# Introduction to Statistics
# 18.05 Spring 2014

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| T | T | T | H | H | T | H | H | H | T |
| H | T | H | T | H | T | H | T | H | T |
| H | T | T | T | H | T | T | T | T | H |
| H | T | T | H | H | T | H | H | T | H |
| T | T | H | H | H | H | T | H | T | H |
| T | T | T | H | T | H | H | H | H | T |
| T | T | T | H | H | H | T | T | T | H |
| H | H | H | H | H | H | H | T | T | T |
| H | T | H | H | T | T | T | H | H | T |
| H | T | H | H | H | T | T | T | H | H |

# Three ‘phases’

- Data Collection:
  Informal Investigation / Observational Study / Formal Experiment
- Descriptive statistics
- Inferential statistics (the focus in 18.05)

*To consult a statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of.*

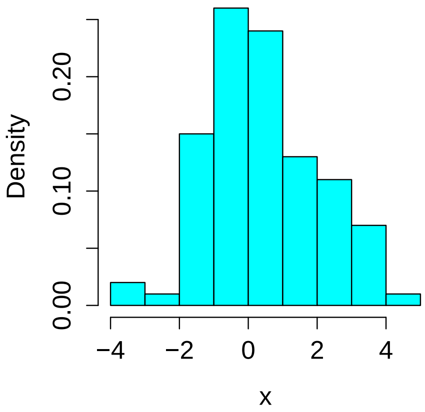
R.A. Fisher

# Is it fair?

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| T | T | T | H | H | T | H | H | H | T |
| H | T | H | T | H | T | H | T | H | T |
| H | T | T | T | H | T | T | T | T | H |
| H | T | T | H | H | T | H | H | T | H |
| T | T | H | H | H | H | T | H | T | H |
| T | T | T | H | T | H | H | H | H | T |
| T | T | T | H | H | H | T | T | T | H |
| H | H | H | H | H | H | H | T | T | T |
| H | T | H | H | T | T | T | H | H | T |
| H | T | H | H | H | T | T | T | H | H |

# Is it normal?

Does it have $\mu = 0$? Is it normal? Is it standard normal?




Sample mean $= 0.38$; sample standard deviation $= 1.59$

# What is a statistic?

**Definition**. A statistic is anything that can be computed from the collected data. That is, a statistic must be observable.

- Point statistic: a single value computed from data, e.g sample average $\overline{x}_n$ or sample standard deviation $s_n$.
- Interval or range statistics: an interval $[a, b]$ computed from the data. (Just a pair of point statistics.) Often written as $\overline{x} \pm s$.
- **Important:** A statistic is itself a random variable since a new experiment will produce new data to compute it.

# Concept question

You believe that the lifetimes of a certain type of lightbulb follow an exponential distribution with parameter $\lambda$. To test this hypothesis you measure the lifetime of 5 bulbs and get data $x_1, \ldots x_5$.

Which of the following are statistics?

**(a)** The sample average $\overline{x} = \frac{x_1+x_2+x_3+x_4+x_5}{5}$.

**(b)** The expected value of a sample, namely $1/\lambda$.

**(c)** The difference between $\overline{x}$ and $1/\lambda$.

1. (a)
2. (b)
3. (c)
4. (a) and (b)
5. (a) and (c)
6. (b) and (c)
7. all three
8. none of them

**answer:** 1. **(a)**. $\lambda$ is a parameter of the distribution it cannot be computed from the data. It can only be estimated.

# Notation

Big letters $X$, $Y$, $X_i$ are random variables.

Little letters $x$, $y$, $x_i$ are data (values) generated by the random variables.

**Example.** Experiment: 10 flips of a coin:

$X_i$ is the random variable for the $i^{\text{th}}$ flip: either 0 or 1.

$x_i$ is the actual result (data) from the $i^{\text{th}}$ flip.
e.g. $x_1, \ldots, x_{10} = 1, 1, 1, 0, 0, 0, 0, 0, 1, 0.$

# Reminder of Bayes' theorem

Bayes's theorem is the key to our view of statistics.
(Much more next week!)

$$P(\mathcal{H}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{H})P(\mathcal{H})}{P(\mathcal{D})}.$$

$$P(\text{hypothesis}|\text{data}) = \frac{P(\text{data}|\text{hypothesis})P(\text{hypothesis})}{P(\text{data})}$$

# Estimating a parameter

**Example.** Suppose we want to know the percentage $p$ of people for whom cilantro tastes like soap.

Experiment: Ask $n$ random people to taste cilantro.

