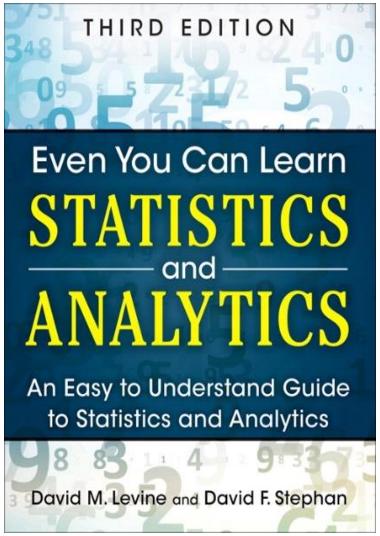
IMGD 2905

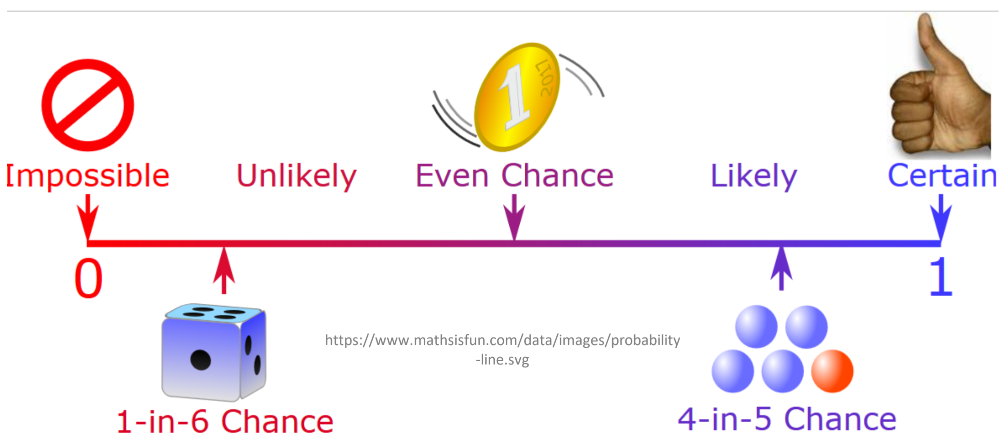
Probability

Chapters 4 & 5

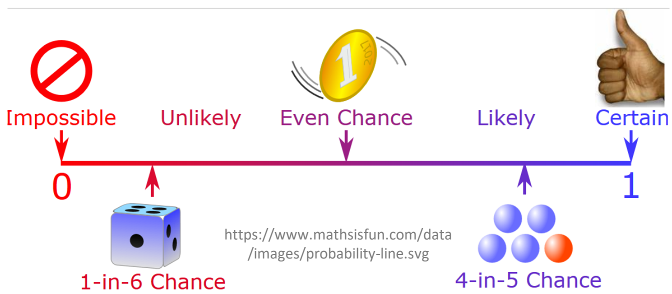


Overview

- Statistics important for game analysis
- Probability important for statistics
- So, understand some basic probability
- Also, probability useful for game development



Groupwork



- What are some examples of probabilities needed for game development?
- Provide a specific example
- Icebreaker, Groupwork, Questions

https://web.cs.wpi.edu/~imgd2905/d22/groupwork/5-probabilities/handout.html

Overview

- Statistics important for game analysis
- Probability important for statistics
- So, understand some basic probability
- Also, probability itself useful for game development

- Probabilities for game development?
- Examples?

Overview

- Statistics important for game analysis
- Probability important for statistics
- So, understand some basic probability
- Also, probability itself useful for game development

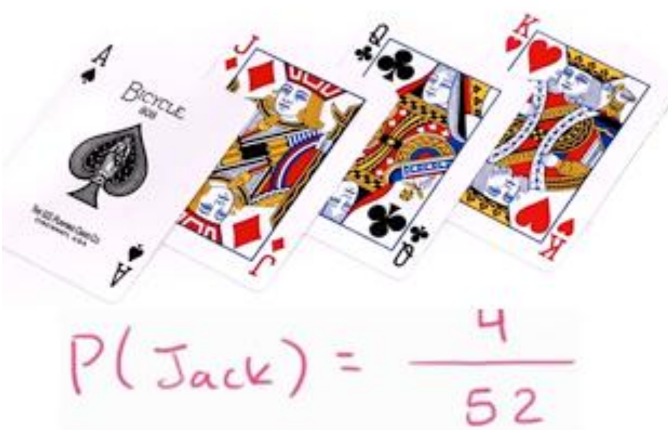
- Probabilities for game development?
- Probability attack will succeed
- Probability loot from enemy contains rare item
- Probability enemy spawns at particular time
- Probability action (e.g., building a castle) takes particular amount of time
- Probability players at server

Outline

- Introduction (done)
- Probability (next)
- Probability Distributions

Probability Definitions (1 of 3)

- Probability – way of assigning numbers to outcomes to express likelihood of event
- Event – outcome of experiment or observation
  - Elementary – simplest type for given experiment. independent
  - Joint/Compound – more than one elementary



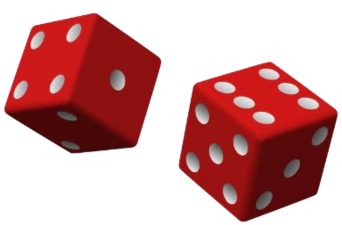
https://cdn.kastatic.org/googleusercontent/Z0TuLq2KolavsrfDXSbLqi0S-wnlCrC13cKGG68wK9ljrTiXzRqvfq7IpWNzcwgzlpEOI8YmMafp4K4zO0sanvXu

- Roll die (d6) and get 6
  - elementary event
- Roll die (d6) and get even number
  - compound event, consists of elementary events 2, 4, and 6
- Pick card from standard deck and get queen of spades
  - elementary event
- Pick card from standard deck and get face card
  - compound event
- Observe players logging in to MMO server and see if two people log in less than 15 minutes apart
  - compound event

We'll treat/compute probabilities of elementary versus compound separately

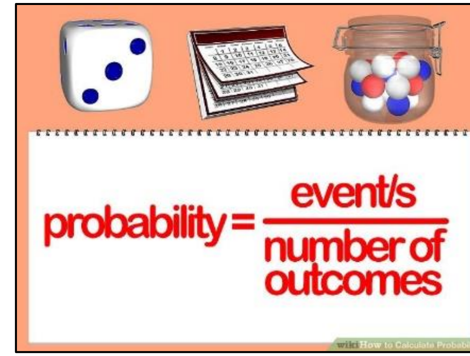
Probability – Definitions (2 of 3)

- Exhaustive set of events – set of all possible outcomes of experiment/observation
- Mutually exclusive sets of events – elementary events that do not overlap
- Roll d6: Events: 1, 2
  - not exhaustive, mutually exclusive



- Roll d6: Events: 1, 2, 3, 4, 5, 6
  - exhaustive, mutually exclusive
- Roll d6: Events: get even number, get number divisible by 3, get a 1 or get a 5
  - exhaustive, but overlap
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 15+ seconds
  - exhaustive, mutually exclusive
- Observe logins: time between arrivals <10 seconds, 10+ and <15 seconds inclusive, or 10+ seconds
  - exhaustive, but overlap

Probability – Definitions (3 of 3)



https://goo.gl/iy3YGr

- Probability – likelihood of event to occur, ratio of favorable cases to all cases
- Set of rules that probabilities must follow
  - Probabilities must be between 0 and 1 (but often written/said as percent)
  - Probabilities of set of *exhaustive, mutually exclusive* events must add up to 1
- e.g., d6: events 1, 2, 3, 4, 5, 6. Probability of $1/6^{th}$ to each, sum of P(1) + P(2) + P(3) + P(4) + P(5) + P(6) = 1

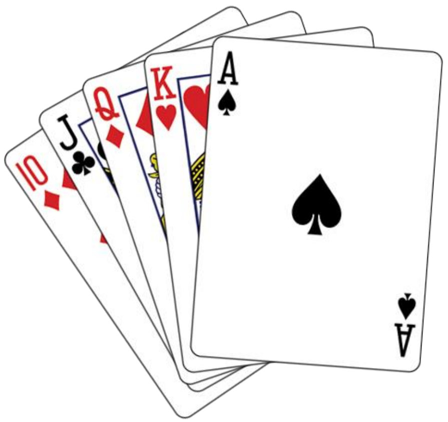
  → legal set of probabilities
- e.g., d6: events 1, 2, 3, 4, 5, 6. Probability of ½ to roll 1, ½ to roll 2, and 0 to all the others sum of P(1) + … + P(6) = 0.5 + 0.5 + 0 … + 0 = 1

  → Also legal set of probabilities
  - Not how honest d6's behave in real life!

Q: how to assign probabilities?

How to Assign Probabilities?

http://static1.squarespace.com/static/5a14961cf14aa1f245bc3942/5a1c5e8d8165f542d6db3b0e/5acecc7f03ce64b9a46d99c6/1529981982981/Michael+Jordan+%2833%29.png?format=1500w



https://newvitruvian.com/images/marbles-clipart-bag-marble-4.png

Q: how to assign probabilities?

Assigning Probabilities

- Classical (by theory)
  - In some cases, exhaustive, mutually exclusive outcomes equally likely → assign each outcome probability of *1/n*
  - e.g., *d6*: 1/6, *Coin*: prob heads ½, tails ½, *Cards*: pick Ace 1/13
- Empirically (by observation)
  - Obtain data through measuring/observing
  - e.g., Watch how often people play PUBG in FL222 versus some other game. Say, 30% PUBG. Assign that as probability
- Subjective (by hunch)
  - Based on expert opinion or other subjective method
  - e.g., eSports writer says probability Fnatic (European LoL team) will win World Championship is 25%

Rules About Probabilities (1 of 2)

- Complement: A an event. Event "Probability A does not occur" called *complement* of A, denoted A'

  P(A') = 1 - P(A) ← Why?

  - e.g., d6: P(6) = 1/6, complement is P(6') and probability of "not 6" is 1-1/6, or 5/6.
  - Note: Value often denoted p, complement is q
- Mutually exclusive: Have no simple outcomes in common – can't both occur in same experiment

  P(A or B) = P(A) + P(B)

  - "Probability either occurs"
  - e.g., d6: P(3 or 6) = P(3) + P(6) = 1/6 + 1/6 = 2/6

Rules About Probabilities (2 of 2)

- Independent: Probability that one occurs doesn't affect probability that other occurs
  - e.g., 2d6: A= die 1 get 5, B= die 2 gets 6. Independent, since result of one roll doesn't affect roll of other
  - "Probability both occur" P(A and B) = P(A) x P(B)
  - e.g., 2d6: prob of "snake eyes" is P(1) x P(1) = 1/6 x 1/6 = 1/36
- Not independent: One occurs affects probability that other occurs
  - Probability both occur P(A and B) = P(A) x P(B | A)
    - Where P(B | A) means prob B given A happened
  - e.g., PUBG chance of getting top 10 is 10%. Chance of using only stock gun 50%. You might think that:
    - P(top 10) x P(stock) = 0.10 x 0.50 = 0.05
  - But likely *not* independent. P(top | stock) < 5%. So, need non-independent formula
    - P(top) * P(top | stock)

(Card example next slide)

Probability Example

- Probability drawing King?

Probability Example

- Probability drawing King?

  $P(K) = \frac{1}{4}$

- Draw, put back. Now?

Probability Example

- Probability drawing King?

  P(K) = ¼

- Draw, put back. Now?

  P(K) = ¼

- Probability *not* King?

Probability Example

- Probability drawing King?

  $P(K) = 1/4$

- Draw, put back. Now?

  $P(K) = 1/4$

- Probability *not* King?

  $P(K') = 1 - P(K) = 3/4$

- Draw, put back. 2 Kings?

Probability Example

- Draw. King or Queen?

- Probability drawing King?

  P(K) = ¼

- Draw, put back. Now?

  P(K) = ¼

- Probability *not* King?

  P(K') = 1-P(K) = ¾

- Draw, put back. Draw. 2 Kings?

  P(K) x P(K) = ¼ x ¼ = 1/16

Probability Example

- Draw. King or Queen?

