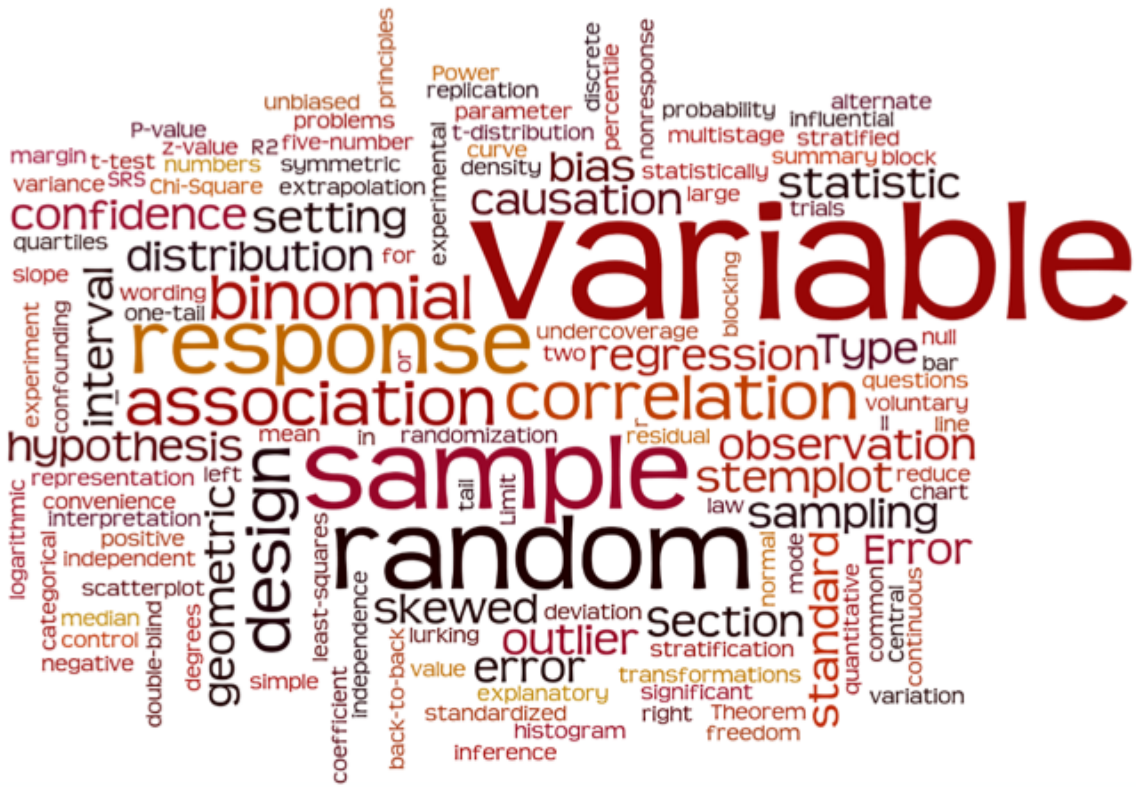
Introduction to Quantitative Methods



Kristin Eccles

Outline

- Deterministic vs. Probabilistic events
- Histograms → Probability of occurrence
- Probability
- Theoretical probability
- Calculating probabilities
  - Examples
- The normal distrubtion (a probability distribution)
- Z-scores
- Calculating z-scores

Recap from last week

Descriptive statistics has 3 components:

- 1) Central tendency
  - Mean
  - Median
  - Mode
- 2) Variation
  - Deviation
  - Sum of squares
  - Variance
  - Standard deviation
  - Co-efficient of variation
- 3) Shape of the distribution
  - Skewedness
  - Kurtosis

Describing Patterns

- In order to describe patterns on the surface of the earth we must develop descriptions and explanations for the phenomena
- We try and understand processes that create these distributions
- We want to make predictions about the future
  - This is how we make policy