Model:
$X_i \sim \text{Bernoulli}(p)$ is whether the $i^{\text{th}}$ person says it tastes like soap.

Data: $x_1, \ldots, x_n$ are the results of the experiment

*Inference*: Estimate $p$ from the data.

# Parameters of interest

**Example.** You ask 100 people to taste cilantro and 55 say it tastes like soap. Use this data to estimate $p$ the fraction of all people for whom it tastes like soap.

So, $p$ is the parameter of interest.

# Likelihood

For a given value of $p$ the probability of getting 55 'successes' is the binomial probability

$$P(55 \text{ soap}|p) = \binom{100}{55} p^{55}(1-p)^{45}.$$

**Definition:**

$$\text{The likelihood } P(\text{data}|p) = \binom{100}{55} p^{55}(1-p)^{45}.$$

**NOTICE:** The likelihood takes the data as fixed and computes the probability of the data for a given $p$.

# Maximum likelihood estimate (MLE)

The maximum likelihood estimate (MLE) is a way to estimate the value of a parameter of interest.

The MLE is the value of $p$ that maximizes the likelihood.

Different problems call for different methods of finding the maximum.

Here are two –there are others:
**1.** Calculus: To find the MLE, solve $\frac{d}{dp}P(\text{data} \mid p) = 0$ for $p$. (We should also check that the critical point is a maximum.)

**2.** Sometimes the derivative is never 0 and the MLE is at an endpoint of the allowable range.

# Cilantro tasting MLE

The MLE for the cilantro tasting experiment is found by calculus.

$$\frac{dP(\text{data} \mid p)}{dp} = \binom{100}{55}(55p^{54}(1-p)^{45} - 45p^{55}(1-p)^{44}) = 0$$

A sequence of algebraic steps gives:

$$\begin{aligned} 55p^{54}(1-p)^{45} &= 45p^{55}(1-p)^{44} \\ 55(1-p) &= 45p \\ 55 &= 100p \end{aligned}$$

Therefore the MLE is $\hat{p} = \frac{55}{100}$.

# Log likelihood

Because the log function turns multiplication into addition it is often convenient to use the log of the likelihood function

$$\text{log likelihood} = \ln(\text{likelihood}) = \ln(P(\text{data} \mid p)).$$

**Example.**

$$\text{Likelihood} \;\; P(\text{data}|p) = \binom{100}{55} p^{55}(1-p)^{45}$$

$$\text{Log likelihood} \;\; = \ln\left(\binom{100}{55}\right) + 55\ln(p) + 45\ln(1-p).$$

(Note first term is just a constant.)

# Board Question: Coins

A coin is taken from a box containing three coins, which give heads with probability $p = 1/3$, $1/2$, and $2/3$. The mystery coin is tossed 80 times, resulting in 49 heads and 31 tails.

**(a)** What is the likelihood of this data for each type on coin? Which coin gives the maximum likelihood?

**(b)** Now suppose that we have a single coin with unknown probability $p$ of landing heads. Find the likelihood and log likelihood functions given the same data. What is the maximum likelihood estimate for $p$?

*See next slide.*

# Solution

**answer:** **(a)** The data $D$ is 49 heads in 80 tosses.
We have three hypotheses: the coin has probability $p = 1/3$, $p = 1/2$, $p = 2/3$. So the likelihood function $P(D|p)$ takes 3 values:

$$P(D|p = 1/3) = \binom{80}{49}\left(\frac{1}{3}\right)^{49}\left(\frac{2}{3}\right)^{31} = 6.24 \cdot 10^{-7}$$

$$P(D|p = 1/2) = \binom{80}{49}\left(\frac{1}{2}\right)^{49}\left(\frac{1}{2}\right)^{31} = 0.024$$

$$P(D|p = 2/3) = \binom{80}{49}\left(\frac{2}{3}\right)^{49}\left(\frac{1}{3}\right)^{31} = 0.082$$

The maximum likelihood is when $p = 2/3$ so this our maximum likelihood estimate is that $p = 2/3$.