  P(K or Q) = P(K) + P(Q)

  = ¼ + ¼ = ½

- Probability drawing King?

  P(K) = ¼

- Draw, put back. Now?

  P(K) = ¼

- Probability *not* King?

  P(K') = 1-P(K) = ¾

- Draw, put back. Draw. 2 Kings?

  P(K) x P(K) = ¼ x ¼ = 1/16

Probability Example

- Probability drawing King?

  P(K) = ¼

- Draw, put back. Now?

  P(K) = ¼

- Probability *not* King?

  P(K’) = 1-P(K) = ¾

- Draw, put back. Draw. 2 Kings?

  P(K) x P(K) = ¼ x ¼ = 1/16

- Draw. King or Queen?

  P(K or Q) = P(K) + P(Q)

  = ¼ + ¼ = ½

- Draw, put back. Draw. Not King either card?

Probability Example

- Probability drawing King?

  $P(K) = \frac{1}{4}$

- Draw, put back. Now?

  $P(K) = \frac{1}{4}$

- Probability *not* King?

  $P(K') = 1 - P(K) = \frac{3}{4}$

- Draw, put back. Draw. 2 Kings?

  $P(K) \times P(K) = \frac{1}{4} \times \frac{1}{4} = 1/16$

- Draw. King or Queen?

  $P(K \text{ or } Q) = P(K) + P(Q)$

  $= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

- Draw, put back. Draw. Not King either card?

  $P(K') \times P(K') = \frac{3}{4} \times \frac{3}{4} = 9/16$

- Draw, *don't* put back. Draw. Not King either card?

Probability Example

- Probability drawing King?

  $P(K) = \frac{1}{4}$

- Draw, put back. Now?

  $P(K) = \frac{1}{4}$

- Probability *not* King?

  $P(K') = 1 - P(K) = \frac{3}{4}$

- Draw, put back. 2 Kings?

  $P(K) \times P(K) = \frac{1}{4} \times \frac{1}{4} = 1/16$

- Draw. King or Queen?

  $P(K \text{ or } Q) = P(K) + P(Q)$

  $= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

- Draw, put back. Draw. Not King either card?

  $P(K') \times P(K') = \frac{3}{4} \times \frac{3}{4} = 9/16$

- Draw, *don't* put back. Draw. Not King either card?

  $P(K') \times P(K' \mid K') = \frac{3}{4} \times (1-1/3)$

  $= \frac{3}{4} \times 2/3$

  $= 6/12 = \frac{1}{2}$

- Draw, don't put back. Draw. King 2^{nd} card?

Probability Example

- Probability drawing King?

  $P(K) = \frac{1}{4}$

- Draw, put back. Now?

  $P(K) = \frac{1}{4}$

- Probability *not* King?

  $P(K') = 1-P(K) = \frac{3}{4}$

- Draw, put back. 2 Kings?

  $P(K) \times P(K) = \frac{1}{4} \times \frac{1}{4} = 1/16$

- Draw. King or Queen?

  $P(K \text{ or } Q) = P(K) + P(Q)$

  $= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

- Draw, put back. Draw. Not King either card?

  $P(K') \times P(K') = \frac{3}{4} \times \frac{3}{4} = 9/16$

- Draw, *don't* put back. Draw. Not King either card?

  $P(K') \times P(K' \mid K') = \frac{3}{4} \times (1-1/3)$

  $= \frac{3}{4} \times 2/3$

  $= 6/12 = \frac{1}{2}$

- Draw, don't put back. Draw. King 2nd card?

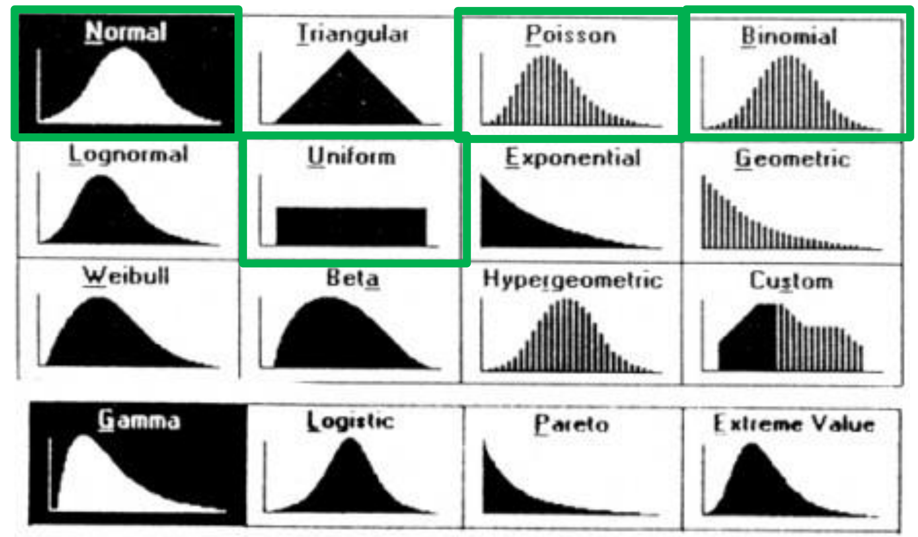
  $P(K') \times P(K \mid K') = \frac{3}{4} \times \frac{1}{3} = 3/12 = \frac{1}{4}$

Outline

- Intro (done)
- Probability (done)
- Probability Distributions (next)

Probability Distributions

- Probability distribution – values and likelihood (expected value) that random variable can take
- Why? If can model mathematically, can use to predict occurrences
  - e.g., probability slot machine pays out on given day
  - e.g., probability game server hosts player today
  - e.g., probability certain game mode is chosen by player
  - Also, some statistical techniques for some distributions only



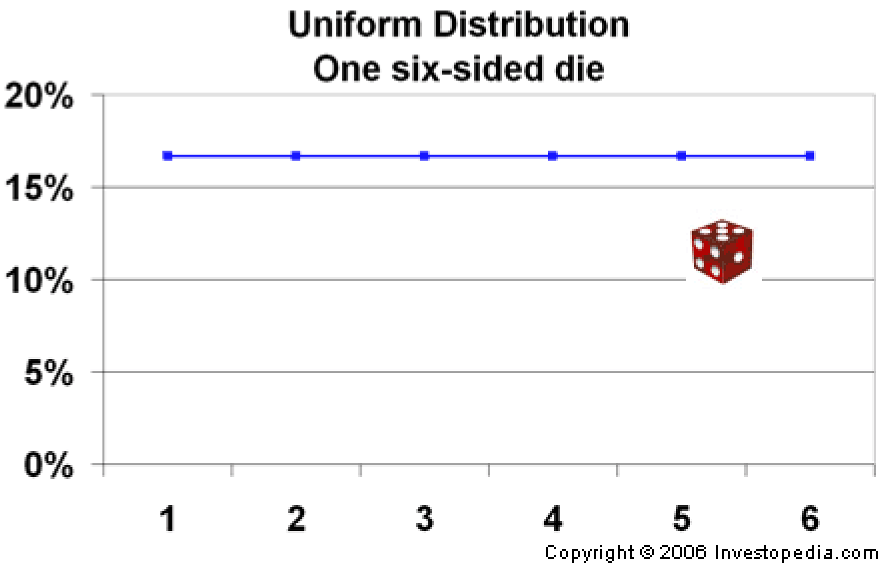
https://goo.gl/jqomFI

Types discussed:

Uniform (discrete)
Binomial (discrete)
Poisson (discrete)
Normal (continuous)

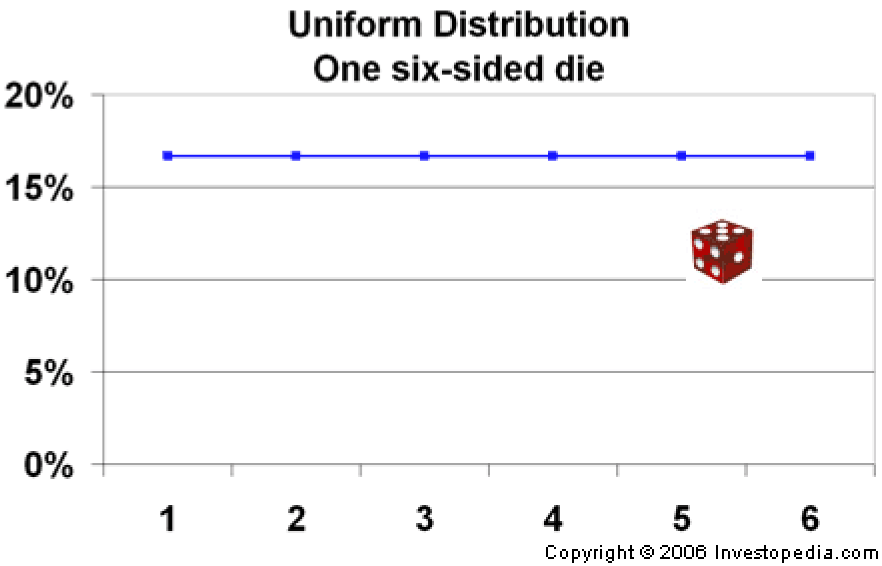
Remember empirical rule?
What distribution did it apply to?

Uniform Distribution



- "So what?"
- Can use known formulas

Uniform Distribution



- "So what?"
- Can use known formulas

Mean	$\frac{a+b}{2}$
Median	$\frac{a+b}{2}$
Mode	N/A
Variance	$\frac{(b-a+1)^2-1}{12}$

Mean = (1 + 6) / 2 = 3.5

Variance = $((6 – 1 + 1)^2 – 1)/12$

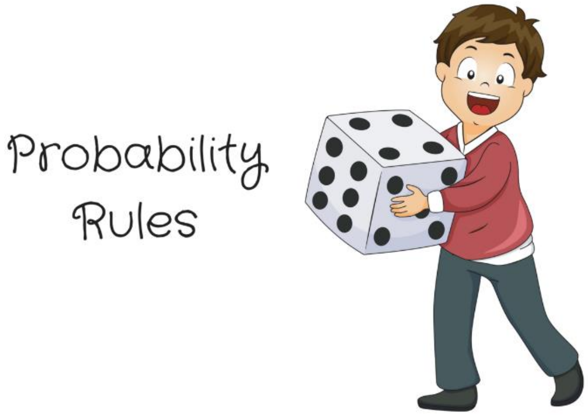
= 2.9

Std Dev = sqrt(Variance) = 1.7

Note – mean is also the expected value
(if you did a lot of trials, would be average result)

Binomial Distribution Example (1 of 3)

How to assign probabilities?



- Suppose toss 3 coins
- Random variable

  X = number of heads
- Want to know probability of *exactly* 2 heads

  P(X=2) = ?

