

Box-and-Whisker Plot

- A box plot is a graphical display showing **S**pread, **O**utliers, **C**enter, and **S**hape (**SOCS**).
- It displays the **5-number summary**: *min*, q_1 , *median*, q_3 , and *max*.

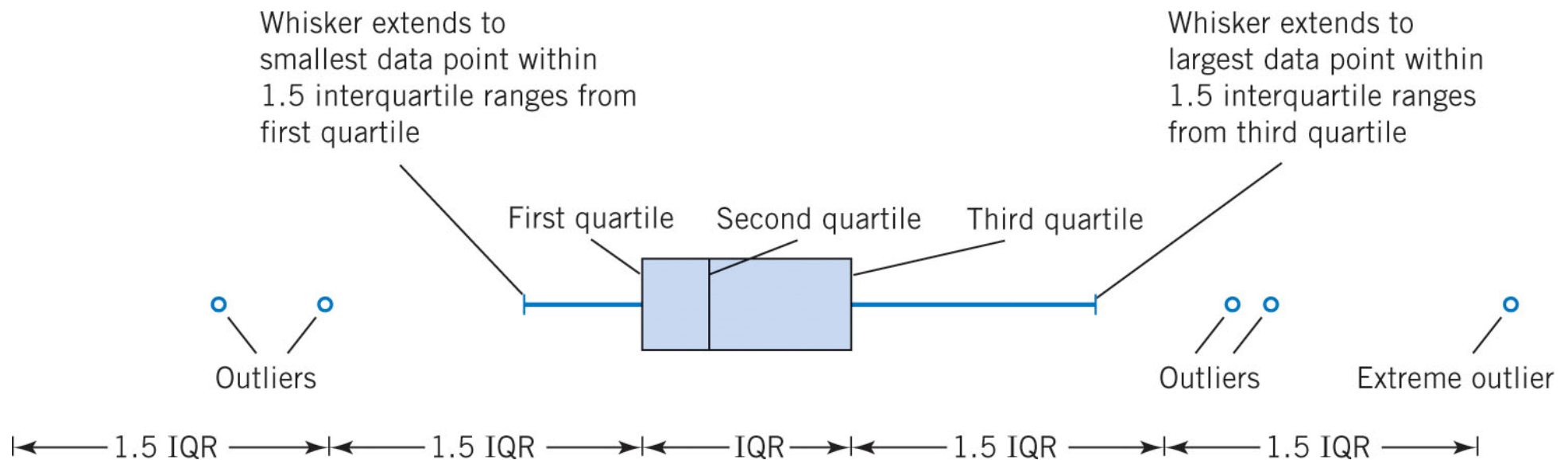
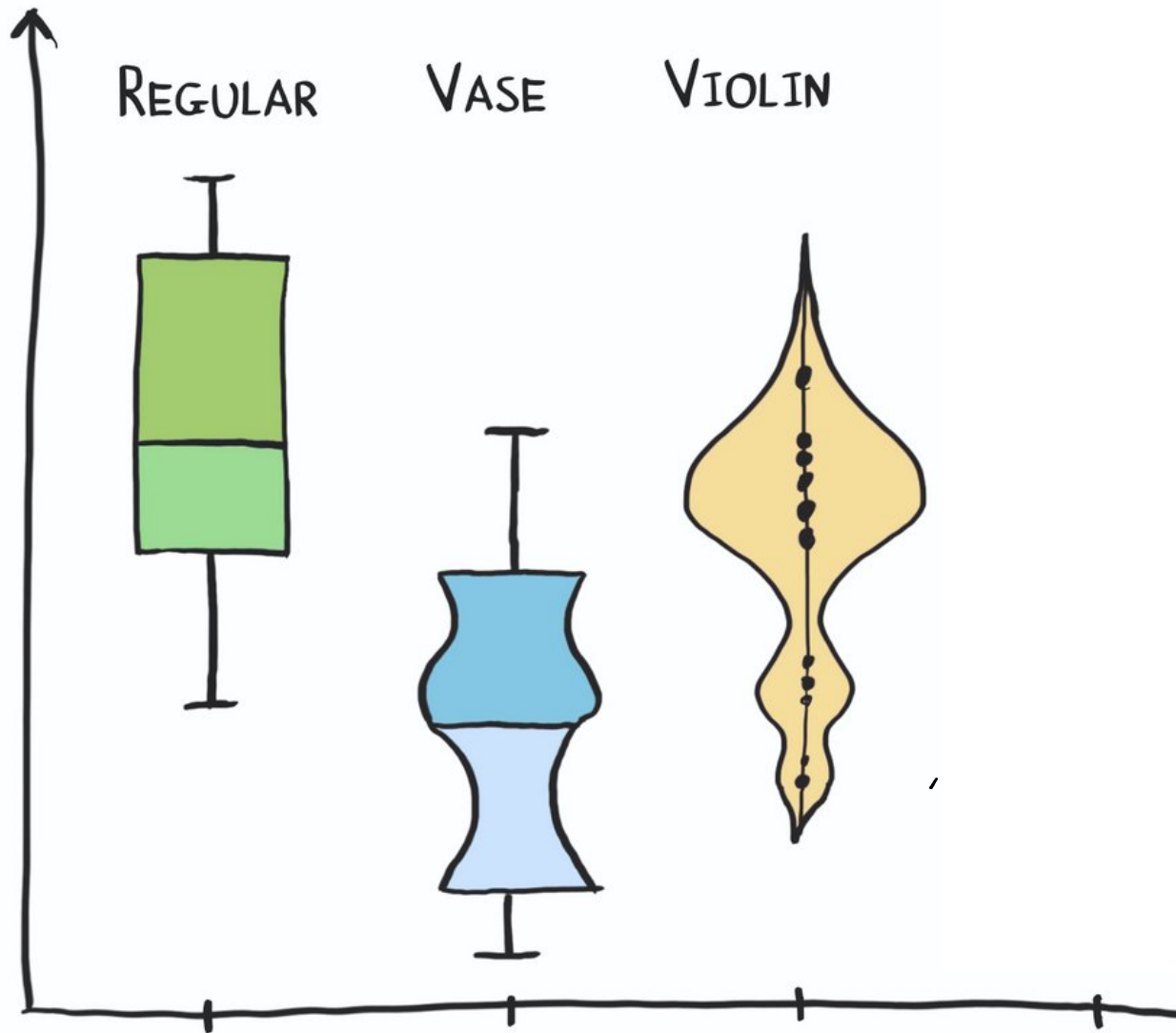


