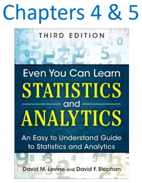
IMGD 2905

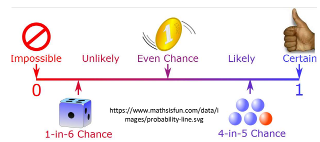
Probability

Chapters 4 & 5



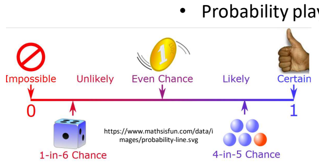
Overview

- Statistics important for game analysis
- Probability important for statistics
- So, understand some basic probability
- Also, probability useful for game development
- What are some examples of probabilities needed for game development?



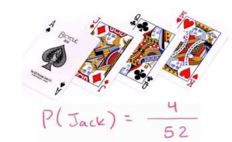
Overview

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



Probability Introduction

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



https://cdn.kastatic.org/googleusercontent/Z0TuLq2KolavsrfDXSbLqiD5-wnlCrC13cKGG68wK9ljrTiXzRqvfq7IpWNzcwgzlpEOl8YmMafp4K4zO0samvXu

- 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 into MMO and see if time between two arrivals is more than 15 seconds

Outline

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

Probability – Definitions

- Exhaustive set of events – set of all possible outcomes of experiment/observation
- Mutually exclusive sets of events – elementary events in each do not overlap
- 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 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 – Definition

- Probability – likelihood of event to occur, measured by ratio of favorable cases to unfavorable 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 1/6th to each
  → legal set of probabilities
- e.g., D6: events 1, 2, 3, 4, 5, 6. Probability of ½ to 1, ½ to 2, and 0 to all the others
  → Also legal set of probabilities
  - Not how honest d6's behave in real life!

So, how to assign probabilities?

Assigning Probabilities

- Classical (by theory)
  - In many 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 League of Legends in lab versus some other game. Say, 30% LoL. Assign that as probability
- Subjective (by hunch)
  - Based on expert opinion or other subjective method
  - e.g., e-sports writer says probability Team SoloMid (League team) will win World Championship is 25%

Rules About Probabilities (1 of 2)

- Complement: A an event, event "A does not occur" called *complement* of A, denoted A'
  P(A') = 1 - P(A)
  - e.g., d6: P(6) = 1/6, complement is P(6') and probability of not 6 is 1-1/6, or 5/6
  - Note: when using p, complement is often q
- Mutually exclusive: Have no simple outcomes in common – can't both occur in same experiment
  P(A or B) = P(A) + P(B)
  - e.g., d6: P(3 or 6) = P(3) + P(6) = 1/6 + 1/6 = 2/6

Rules About Probabilities (2 of 2)

- Independence: 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 the prob B given A happened
  - e.g., MMO has 10% mages, 40% warriors, 80% Boss defeated. Probability Boss fights mage and is defeated?
  - You might think that = P(mage) x P(defeat B) = .10 * .8 = .08
  - But likely not independent. P(defeat B | mage) < 80%. So, need not-independent formula P(mage)* P(defeat B | mage)

Probability Example

- Probability drawing King?

Probability Example

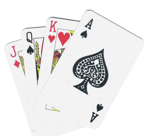
- Probability drawing King?
  P(K) = ¼
- Draw, put back. Now?

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) = ¼
- Draw, put back. Now?
  P(K) = ¼
- Probability *not* King?
  P(K') = 1-P(K) = ¾
- Draw, put back. 2 Kings?



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. 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. 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. 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
- Draw, put back. Not 2 Kings?

- 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. 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. 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. Not 2 Kings?
  P(K') x P(K') = ¾ x ¾ = 9/16
- Draw, *don't* put back. Not 2 Kings?

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. 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. Not 2 Kings?
  P(K') x P(K') = ¾ x ¾ = 9/16
- Draw, *don't* put back. Not 2 Kings?
  P(K') x P(K' | K') = ¾ x 2/3
  = 6/12 = ½
- Draw, don't put back. King 2nd card?