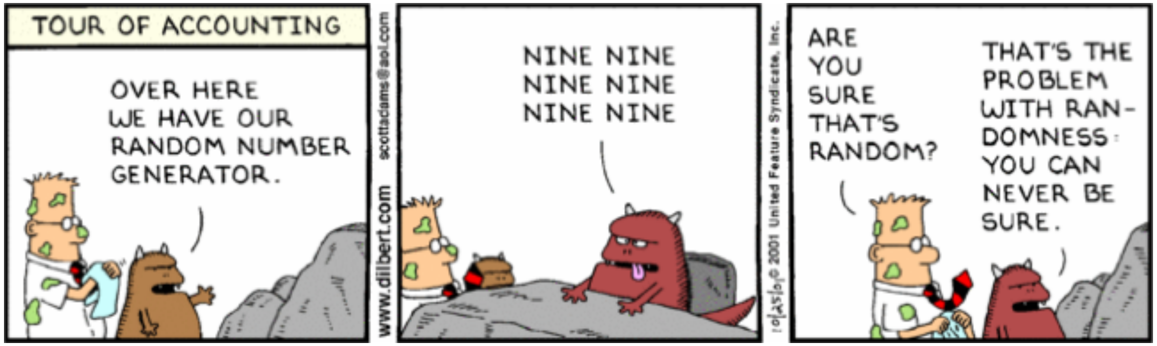
Describing processes

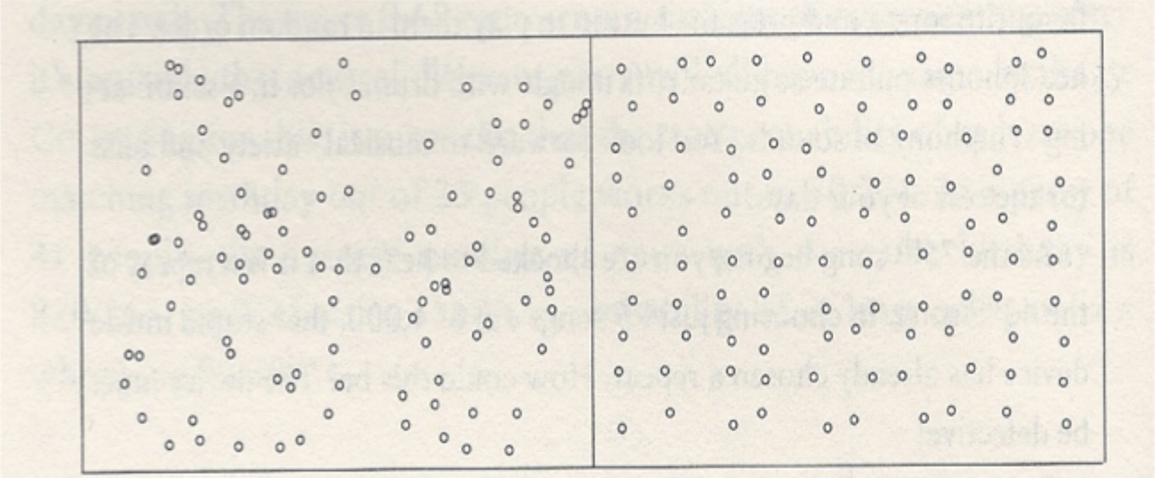
- How do we describe a processes on the surface of the earth?
  - **Deterministic**
    - Create patterns with total certainty
    - Outcomes can be specified with 100% certainty
    - E.g. Rolling a fair die: each number on a six-sided die has the same odds (1/6) of coming up.
    - E.g. The length of time solar radiation hits the surface of the earth

  - **Probabilistic**
    - Processes that cannot be determined with complete certainty
      - **Random**
        - Equal probability assigned to each outcome
        - E.g. amount of solar radiation actually reaching the earth's surface
        - E.g. the touch down location of a tornado
      - **Stochastic**
        - The likelihood/ probability of any particular outcome can be specified but not all outcomes are equal

Random

- An equal probability of being selected...
- Not haphazard or arbitrary
- Not equally-spaced either
- Clumping can occur (this can freak people out)
- Hard to know if numbers are truly random





Refs: Rosenthal, J. 2005. Struck by Lightning. Fig 2.1

Suppose this is 100% accurate...