Figure 6-13 Description of a box plot.

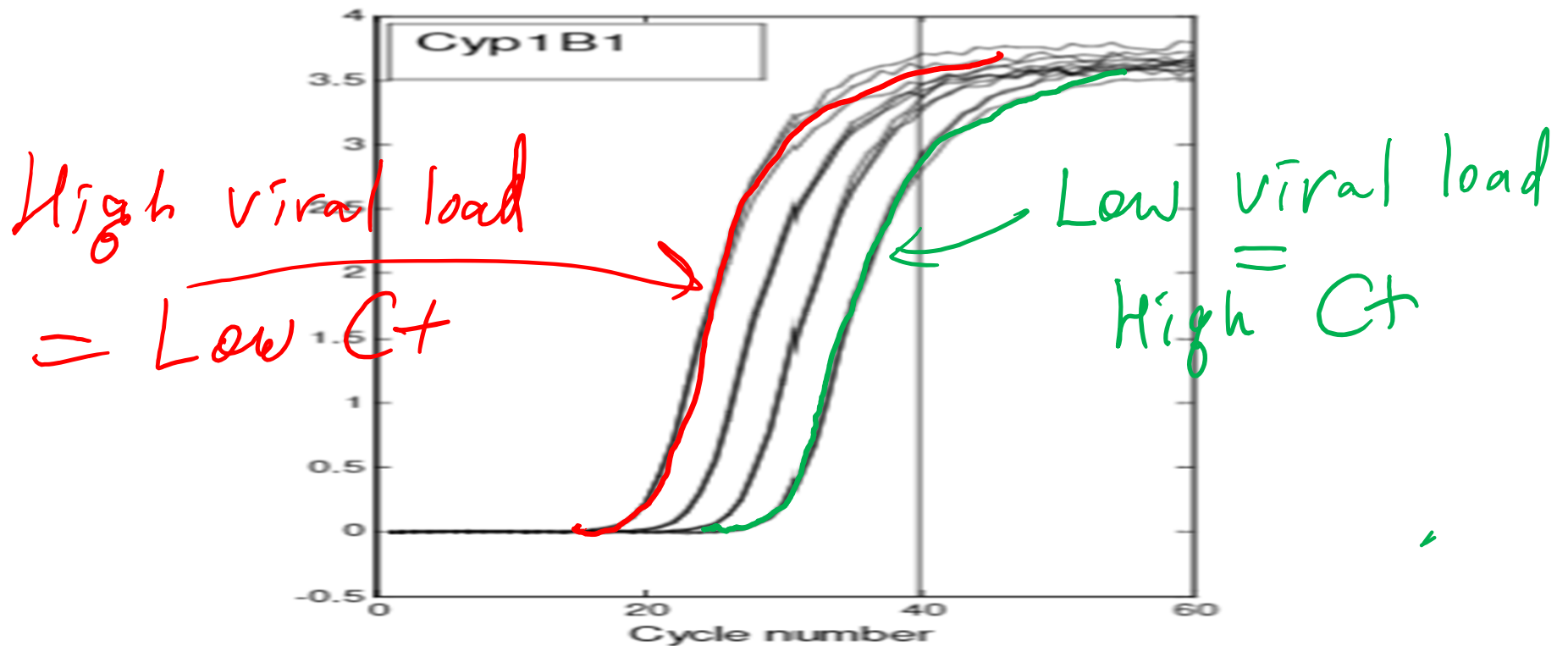
TYPES OF BOX AND WHISKER PLOT



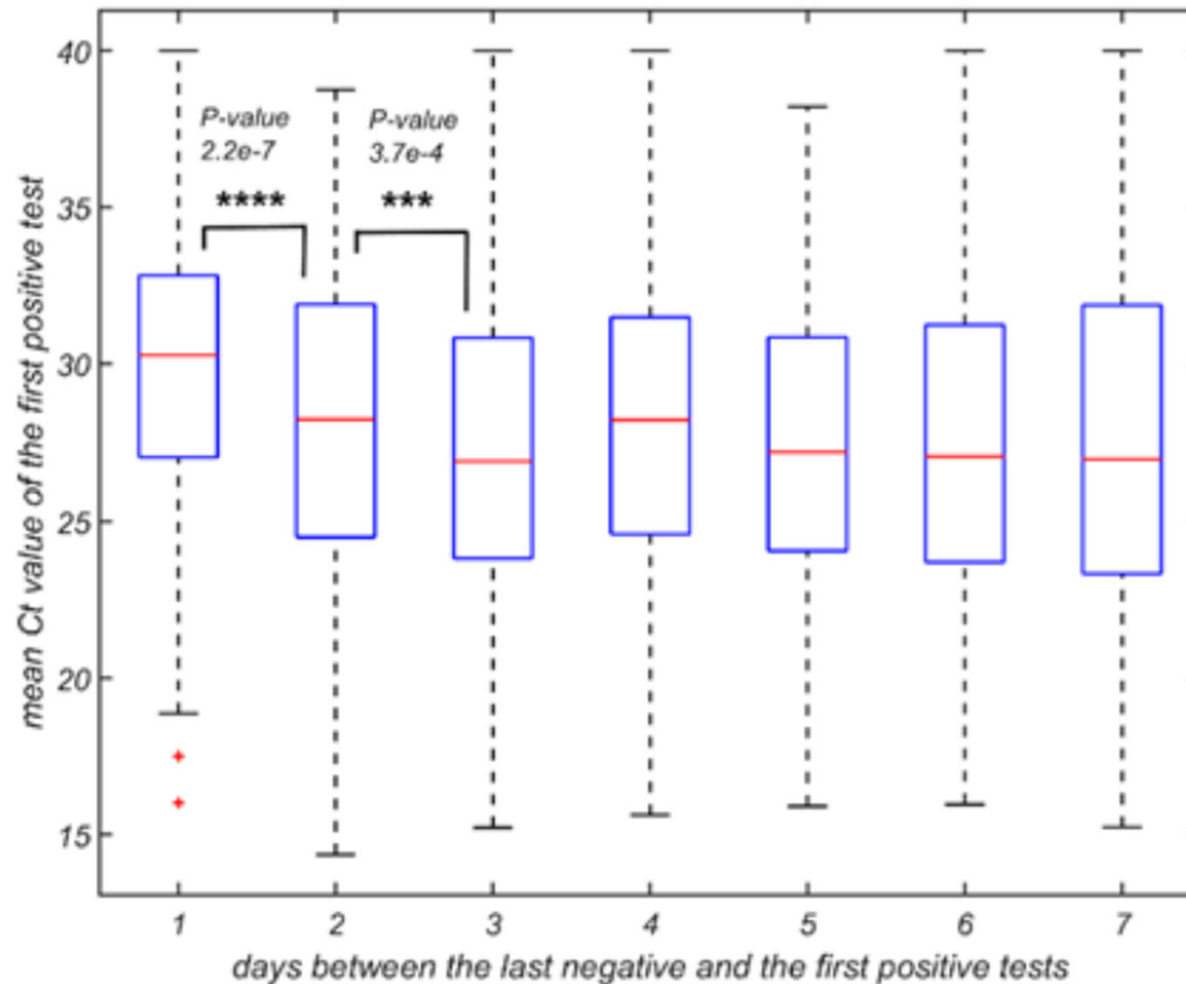
Reminder

What is the Cycle threshold (Ct)
value of a PCR test?

$$Ct = \text{const} - \log_2(\text{viral DNA concentration})$$



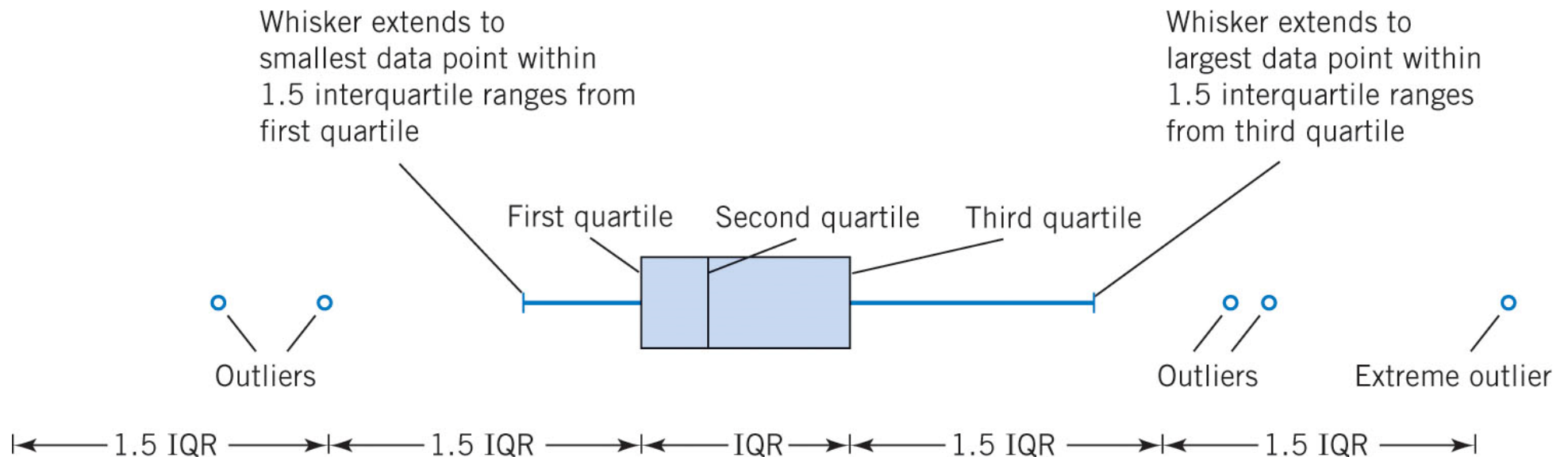
Bar plot based on COVID-19 tests at UIUC



Ranoa, D. R. E. et al. Mitigation of SARS-CoV-2 transmission at a large public university. Nat Commun 13, 3207 (2022)

Matlab exercise #2:

- Generate a sample with $n= 1000$ following **standard normal distribution**
- Calculate **median, first, and third quartiles**
- Calculate **IQR** and find ranges shown below
- Find and count **left and right outliers**
- **Do not use built-in Matlab functions for this!**
- Make box and whisker plot: use **boxplot**



How many right outliers one expects in a sample of $n=1000$ following normal distribution?