Binomial Distribution Example (1 of 3)

- Suppose toss 3 coins
- Random variable

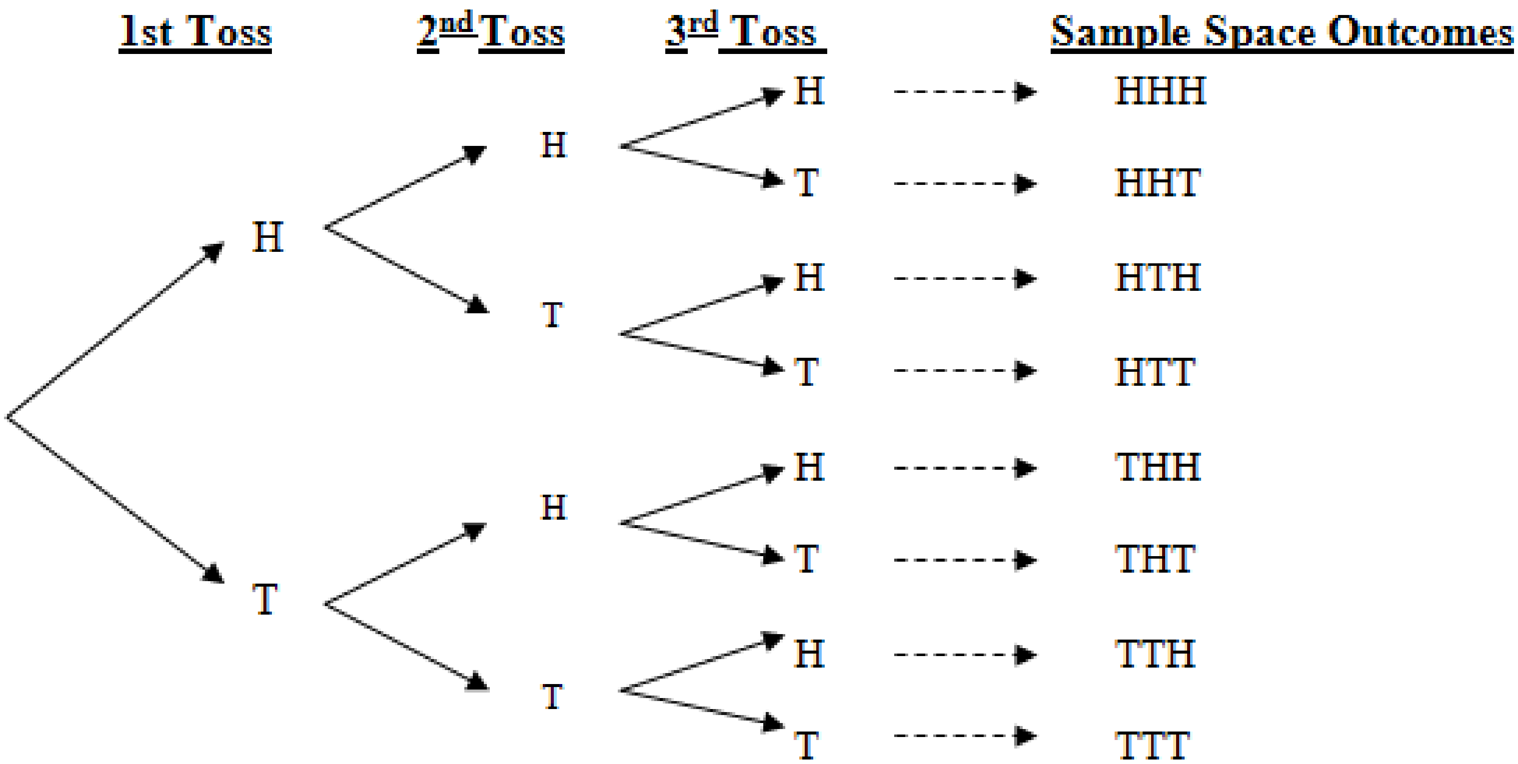
  X = number of heads
- Want to know probability of *exactly* 2 heads

  P(X=2) = ?

How to assign probabilities?

- Could *measure* (empirical)
  - *Q: how?*
- *Could use "hunch"* (subjective)
  - *Q: what do you think?*
- *Could use theory* (classical)
  - *Math using our probability* rules (not shown)
  - Enumerate (next)

Binomial Distribution Example (2 of 3)

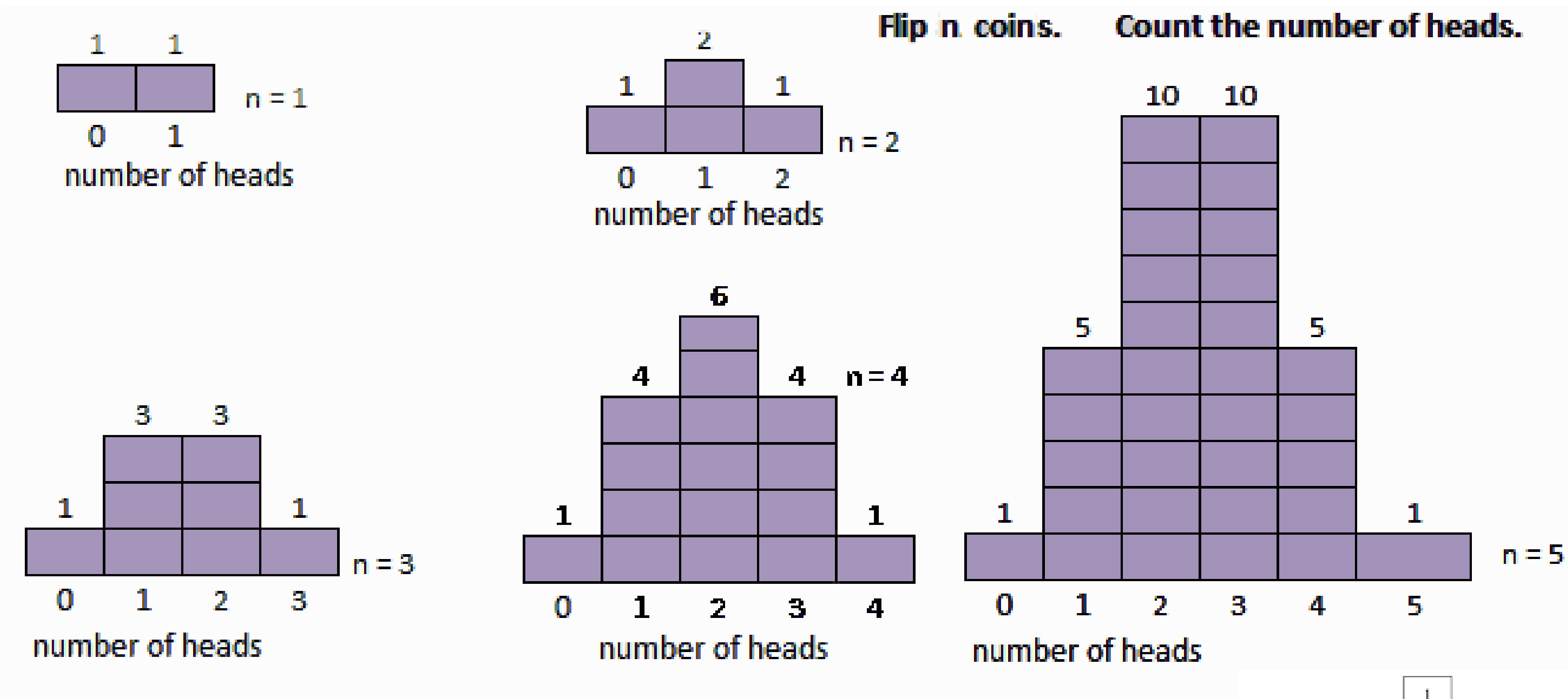


http://web.mnstate.edu/peil/MDEV102/U3/S25/Cartesian3.PNG

All equally likely (p is 1/8 for each)
→ P(HHT) + P(HTH) + P(THH) = 3/8

Can draw histogram of number of heads

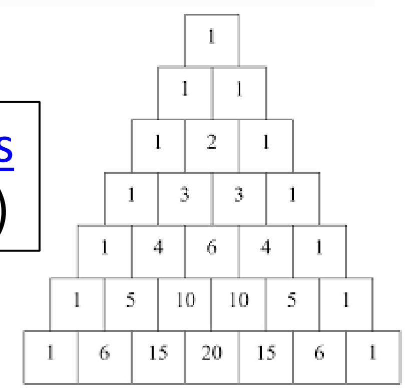
Binomial Distribution Example (3 of 3)



http://www.mathnstuff.com/math/spoken/here/2class/90/binom2.gif

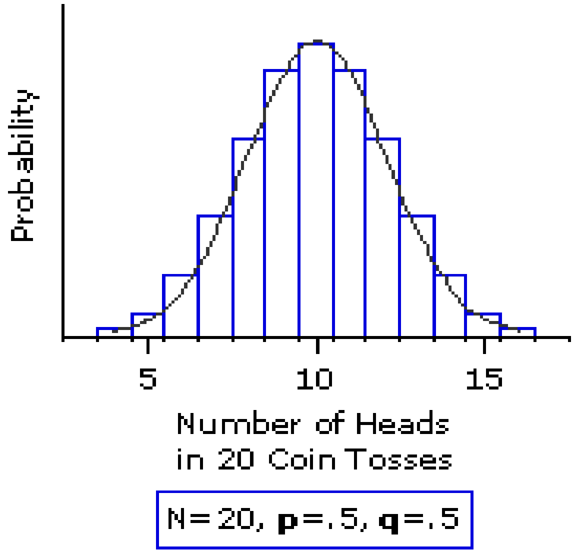
These are *all* binomial distributions

(Pascal's Triangle)



Binomial Distribution (1 of 2)

- In general, any number of trials (n) & any probability of successful outcome (p) (e.g., heads)



http://www.vassarstats.net/textbook/f0603.gif

- Characteristics of experiment that gives random number with binomial distribution:
  - Experiment of n identical trials.
  - Trials are independent
  - Each trial only two possible outcomes, Success or Fail
  - Probability of Success each trial is same, denoted p
  - Random variable of interest (X) is number of Successes in n trials

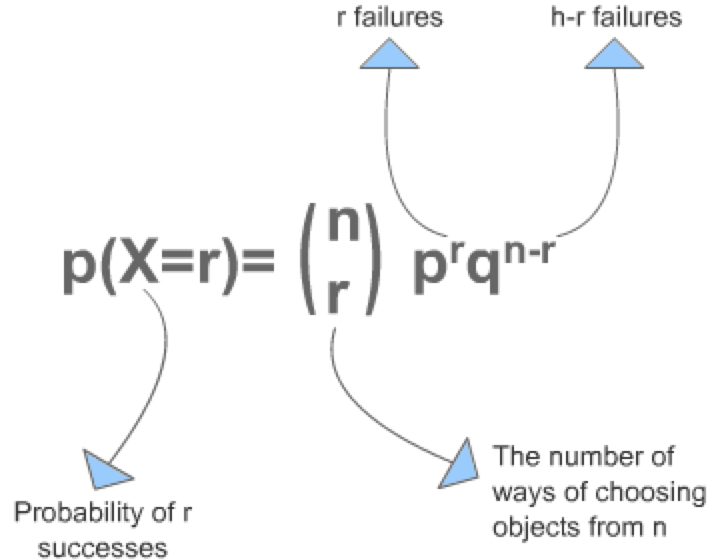
Binomial Distribution (2 of 2)

- "So what?"
- Can use known formulas

$$MEAN: \quad \mu = np$$

$$Variance: \; \sigma^2 = npq$$

$$SD: \; \sigma = \sqrt{npq}$$



http://www.s-cool.co.uk/gifs/a-mat-sdisc-dia08.gif

Excel: `binom.dist()`

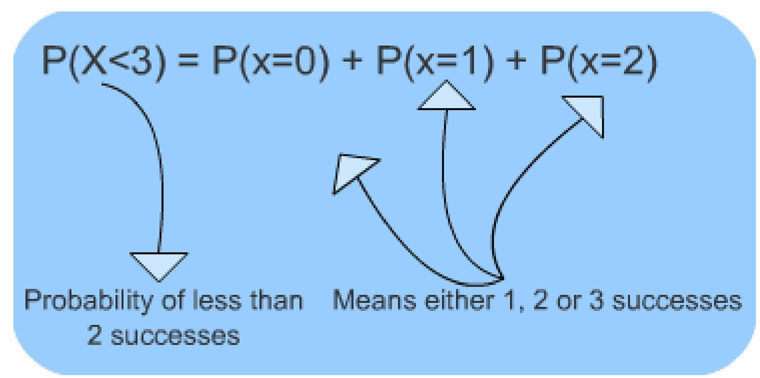
`binom.dist(x,trials,prob,cumulative)`

– 2 heads, 3 flips, coin, discrete

`=binom.dist(2, 3, 0.5, FALSE)`

`=0.375` (i.e., 3/8)

If "true"?