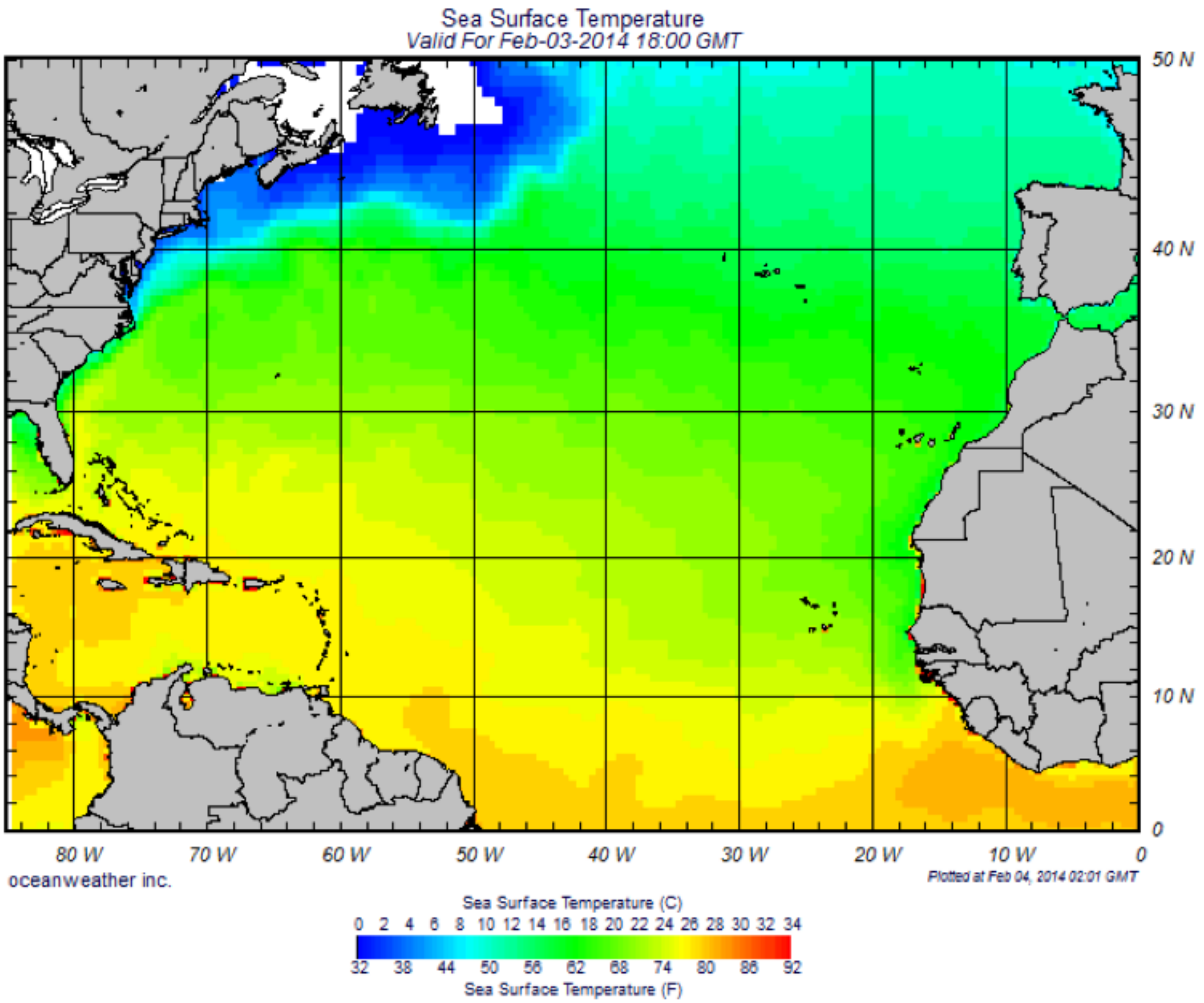


Figure 1. Sea surface temperature for part of the Atlantic Ocean on Feb 3, 2014.
Source: http://www.oceanweather.com/data/

Suppose we go out and sample...

Sample the temperature of the ocean in many locations and record values....

100 sea surface temperatures (^{o}C) from all over... :

7.5,10.1,25.3,24.8,9.9,7.9,9.1,22.5,12.2,16.4,19.8,12.7,29.0,11.1,3.7,17.9,24.5,9.8,19.8,13.6,16.8,10.1,12.9,17.2,17.0,19.9,22.7,16.0,7.3,15.7,27.1,18.1,10.7,19.0,8.2,12.6,18.5,19.0,18.6,12.9,19.1,14.0,15.4,19.9,17.3,16.7,8.1,16.6,3.6,9.1,12.1,18.7,25.4,7.9,10.4,12.5,14.0,12.7,10.6,15.9,25.6,16.0,14.5,0.9,20.7,4.3,33.0,24.6,19.5,15.6,19.2,8.4,12.8,15.0,21.0,12.6,10.7,15.8,18.3,25.5,15.2,25.6,19.1,17.9,20.7,23.4,10.5,17.7,10.2,11.4,17.1,8.4,23.3,12.8,14.6,7.6,8.9,18.0,21.7,8.8

Mean: 15.6^{o}C

Median: 15.7^{o}C

Standard deviation: 6.1^{o}C

Min.: 0.89^{o}C

Max.: 33.0^{o}C

Histogram – frequency distribution

Notice the shape:

Quasi-symmetrical

Mean, median, mode near the middle

This is a typical distribution for data

= 'normal'

Can change frequency count to *proportion of total counts* (count / *n*)