- **% find the third quartile of a standard distribution**
- **norminv(0.75) %ans = 0.6745**
- **% Calculate IQR - Inter Quartale Range**
- **IQR=2.*norminv(0.75) % 1.3490**
- **% Calculate $0.5*IQR+1.5*IQR$ - the right whisker position**
- **whisker=0.5.*IQR+1.5*IQR %ans = 2.6980**
- **% Find the probability to be above the right whisker**
- **1-normcdf(whisker) %ans = 0.00349**
- **% Find number of right outliers in a sample of 1000 points**
- **1000.*(1-normcdf(whisker)) %ans = 3.49**

Credit: XKCD
comics

WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS

WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT

QUESTIONS

FOUND IN GOOGLE AUTOCOMPLETE



WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY

WHY AREN'T ECONOMISTS RICH

WHY DO AMERICANS CALL IT SOCCER

WHY ARE MY EARS RINGING

WHY ARE THERE SO MANY AVENGERS

WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS

WHY ARE THERE ANTS IN MY LAPTOP

WHY IS EARTH TILTED

WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN



WHY IS THERE AN OWL IN MY BACKYARD

WHY IS THERE AN OWL OUTSIDE MY WINDOW

WHY IS THERE AN OWL ON THE DOLLAR BILL

WHY DO OWLS ATTACK PEOPLE

WHY ARE AK 47s SO EXPENSIVE

WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE

WHY ARE THERE GODS

WHY ARE THERE TWO SPOCKS

WHY IS MT VESUVIUS THERE

WHY DO THEY SAY T MINUS

WHY ARE THERE OBELISKS

WHY ARE WRESTLERS ALWAYS WET

WHY ARE OCEANS BECOMING MORE ACIDIC

WHY IS ARWEN DYING

WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING

WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY IS LIFE SO BORING

WHY ARE CIGARETTES LEGAL

WHY ARE THERE DUCKS IN MY POOL

