Probability and Z-tests

Rafael Garcia

PROBABILITY BASICS

What are they

- Proportion of times/chance that an outcome ought to occur given many, many, many observations

- Denoted using P(A) or simply p
  – Probability of event/set/outcome A
  – (# of an occurance)/(total possible occurances)
    - f/N   or   f/n

Set & Sample Space

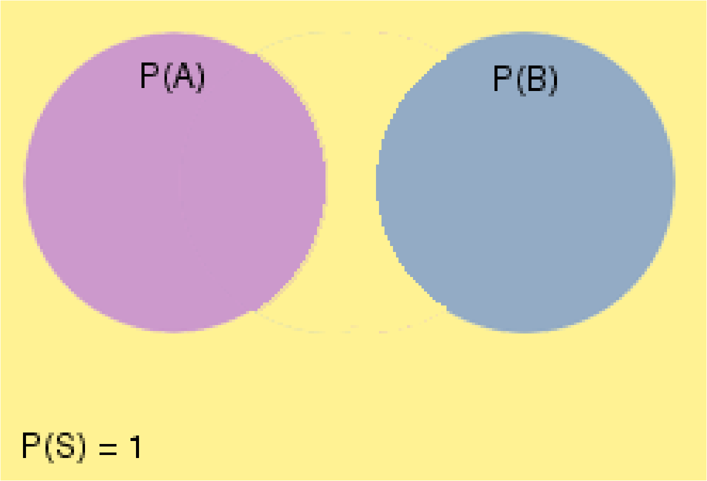
- Ex: If I'm interested in P(Even) on the roll of a single die…
  - Set: Even = {2, 4, 6}
  - Sample Space: {1, 2, 3, 4, 5, 6}

  - P(Even) = 3/6 = 0.50 or 50%

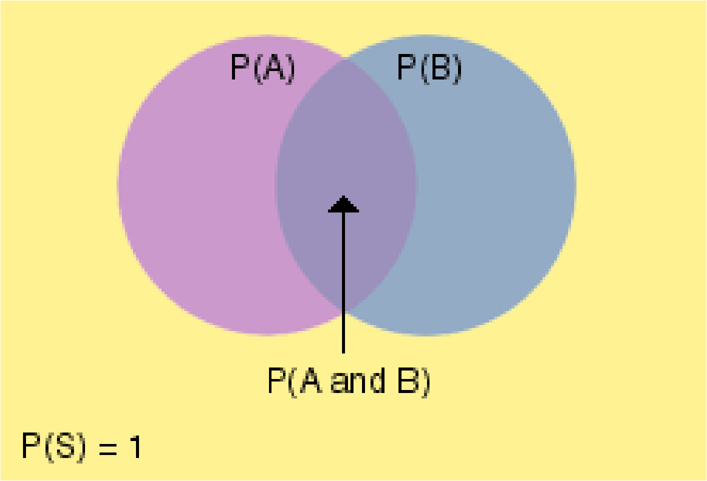
Properties

- $P(A^C)$ [not A]
  - $P(A) = 1 - P(A^C)$
    - Ex: $P(N) = 0.8$ $\underline{P(Y)} = 1 - 0.8 = \underline{0.2}$
- $P(A \cup B)$ [OR... mutually exclusive]
  - $P(A \cup B) = P(A) + P(B)$
    - Ex: $P(A) = 0.2$; $P(B) = 0.3$ $P(A \cup B) = 0.2+0.3 = 0.5$
- $P(A \cap B)$ [AND... assuming independence]
  - $P(A \cap B) = P(A) * P(B)$
    - Ex: $P(A) = 0.2$; $P(B) = 0.3$ $P(A \cap B) = 0.2*0.3 = 0.06$

Mutually Exclusive



Not M.E.



Independence

- Independence:
  - Event A does not affect the likelihood of Event B
    - Ex: Coin flip, rolling dice, drawing (with replacement)
  - By definition, two events are independent **if and only if** $P(A \cap B) = P(A) * P(B)$
  - If $P(A \cap B) \neq P(A) * P(B)$, then two events are dependent
    - Ex: drawing (without replacement), events with a common cause

PROBABILITY DISTRIBUTIONS

Probability Distribution

- **Probability Distribution**: Lists the possible outcomes and their probabilities
  - We list each possible response, and then assign a probability to its occurrence

Example

- Please select whether you agree (1) or disagree (0) for the following questions:
  - Chocolate is the best flavor of ice cream.

    Disagree Agree
  - Cats are better pets than dogs.

    Disagree Agree
  - Pepsi tastes better than Coke.

    Disagree Agree
  - Mexican food is better than Italian.

    Disagree Agree

Example

- Possible combination of responses ($2^4 = 16$)

(1,1,1,1)	(1,1,1,0)	(1,1,0,1)	(1,0,1,1)
(0,1,1,1)	(1,1,0,0)	(1,0,1,0)	(1,0,0,1)
(0,1,1,0)	(0,1,0,1)	(0,0,1,1)	(1,0,0,0)
(0,1,0,0)	(0,0,1,0)	(0,0,0,1)	(0,0,0,0)

Proportion of yeses (1) = $\dfrac{\text{Number of yeses}}{\text{Number of questions}}$

Example

(1,1,1,1)	(1,1,1,0)	(1,1,0,1)	(1,0,1,1)
(0,1,1,1)	(1,1,0,0)	(1,0,1,0)	(1,0,0,1)
(0,1,1,0)	(0,1,0,1)	(0,0,1,1)	(1,0,0,0)
(0,1,0,0)	(0,0,1,0)	(0,0,0,1)	**(0,0,0,0)**