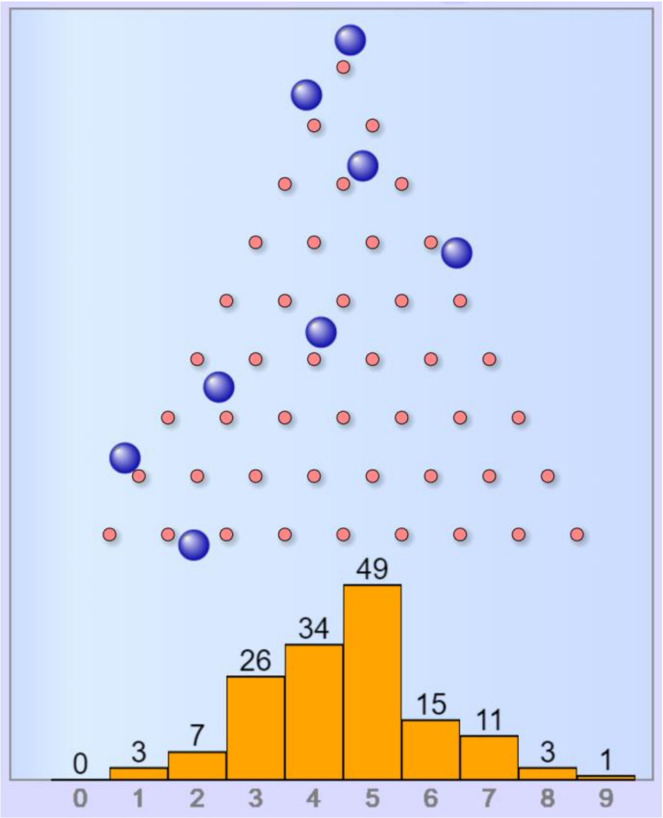


http://www.s-cool.co.uk/gifs/a-mat-sdisc-dia12.gif

Binomial Distribution Example

- Each row is like a coin flip
  - right = "heads"
  - left = "tails"
- Bottom axis is number of heads
- Gives and "empirical" way to estimate P(X)

bin(X) ÷
sum(bin(0) + bin(1) + …)



https://www.mathsisfun.com/data/quincunx.html

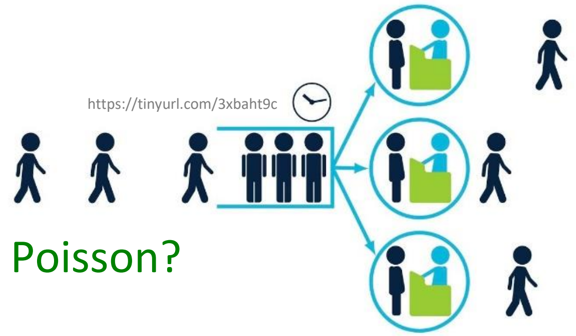
Poisson Distribution

- Distribution of probability of x events occurring in certain interval (broken into units)
  - Interval can be time, area, volume, distance
  - e.g., number of players arriving at server lobby in 5-minute period between noon-1pm
- Requires
  1. Probability of event same for all time units
  2. Number of events in one time unit independent of number of events in any other time unit
  3. Events occur singly (not simultaneously). In other words, as interval unit gets smaller, probability of two events occurring approaches 0

Poisson Distributions?

Could Be Poisson

- Number of groups arriving at restaurant during dinner hour
- Number of logins to MMO during prime time
- Number of defects (bugs) per 100 lines of code
- People arriving at cash register (if they shop individually)



Not Poisson

- Number of people arriving at restaurant during dinner hour
  - People frequently arrive in groups
- Number of students registering for course in Workday per hour on first day of registration
  - Prob not equal – most register in first few hours
  - Not independent – if too many register early, system crashes

Phrase people use is
random arrivals

Poisson Distribution

- Distribution of probability of x events occurring in certain interval

$$P(X=x)=e^{-\lambda}\frac{\lambda^{x}}{x!}$$

- X = a Poisson random variable
- x = number of events whose probability you are calculating
- λ = the Greek letter "lambda," which represents the average number of events that occur per time interval
- e = a constant that's equal to approximately 2.71828

http://www.dummies.com/education/math/business-statistics/how-to-compute-poisson-probabilities/

Poisson Distribution Example

1. Number of games student plays per day averages 1 per day
2. Number of games played per day independent of all other days
3. Can only play one game at a time

What's probability of playing 2 games tomorrow?

In this case, the value of $\lambda = 1$, want $P(X=2)$

$$P(X=2) = e^{-1}\frac{1^2}{2!} = 0.1839$$

Current Poisson Distribution Example

- New England city
- Average new COVID-19 cases 50/day
- Local hospital has 60 free beds
- What is the probability more than 60 in one day?

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad = ???$$

Current Poisson Distribution Example

- New England city
- Average new COVID-19 cases 50/day
- Local hospital has 60 free beds
- What is the probability more than 60 in one day?

60

$$P(X=x)=e^{-\lambda}\frac{\lambda^{x}}{x!} = ???$$

50

60

Current Poisson Distribution Example

https://stattrek.com/online-calculator/poisson.aspx

- New England city
- Average new COVID-19 cases 50/day
- Local hospital has 60 free beds
- What is the probability more than 60 in one day?

Poisson random variable (x)	60
Average rate of success	50
Poisson Probability: P(X = 60)	0.02010
Cumulative Probability: P(X < 60)	0.90774
Cumulative Probability: P(X $\leq$ 60)	0.92784
Cumulative Probability: P(X > 60)	0.07216
Cumulative Probability: P(X $\geq$ 60)	0.09226

60

$$P(X=x)=e^{-\lambda}\frac{\lambda^{x}}{x!} = ???$$

50

60

Current Poisson Distribution Example

- New England city
- Average new COVID-19 cases 50/day
- Local hospital has 60 free beds
- What is the probability more than 60 in one day?

60

$$P(X = x) = e^{-\lambda}\frac{\lambda^{x}}{x!} = 0.02$$

50 60

https://stattrek.com/online-calculator/poisson.aspx

Poisson random variable (x)	60
Average rate of success	50
Poisson Probability: P(X = 60)	0.02010
Cumulative Probability: P(X < 60)	0.90774
Cumulative Probability: P(X $\leq$ 60)	0.92784
Cumulative Probability: P(X > 60)	0.07216
Cumulative Probability: P(X $\geq$ 60)	0.09226

Q: How do we get greater than 60?

Current Poisson Distribution Example

https://stattrek.com/online-calculator/poisson.aspx

- New England city
- Average new COVID-19 cases 50/day
- Local hospital has 60 free beds
- What is the probability more than 60 in one day?

Poisson random variable (x)	60
Average rate of success	50
Poisson Probability: P(X = 60)	0.02010
Cumulative Probability: P(X < 60)	0.90774
Cumulative Probability: P(X $\leq$ 60)	0.92784
Cumulative Probability: P(X > 60)	0.07216
Cumulative Probability: P(X $\geq$ 60)	0.09226

60

$$P(X = x) = e^{-\lambda}\frac{\lambda^{x}}{x!} = 0.02$$

50 60

Q: How do we get greater than 60?

P(0) + P(1) + … + P(60) → P(≤60)

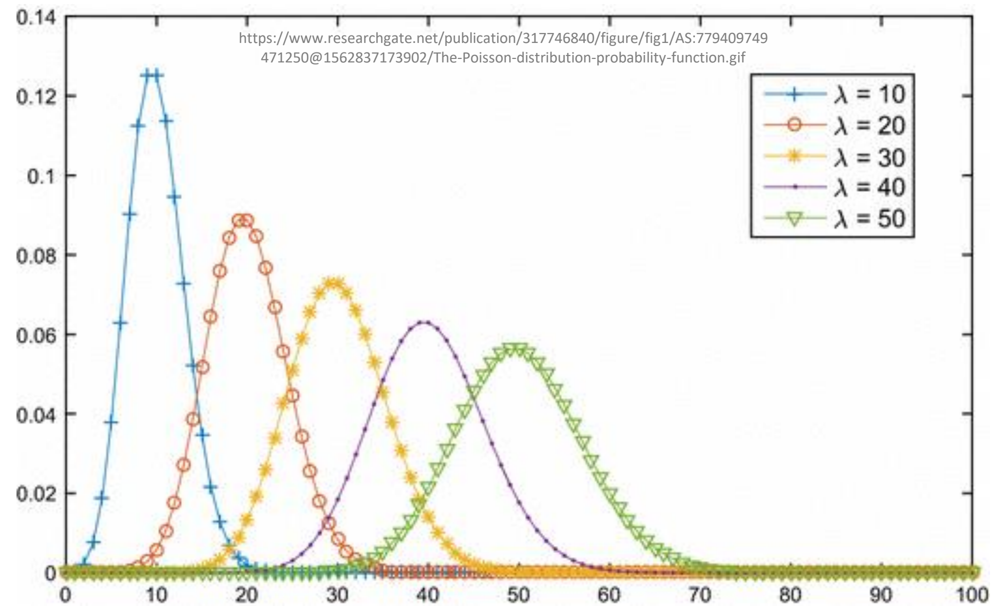
P(>60) = 1 - P(≤ 60)

Poisson Distribution

- "So what?" → Known formulas

$$P(X=x)=e^{-\lambda}\frac{\lambda^{x}}{x!}$$

- Mean = λ
- Variance = λ
- Std Dev = sqrt (λ)



Excel: `poisson.dist()`

`poisson.dist(x,mean,cumulative)`

mean 50 per day, 60 beds, chance > 60?

= `1 - POISSON.DIST(60, 50, TRUE)`

= 0.07216

e.g., Games → may want to know likelihood of 1.5x average people arriving at server

Expected Value – Formulation

- Expected value of discrete random variable is value you'd *expect* after many experimental trials. i.e., mean value of population

Value:	x_1	x_2	x_3	…	x_n
Probability:	$P(x_1)$	$P(x_2)$	$P(x_3)$	…	$P(x_n)$

- Compute by multiplying each value by probability and summing

$$\mu_x = E(X) = x_1P(x_1) + x_2P(x_2) + \ldots + x_nP(x_n)$$
$$= \Sigma\, x_iP(x_i)$$

Expected Value Example – Gambling Game

- Pay $3 to enter
- Roll 1d6 → 6? Get $7   1-5? Get $1
- What is expected payoff?

Outcome	Payoff	P(x)	xP(x)
1-5	$1		
6	$7		

Expected Value Example – Gambling Game

- Pay $3 to enter
- Roll 1d6 → 6? Get $7 1-5? Get $1
- What is expected payoff?

Outcome	Payoff	P(x)	xP(x)
1-5	$1	5/6	$5/6
6	$7	1/6	$7/6

E(X) =

Expected Value Example – Gambling Game

- Pay $3 to enter
- Roll 1d6 → 6? Get $7   1-5? Get $1
- What is expected payoff? Expected net?

Outcome	Payoff	P(x)	xP(x)
1-5	$1	5/6	$5/6
6	$7	1/6	$7/6

E(X) = $5/6 + $7/6 = $12/6 = $2

E(net) =

Expected Value Example – Gambling Game

- Pay \$3 to enter
- Roll 1d6 → 6? Get \$7   1-5? Get \$1
- What is expected payoff? Expected net?

Outcome	Payoff	P(x)	xP(x)
1-5	\$1	5/6	\$5/6
6	\$7	1/6	\$7/6

$E(X) = \$5/6 + \$7/6 = \$12/6 = \$2$

$E(net) = E(X) - \$3 = \$2 - \$3 = \$-1$

Outline

- Intro (done)
- Probability (done)
- Probability Distributions
  - Discrete (done)

So far random variable could take only discrete set of values

Q: What does that mean?

Q: What *other* distributions might we consider?

Outline

- Intro (done)
- Probability (done)
- Probability Distributions
  - Discrete (done)
  - Continuous (next)

Continuous Distributions

- Many random variables are continuous
  - e.g., recording *time* (time to perform service) or measuring something (*height*, *weight*, *strength*)
- For continuous, doesn't make sense to talk about P(X=x) → continuum of possible values for X
  - Mathematically, if all non-zero, total probability infinite (this violates our rule)

- So, continuous distributions have probability density, f(x)

→ How to use to calculate probabilities?