WHY IS JESUS WHITE

WHY IS THERE LIQUID IN MY EAR

WHY DO Q TIPS FEEL GOOD

WHY DO GOOD PEOPLE DIE



WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG

WHY DO WHALES JUMP

WHY ARE WITCHES GREEN

WHY ARE THERE MIRRORS ABOVE BEDS

WHY DO I SAY UH

WHY IS SEA SALT BETTER

WHY ARE THERE TREES IN THE MIDDLE OF FIELDS

WHY IS THERE NOT A POKEMON MMO

WHY IS THERE LAUGHING IN TV SHOWS

WHY ARE THERE DOORS ON THE FREEWAY

WHY ARE THERE SO MANY SVCHOST.EXE RUNNING

WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA

WHY ARE THERE SCARY SOUNDS IN MINECRAFT

WHY IS THERE KICKING IN MY STOMACH

WHY ARE THERE TWO SLASHES AFTER HTTP

WHY ARE THERE CELEBRITIES

WHY DO SNAKES EXIST

WHY DO OYSTERS HAVE PEARLS

WHY ARE DUCKS CALLED DUCKS

WHY DO THEY CALL IT THE CLAP

WHY ARE KYLE AND CARTMAN FRIENDS

WHY IS THERE AN ARROW ON AANG'S HEAD

WHY ARE TEXT MESSAGES BLUE

WHY ARE THERE MUSTACHES ON CLOTHES

WHY ARE THERE MUSTACHES ON CARS

WHY ARE THERE MUSTACHES EVERYWHERE

WHY ARE THERE SO MANY BIRDS IN OHIO

WHY IS THERE SO MUCH RAIN IN OHIO

WHY IS OHIO WEATHER SO WEIRD

WHY ARE THERE BRIDESMAIDS

WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD KUNGONS DIFFERENT



WHY IS PROGRAMMING SO HARD

WHY IS THERE A 0 OHM RESISTOR

WHY DO AMERICANS HATE SOCCER

WHY DO RHYMES SOUND GOOD

WHY DO TREES DIE

WHY ARE THERE TINY SPIDERS IN MY HOUSE

WHY DO SPIDERS COME INSIDE

WHY ARE THERE HUGE SPIDERS IN MY HOUSE

WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE

WHY ARE THERE SPIDERS IN MY ROOM

WHY ARE THERE SO MANY SPIDERS IN MY ROOM

WHY DO SPIDER BITES ITCH

WHY IS DYING SO SCARY