Saying YES (1) to NO questions: Proportion = 0/4 = 0.00

Occurs 1 out of 16 times
Probability of a completely YES response is 1/16 or 0.0625

Example

(1,1,1,1)	(1,1,1,0)	(1,1,0,1)	(1,0,1,1)
(0,1,1,1)	(1,1,0,0)	(1,0,1,0)	(1,0,0,1)
(0,1,1,0)	(0,1,0,1)	(0,0,1,1)	**(1,0,0,0)**
(0,1,0,0)	**(0,0,1,0)**	**(0,0,0,1)**	(0,0,0,0)

Saying YES (1) to ONE question: Proportion = 1/4 = 0.25

Occurs 4 out of 16 times

Probability of a completely YES response is 4/16 or 0.25

Example

(1,1,1,1)	(1,1,1,0)	(1,1,0,1)	(1,0,1,1)
(0,1,1,1)	**(1,1,0,0)**	**(1,0,1,0)**	**(1,0,0,1)**
(0,1,1,0)	**(0,1,0,1)**	**(0,0,1,1)**	(1,0,0,0)
(0,1,0,0)	(0,0,1,0)	(0,0,0,1)	(0,0,0,0)

Saying YES (1) to TWO questions: Proportion = 2/4 = 0.50

Occurs 6 out of 16 times
Probability of a completely YES response is 6/16 or 0.375

Example

(1,1,1,1)	**(1,1,1,0)**	**(1,1,0,1)**	**(1,0,1,1)**
(0,1,1,1)	(1,1,0,0)	(1,0,1,0)	(1,0,0,1)
(0,1,1,0)	(0,1,0,1)	(0,0,1,1)	(1,0,0,0)
(0,1,0,0)	(0,0,1,0)	(0,0,0,1)	(0,0,0,0)

Saying YES (1) to THREE questions: Proportion = 3/4 = 0.75

Occurs 4 out of 16 times

Probability of a completely YES response is 4/16 or 0.25

Example

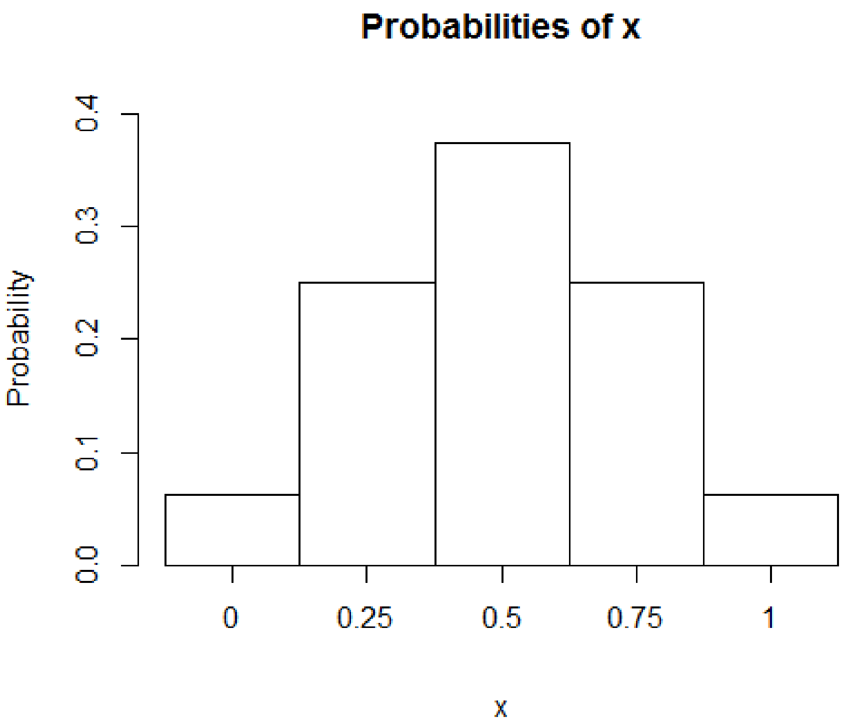
(1,1,1,1)	(1,1,1,0)	(1,1,0,1)	(1,0,1,1)
(0,1,1,1)	(1,1,0,0)	(1,0,1,0)	(1,0,0,1)
(0,1,1,0)	(0,1,0,1)	(0,0,1,1)	(1,0,0,0)
(0,1,0,0)	(0,0,1,0)	(0,0,0,1)	(0,0,0,0)

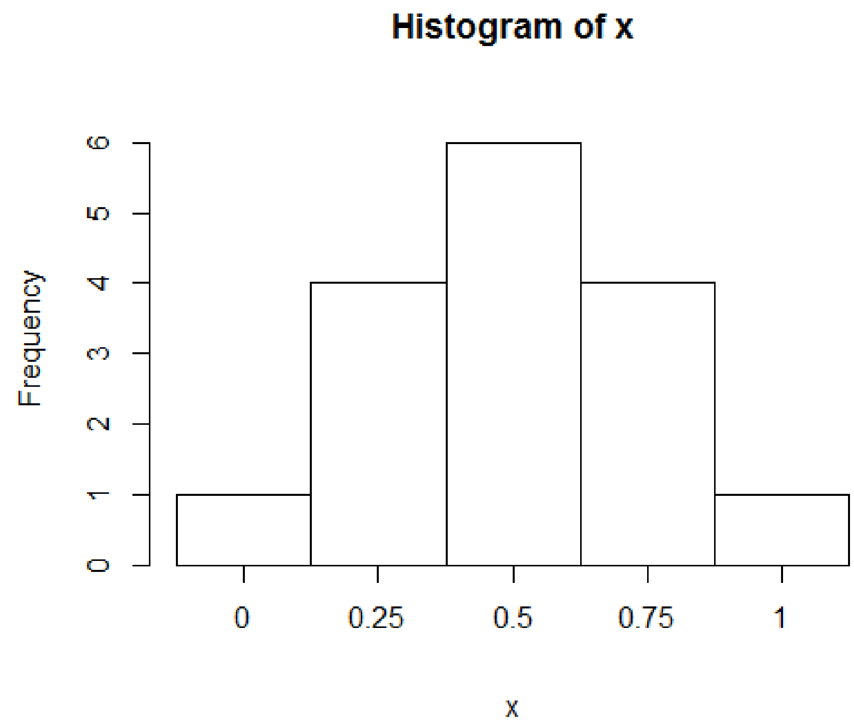
Saying YES (1) to ALL FOUR questions: Proportion = 4/4 = 1.00

