

How many times you met for group exercises **during regular class hours?**

- A. 0
- B. 1
- C. 2
- D. 3
- E. 4

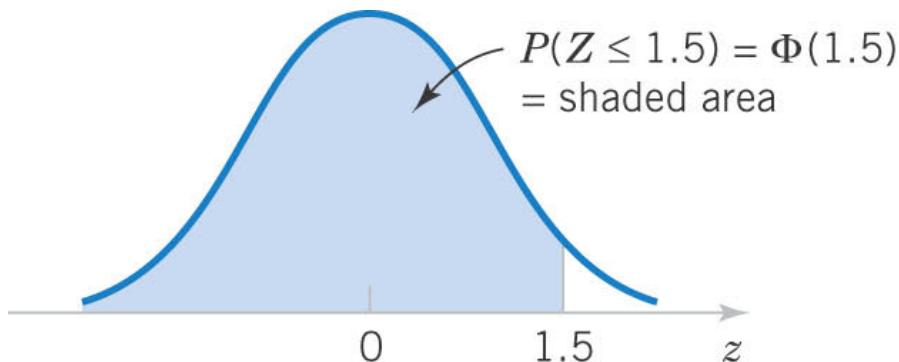
Get your i-clickers

<i>z</i>	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

# Standard Normal Distribution Tables

Assume  $Z$  is a standard normal random variable.

Find  $P(Z \leq 1.50)$ . Answer: 0.93319



$z$	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
1.5	0.93319	0.93448	0.93574	0.93699

Figure 4-13 Standard normal PDF

Find  $P(Z \leq 1.53)$ .

Answer: 0.93699

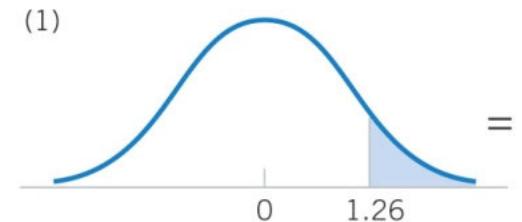
Find  $P(Z \leq 0.02)$ .

Answer: 0.50398

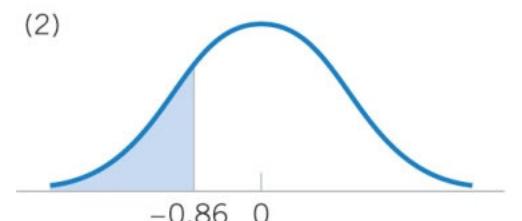
Table III from,  
Appendix A in  
Montgomery  
& Runger

# Standard Normal Exercises

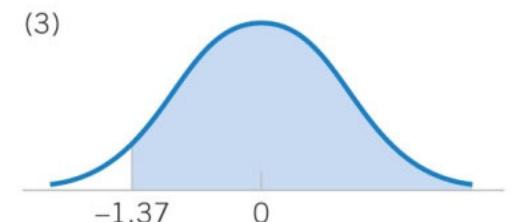
1.  $P(Z > 1.26) = 1 - P(Z < 1.26) = 1 - \textcolor{red}{0.8962} =$   
 $= 0.1038$



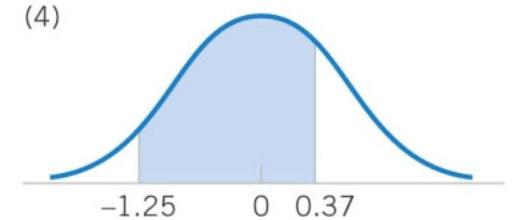
2.  $P(Z < -0.86) = P(Z > 0.86) = 1 - P(Z < 0.86) =$   
 $\textcolor{red}{1 - 0.815} = \textcolor{blue}{0.195}$



3.  $P(Z > -1.37) = P(Z < 1.37) = \textcolor{red}{0.915}$



4.  $P(-1.25 < Z < 0.37) = P(Z < 0.37) - P(Z < -1.25)$   
 $= P(Z < 0.37) - P(Z > 1.25) = P(Z < 0.37) -$   
 $(1 - P(Z < 1.25)) = \textcolor{red}{0.6443} - (1 - \textcolor{red}{0.8944}) = \textcolor{blue}{0.5387}$



# X and Y are Bernoulli variables

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

What is the marginal  $P_Y(Y=0)$ ?

- A. 1/6
- B. 2/6
- C. 3/6
- D. 4/6
- E. I don't know

Get your i-clickers

# X and Y are Bernoulli variables

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

What is the marginal  $P_Y(Y=0)$ ?

- A. 1/6
- B. 2/6
- C. 3/6
- D. 4/6
- E. I don't know

Get your i-clickers

# X and Y are Bernoulli variables

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

What is the conditional  $P(X=0 | Y=1)$ ?