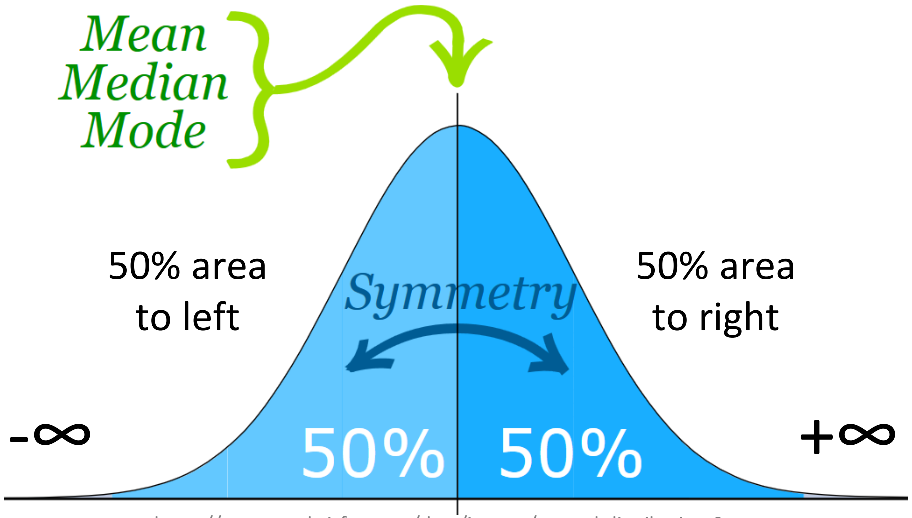
- Don't care about specific values
  - e.g., P(Height = 60.1946728163 inches)
- Instead, ask about *range* of values
  - e.g., P(59.5" < X < 60.5")
- Uses calculus (integrate area under curve) (not shown here)

Q: What continuous distribution is **especially** important?

→ the Normal Distribution

Normal Distribution (1 of 2)

- "Bell-shaped" or "Bell-curve"
  - Distribution from -∞ to +∞
- Symmetric
- Mean, median, mode all same
  - Mean determines location, standard deviation determines "width"
- Super important!
  - Lots of distributions follow a normal curve
  - Basis for inferential statistics (e.g., statistical tests)
  - "Bridge" between probability and statistics

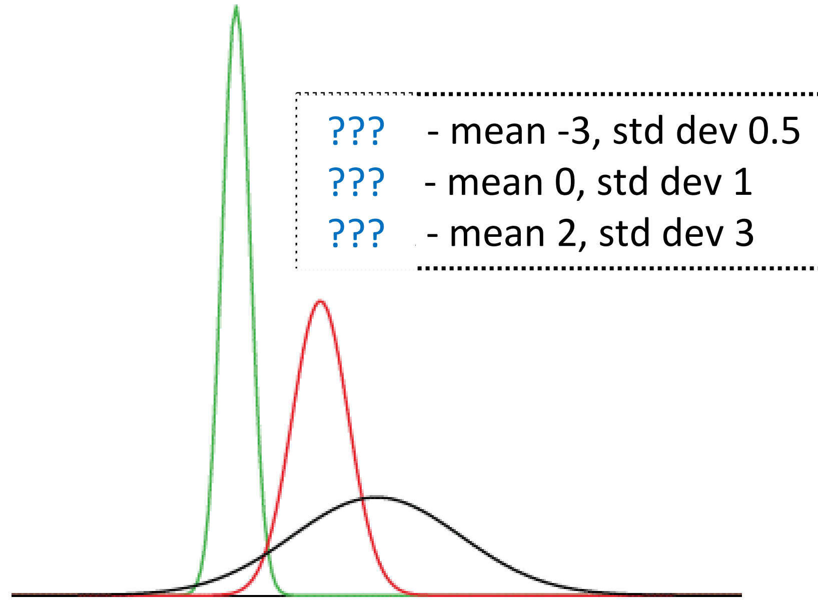


https://www.mathsisfun.com/data/images/normal-distribution-2.svg

Aka "Gaussian" distribution

Normal Distribution (2 of 2)

- *Many* normal distributions (see right)
- However, "the" normal distribution refers to standard normal
  - Mean (μ) = 0
  - Standard deviation (σ) = 1
- Can *convert* any normal to the standard normal
  - Given sample mean ($\bar{x}$)
  - Sample standard dev. (s)



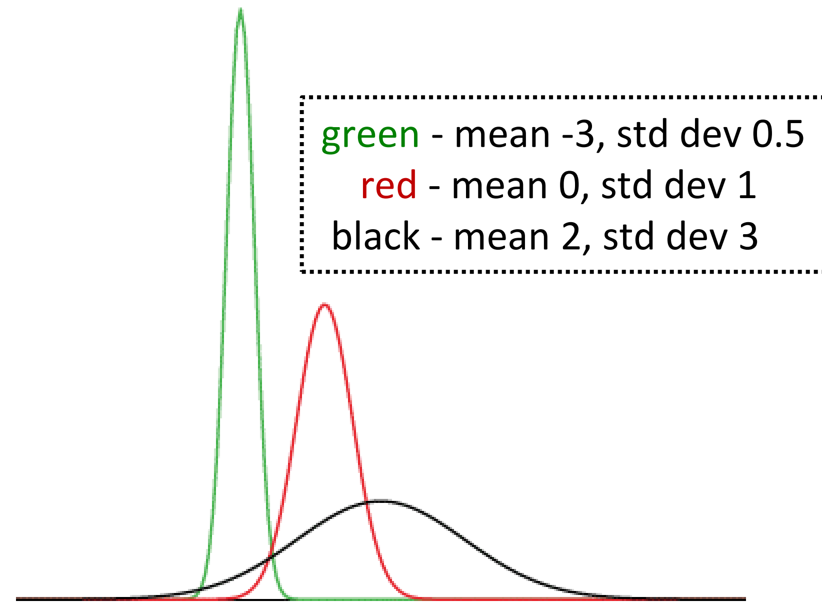
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

(Next)

`=norm.dist()`

Normal Distribution (2 of 2)

- *Many* normal distributions (see right)
- However, "the" normal distribution refers to standard normal
  - Mean (μ) = 0
  - Standard deviation (σ) = 1
- Can *convert* any normal to the standard normal
  - Given sample mean (x̄)
  - Sample standard dev. (s)



$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

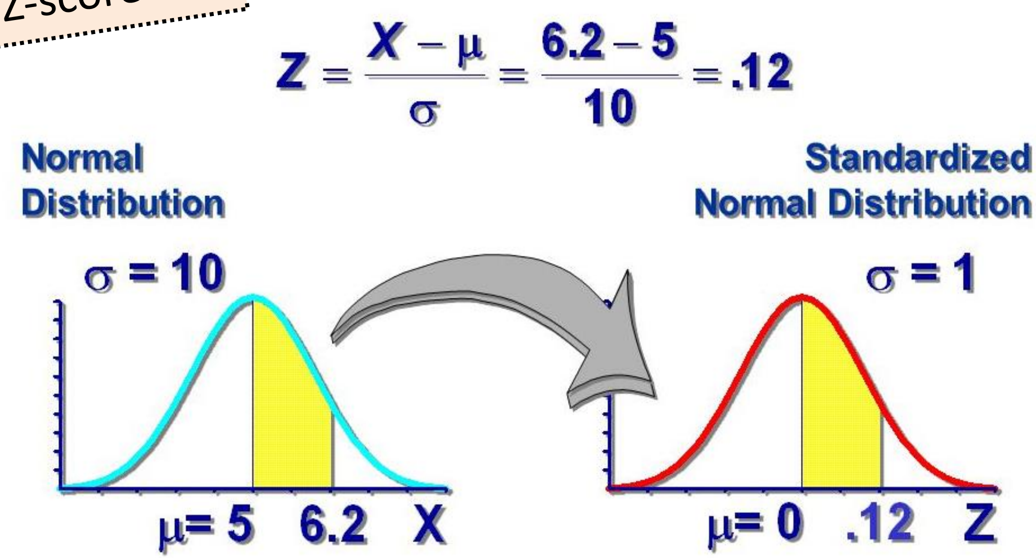
(Next)

=norm.dist()

Standard Normal Distribution

- Standardize
  - Subtract sample mean ($\bar{x}$)
  - Divide by sample standard deviation (s)
- Mean μ = 0
- Standard Deviation σ = 1
- Total area under curve = 1
  - Sounds like probability!

Remember the Z-score?



http://images.slideplayer.com/10/2753952/slides/slide_2.jpg

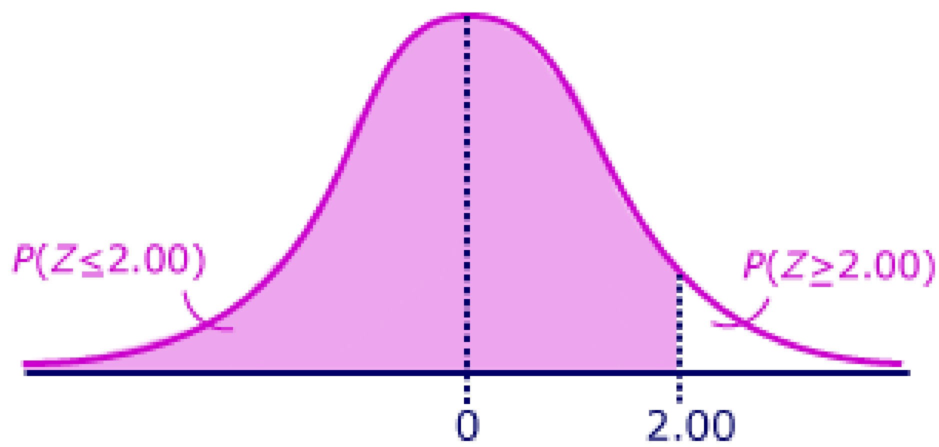
Use to predict how likely an observed sample is given population mean (next)

Using the Standard Normal

- Suppose *League of Legends* Champion released once every 24 days on average, standard deviation of 3 days
- What is the probability Champion released 30+ days?
- $x = 30$, $\bar{x} = 24$, $s = 3$

$$Z = (x - \bar{x}) / s$$
$$= (30 - 24) / 3$$
$$= 2$$

- Want to know $P(Z > 2)$



http://ci.columbia.edu/ci/premba_test/c0331/s6/s6_4.html

Q: how? Hint: what rule might help?

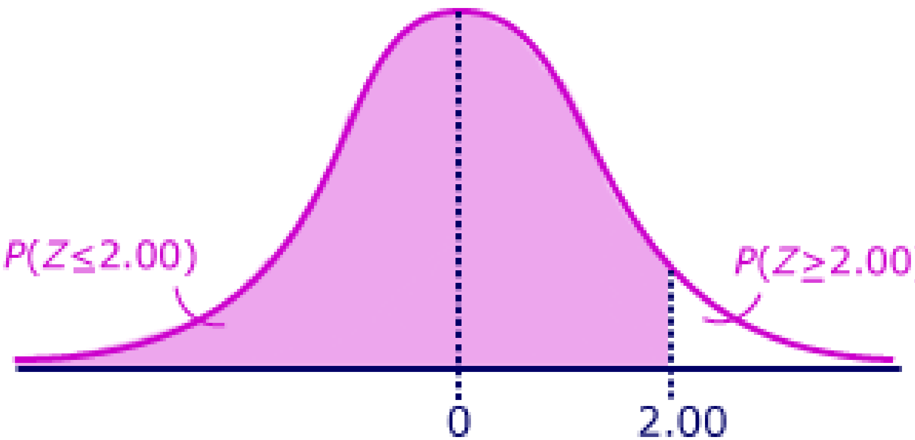
Using the Standard Normal

- Suppose *League of Legends* Champion released once every 24 days on average, standard deviation of 3 days
- What is the probability Champion released 30+ days?
- $x = 30$, $\bar{x} = 24$, $s = 3$

```
=norm.dist(x,mean,stddev,cumulative)
=norm.dist(30,24,3,false)
```

$$Z = (x - \bar{x}) / s$$
$$= (30 - 24) / 3$$
$$= 2$$

- Want to know $P(Z > 2)$



http://ci.columbia.edu/ci/premba_test/c0331/s6/s6_4.html

Empirical Rule. Or use table (Z-table)