Occurs 1 out of 16 times

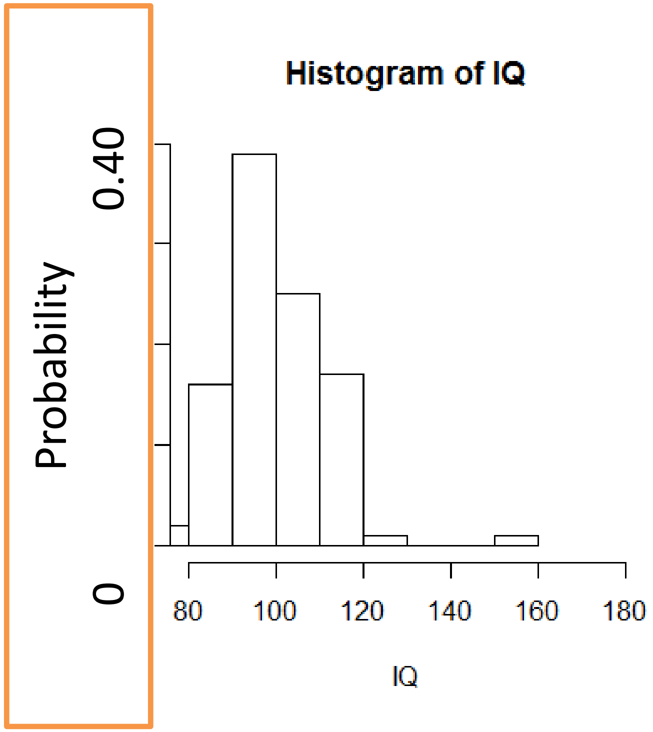
Probability of a completely YES response is 1/16 or 0.0625

Probability vs. Frequency distribution





X	f	cf	p	cp	C%
150-160	1	101	0.01	1.00	100
140-150	0	100	0.00	0.99	99
130-140	0	100	0.00	0.99	99
120-130	1	100	0.01	0.99	99
110-120	17	99	0.17	0.98	98
100-110	25	82	0.25	0.81	81
90-100	39	57	0.39	0.56	56
80-90	16	18	0.16	0.18	18
70-80	2	2	0.02	0.02	2

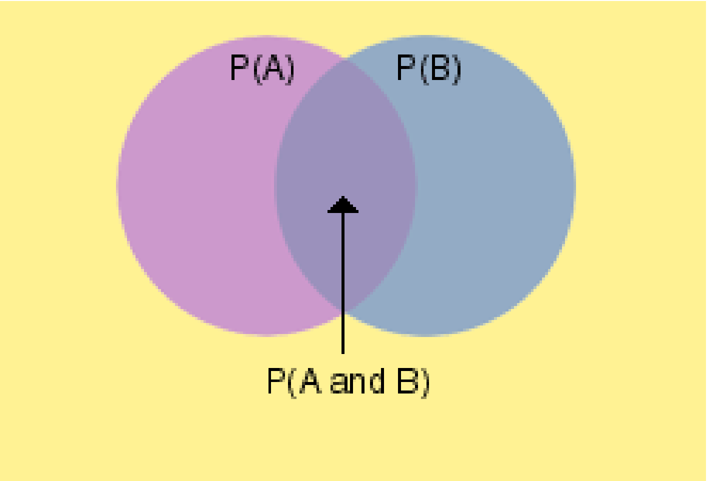


CONDITIONAL PROBABILITIES

- $P(A|B) = P(A \cap B)/P(B)$
  - The probability of A given the probability of B

- Ex: What is the probability of drawing a '9' given that you've drawn a 'red card'?
  - $P(N|R) = P(N \cap R)/P(R)$
    - $P(N \cap R) = 4/52 * 26/52 = 2/52$
    - $P(R) = 26/52$
    - So… 2/26 or 2 '9s' out of 26 red cards

- Ex2: A teacher gives two tests. 25% passed both tests and 50% passed the first test. What % who passed the first test also passed the second test?
  - $P(S|F) = P(S \cap F)/P(F) = 0.25/0.50 = 0.50 = 50\%$