- A. 2/6
- B. 1/2
- C. 1/6
- D. 4/6
- E. I don't know

Get your i-clickers

# X and Y are Bernoulli variables

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

What is the conditional  $P(X=0 | Y=1)$ ?

- A. 2/6
- B. 1/2
- C. 1/6
- D. 4/6
- E. I don't know

Get your i-clickers

# X and Y are Bernoulli variables

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

Are they independent?

- A. yes
- B. no
- C. I don't know

Get your i-clickers

# X and Y are Bernoulli variables

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

Are they independent?

- A. yes
- B. no
- C. I don't know

Get your i-clickers

# X and Y are Bernoulli variables

	Y=0	Y=1
X=0	1/2	0
X=1	0	1/2

Are they independent?

- A. yes
- B. no
- C. I don't know

Get your i-clickers

# X and Y are Bernoulli variables

	Y=0	Y=1
X=0	1/2	0
X=1	0	1/2

Are they independent?

- A. yes
- B. no
- C. I don't know

Get your i-clickers

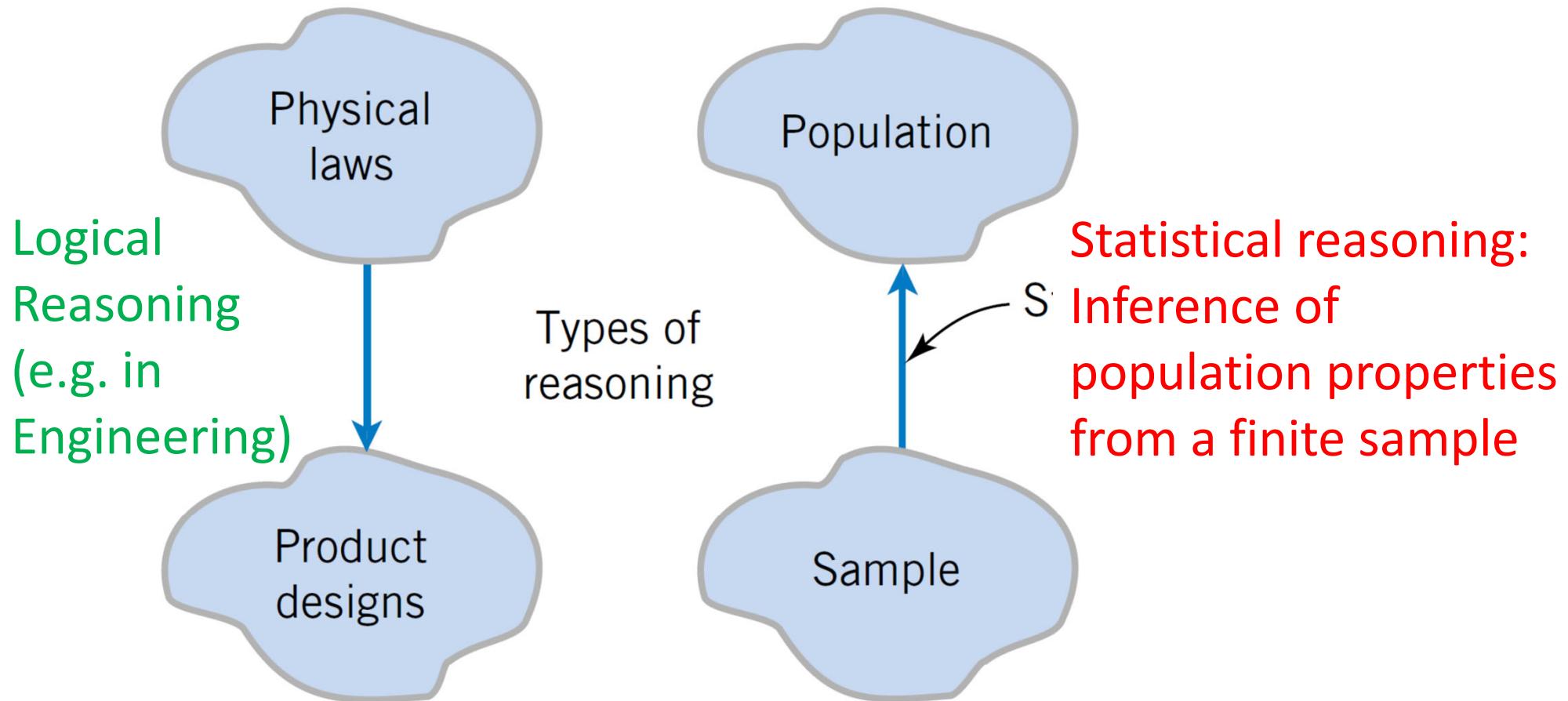
## **Descriptive statistics:**

**Populations, Samples**

**Histograms, Quartiles**

**Sample mean and  
variance**

# Two types of reasoning



# Numerical Summaries of Data

- Data are the **numerical observations** of a **phenomenon of interest**.
- The totality of all observations is a **population**.
  - **Population can be infinite** (e.g. abstract random variables)
  - **It can be very large** (e.g. 7 billion humans or all patients who have cancer of a given type)
- A (usually small) portion of the population collected for analysis is a random **sample**.
- We want to **use sample to infer facts about populations**
- The **inference** is not perfect but **gets better and better as sample size increases**.