Probability Example

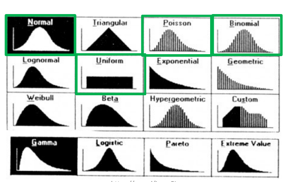
- Probability drawing King?
  P(K) = ¼
- Draw, put back. Now?
  P(K) = ¼
- Probability *not* King?
  P(K') = 1-P(K) = ¾
- Draw, put back. 2 Kings?
  P(K) x P(K) = ¼ x ¼ = 1/16
- Draw. King or Queen?
  P(K or Q) = P(K) + P(Q)
  = ¼ + ¼ = ½
- Draw, put back. Not 2 Kings?
  P(K') x P(K') = ¾ x ¾ = 9/16
- Draw, *don't* put back. Not 2 Kings?
  P(K') x P(K' | K') = ¾ x 2/3
  = 6/12 = ½
- Draw, don't put back. King 2nd card?
  P(K') x P(K | K') = ¾ x ⅓ = 3/12 = ¼

Outline

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

Probability Distributions

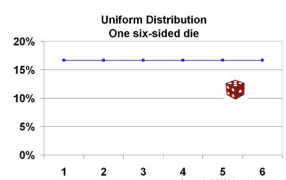
- Probability distribution – values and likelihood of those values that random variable can take
- Why? If can model mathematically, can use to predict occurrences
  - e.g., probability slot machine pays out on given day
  - e.g., probability game server hosts player today
  - e.g., probability certain 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)

Uniform Distribution



Mean = (1 + 6) / 2 = 3.5
Variance = $((6 - 1 + 1)^2 - 1)/12$
= 2.9
Std Dev = sqrt(Variance) = 1.7

- "So what?"
- Can use known formulas

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

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) = ?

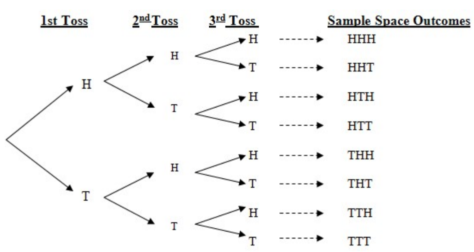
Binomial Distribution Example (1 of 3)

- 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)

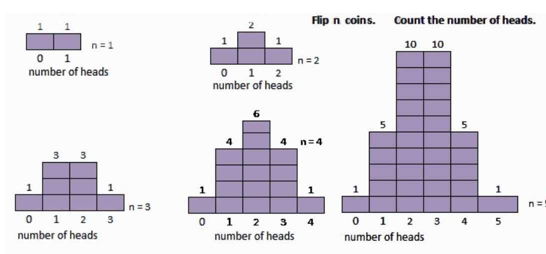
Binomial Distribution Example (2 of 3)



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

Can draw histogram of number of heads

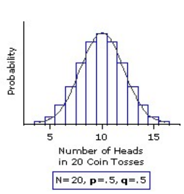
Binomial Distribution Example (3 of 3)



These are all binomial distributions

Binomial Distribution (1 of 2)

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



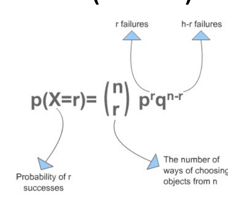
- Characteristics of experiment that gives random number with binomial distribution:
  - Experiment consists of n identical trials.
  - Each trial results in only two possible outcomes, S or F
  - Probability of S each trial is same, denoted p
  - Trials are independent
  - Random variable of interest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

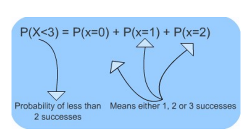
Binomial Distribution (2 of 2)

- "So what?"
- Can use known formulas

$$MEAN: \ \mu = np$$
$$Variance: \ \sigma^2 = npq$$
$$SD: \ \sigma = \sqrt{npq}$$

Excel: `binom.dist()`
– 2 heads, 3 flips
`=binom.dist(2,3,0.5,FALSE)`
`=0.375` (i.e., 3/8)





Poisson Distribution

- Distribution of probability of 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 units
  2. Number of events in one unit independent of number of events in any other unit
  3. Events occur singly (not simultaneously). In other words, as unit gets smaller, probability of two events occurring approaches 0

Poisson Distributions?