WHY IS THERE NO GPS IN LAPTOPS

WHY DO KNEES CLICK

WHY AREN'T THERE E GRADES

WHY IS SEX SO IMPORTANT



WHY IS ISOLATION BAD

WHY DO BOYS LIKE ME

WHY DON'T BOYS LIKE ME

WHY IS THERE ALWAYS A JAVA UPDATE

WHY ARE THERE RED DOTS ON MY THIGHS

WHY IS LYING GOOD

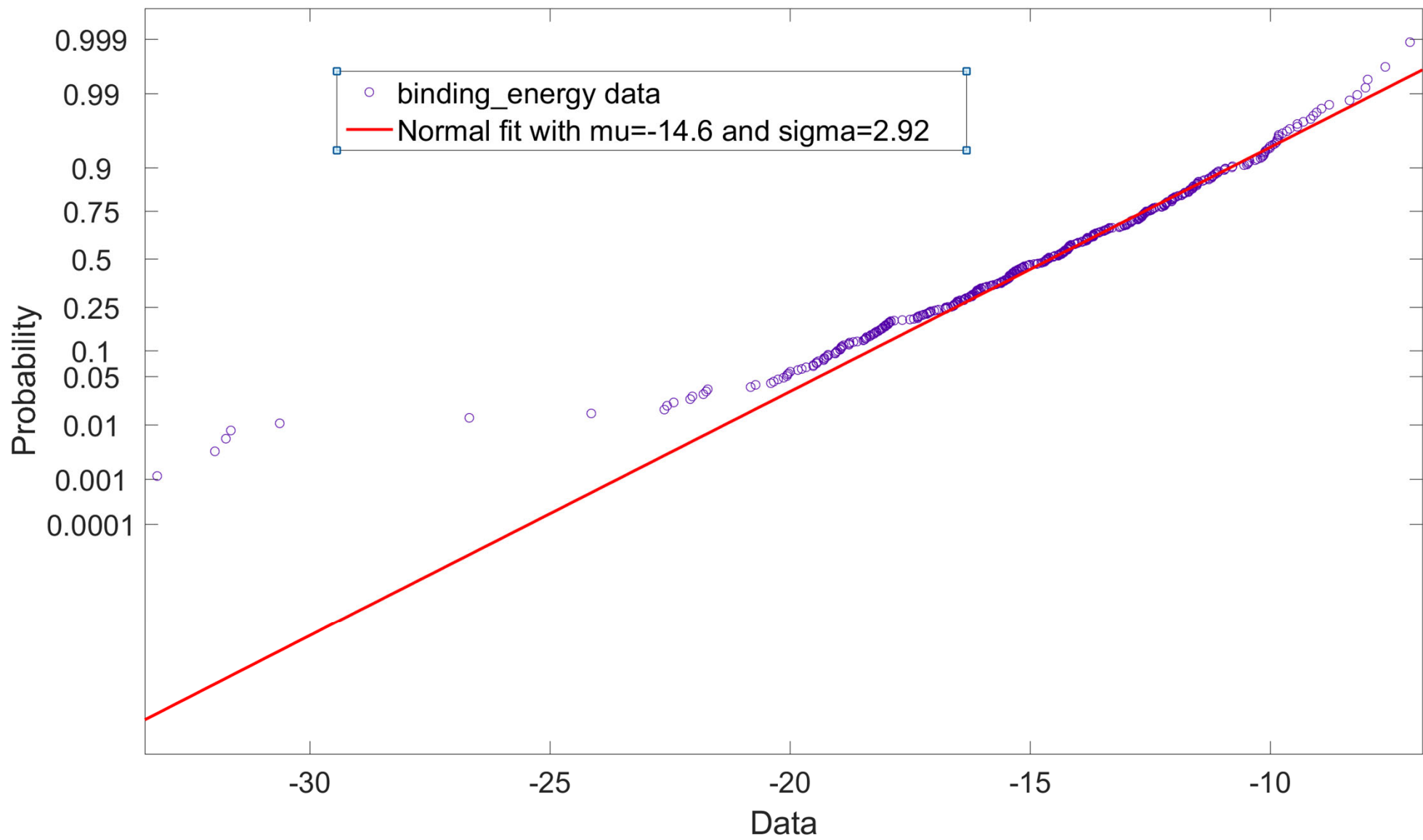
WHY AREN'T THERE DINOSAUR GHOSTS

WHY ARE THERE FEMALE MR NIMES

WHY IS GPS FREE

Probability Plots

- How do we know if a particular probability distribution is a reasonable model for a data set?
- A **histogram** of a large data set **reveals the shape of a distribution**. The histogram of a small data set **does not provide a clear picture**.
- A **probability plot** is helpful for all data set size. How good is the model based on a particular probability distribution can be verified using a **subjective visual examination**.



How To Build a Probability Plot

- Sort the data observations in ascending order:
 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$.
- Empirically determined cumulative frequency
 $\text{Prob}(x \leq x_{(j)}) = j/n$. To correct for discreteness of $x_{(j)}$ better use $\text{Prob}(x \leq x_{(j)}) = (j-0.5)/n$
- If you believe that $\text{CDF}(x)$ describes your random variable $(j-0.5)/n$ should be close to $\text{CDF}(x_{(j)})$
- Probability plot is $x_{(j)} \cdot [(j-0.5)/n] / \text{CDF}(x_{(j)})$ plotted versus the observed value $x_{(j)}$.
- If the fit is good - one gets a straight line
- Deviations can be seen especially at tails.

Probability Plot Variations

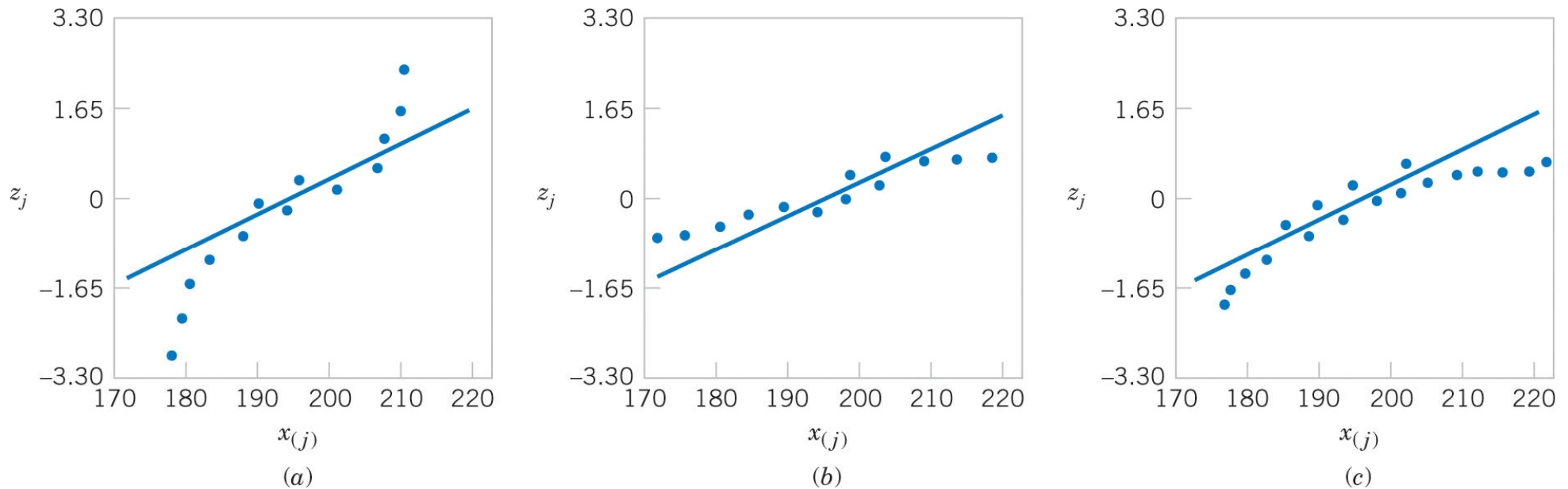


Figure 6-21 Normal probability plots indicating a non-normal distribution.