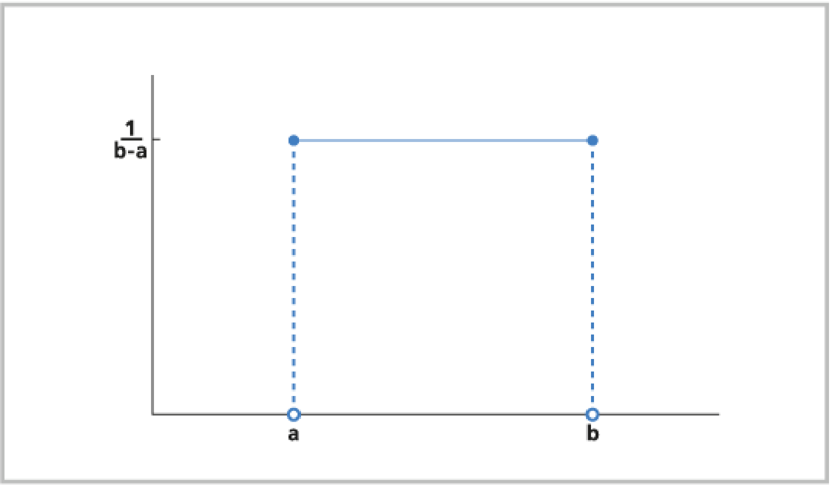


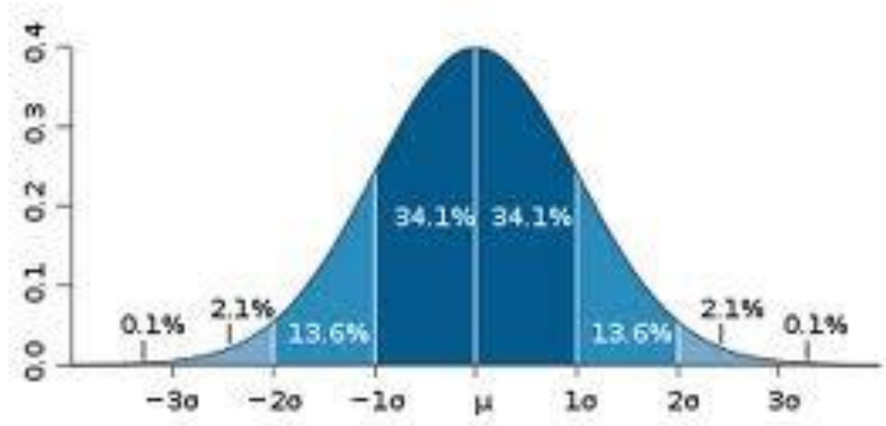
MISC. PROBABILITY STUFF

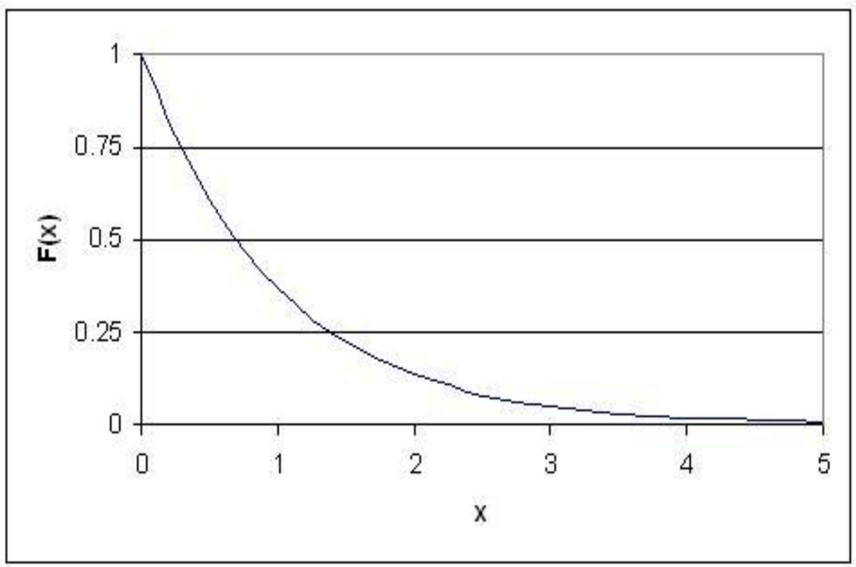
Sure things and Impossibilities

- The sum of all possible outcomes
  - $P(A) + P(B) + P(C) + \ldots + P(Z) = 1.0$
- When an event is guaranteed to happen
  - $P(A) = 1.0$
- When an event is impossible
  - $P(A) = 0.0$

Types of probability distributions

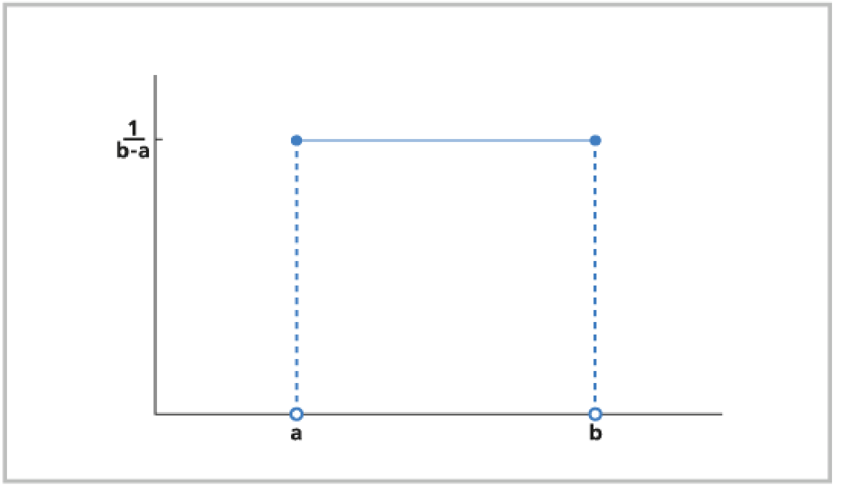






Randomness

- Random :
  - there is an equal probability that any possible outcomes will happen.
  - Ex: Rolling dice is random, flipping a coin is random
    - What decision someone will make is usually NOT random