*Answer to part (b) is on the next slide*

# Solution to part (b)

**(b)** Our hypotheses now allow $p$ to be any value between 0 and 1. So our likelihood function is

$$P(D|p) = \binom{80}{49} p^{49}(1-p)^{31}$$

To compute the maximum likelihood over all $p$, we set the derivative of the log likelihood to 0 and solve for $p$:

$$\begin{aligned}
\frac{d}{dp}\ln(P(D|p)) = \frac{d}{dp}\left(\ln\left(\binom{80}{49}\right) + 49\ln(p) + 31\ln(1-p)\right) &= 0 \\
\Rightarrow \ \frac{49}{p} - \frac{31}{1-p} &= 0 \\
\Rightarrow \ p &= \frac{49}{80}
\end{aligned}$$

So our MLE is $\hat{p} = 49/80$.

# Continuous likelihood

Use the pdf instead of the pmf

**Example. Light bulbs**

Lifetime of each bulb $\sim$ exp($\lambda$).

Test 5 bulbs and find lifetimes of $x_1, \ldots, x_5$.

**(i)** Find the likelihood and log likelihood functions.

**(ii)** Then find the maximum likelihood estimate (MLE) for $\lambda$.

**answer:** *See next slide.*

# Solution

**(i)** Let $X_i \sim \exp(\lambda) =$ the lifetime of the $i^{\text{th}}$ bulb.

Likelihood $=$ joint pdf (assuming independence):

$$f(x_1, x_2, x_3, x_4, x_5|\lambda) = \lambda^5 \mathrm{e}^{-\lambda(x_1+x_2+x_3+x_4+x_5)}.$$

Log likelihood

$$\ln(f(x_1, x_2, x_3, x_4, x_5|\lambda)) = 5\ln(\lambda) - \lambda(x_1 + x_2 + x_3 + x_4 + x_5).$$

**(ii)** Using calculus to find the MLE:

$$\frac{d\ \ln(f(x_1, x_2, x_3, x_4, x_5|\lambda))}{d\,\lambda} = \frac{5}{\lambda} - \sum x_i = 0 \ \Rightarrow \ \boxed{\hat{\lambda} = \frac{5}{\sum x_i}}.$$

# Board Question

Suppose the 5 bulbs are tested and have lifetimes of 2, 3, 1, 3, 4 years respectively. What is the maximum likelihood estimate (MLE) for $\lambda$?

*Work from scratch. Do not simply use the formula just given.*

Set the problem up carefully by defining random variables and densities.

*Solution on next slide.*

# Solution

**answer:** We need to be careful with our notation. With five different values it is best to use subscripts. So, let $X_j$ be the lifetime of the $i^{\text{th}}$ bulb and let $x_i$ be the value it takes. Then $X_i$ has density $\lambda e^{-\lambda x_i}$. We assume each of the lifetimes is independent, so we get a joint density

$$f(x_1, x_2, x_3, x_4, x_5 | \lambda) = \lambda^5 e^{-\lambda(x_1+x_2+x_3+x_4+x_5)}.$$

Note, we write this as a conditional density, since it depends on $\lambda$. This density is our likelihood function. Our data had values

$$x_1 = 2,\ x_2 = 3,\ x_3 = 1,\ x_4 = 3,\ x_5 = 4.$$

So our likelihood and log likelihood functions with this data are

$$f(2,3,1,3,4 \,|\, \lambda) = \lambda^5 e^{-13\lambda}, \qquad \ln(f(2,3,1,3,4 \,|\, \lambda)) = 5\ln(\lambda) - 13\lambda$$

*Continued on next slide*

# Solution continued

Using calculus to find the MLE we take the derivative of the log likelihood

$$\frac{5}{\lambda} - 13 = 0 \ \Rightarrow \ \boxed{\hat{\lambda} = \frac{5}{13}}.$$

MIT OpenCourseWare
https://ocw.mit.edu

18.05 Introduction to Probability and Statistics
Spring 2014

For information about citing these materials or our Terms of Use, visit: https://ocw.mit.edu/terms.