(a) Light tailed distribution (squeezed together)

(b) Heavy tailed distribution (stretched out)

(c) Right skewed distribution (left end squeezed, right end stretched)

Credit: XKCD
comics

WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS

WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT

QUESTIONS FOUND IN GOOGLE AUTOCOMLETE



WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY

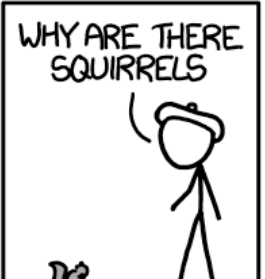
WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD

WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS

WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE

WHY ARE THERE ANTS IN MY LAPTOP

WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD KUNGONS DIFFERENT



WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN



WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS

WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND

WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL

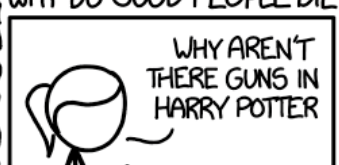
WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY

WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD



WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE



WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG

Descriptive statistics:

Sample mean and
its variance

Standard error vs
Standard deviation

Some Definitions

- The random variables X_1, X_2, \dots, X_n are a **random sample** of **size n** if:
 - a) The X_i are **independent** random variables.
 - b) Every X_i has **the same probability distribution**.

Such X_1, X_2, \dots, X_n are also called independent and identically distributed (or **i. i. d.**) random variables

- A **statistic** is any function of the observations in a random sample.
- The probability distribution of a statistic is called a **sampling distribution**.

Statistic #1: Sample Mean

If the values of n observations in a random sample are denoted by x_1, x_2, \dots, x_n , the **sample mean** is

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (6-1)$$

New random variable \bar{X} is a linear combination of n independent identically distributed variables X_1, X_2, \dots, X_n

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Mean & Variance of a Linear Function

$$Y = c_1X_1 + c_2X_2 + \dots + c_pX_p$$

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \dots + c_pE(X_p) \quad (5-25)$$

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \dots + c_p^2V(X_p) + 2 \sum_{i < j} \sum c_i c_j \text{cov}(X_i X_j) \quad (5-26)$$

If X_1, X_2, \dots, X_p are **independent**, then $\text{cov}(X_i X_j) = 0$,

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \dots + c_p^2V(X_p) \quad (5-27)$$

IMPORTANT:

Sample mean \bar{X} is drawn from a random variable

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$E(\bar{X}) = \frac{n \cdot E(X_i)}{n} = \frac{n \cdot \mu}{n} = \mu$$

$$V(\bar{X}) = \frac{n \cdot V(X_i)}{n^2} = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$$\text{Stand. dev. } (\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Central Limit Theorem

If X_1, X_2, \dots, X_n is a random sample of size n is taken from a population with mean μ and **finite variance σ^2** , and **any distribution**. If \bar{X} is the sample mean, then the **limiting form of the distribution** of

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (7-1)$$

for **large n** , is the **standard normal distribution**.

If X_1, X_2, \dots, X_n are themselves normally distributed – for any n

Test CLT for your own random variable

- Go to:
https://onlinestatbook.com/stat_sim/sampling_dist/
- Select “Custom” at the top and use mouse to sketch the PMF of your own random variable
- Select “mean” and $n=5$ in the third panel
- Choose “Animated” in the second panel and use `number_of_experiments=5` to see one sample being generated
- Repeat with `number_of_experiments =10,000`
- Now select “mean” and $n=25$ in the fourth panel
- Skewness and Kurtosis are measures of how good is the normal (Gaussian) fit (choose “fit normal”)