Random number generators

- How do we intentionally select a random number?
  - We don't really…

- The closest thing we have
  - www.random.org

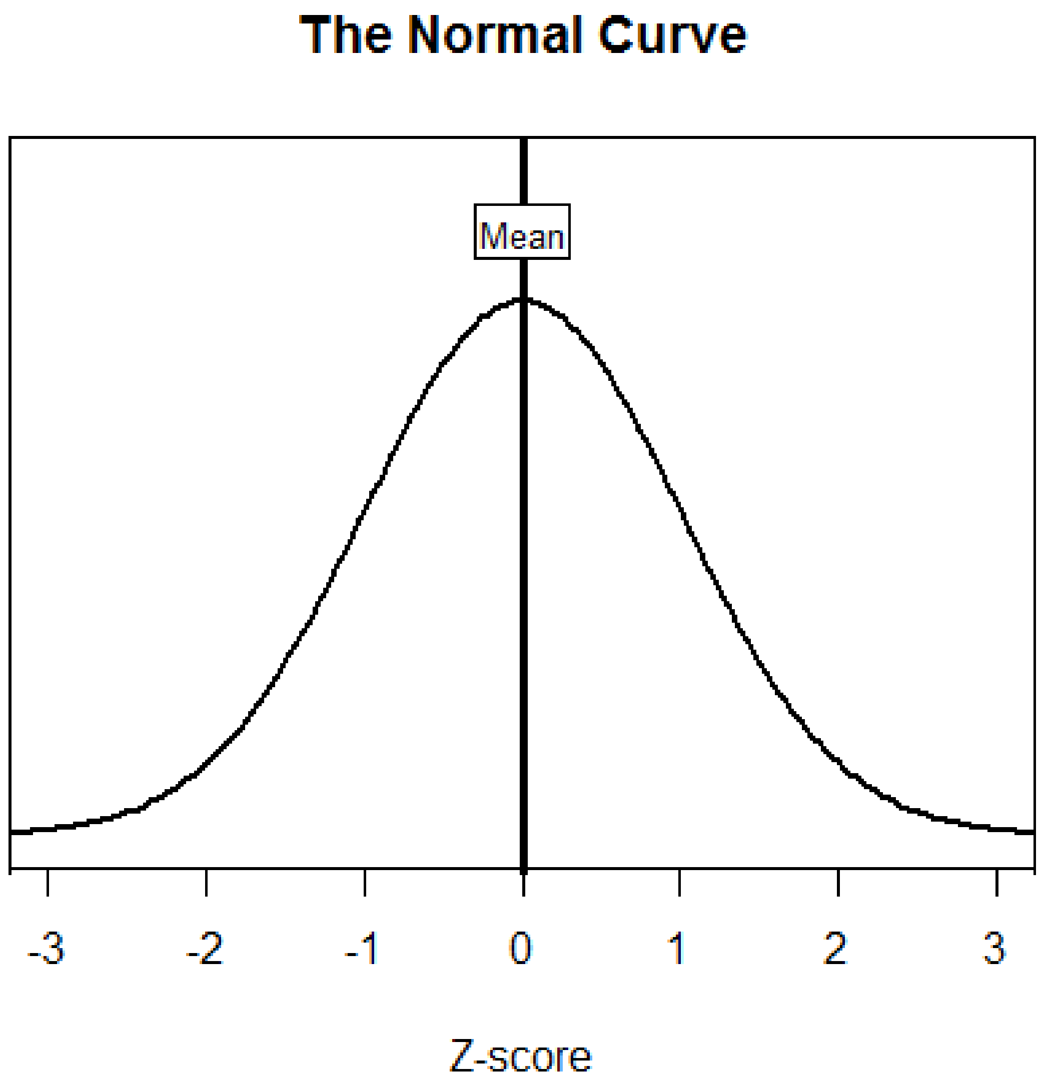
Additional Resources

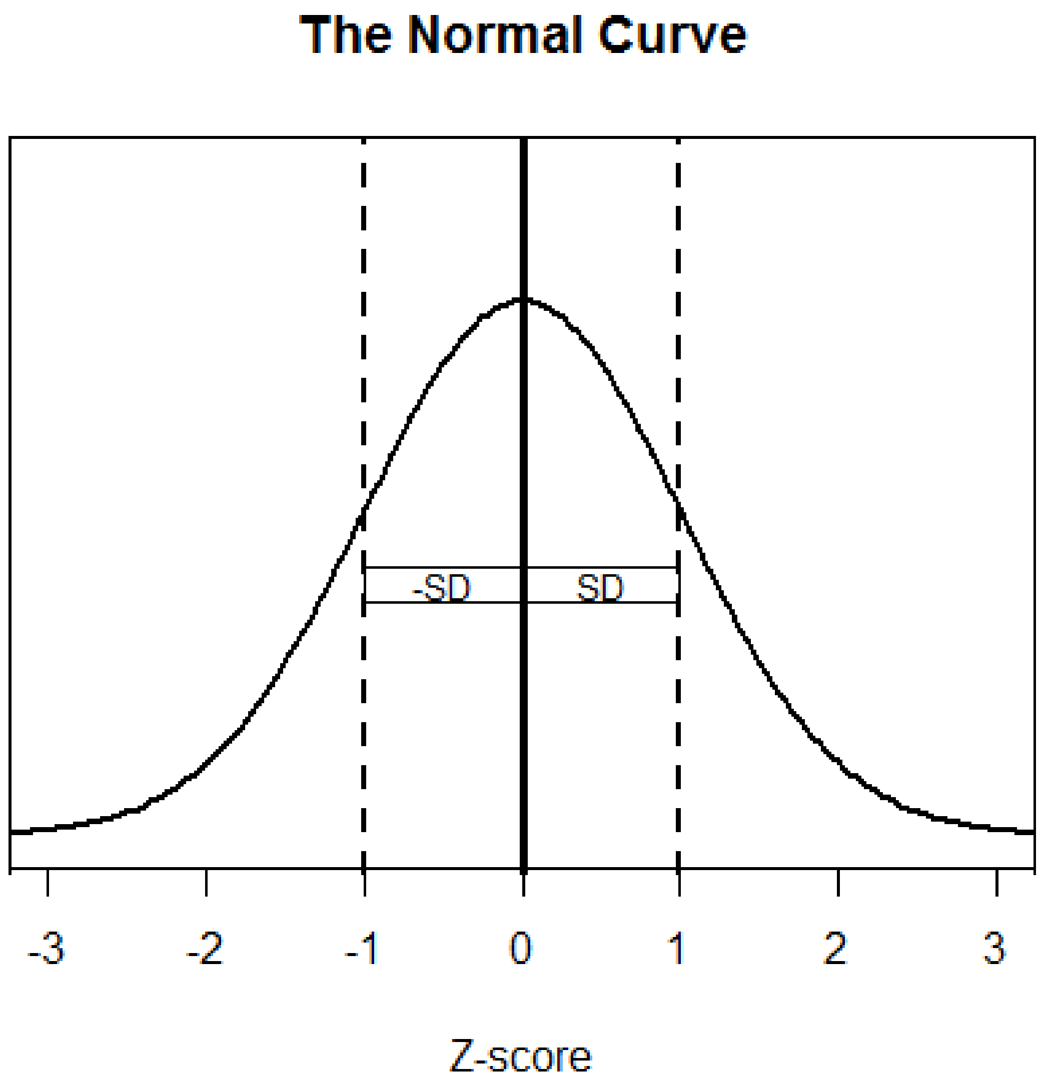
- Decent Tutorial: http://www.mathgoodies.com/lessons/toc_vol6.html
    - Practice Problems: http://www.mathgoodies.com/lessons/vol6/practice_unit6.html
    - http://www.mathgoodies.com/lessons/vol6/challenge_vol6.html
    - http://www.mathgoodies.com/worksheets/probability_wks.html
- A list of relevant notation and symbols
  – http://www.rapidtables.com/math/probability/basic_probability.htm