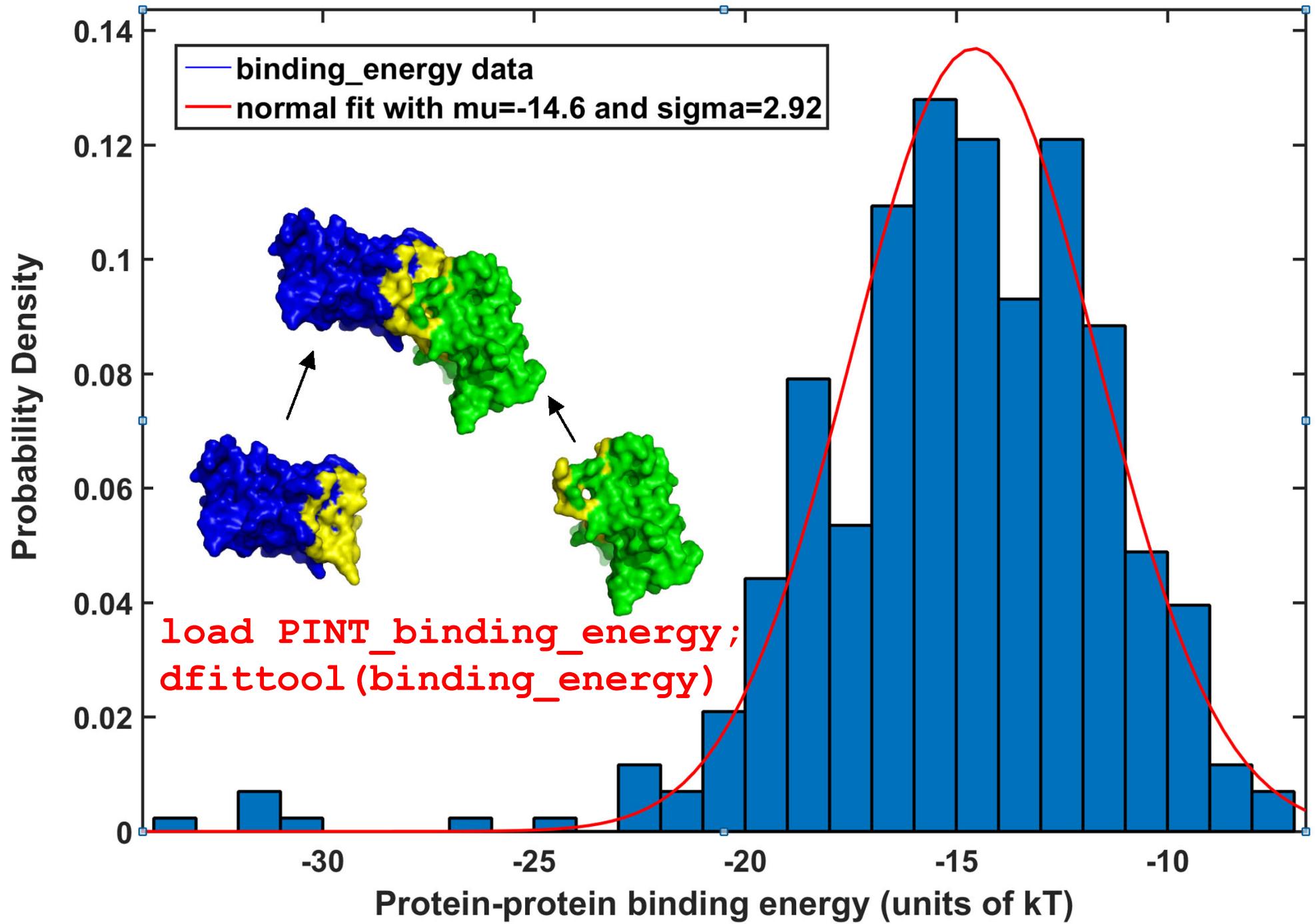
# Some Definitions

- 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