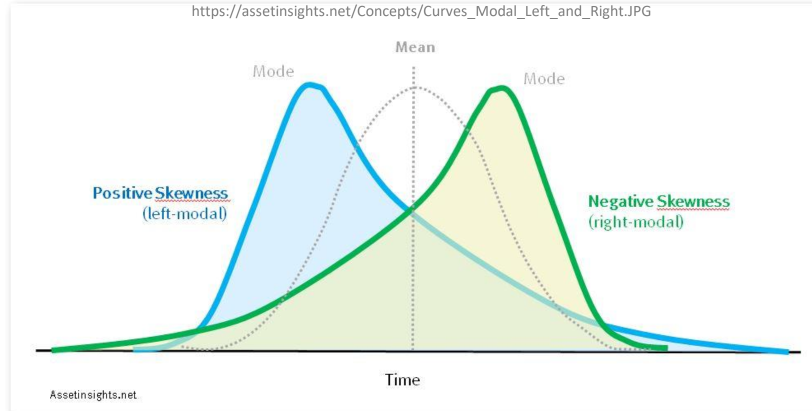
→ 5% / 2 = 2.5% likely (Z-table is 2.28%)

Test for Normality

- Why?
  - Can use Empirical Rule
  - Use some inferential statistics (parametric tests)
- How?
  1. Measure skewness (*next*)
  2. Looks normal
     - Histogram
     - Normal probability plot (QQ plot) – graphical technique to see if data set is approximately normally distributed
  3. Statistical test
     - Kolmogorov-Smirnov test (K-S) or Shapiro-Wilk (S-W) that compare to normal (won't do, but ideas in next slide deck)

Measuring Skewness

- Measure of symmetry of distribution
  - Normal is perfectly symmetric, skewness 0
- Easy equations:



$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

$$\frac{\frac{Q_3 + Q_1}{2} - Q_2}{\frac{Q_3 - Q_1}{2}}$$

=skew(A1:A10)

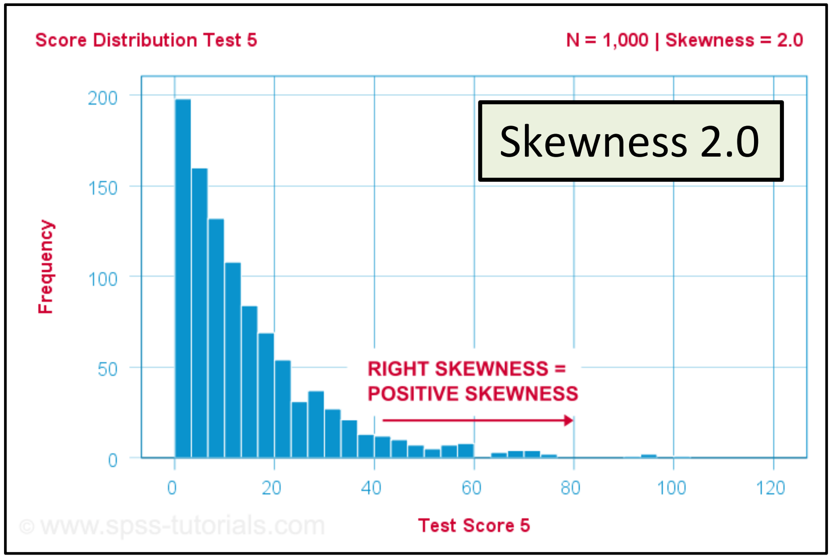
$$\frac{n}{(n-1)(n-2)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^3$$

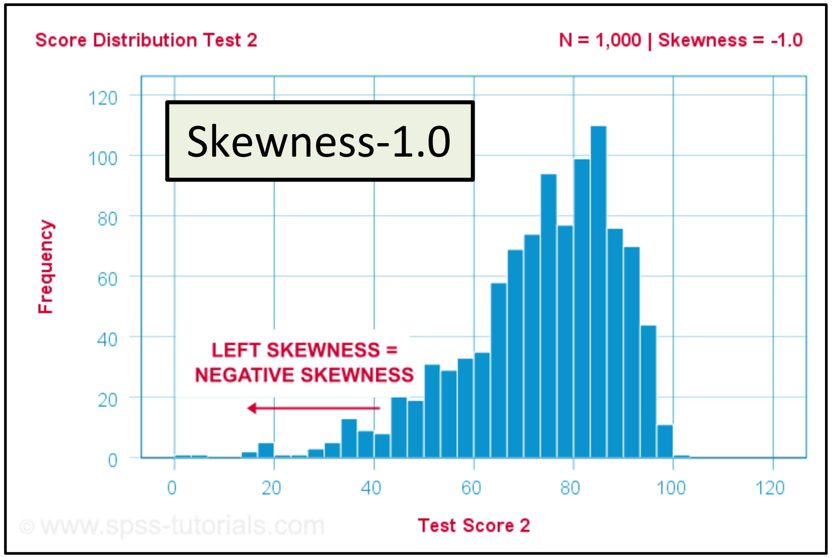
"Fisher–Pearson standardized moment"

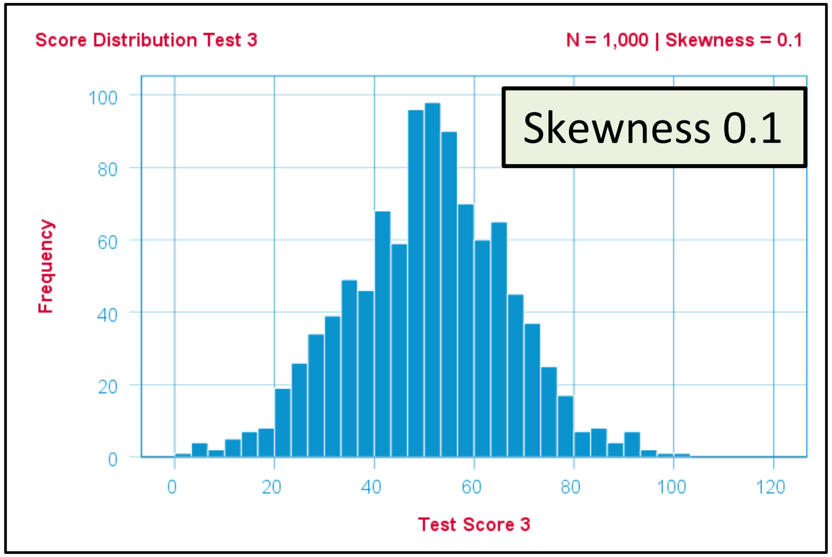
- "How much" is non-normal?
  - Somewhat arbitrary
  - Less than -1 or greater than +1
    - Highly skewed
  - Between [-1, -0.5] or [0.5, +1]
    - Moderately skewed
  - Between -0.5 and 0.5
    - Symmetric

[Note, related "Kurtosis" is how clumped]

Skewness Examples

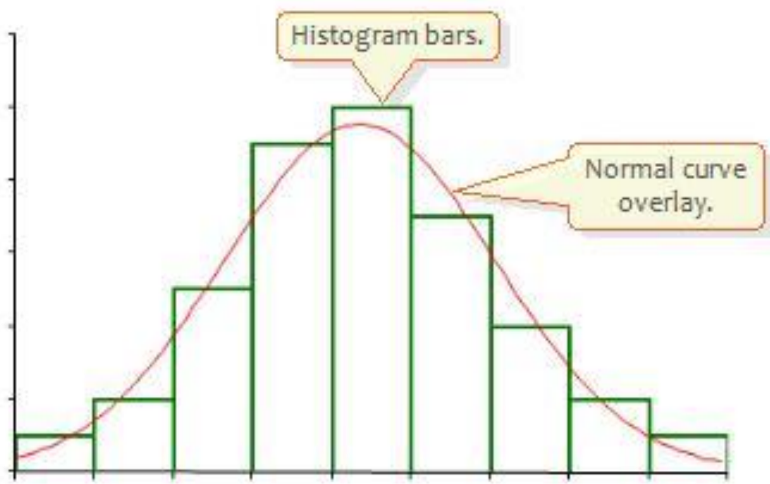




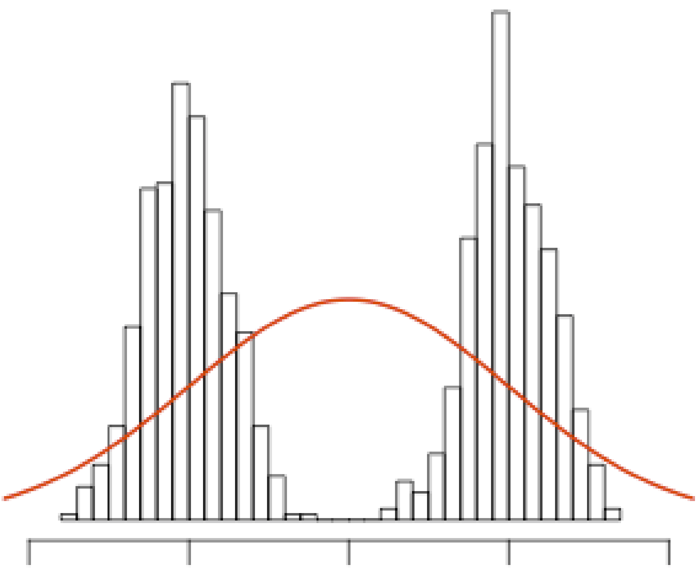


Normality Testing with a Histogram

- Use histogram shape to look for "bell curve"



http://2.bp.blogspot.com/_g8gh7I4zSt4/TR85eGJIMfI/AAAAAAAAAQs/PaOHJsjonPM/s1600/histo.JPG