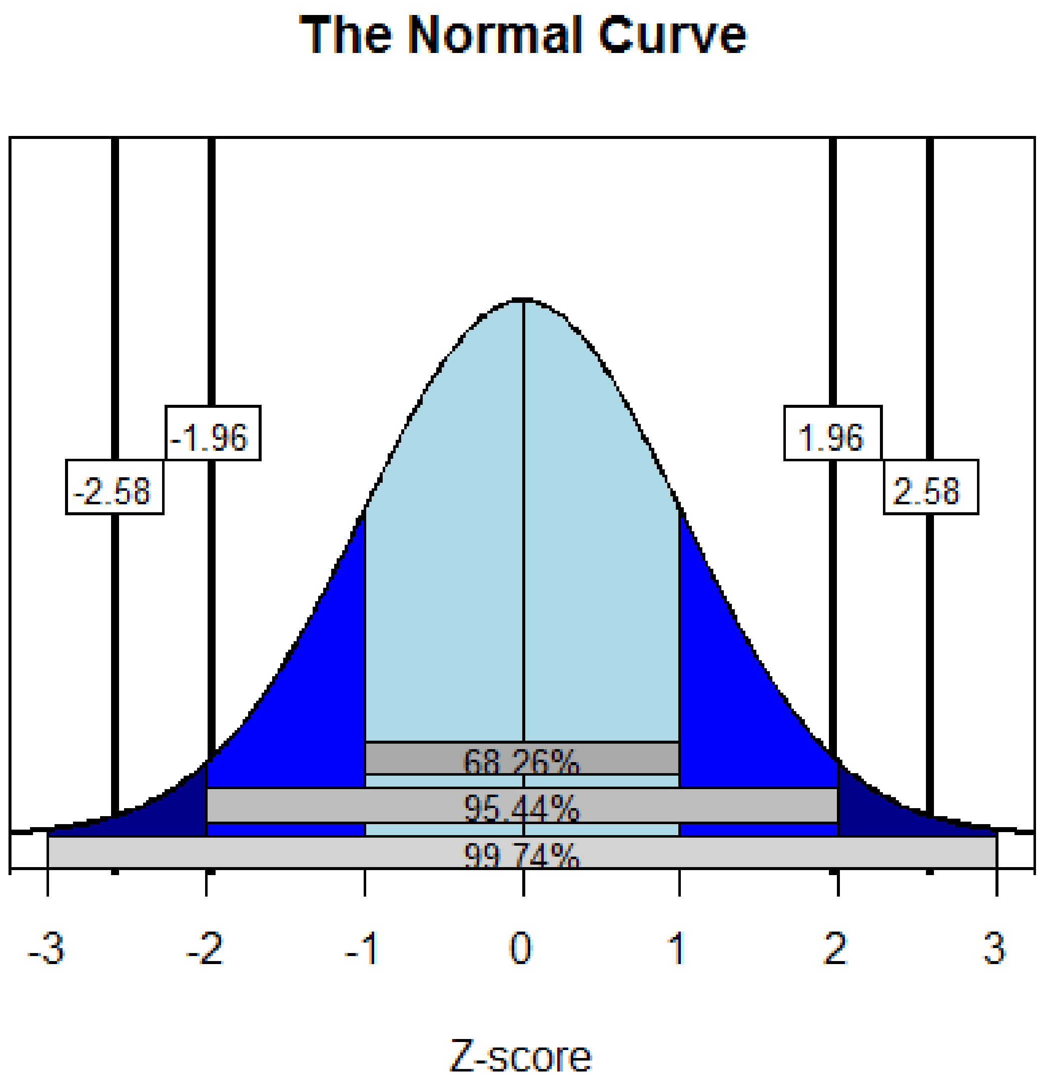
THE NORMAL DISTRIBUTION

- Distribution of Error

- We know the properties of the Normal Distribution (we made it)


The Normal Curve
Mean
-3
-2
-1
0
1
2
3
Z-score


The Normal Curve
-SD
SD
-3
-2
-1
0
1
2
3
Z-score


The Normal Curve
-2.58
