Statistical Inference

Inference for Categorical Variables

Behnam Bahrak
Spring 2020

Small Sample Proportion

Paul the Octopus

- Paul the Octopus predicted 8 World Cup 2010 games, and predicted them all correctly.
- Does this provide convincing evidence that Paul actually has psychic powers, i.e. that he does better than just randomly guessing?

Small Sample Proportion

- Paul the Octopus predicted 8 World Cup games, and predicted them all correctly. Does this provide convincing evidence that Paul actually has psychic powers, i.e. that he does better than just randomly guessing?

$$H_0: p = 0.5 \qquad n = 8 \;\;,\;\; \hat{p} = 1$$
$$H_A: p > 0.5$$

1. Independence:
   we can assume that his guesses are independent
2. Sample size / skew: $8 \times 0.5 \;=\; 4$ $\rightarrow$ not met
   distribution of sample proportions cannot be assumed to be nearly normal

Revisit: inference via simulation

- The ultimate goal of a hypothesis test is a p-value
- p-value = P(observed or more extreme outcome | $\boldsymbol{H_0}$ **true**)
- Devise a simulation scheme that assumes the null hypothesis is true
- Repeat the simulation many times and record relevant sample statistic
- Calculate p-value as the proportion of simulations that yield a result favorable to the alternative hypothesis

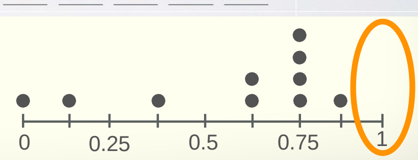
Small Sample Proportion

- Paul the Octopus predicted 8 World Cup games, and predicted them all correctly. Does this provide convincing evidence that Paul actually has psychic powers, i.e. that he does better than just randomly guessing? $H_0: p = 0.5$ $H_A: p > 0.5$
- Use a fair coin, and label head as success (correct guess)
- One simulation: flip the coin 8 times and record the proportion of heads (correct guesses): $\hat{p}_{sim}$
- Repeat the simulation many times, recording the proportion of heads at each iteration: $\hat{p}_{sim,1}, \hat{p}_{sim,2}, \cdots, \hat{p}_{sim,N}$
- Calculate the percentage of simulations where the simulated proportion of heads is at least as extreme as the observed proportion

Inference via Simulation

									$\hat{p}$
simulation 1:	H	H	H	T	H	H	H	H	7 / 8 = 0.875
simulation 2:	T	H	H	T	H	T	T	T	3 / 8 = 0.375
simulation 3:	T	T	H	H	H	H	T	H	5 / 8 = 0.625
...									...
simulation 10:	T	H	T	H	H	H	H	H	6 / 8 = 0.75

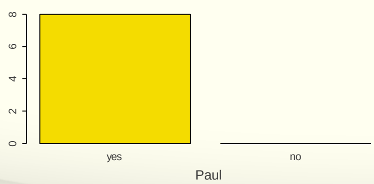
p-value $= P(\hat{p}_{sim} \geq 1 \mid p = 0.5)$
$= 0$

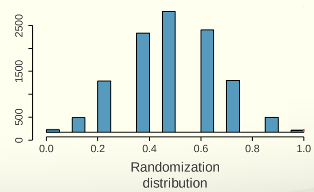


Inference via Simulation

R

```
> source("inference.R")
> paul = factor(c(rep("yes", 8), rep("no", 0)), levels = c("yes","no"))
> inference(paul, est = "proportion", type = "ht", method =
"simulation", success = "yes", null = 0.5, alternative = "greater")
Single proportion -- success: yes
Summary statistics: p_hat = 1 ; n = 8
H0: p = 0.5
HA: p > 0.5
p-value = 0.0037
```





Example

- There is a saying in English *"to know something like the back of your hand"*, which means to know something very well. MythBusters (a popular TV show) put to test the validity of this saying.
- They recruited 12 volunteers, each of whom were shown 10 pictures of backs of hands (while wearing gloves so they couldn't see their own hands), and asked them to identify their own hand among the 10 pictures. 11 out of 12 people completed the task successfully.
- What are the hypotheses for evaluating whether these data provide convincing evidence of the validity of the saying, i.e. that people do better than random guessing when it comes to recognizing the back of their own hand?

$$H_0: p = 0.1$$
$$H_A: p > 0.1$$

Example

Fill in the blanks below:

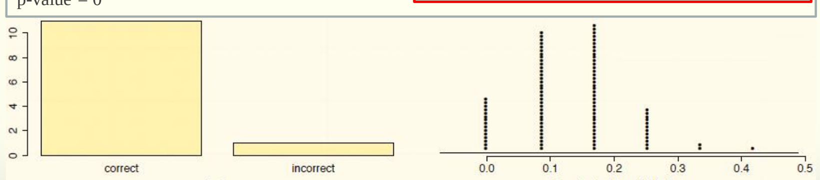
1. Use a _10_-sided fair die to represent the sampling space, and call 1 a success (guessing correctly), and all other outcomes failures (guessing incorrectly).
2. Roll the die _12_ times (each representing one of _12_ people in the experiment), count the number rolls that resulted in __ones__, and calculate the proportion of correct guesses in one simulation of _12_ rolls.
3. Repeat step (2) 100 times, each time recording the proportion of simulated 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 is ____11/12 or greater____ (the ___observed___ proportion).

Example

R

```
> source("inference.R")
> back = factor(c(rep(“correct", 11), rep(“incorrect", 1)))
> inference(back, est = "proportion", type = "ht", method =
"simulation", success = “correct", null = 0.1, alternative =
"greater", 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
```

p-value = $P(\hat{p} \geq 0.9167 | p = 0.1) = 0$



Comparing two small sample proportions

"to know something like the back of your hand"

	 back	 palm	total
correct	11	7	18
incorrect	1	5	6
total	12	12	24
$\hat{p}$	0.9167	0.5833	0.75

Example

- Do these data provide convincing evidence that there is a difference in how good people are at recognizing the backs and the palms of their hands?

$$H_0: p_{back} - p_{palm} = 0 \qquad H_A: p_{back} - p_{palm} \neq 0$$

1. Independence:
   - ✓ within groups: within each group we can assume that the guess of one subject is independent of another.
   - ✓ between groups: no, same people guessing – assume to be met for illustrative purposes
2. Sample size / skew: $12 \times 0.75 = 9$ and $12 \times 0.25 = 3$ - not met, use simulation methods

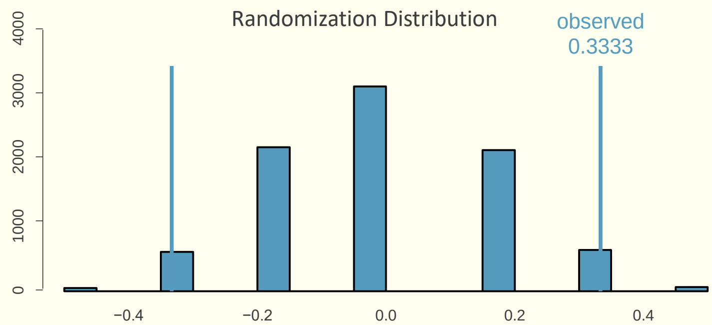
Simulation Scheme

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.

Interpreting the simulation results

- Simulate the experiment under the assumption of independence, i.e. leaving things up to chance.
- Results from the simulations look like the data → the difference between the proportions of correct guesses in the two groups was **due to chance**.
- Results from the simulations do not look like the data → 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**.

Computing p-value



p-value = 0.1566