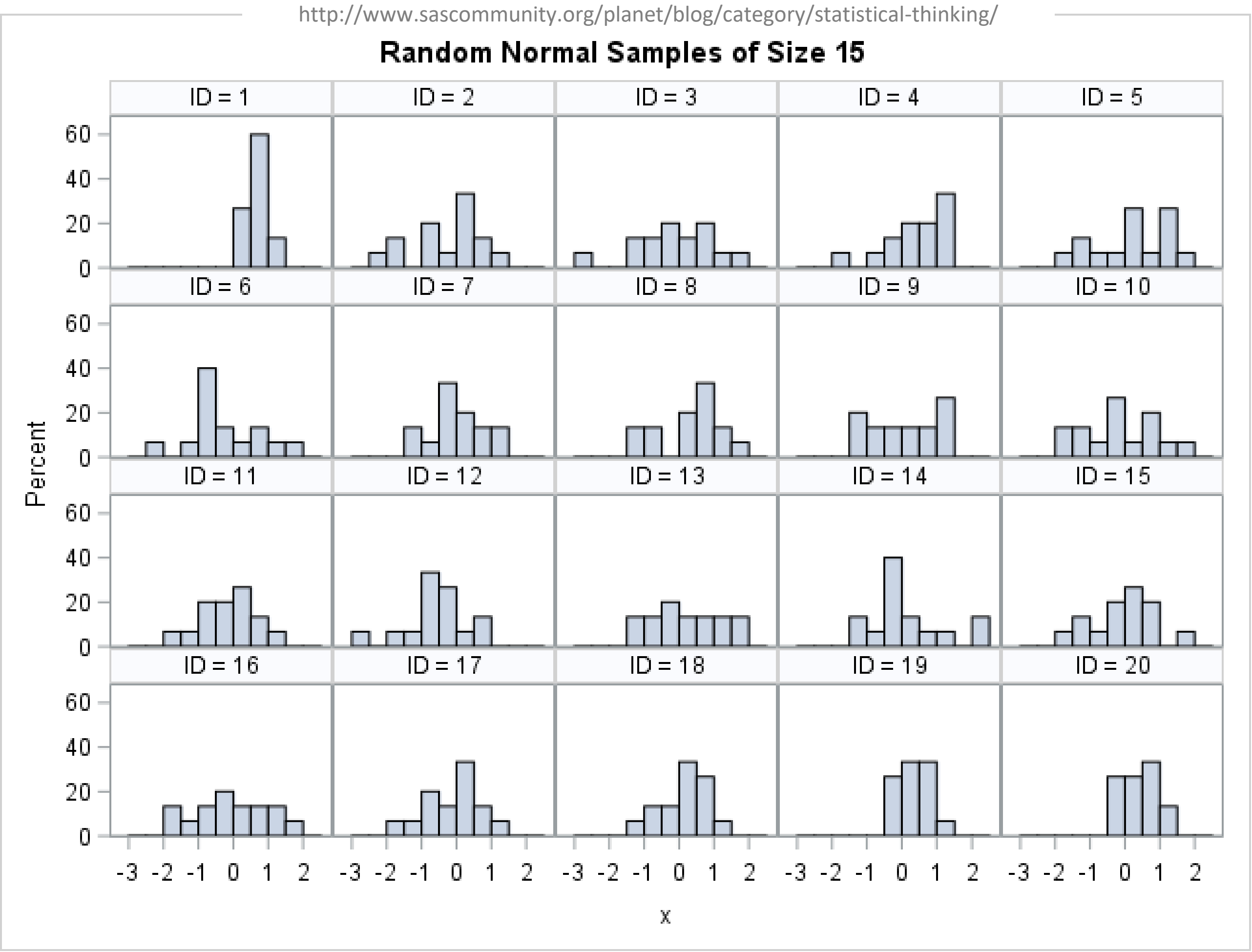


http://seankross.com/img/biqq.png

Yes

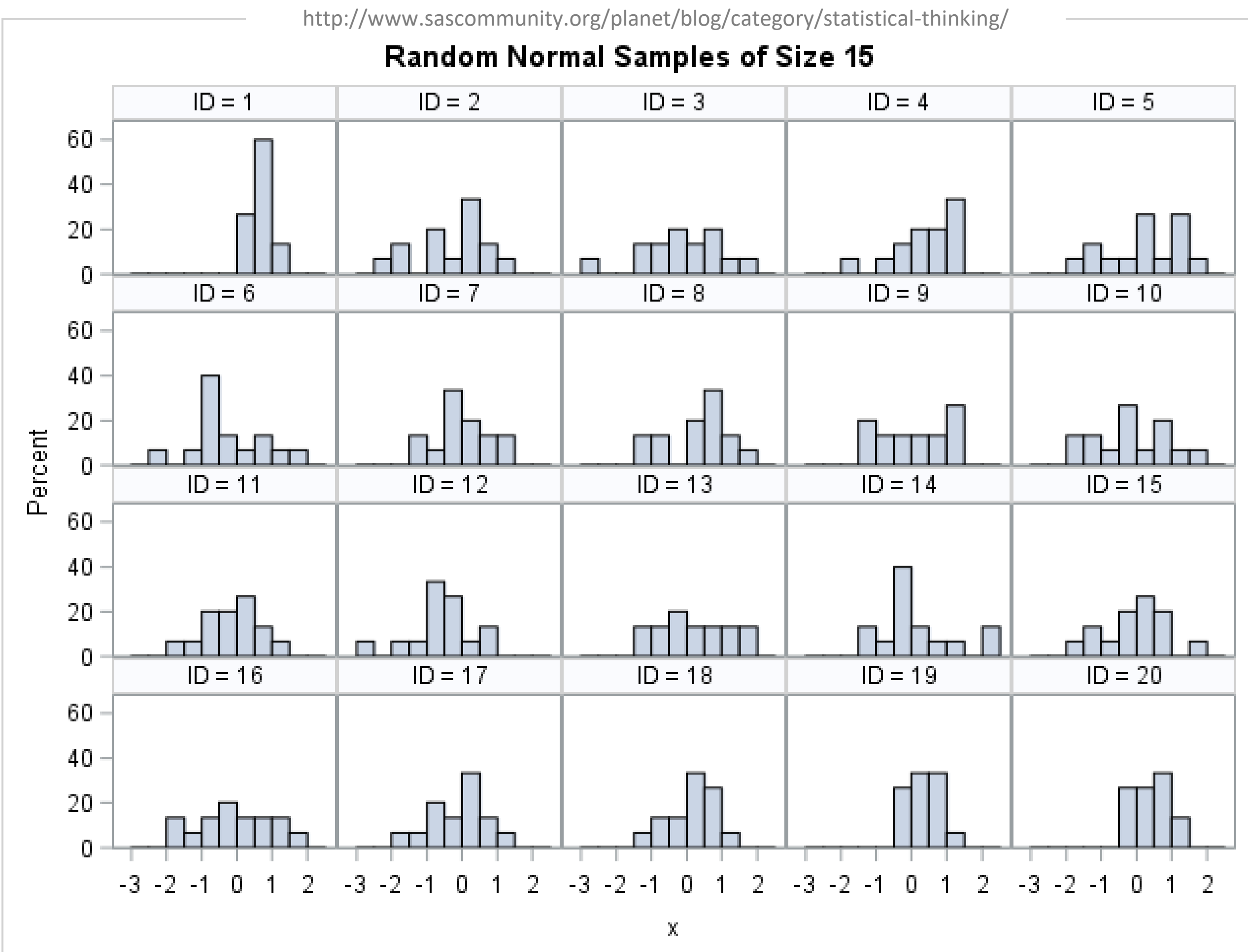
No

Normality Testing with a Histogram



Q: What distributions are these from? Any normal?

Normality Testing with a Histogram



They are *all* from normal distribution! Suffer from:

- Binning (not continuous)
- Few samples (15)

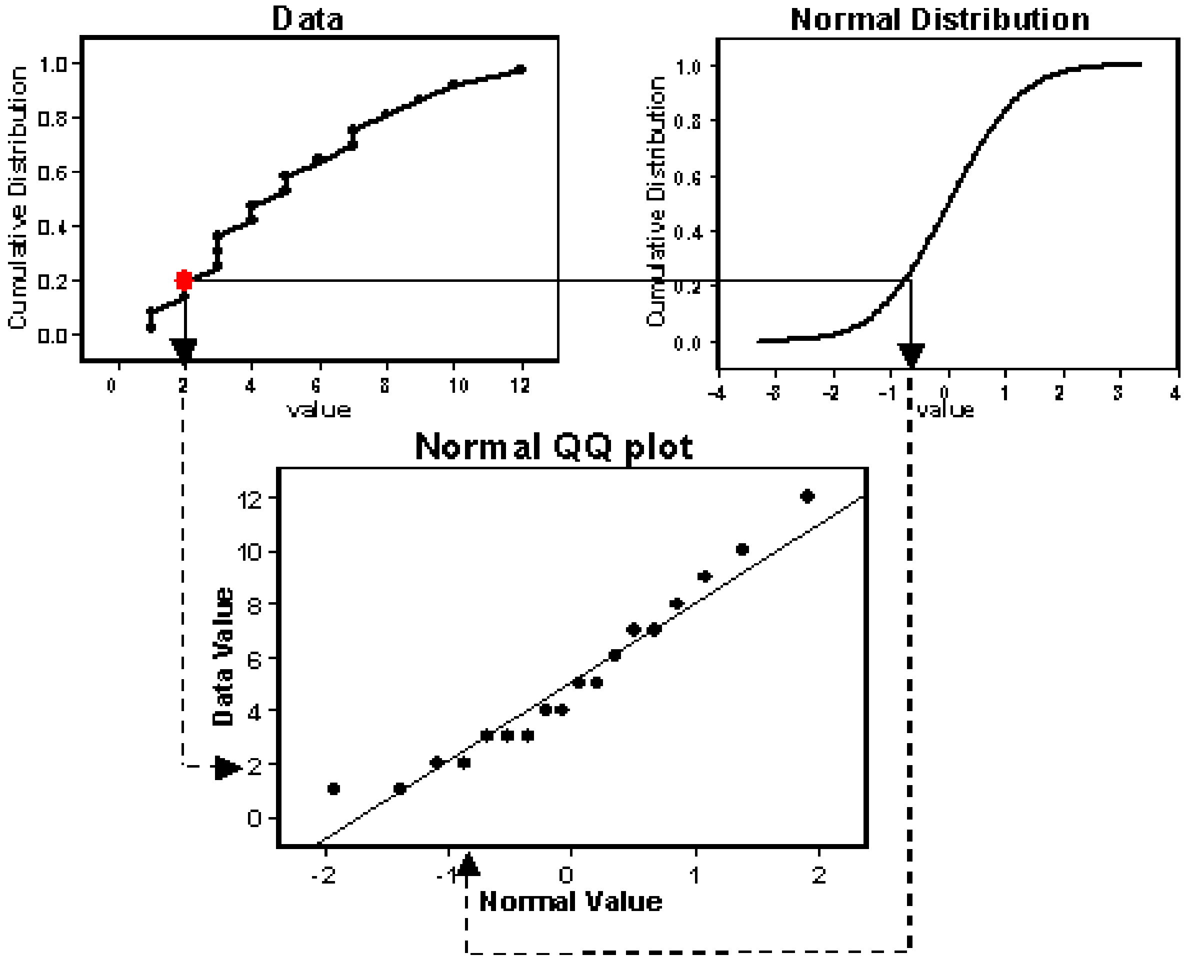
Normality Testing with a Quantile-Quantile Plot

- Percentiles (quantiles) of one versus another
- If line → same distribution

1. Order data
2. Compute Z scores (normal)
3. Plot data (y-axis) versus Z (x-axis)

- Normal? → line

https://goo.gl/rLLSIQ



Quantile-Quantile Plot Example

- Do the following values come from a normal distribution?

7.19, 6.31, 5.89, 4.5, 3.77, 4.25, 5.19, 5.79, 6.79

1. Order data
2. Compute Z scores
3. Plot data versus Z

Show each step, next

http://www.statisticshowto.com/q-q-plots/

Quantile-Quantile Plot Example – Order Data

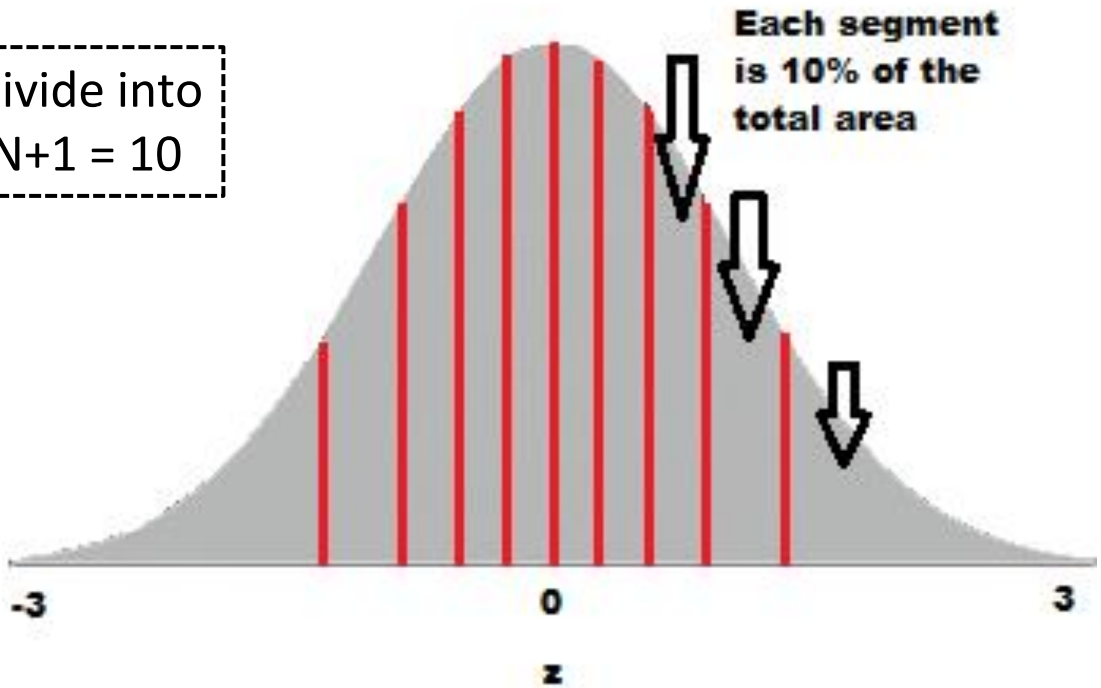
Unordered		Ordered (low to high)
7.19		3.77
6.31		4.25
5.89		4.50
4.50		5.19
3.77	→	5.89
4.25		5.79
5.19		6.31
5.79		6.79
6.79		7.19

N = 9 data points

http://www.statisticshowto.com/q-q-plots/

Quantile-Quantile Plot Example – Compute Z scores

Divide into N+1 = 10



10% = ?
20% = ?
30% = ?
40% = ?
50% = 0
60% = ?
70% = ?
80% = ?
90% = ?