-1.96
1.96
2.58
68.26%
95.44%
99.74%
-3
-2
-1
0
1
2
3
Z-score

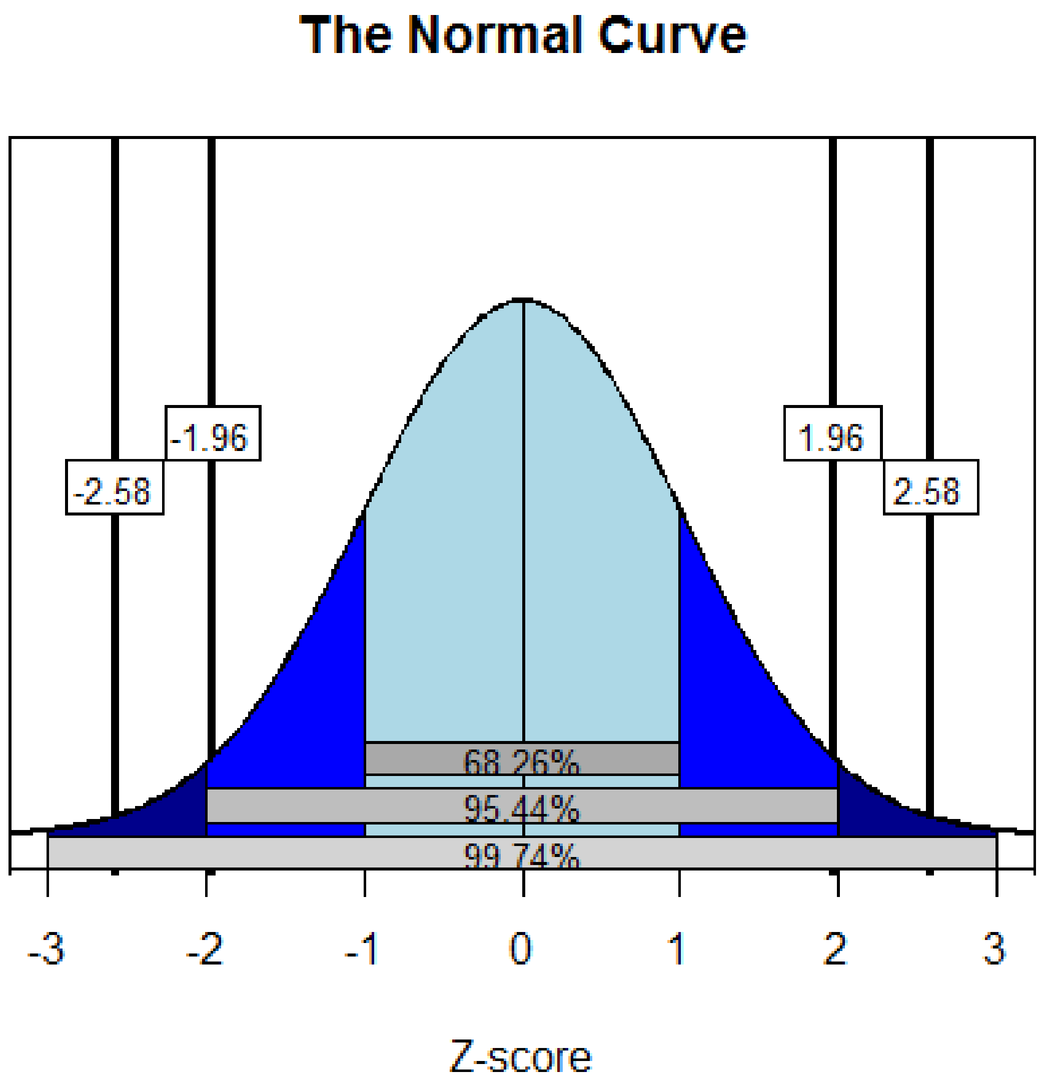
- Distribution of Error
- We know the properties of the Normal Distribution (we made it)
- If we can get our raw deviations into the same metric, we can begin to compare them.

