \times 0.1 = 1.2$$

So what do we do? Since the sample size isn't large enough to use CLT based methods, we use a simulation method instead.

Randomization HT for a proportion

Describe how you would test if people are capable of recognizing the back of their hand at a rate that is better than random guessing.

$$H_0 : p = 0.10 \qquad H_A : p > 0.10 \qquad \hat{p} = 11/12 = 0.9167$$

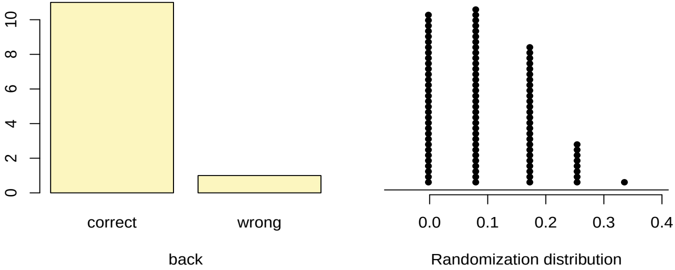
1. Use a 10-sided fair die to represent the sampling space, and call 1 a success (guessing correctly), and all other outcomes failures (guessing incorrectly).
2. Roll the die 12 times (representing 12 people in the experiment), count the number of 1s, and calculate the proportion of correct guesses in one simulation of 12 rolls.
3. Repeat step (2) many times, each time recording the proportion of successes in a series of 12 rolls of the die.
4. Create a dot plot of the simulated proportions from step (3) and count the number of simulations where the proportion was at least as high as 0.9167 (the observed proportion).

Simulation results

- In the next slide you can see the results of a hypothesis test (using only 100 simulations to keep things simple).
- Each dot represents a simulation proportion of success. There were 25-30 simulations where the success rate ($\hat{p}$) was 10%, 40-45 simulations where the success rate was slightly less than 10%, about 20 simulations where the success rate was slightly less than 20% and 1 simulation where the success rate was more than 30%.
- There are no simulations where the success rate is as high as the observed success rate of 91.67%.
- Therefore we conclude that the observed result is near impossible to have happened by chance (p-value = 0).
- And hence this data provide significant evidence that people are capable of recognizing the back of their hand at a rate that is better than random guessing.

```
back = as.factor(c(rep("correct", 11), rep("wrong", 1)))
inference(back, est = "proportion", type = "ht", method = "simulation",
^^Isuccess = "correct", null = 0.1, alternative = "greater", seed = 654, nsim = 100)
```

```
Single proportion -- success: correct
Summary statistics: p_hat = 0.9167 ;  n = 12
H0: p = 0.1
HA: p > 0.1
p-value =  0
```



MythBusters also asked these people to guess the palms of their hands. This time 7 out of the 12 people guesses correctly. The data are summarized below.

	Back	Palm	Total
Correct	11	7	18
Wrong	1	5	6
Total	12	12	24

Proportion of correct guesses

- Proportion of correct in the back group: $\frac{11}{12} = 0.916$
- Proportion of correct in the palm group: $\frac{7}{12} = 0.583$
- Difference: 33.3% more correct in the back of the hand group.

What are the hypotheses for comparing if the proportion of people who can guess the backs of their hands correctly is different than the proportion of people who can guess the palm of their hands correctly?

H_0: $p_{back} = p_{palm}$

H_0: $p_{back} \neq p_{palm}$

Conditions?

- Independence - within groups, between groups?
  - Within each group we can assume that the guess of one subject is independent of another.
  - Between groups independence is not satisfied - we have the same people guessing.
- Sample size?
  - $\hat{p}_{pool} = \frac{11+7}{12+12} = \frac{18}{24} = 0.75$
  - Expected successes in back group: $12 \times 0.75 = 9$, failures = 3 *<10*
  - Expected successes in palm group: $12 \times 0.75 = 9$, failures = 3 *<10*
- Since independence and S/F conditions fail, we need to use simulation to compare the proportions.

1. Use 24 index cards, where each card represents a subject.
2. Mark 18 of the cards as “correct” and the remaining 6 as “wrong”.
3. Shuffle the cards and split into two groups of size 12, for back and palm.
4. Calculate the difference between the proportions of “correct” in the back and palm decks, and record this number.
5. Repeat steps (3) and (4) many times to build a randomization distribution of differences in simulated proportions.

If results from the simulations based on the *null model* look like the data, then we can determine that the difference between the proportions of correct guesses in the two groups was simply *due to chance*.

If the results from the simulations based on the *null model* do not look like the data, then we can determine that the difference between the proportions of correct guesses in the two groups was *not due to chance*, but *because people actually know the backs of their hands better*.

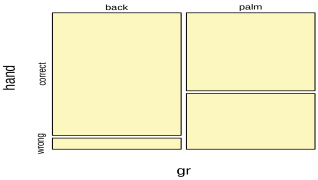
Simulation results

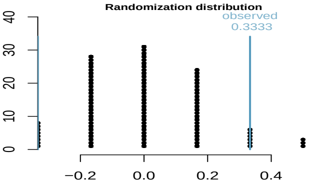
- In the next slide you can see the result of a hypothesis test (using only 100 simulations to keep the results simple).

- Each dot represents a difference in simulated proportion of successes. We can see that the distribution is centered at 0 (the null value).
- We can also see that 18 out of the 100 simulations yielded simulated differences (absolute value) at least as large as the observed difference (p-value = 0.18).

```
hand = as.factor(c(rep("correct", 7), rep("wrong", 5), c(rep("correct", 11), rep("wrong", 1))))
gr = c(rep("palm",12),rep("back",12))
inference(hand, gr, est = "proportion", type = "ht", null = 0, alternative = "twosided",
^^Iorder = c("back","palm"), success = "correct", method = "simulation", seed = 879,
^^Insim = 100)
```

```
Response variable: categorical, Explanatory variable: categorical
Difference between two proportions -- success: correct
Summary statistics:
         x
y          back palm Sum
  correct    11    7  18
  wrong       1    5   6
  Sum        12   12  24
Observed difference between proportions (back-palm) = 0.3333
H0: p_back - p_palm = 0
HA: p_back - p_palm != 0
p-value =  0.18
```





Conclusion

There were 33.3% more correct in the back group in the data. Do the simulation results suggest that people know the backs of their hands significantly better?

(a) Yes

(b) No

p-value = 0.18 > 0.05, fail to reject H_0. The data do not provide convincing evidence that people know the backs of their hands differently than the palms of their hands.

Simulation scheme

- create a randomized contingency table under the *null hypothesis*, then compute a chi-square test statistic χ^2_{sim}
- repeat this many times and examine the distribution of these simulated test statistics – *null distribution*
- As before, we can use the upper tail of this null distribution to calculate the p-value.

Remark

- This randomization approach is valid for any sized sample, especially for cases where one or more expected cell counts do not meet the minimum threshold of *5*.
- When the minimum threshold is met, the simulated null distribution will very closely resemble the chi-square distribution.