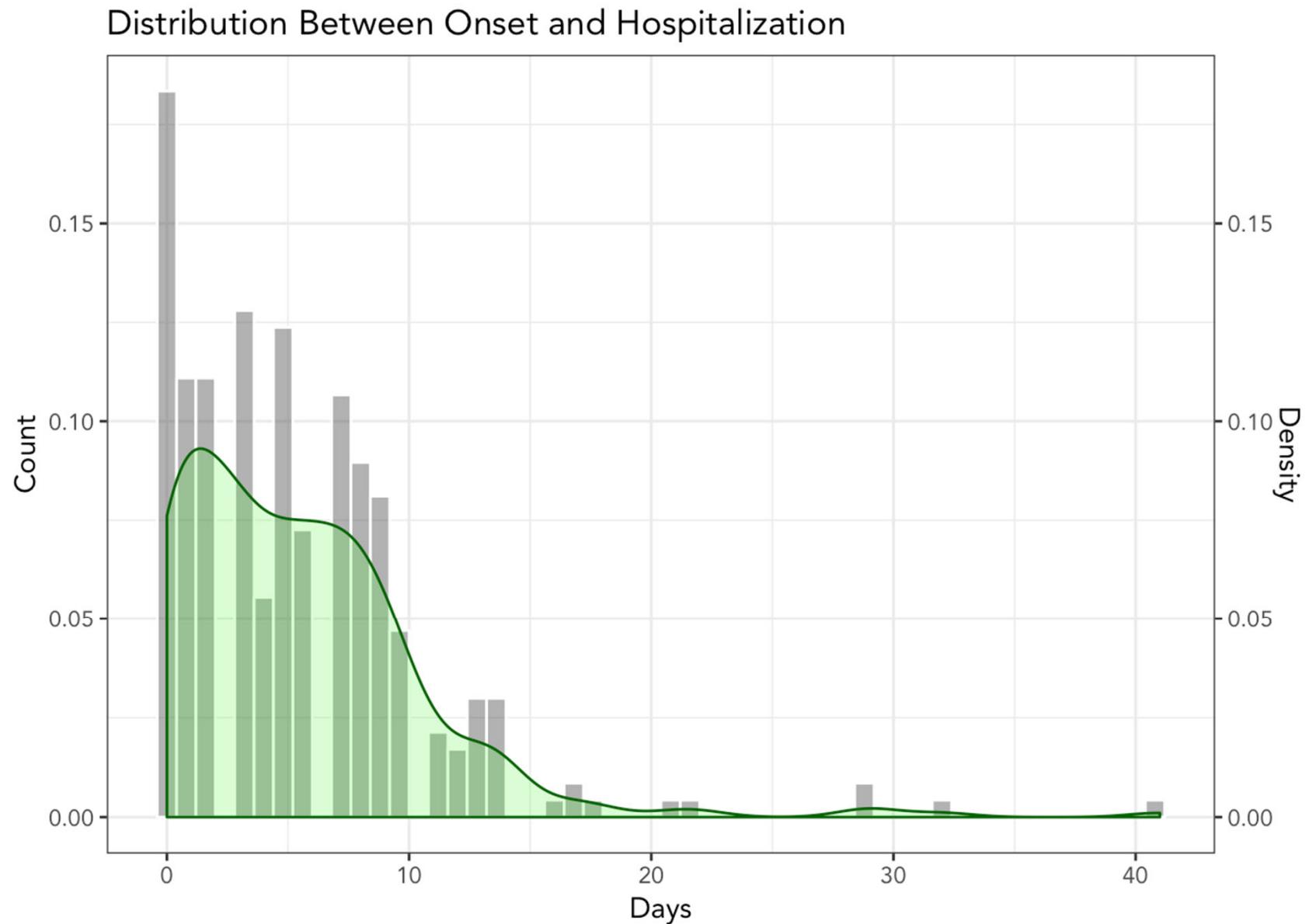
Ways to describe a sample:

Histogram

approximates PDF  
(or PMF)