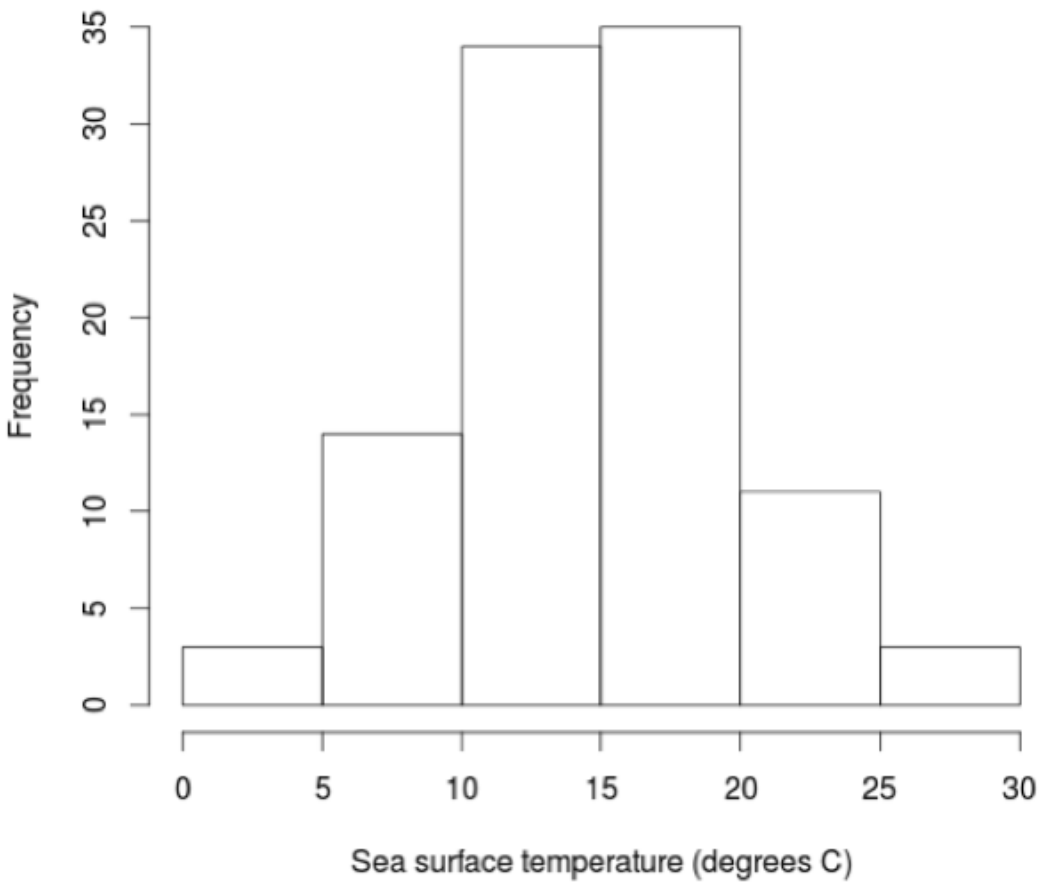
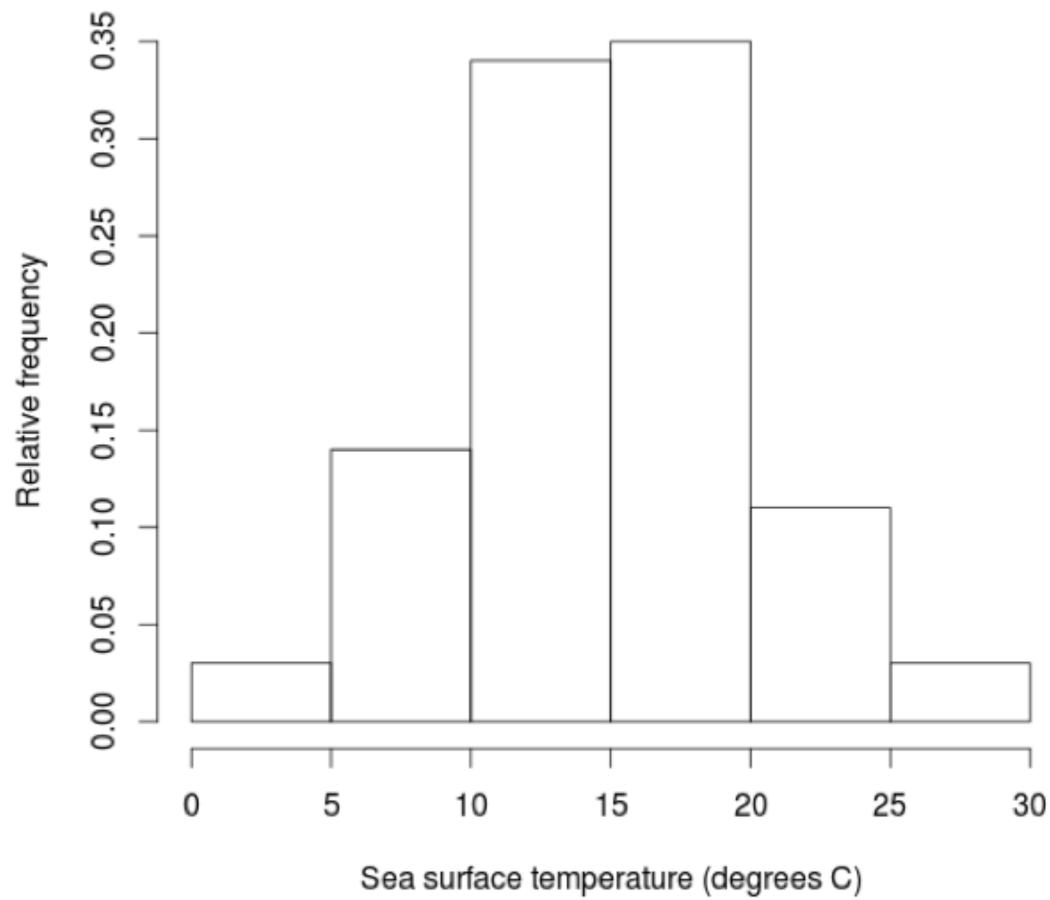


Figure 2: Histogram of sea surface temperature for the Atlantic Ocean from our sample (n=100)

Histogram – Relative frequencies

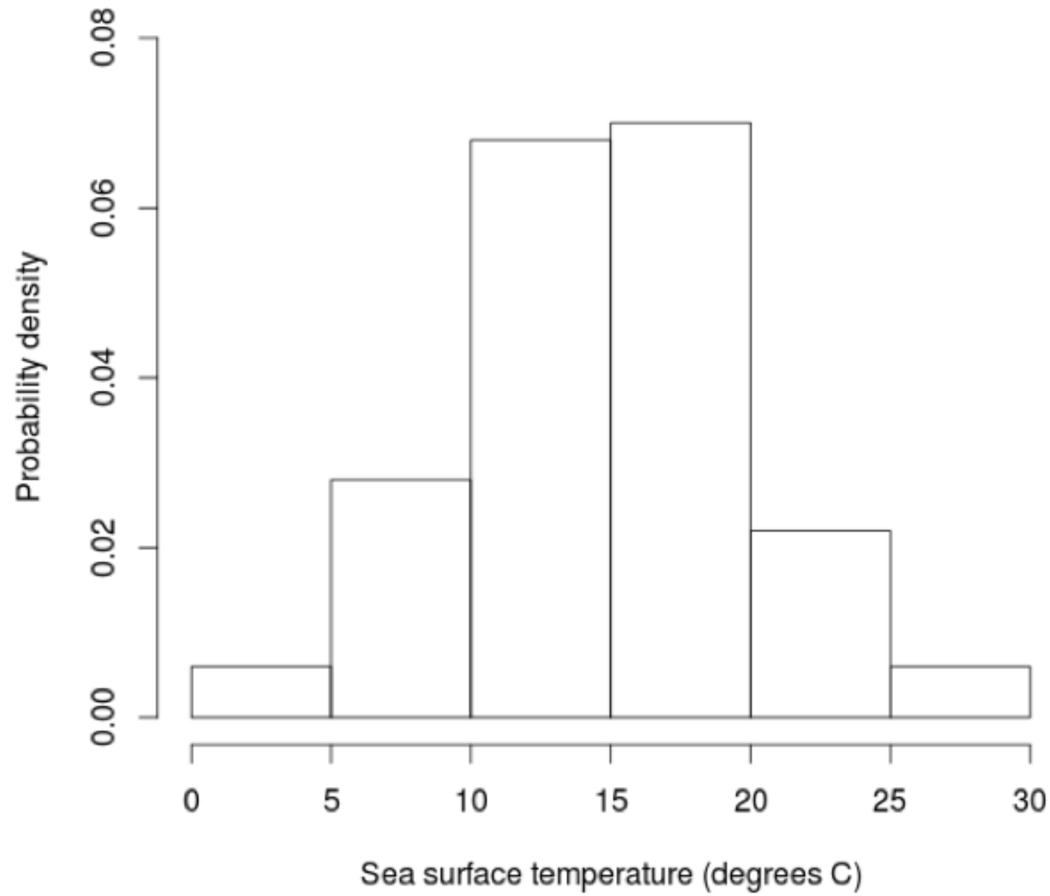
- New y-axis: proportion = *relative frequencies*
- Proportion in all the bins together adds to 1
- And recall relative frequency is the *probability* too