Not Poisson

- Number of people arriving at restaurant during dinner hour
  - People frequently arrive in groups
- Number of students register for course in BannerWeb 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

Probably Poisson

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

Phrase people use is "random arrivals"

Poisson Distribution

- Distribution of probability of 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

- Number of games student plays per day averages one per day
- Number of games played per day independent of all other days
- Can only play 1 game at a time
- What's probability of playing two games next day?
- In this case, the value of λ = 1

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

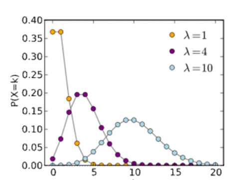
Poisson Distribution

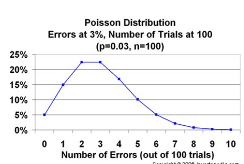
- "So what?" → Known formulas

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

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

Excel: `poisson.dist()`
1 game per day, chance for 2
`= poisson.dist(2,1,false)`
`= 0.18394`





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)

What continuous distribution is **especially** important?

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)

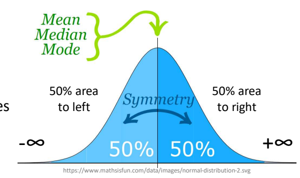
→ How to use to calculate probabilities?

- 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)

What continuous distribution is **especially** important? → The Normal Distribution

Normal Distribution (1 of 2)

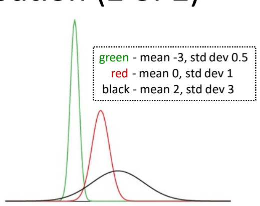
- "Bell-shaped" or "Bell-curve"
  - Distribution from -∞ to +∞
- Symmetric
- Mean, median, mode all same
  - Mean determines location, standard deviation determines "width"
- Super important!
  - Lots of distributions follow normal ("bell curve")
  - Basis for inferential statistics (e.g., statistical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



Aka "Gaussian" distribution

Normal Distribution (2 of 2)

- *Many* normal distributions
- However, "the" normal distribution refers to standard normal
  - Subtract mean (μ)
  - Divide by standard deviation (σ)

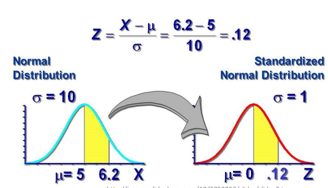


$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

```
=norm.dist()
```

Standard Normal Distribution

- Standardize
  - Subtract mean
  - Divide by standard deviation
- Mean μ = 0
- Standard Deviation σ = 1
- Total area under curve = 1
  - Sounds like probability!



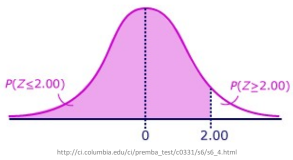
Use to predict how likely an observed sample is given a population mean

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?
- $x = 30$, $\mu = 24$, $\sigma = 3$

$Z = (x - \mu) / \sigma$
$= (30 - 24) / 3$
$= 2$

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



Use table (Z-table). Or Empirical Rule?

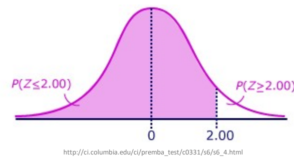
Using the Standard Normal

- 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?
- $x = 30$, $\mu = 24$, $\sigma = 3$

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

$Z = (x - \mu) / \sigma$
$= (30 - 24) / 3$
$= 2$

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



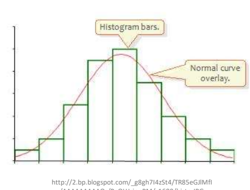
Use table (Z-table). Or Empirical Rule?
→ 5% / 2 = 2.5% likely (actual is 2.28%)

Test for Normality

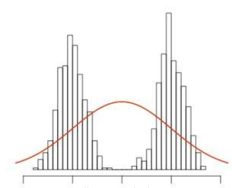
- Why?
  - Use some inferential statistics (parametric tests)
  - Can use Empirical Rule
- How? Several ways. One:
  - Normal probability plot – graphical technique to see if data set is approximately normally distributed