DESCRIPTIVE Z-SCORES

- (Score Deviation)/SD
  - # of SDs away
    - Unitless

- Can also give probability of a score's occurrence
  - Z-score → p-value
    - Greater the z-score, the smaller the p-value
  - Z-table

$$z = \frac{X - \mu}{\sigma}$$


The Normal Curve
-2.58
-1.96
1.96
2.58
68.26%
95.44%
99.74%
-3
-2
-1
0
1
2
3
Z-score

Calculating

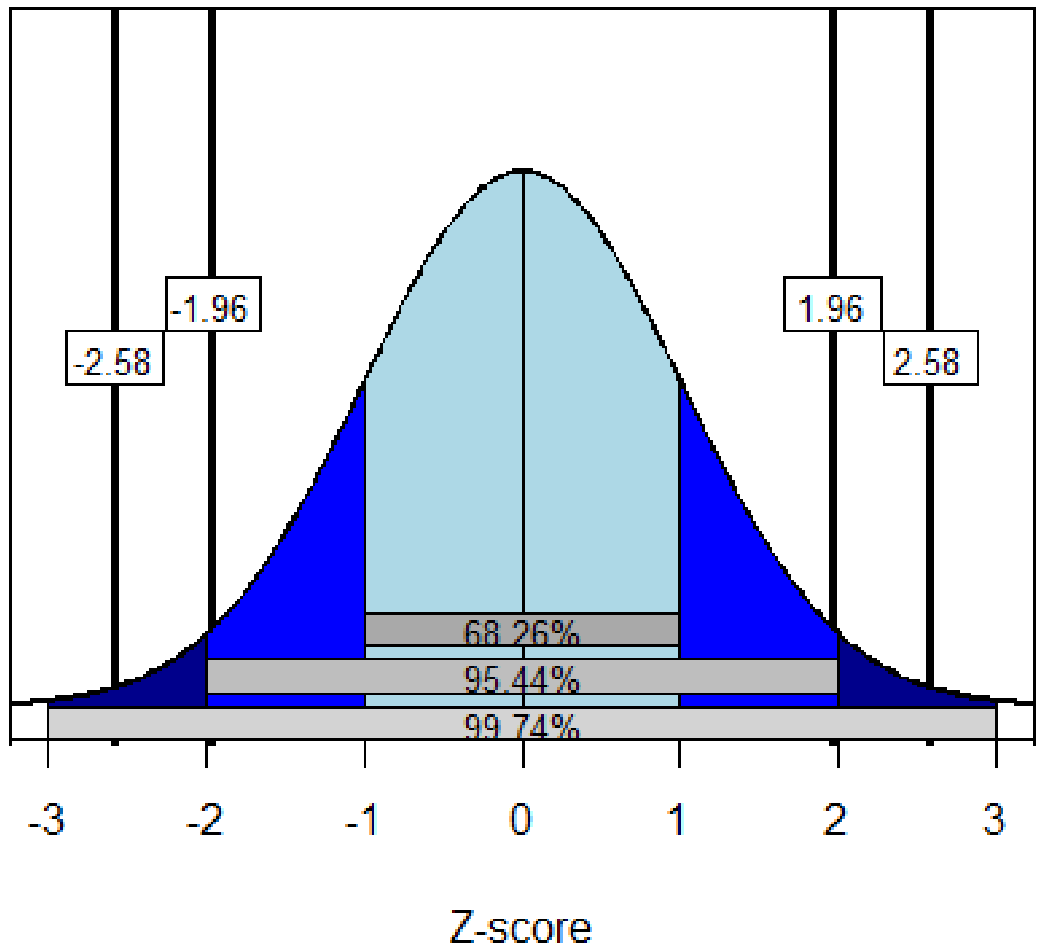
$$z = \frac{X - M}{s}$$

$$Sample$$

$$z = \frac{X - \mu}{\sigma}$$

$$Population$$

The Normal Curve



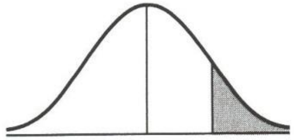
Using

- I scored a 135 on an IQ test, how many SDs above/below the mean did I score?
  - $\mu_{IQ} = 100$; $\sigma_{IQ} = 15$
  - $z_{IQ} = (135\text{-}100)/15 = 2.33$

- What percentage of people am I smarter than?

TABLE 1 Normal distribution, right-hand tail probabilities

http://www.rochester.edu/college/psc/clarke/201/ztable2.jpg



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002

- $z_{IQ} = 2.33$
- ~0.01
  - The top 1%
- 1.00 - 0.01 = 0.99
  - Outscored 99% of others
- An IQ score of 135 would mean that I was smarter than 99% of people that took the test

**I do not know my IQ score. I've never taken an IQ test.

- Compare apples to oranges
  - Which do I have more of relative to my neighbors?
  - I have 50 apples and 50 oranges
    - $\mu_{apples} = 50$; $\sigma_{apples} = 15$
    - $\mu_{oranges} = 30$; $\sigma_{oranges} = 10$
  - $z_{apples} = (50-50)/15 = 0 = 50^{th}$ percentile
  - $z_{oranges} = (50-30)/10 = 2 =$ approximately the 98^{th} percentile

- I have more oranges relative to my neighbors, but not more apples.

Going the other way…

- Knowing the z-score, the M or μ and the s or σ you can find the original raw score

— OR—

- You can put the original distribution (the same shape) into different numbers

$$X = M + z * s$$

- Examples:
- $z = 1.5$; $M = 52$; $s = 14$
  - $X = 52 + (1.5)*(14) = 73$
  - OR-
  - Transform to $\mu = 100$; $\sigma = 15$
  - $X = 100 + 1.5*(15) = 122.5$

- The absolute scores are different, but the relative scores are the same.