Technically, this assumes you have a large number of *independent* observations

Histogram – Probability density

- Now consider that the bins may not always be the same width... (you have 10 bins, I have 11 bins...)
- This is a problem for inter-comparison
  - Divide the relative frequency by the bin width
  - This makes the *area* under each bar proportional to the probability of finding data inside the bin



- The total area for all bars = 1

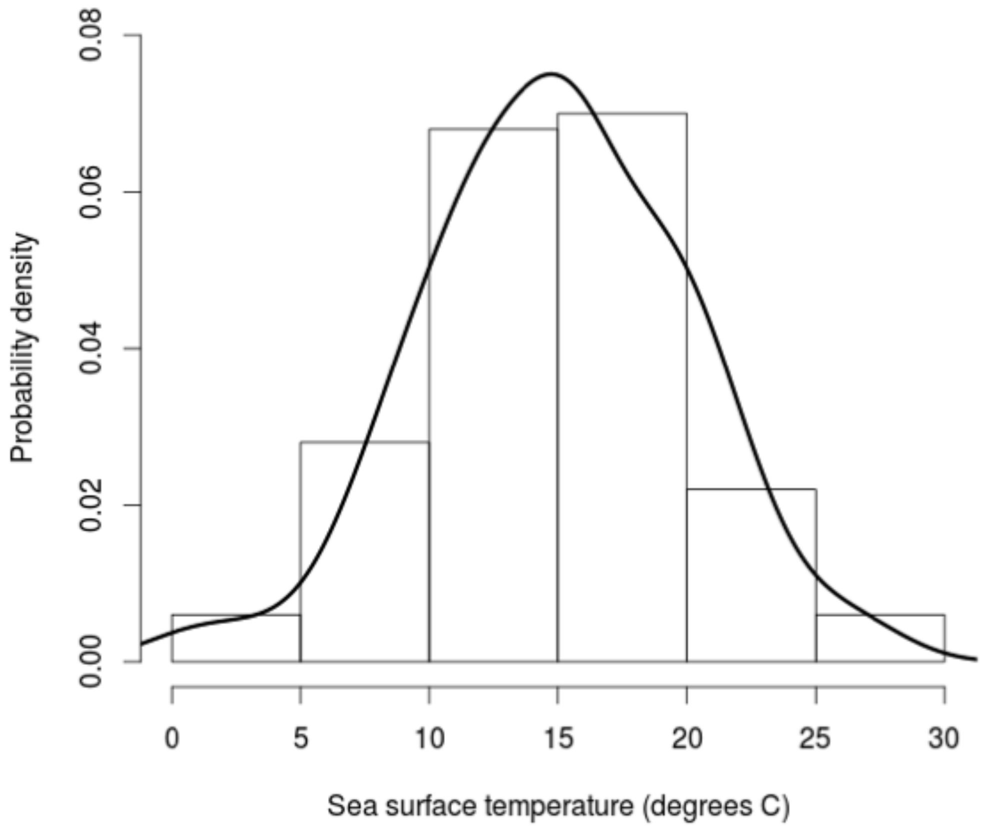
Histogram – Continuous probability density

Imagine instead of discontinuous bins there is a *smooth curve* that represents the distribution...

This is the **probability density function**

Total area under the curve = 1

See how the line follows the bars (more or less)