# PDF of time between COVID-19 symptoms onset and hospitalization in IL, April 2020

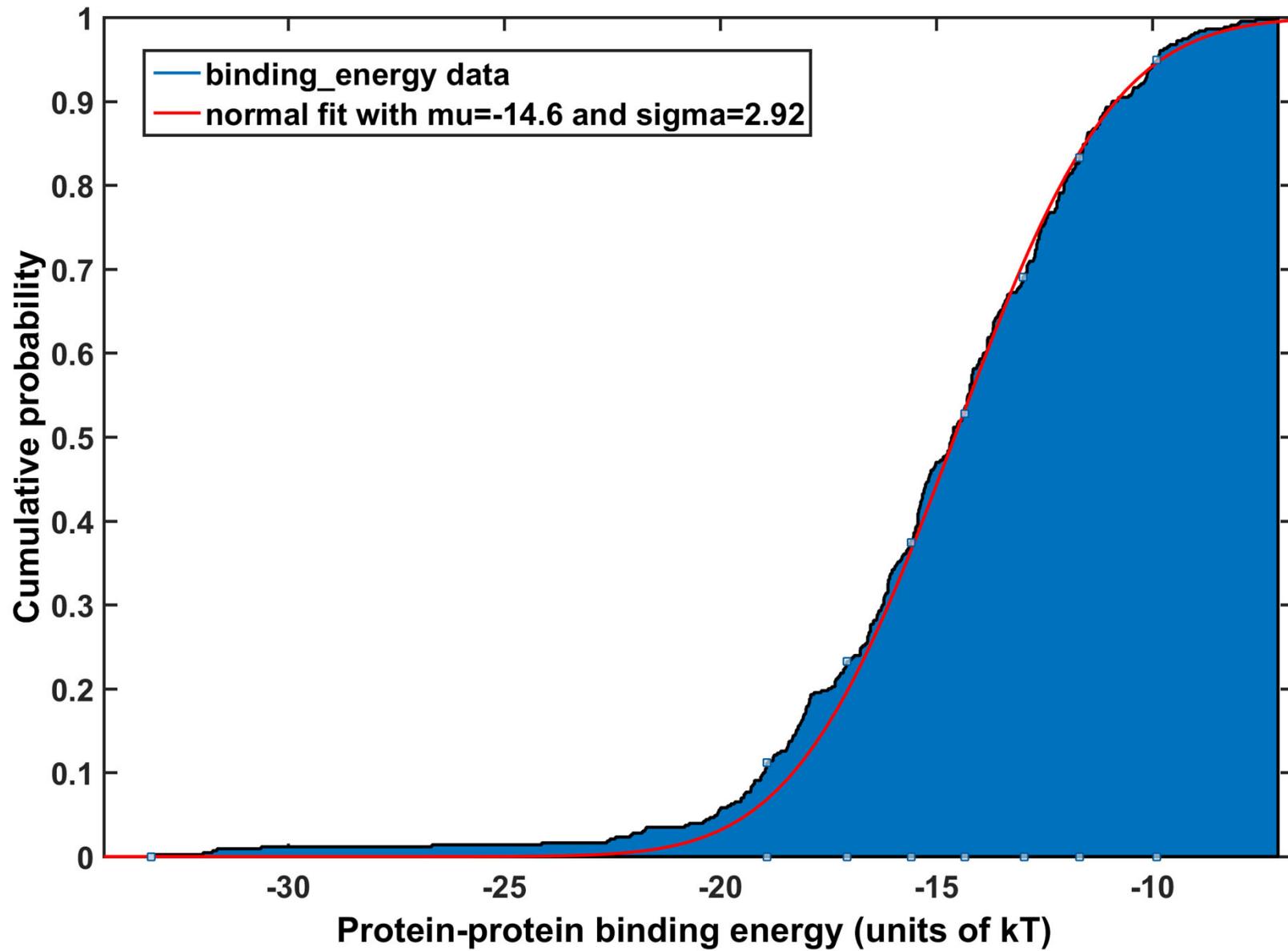


# Histograms with Unequal Bin Widths

- If the data is tightly clustered in some regions and scattered in others, it is visually helpful to use **narrow bin widths** in the **clustered region** and **wide bin widths** in the **scattered areas**.
- To approximate the PDF, the **rectangle area**, not the height, must be proportional to the **bin relative frequency**.

$$\text{Rectangle height} = \frac{\text{bin relative frequency}}{\text{bin width}}$$

# Cumulative Frequency Plot



# Median, Quartiles, Percentiles

- The **median**  $q_2$  divides the sample into two equal parts: 50% ( $n/2$ ) of sample points below  $q_2$  and 50% ( $n/2$ ) points above  $q_2$
- The **three quartiles** partition the data into four equally sized counts or segments.
  - 25% of the data is less than  $q_1$ .
  - 50% of the data is less than  $q_2$ , the median.
  - 75% of the data is less than  $q_3$ .
- There are **100 percentiles**.  $n$ -th percentile  $p_n$  is defined so that  $n\%$  of the data is less than  $p_n$

# Matlab exercise #1

- Find the median and lower & upper quartiles of a n=100 sample drawn from a continuous uniform distribution in [0,1]
- Do not use built-in Matlab functions for this exercise!
- Hint: use `[a,b]=sort(r1);` to rank order your sample. The variable `a` returns `r1` sorted in the increasing order.
- How to find the median and both quartiles from `a`?