Normality Testing with a Histogram

- Use histogram shape to look for "bell curve"

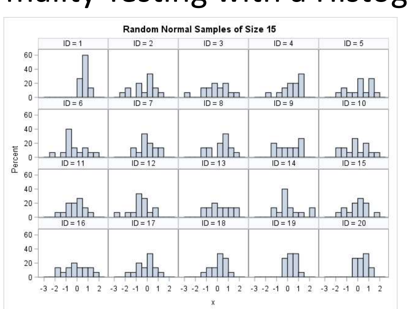


Yes



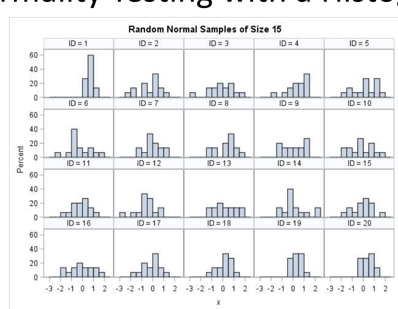
No

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:

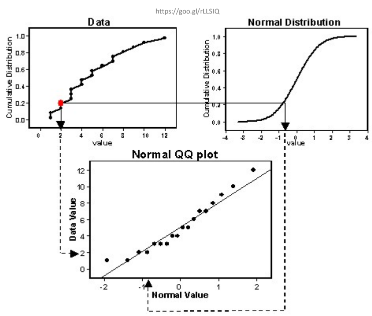
- Binning (not continuous)
- Few samples (15)

Normality Testing with a Quantile-Quantile Plot

- 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



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

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

Quantile-Quantile Plot Example – Order Data

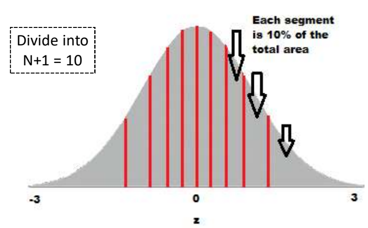
Unordered	Ordered (low to high)
7.19	3.77
6.31	4.25
5.89	4.50
4.50	5.19
3.77	5.89
4.25	5.79
5.19	6.31
5.79	6.79
6.79	7.19

N = 9 data points

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

Quantile-Quantile Plot Example – Compute Z scores

Divide into N+1 = 10



Want Z-score for that segment

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

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(-∞ to Z)

e.g., 80%?

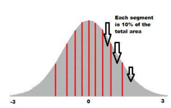
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.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

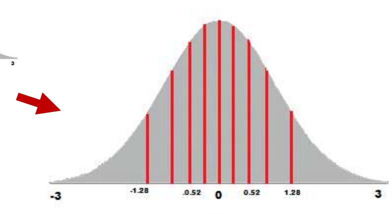
(Note: Above for positive Z-scores – also negative tables, or subtract 0.5)

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

Quantile-Quantile Plot Example – Compute Z scores

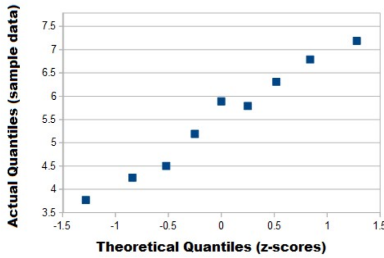


10% = -1.28
20% = -0.84
30% = -0.52
40% = -0.25
50% = 0
60% = 0.25
70% = 0.52
80% = 0.84
90% = 1.28



(Only some points shown)

Quantile-Quantile Plot Example – Plot



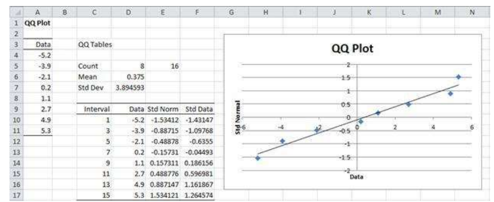
Linear? → Normal

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

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