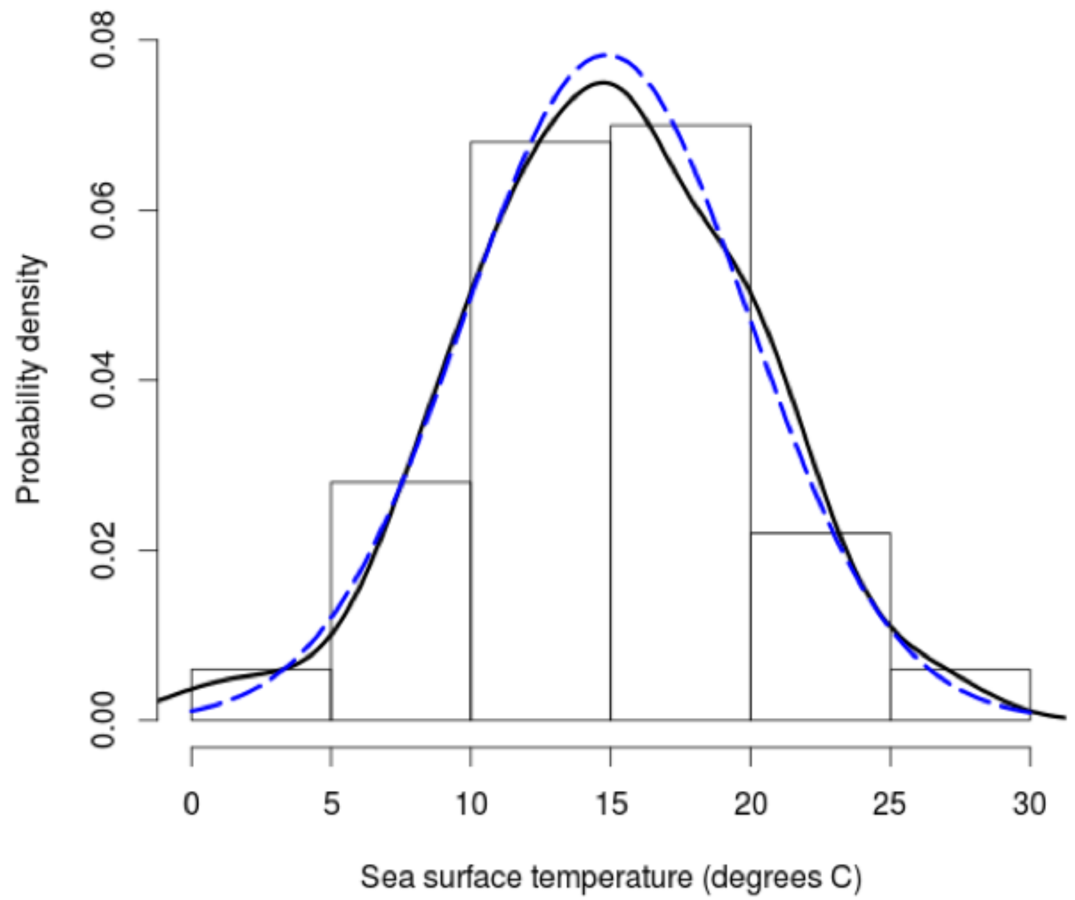
Histogram – probability distribution

Can compare your data to a known **probability distribution**:

Certain data will tend to follow a given *theoretical distribution*

In this case the **normal distribution** (blue curve) looks like a good match!

There are other probability distributions too (t-distribution, F-distribution, X^2-distribution, etc... for later)



Probability

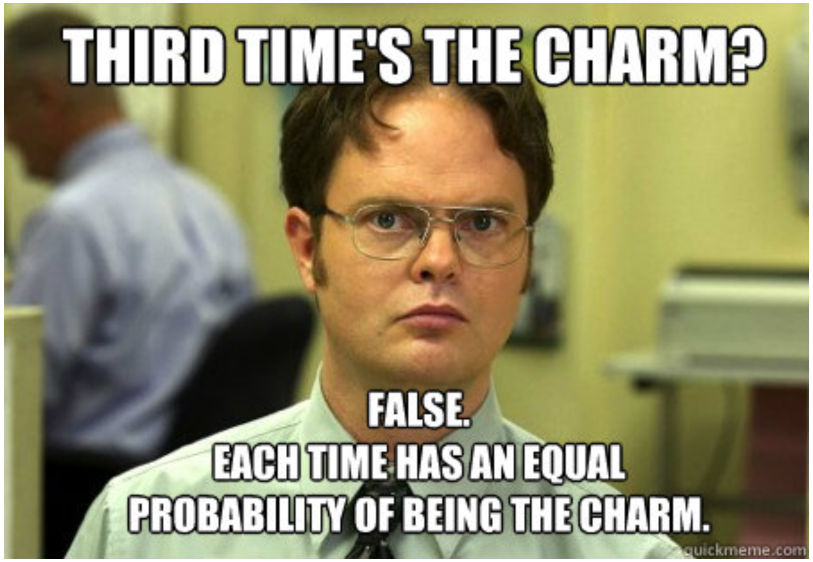
- What is probability?
  - How **likely** something is to happen
  - Can be thought of as a relative frequency
    - Ratio between frequency of a particular outcome and the frequency of all outcomes
- The analyses of many problems are based on probabilities:
  - What are the chances that the 100-year flood will occur within the next ten years?
  - How likely is it that a pixel on a satellite image is correctly classified or misclassified?

(Theoretical) probability

- A random phenomenon
- There is more than one possible outcome that are:
  - Equally likely to occur, mutually exclusive, and exhaustive
- How many possible outcomes are there?
- Known as the sample space (S)
- Coin: S= {Heads, Tails}
- Die: S= {1,2,3,4,5,6}
- Cards: S={A♠,2♠,3♠… A♣,2♣,3♣…A♦… …K♥}
- Simple lottery (with one winner/draw): S={picture several million ticket holder names here…}

(Theoretical) probability

- An event is a combination of outcomes in a trial
- Can be a specific unique outcome (drawing the A♠) or a set of outcomes (drawing any ♠)
- The (theoretical) probability of an event is:
  - The number of outcomes in A divided by:
  - The number of possible outcomes (S)
  - P(A) = (A)/(S)
    - where P = probability,
    - A = specific event,
    - S = sample space of all possible outcomes



Probability *Rules!*

- Probability is a number between zero and one (but can also be expressed as a percent)
  - $0 \leq P(A) \leq 1$
- Together, the probability of all outcomes is one
  - $P(S)=1$
- The probability of A not occurring (the complement of A aka A^c) is one minus the probability that it does occur
  - $P(A^c) = 1-P(A)$
- The probability of two events that have no outcome in common is the sum of the individual probabilities
  - $P(A \text{ or } B) = P(A) + P(B)$

P(S)=1

- Together, the probability of all outcomes is one

Ex. Using a single 6-sided die is rolled, what is the probability of each outcome? What is the probability of rolling an even number? of rolling an odd number?