# How to find the median & quartiles

- % Example: find median and lower quartile of
- % a sample with n=100 drawn from uniform
- $r1=rand(100,1);$
- $[a,b]=sort(r1);$
- $med=(a(50)+a(51))./2$
- $sum(r1<med)$  % verify
- $q1=(a(25)+a(26))./2$
- $sum(r1<q1)$  % verify

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 SVHOST.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 MALE AND FEMALE BIKES**

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

**WHY ARE THERE SQUIRRELS**



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

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 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 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**

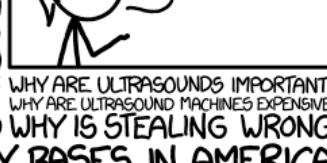
WHY IS EARTH TILTED  
WHY IS SPACE BLACK  
WHY IS OUTER SPACE SO COLD  
WHY ARE THERE PYRAMIDS ON THE MOON  
WHY IS NASA SHUTTING DOWN  
**WHY ARE THERE GHOSTS**  
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE  
WHY ARE THERE SPIDERS IN MY ROOM  
WHY ARE THERE SO MANY SPIDERS IN MY ROOM  
**WHY ARE THERE FEMALE**  
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 ARE ULTRASOUNDS IMPORTANT  
WHY ARE ULTRASOUND MACHINES EXPENSIVE  
WHY IS STEALING WRONG  
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY AREN'T MY ARMS GROWING  
WHY DO I FEEL DIZZY  
WHY ARE THERE WEEKS  
WHY ARE THERE DOGS AFRAID OF FIREWORKS  
WHY ARE THERE NO KING IN ENGLAND

WHY IS SEX SO IMPORTANT



WHY AREN'T THERE GUNS IN HARRY POTTER



# Box-and-Whisker Plot

- A box plot is a graphical display showing **Spread**, **Outliers**, **Center**, and **Shape (SOCS)**.
- It displays the **5-number summary**: *min,  $q_1$ , median,  $q_3$ , and max.*