Want Z-score for that segment

Lookup in Z-table

http://www.statisticshowto.com/q-q-plots/

Z-Table

- Tells what cumulative percentage of the standard normal curve is under any point (Z-score). Or, $P(-\infty \text{ to } Z)$

e.g., 80%?

Find closest value in table to desired percent

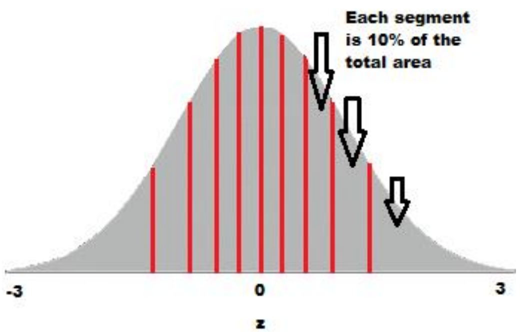
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

10% = -1.28
20% = -0.84
30% = -0.52
40% = -0.25
50% = 0
60% = 0.25
70% = 0.52
80% = **0.84**
90% = 1.28

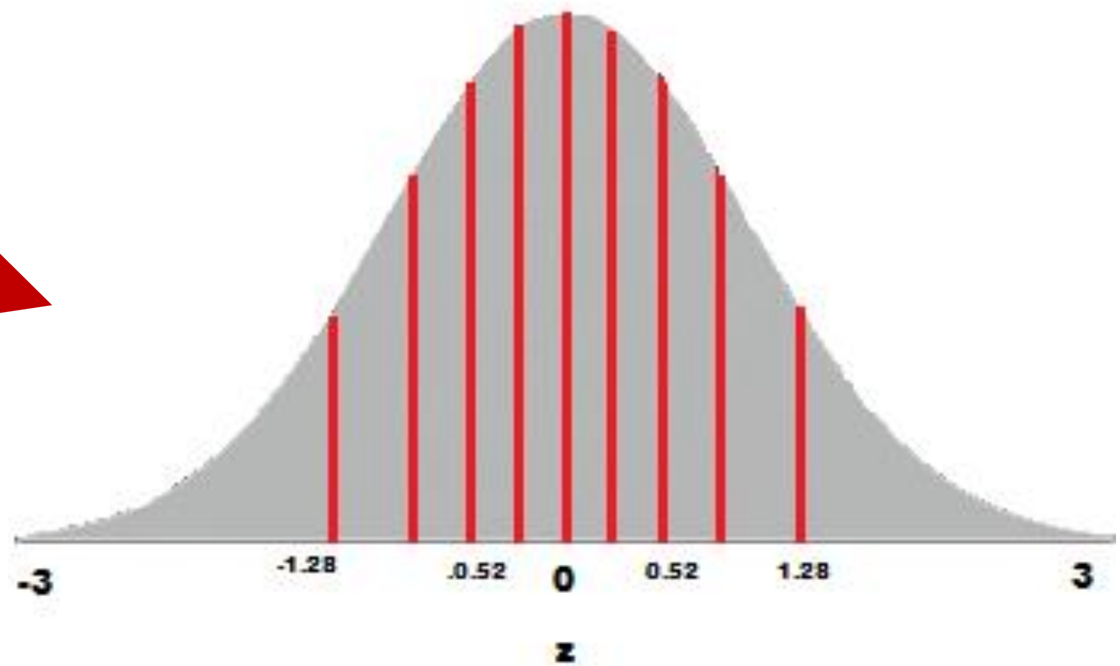
(Note: Above for positive Z-scores – also negative tables, or diff from 50%)

`=NORMSINV(area)` – provide Z for area under standard normal curve
=NORMSINV(.80)
=0.841621

Quantile-Quantile Plot Example – Compute Z scores

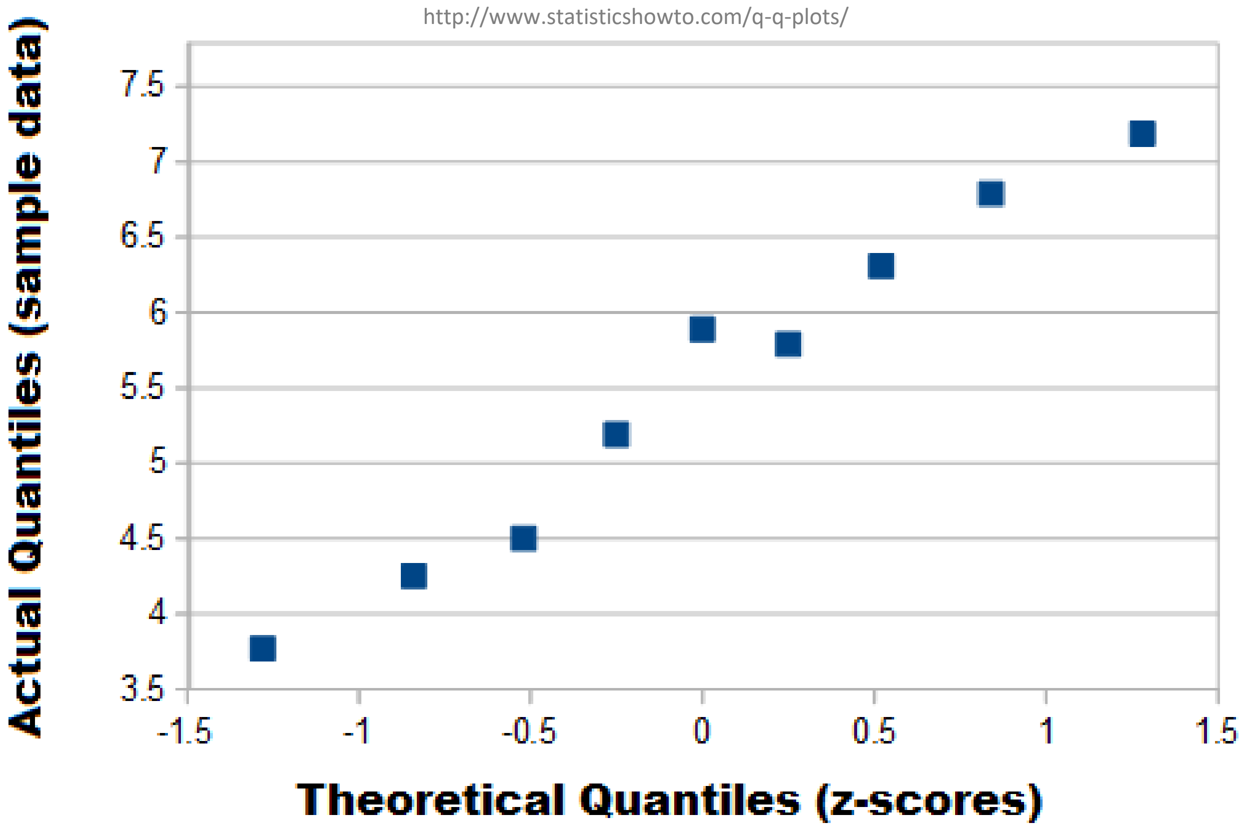


10% = -1.28
20% = -0.84
30% = -0.52
40% = -0.25
50% = 0
60% = 0.25
70% = 0.52
80% = 0.84
90% = 1.28



(Only some points shown)

Quantile-Quantile Plot Example – Plot



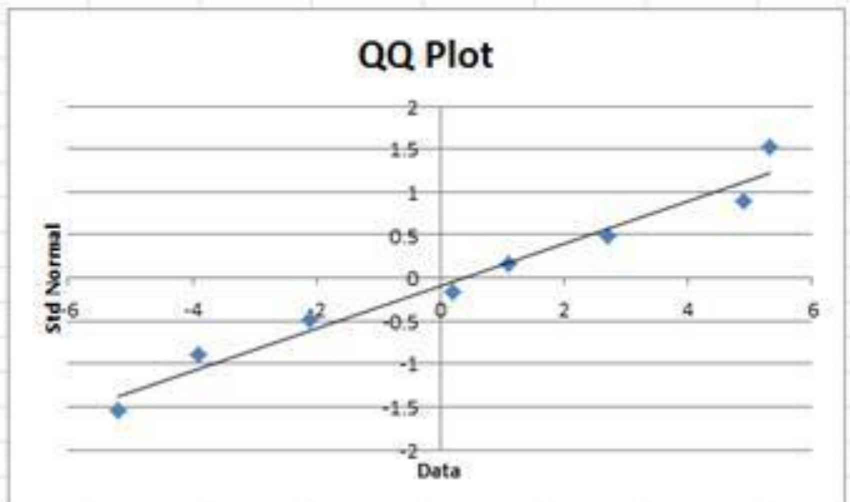
Linear? → Normal

Quantile-Quantile Plots in Excel

- Mostly, a manual process. Do as per above.
- Example of step by step process (with spreadsheet):

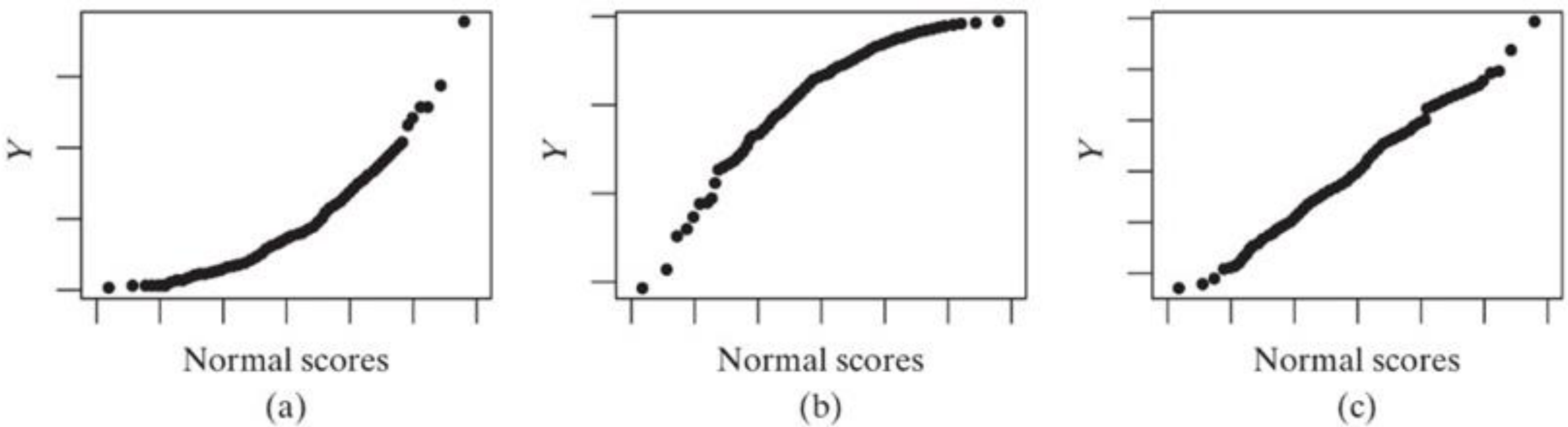
http://facweb.cs.depaul.edu/cmiller/it223/normQuant.html

	A	B	C	D	E	F
1	**QQ Plot**					
2						
3	Data		QQ Tables			
4	-5.2					
5	-3.9		Count	8	16	
6	-2.1		Mean	0.375		
7	0.2		Std Dev	3.894593		
8	1.1					
9	2.7		Interval	Data	Std Norm	Std Data
10	4.9		1	-5.2	-1.53412	-1.43147
11	5.3		3	-3.9	-0.88715	-1.09768
12			5	-2.1	-0.48878	-0.6355
13			7	0.2	-0.15731	-0.04493
14			9	1.1	0.157311	0.186156
15			11	2.7	0.488776	0.596981
16			13	4.9	0.887147	1.161867
17			15	5.3	1.534121	1.264574



https://i2.wp.com/www.real-statistics.com/wp-content/uploads/2012/12/qq-plot-normality.jpg

Examples of Normality Testing with a Quantile-Quantile Plot



http://d2vlcm61l7u1fs.cloudfront.net/media%2Fb95%2Fb953e7cd-31c3-45b0-a8ec-03b0e81c95d1%2Fphp2Y86od.png