- P(1)=1/6=0.166
- P(2)=1/6=0.166
- P(3)=1/6=0.166
- P(4)=1/6=0.166
- P(5)=1/6=0.166
- P(6)=1/6=0.166

$P(A^c) = 1\text{-}P(A)$

- The probability of A not occurring (the complement of A aka A^c) is one minus the probability that it does occur

Ex. A glass jar contains 6 red, 5 green, 8 blue and 3 yellow marbles. If a single marble is chosen at random from the jar, what is the probability of choosing a red marble? a green marble? a blue marble? a yellow marble?

- P(Yellow)= 1- (0.27+0.23+ 0.36)= 0.13

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

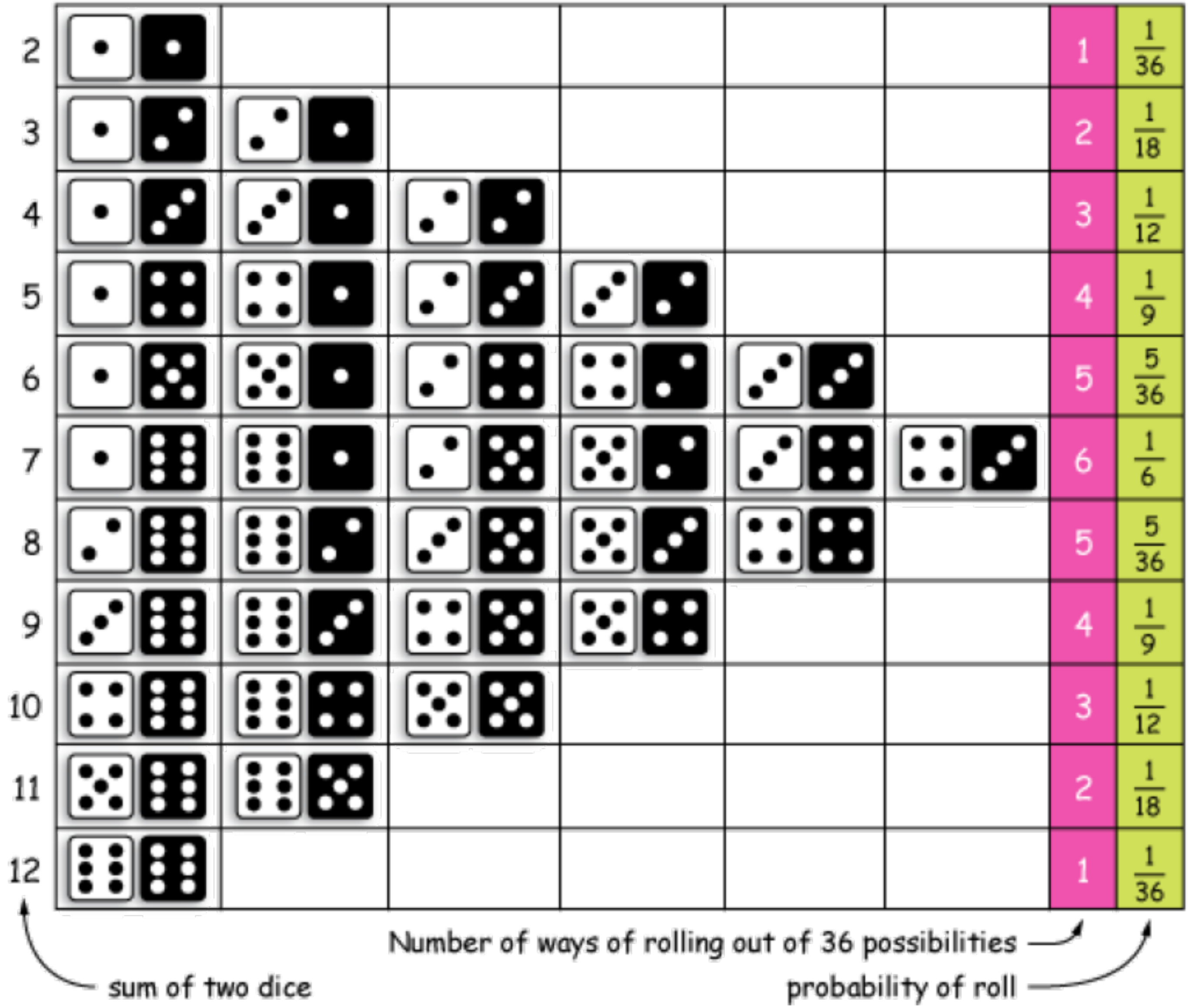
- The probability of two events that have no outcome in common is the sum of the individual probabilities

  - Ex. You have a new drug that will only work on people who have blood types A and O. What is the likelihood of you sampling someone of these blood types from a population?
    - Likelihood of blood types:
      - P(O)= 0.45
      - P(A)= 0.40
      - P(B)= 0.11
      - P(AB)= 0.04

      - P(A or O) = P(A) + P(0)= 0.45+ 0.40= 0.80

The probability of rolling a three is just the number of ways of rolling a four (3) divided by the total number of possibilities of a two-dice roll (36), so 3/36 = 1/12. It's the same for all other rolls. Notice that there are 6 ways of rolling a 7, making it the most probable roll.