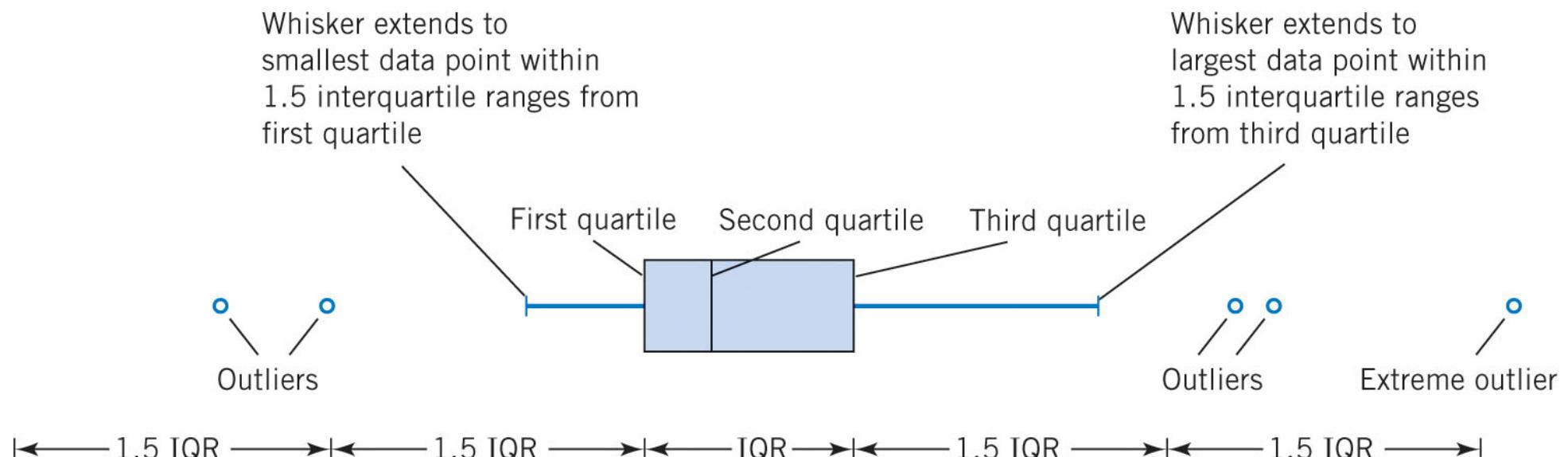
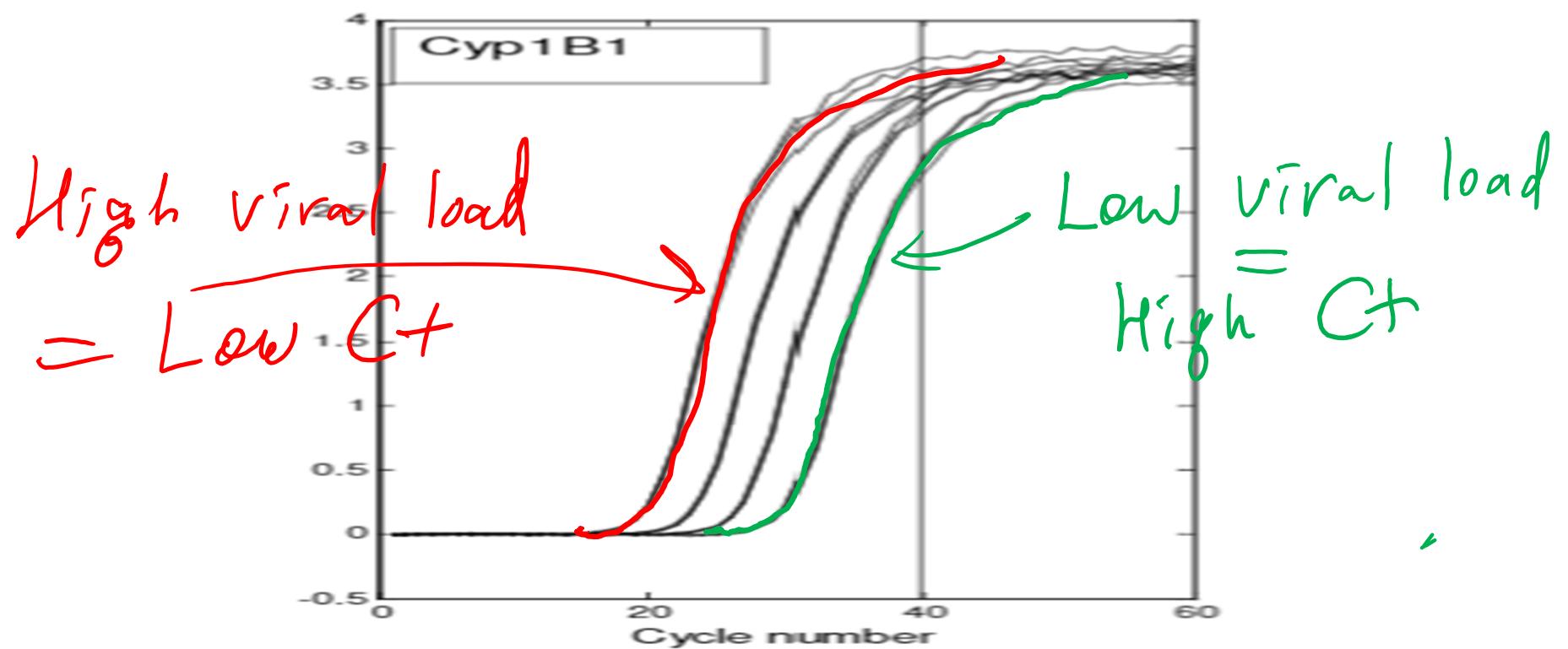


Figure 6-13 Description of a box plot.

