P-Values

SML201: Introduction to Data Science, Spring 2019

Michael Guerzhoy

P-Value

- Assuming the Null Hypothesis is true, the probability of observing a value that is as extreme or more extreme than what we observe
- Informally: if nothing is actually going on, how weird would it be to observe the data we do?

Coin flip example

- A fair coin has a probability of 0.5 of coming up heads
- We observe that the coin came up heads 60 times out of 100
- Assuming the coin is actually fair, the probability of observing a value that's as extreme or more extreme than 60 is

$$P(x \geq 60 \ or \ x \leq 40), x \sim Binomial(100, 0.5)$$

$x \sim Binomial(100, 0.5)$ means that x is distributed according to a Binomial distribution, with a probability of 1 equal to 0.5

Coin flip example

$$P(x \geq 60\ or\ x \leq 40) = P(x \geq 60) + P(x \leq 40)$$

- Compute this using

```
pbinom(q = 40, size = 100, prob = 0.5) +
(1 - pbinom(q = 59, size = 100, prob = 0.5)  = 0.057
```

- We'd see data that's as weird or weirder than what we observe more than 1 time out of 20
  - Some, but not overwhelming evidence that the coin is not fair

Coin flip – fake data approach

- Idea: generate fake datasets where the null hypothesis is true, and see how often we observe data that's as weird or weirder than what we actually observe
- Generate 10,000 datasets where we flipped a coin 100 times

```
fake.data <- rbinom(10000, size = 100, prob = 0.5)
> fake.data
55 51 42 44 44 54 49 ....

> mean((fake.data <= 40) | (fake.data >= 60))

0.05833
```

Darwin’s Finches example

- Beak size information for 89 finches in each of 1977 and 1979
- Want to know whether the difference between the means could be observed often if there were no difference between the populations in 1977 and 1979
- Generate fake data where there is no difference in mean beak depth, measure the difference every time
    - How often would be observe a difference greater or equal to what we observe?

Notation

- $x \sim N(\mu, \sigma^2)$ means x is normally distributed with mean μ and standard deviation σ
  - Samples from x will usually be within at most 3σ of the mean μ
- The *sample mean* is the mean of the individual measurements in the sample
  - Suppose out samples from $N(5, 2^2)$ are 5.1, 4.95, and 4.9. The sample mean is $\bar{x} = \frac{5.1+4.95+4.9}{3}$
  - The sample mean is often denoted $\bar{x}$, and the individual samples $x_1, x_2, \ldots, x_n$

Distribution of the sample mean

- Suppose $X \sim N(\mu, \sigma^2)$, and we take n (independent) measurements $x_1, \ldots, x_n$. Now $x_1 \sim N(\mu, \sigma^2), x_2 \sim N(\mu, \sigma^2), \ldots$
- We can also obtain the distribution of the sample mean. In fact,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- If we take n samples, $\bar{x}$ will almost always be within $3\sigma/\sqrt{n}$ of the mean μ
- The distribution of $\bar{X}$ is centered around μ
  - Makes sense!

Is the mean 0?

- Suppose we sample $n = 100$ measurements from $N(\mu, 2^2)$. The sample mean is 0.2. Is there evidence against the null hypothesis that $\mu = 0$?
- Assuming the null hypothesis is true, $\bar{X} \sim N\left(0, \frac{2^2}{100}\right)$

```
pnorm(q = -0.2, mean = 0, sd = 2/10) + (1 - pnorm(q = 0.2, mean = 0, sd =
2/10)) = 0.32
```

- No strong evidence against the null hypothesis

Is the mean 0?

- Suppose we sample $n = 100$ measurements from $N(\mu, \sigma^2)$. The sample mean is 0.2. Is there evidence the null hypothesis that $\mu = 0$?
- This time, we don't know the σ^2
  - But we can estimate it using $s^2 = \frac{(x_1-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n}$
- Fact: $\bar{x}/(\frac{s}{\sqrt{n}})$ is distributed according to $t(n-1)$, the Student t-distribution with n-1 degrees of freedom
  - Looks very much like a normal distribution $N(0, 1)$unless n is relatively small (< 30)
  - Assuming $\mu = 0$!!

Is the mean 0?

- P-value:

```
pt(q = -0.2/(sd(sample)/10), df = 99) +
(1 - pt(q = 0.2/(sd(sample)/10), df = 99))
```