Central Limit Theorem and Hypothesis Testing

M. Drew LaMar
September 05, 2016

> "Statistics are used much like a drunk uses a lamppost: for support, not illumination."
>
> - Andrew Lang

Introduction to Quantitative Biology, Fall 2016

Class annou

Answer:

$$P = 2 \times (0.02) = 0.04$$

- Blog #1
  - Blog Prompt: Write about an area of biology that you are particularly interested in learning about through a quantitative lens.
  - Original post due date: Wednesday, September 7, 11:59 pm
  - Comment due date: Friday, September 9, 11:59 pm
- Reading assignment for Wednesday (posted on BB): **OpenIntro Stats, Chapter 7: Introduction to Linear Regression**

Central Limit Theorem (informal)

If a sample consists of at least 30 independent observations and the data are not strongly skewed, then the sampling distribution for the mean is well approximated by a normal model.

http://www.zoology.ubc.ca/~whitlock/kingfisher/CLT.htm

Hypothesis testing (general)

Definition: ***Hypothesis testing*** compares data to what we would expect to see if a specific null hypothesis were true. If the data are too unusual, compared to what we would expect to see if the null hypothesis were true, then the null hypothesis is rejected.

Definition: A ***null hypothesis*** is a specific statement about a population parameter made for the purpose of argument.

Definition: The ***alternative hypothesis*** includes all other feasible values for the population parameter besides the value stated in the null hypothesis.

Hypothesis testing

Can parents distinguish their own children by smell alone? To investigate, Porter and Moore (1981) gave new T-shirts to children of nine mothers. Each child wore his or her shirt to bed for three consecutive nights. During the day, from waking until bedtime, the shirts were kept in individually sealed plastic bags. No scented soaps or perfumes were used during the study. Each mother was then given the shirt of her child and that of another, randomly chosen child and asked to identify her own by smell.

Discuss: What is the ***null hypothesis***? ***alternative hypothesis***?

Hypothesis testing

Can parents distinguish their own children by smell alone? To investigate, Porter and Moore (1981) gave new T-shirts to children of nine mothers. Each child wore his or her shirt to bed for three consecutive nights. During the day, from waking until bedtime, the shirts were kept in individually sealed plastic bags. No scented soaps or perfumes were used during the study. Each mother was then given the shirt of her child and that of another, randomly chosen child and asked to identify her own by smell.

Discuss: What is the ***null hypothesis***? ***alternative hypothesis***?

Answer: With p the probability of choosing correctly,

$$H_0 : \ p = 0.5$$
$$H_A : \ p \neq 0.5$$

Hypothesis testing (how it's done)

Definition: The ***test statistic*** is a number calculated from the data that is used to evaluate how compatible the data are with the result expected under the null hypothesis.

Definition: The ***null distribution*** is the sampling distribution of outcomes for a test statistic under the assumption that the null hypothesis is true.

Definition: A ***P-value*** is the probability of obtaining the data (or data showing as great or greater difference from the null hypothesis) if the null hypothesis were true.

Hypothesis testing (how it's done)

- State the hypotheses.
- Compute the test statistic.
- Determine the P-value.
- Draw the appropriate conclusions.

Hypothesis testing

Can parents distinguish their own children by smell alone? To investigate, Porter and Moore (1981) gave new T-shirts to children of nine mothers. Each child wore his or her shirt to bed for three consecutive nights. During the day, from waking until bedtime, the shirts were kept in individually sealed plastic bags. No scented soaps or perfumes were used during the study. Each mother was then given the shirt of her child and that of another, randomly chosen child and asked to identify her own by smell. **Eight of nine mothers identified their children correctly.**

Discuss: What ***test statistic*** should you use?

Answer: The number of mothers with correct identifications.

Hypothesis testing

So, P = 0.04. Is that good?

https://youtu.be/7jSE3JANx14?t=4m29s

Hypothesis testing

So, P = 0.04. Is that good?

Definition: The ***significance level***, α, is the probability used as a criterion for rejecting the null hypothesis. If the P-value is less than or equal to α, then the null hypothesis is rejected. If the P-value is greater than α, then the null hypothesis is ***not*** rejected

Definition: A result is considered ***statistically significant*** when P-value $< \alpha$.

Definition: A result is considered ***not statistically significant*** when P-value $\geq \alpha$.

Hypothesis testing

Can parents distinguish their own children by smell alone? To investigate, Porter and Moore (1981) gave new T-shirts to children of nine mothers. Each child wore his or her shirt to bed for three consecutive nights. During the day, from waking until bedtime, the shirts were kept in individually sealed plastic bags. No scented soaps or perfumes were used during the study. Each mother was then given the shirt of her child and that of another, randomly chosen child and asked to identify her own by smell. **Eight of nine mothers identified their children correctly.**

Discuss: Given $\alpha = 0.05$, $\{H_0 : \ p = 0.5\}$, and P-value of 0.04, what is the appropriate conclusion?

Answer: Reject H_0. There is evidence that mothers consistently identify own children correctly by smell.

LOTS of confusion about P-values

"We want to know if results are right, but a p-value doesn't measure that. It can't tell you the magnitude of an effect, the strength of the evidence or the probability that the finding was the result of chance."

Christie Aschwanden

http://fivethirtyeight.com/pvalue

"Belief that "statistical significance" can alone discriminate between truth and falsehood borders on magical thinking."

Cohen

LOTS of confusion about P-values

Table 1 Twelve *P*-Value Misconceptions

1	*If P = .05, the null hypothesis has only a 5% chance of being true.*
2	*A nonsignificant difference (eg, P ≥.05) means there is no difference between groups.*
3	*A statistically significant finding is clinically important.*
4	*Studies with P values on opposite sides of .05 are conflicting.*
5	*Studies with the same P value provide the same evidence against the null hypothesis.*
6	*P = .05 means that we have observed data that would occur only 5% of the time under the null hypothesis.*
7	*P = .05 and P ≤.05 mean the same thing.*
8	*P values are properly written as inequalities (eg, "P ≤.02" when P = .015)*
9	*P = .05 means that if you reject the null hypothesis, the probability of a type I error is only 5%.*
10	*With a P = .05 threshold for significance, the chance of a type I error will be 5%.*
11	*You should use a one-sided P value when you don't care about a result in one direction, or a difference in that direction is impossible.*
12	*A scientific conclusion or treatment policy should be based on whether or not the P value is significant.*

A Dirty Dozen: Twelve P-Value Misconceptions, by Steven Goodman [http://dx.doi.org/10.1053/j.seminhematol.2008.04.003]

Recommended practice

Measure and report precision and effect size separately (the P-value is a summary measure that mixes them):

- Present the magnitude of effect through the use of measures such as rates, risk differences, and odds ratios.
- Report precision with standard errors or confidence intervals.

Caveats

- Statistical significance is NOT the same as biological importance.
- Effect sizes are important. Large sample sizes can lead to statistically significant results, even though the effect size is small!