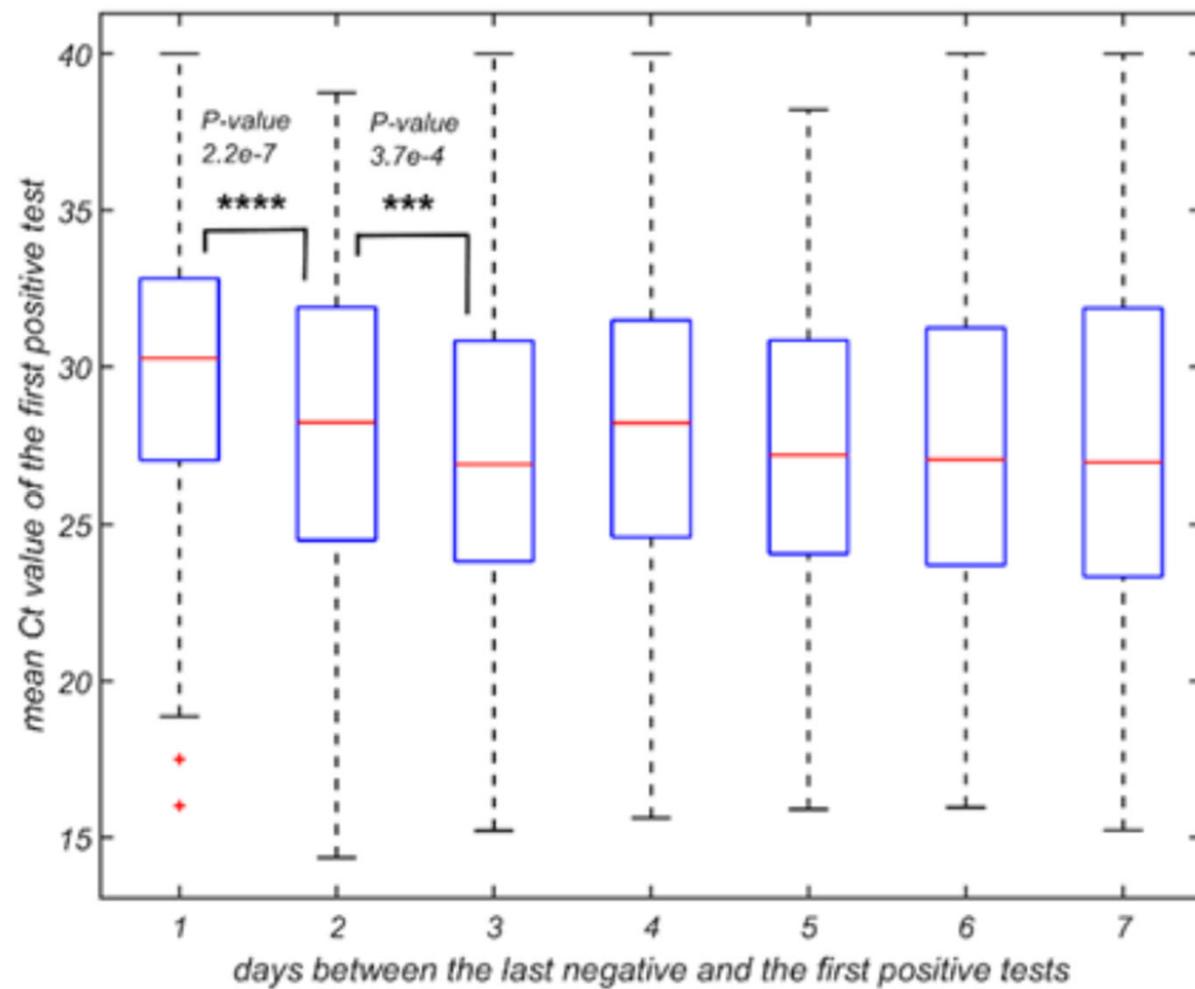
# Reminder

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

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



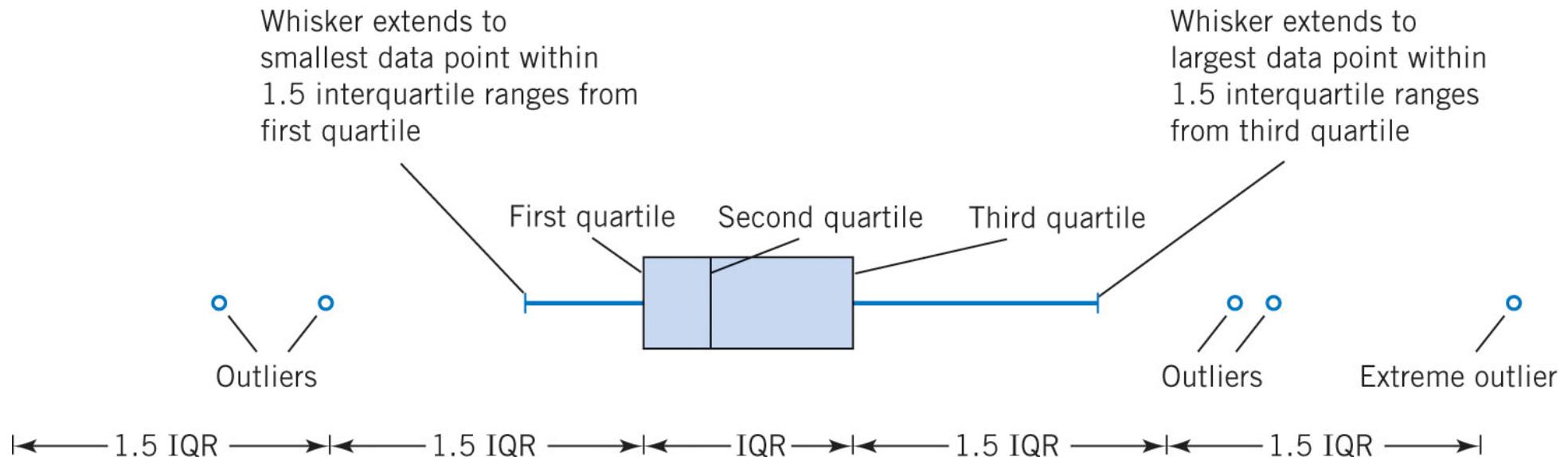
# Bar plot based on COVID-19 tests at UIUC



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