If we insist that in rolling a 7, we roll a 1 on the white die and a 6 on the black (we call that a case with ***distinguishable*** *dice), then the probability just "collapses" to 1/6 · 1/6 = 1/36*



Ok so why do we care about probability?

- Central Limit Theorem (CLT)
  - In probability theory the CLT states that when a variable is **independent and random** the sample average will tend to form a normal curve
  - This implies that probabilistic and statistical methods work for normal distributions

  - https://www.youtube.com/watch?v=JNm3M9cqWyc

Statistical independence

- Outcomes of one event have no influence on the probability of another event
- Very important in statistics
- For example, dice/coins have no memory:
  - An outcome for each individual toss has the same probability of occurring
- Termed unconditional probability

Dependent events

- If an event is not statistically independent:
  - Probability of an event changes depending on the outcome of a previous event
- Winning a top prize on a scratch ticket
- This is called conditional probability
- You should be able to tell the difference between an independent and a dependent event!
- But you don't need to know any more than that (conditional probability is harder to understand)
  - E.g. The Monty Hall Problem (http://www.youtube.com/watch?v=mhlc7peGlGg)

Applying probability theory to real data

(Empirical) probability

- You can use experimental data or look in the past to find probabilities!
- Determine the relative frequency of an event:
  - count how many times A occurs and divide by the total number of trials
- The law of large numbers:
  - Relative frequency will converge on the probability as n increases
- This gives the probability of A:
  - P(A) = Count of A / Count of trials
  - BUT, you have to have a large number of trials
  - AND the events must be independent

The Normal Distribution

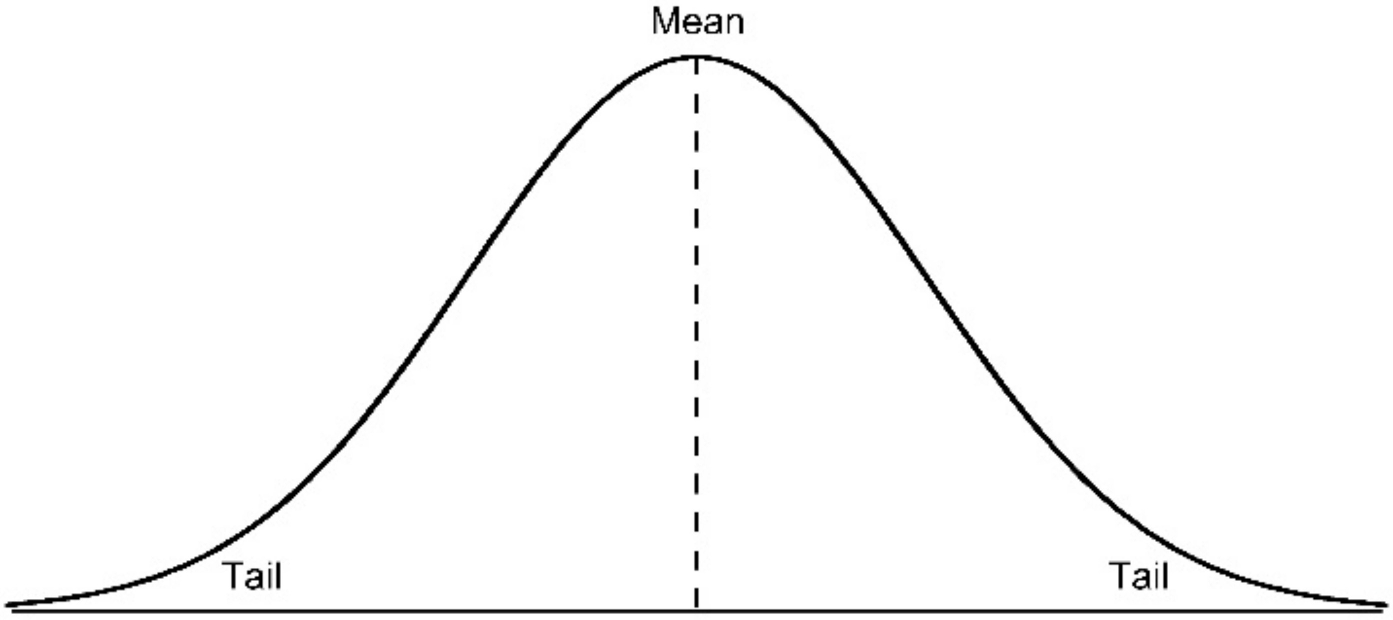


FIGURE 6.1
The Normal Distribution

The normal distribution

- Characteristics:
- Mean, median, 'mode' are all the same & in the middle of the distribution
- Symmetrical, smooth, continuous distribution
- Decline in probability/occurrence of values from the centre to the tails
- Total area under the curve is = 1 (probability is 1)

$$\Sigma_i (P(X_i) = 1$$

This means that a probability can range from zero to one $0 \leq P \leq 1$

- A theoretical distribution – this is not real data, but an approximation of what data distributions are often like (CLT)
- Sometimes variables have histograms that look like the normal distribution

Properties of the normal distribution

- The area under the curve gives the probability of finding a given value in that range
- This gives you an idea of where the data are distributed
  - What the probability is of getting an extreme value (the tail)
- 68% of the values are within one standard deviation of the mean
  - Or, there is a probability of 0.68 that a given value is within 1 s.d. of the mean
- 95% of the area under the curve is within 1.96 standard deviations of the mean
- 99.7% of the area is within 3 standard deviations of the mean
- As with any empirical probability, this relies on the law of large numbers, the relative frequency will converge on the true probability in the long run

Graphical (Properties of the normal distribution)- Proportions

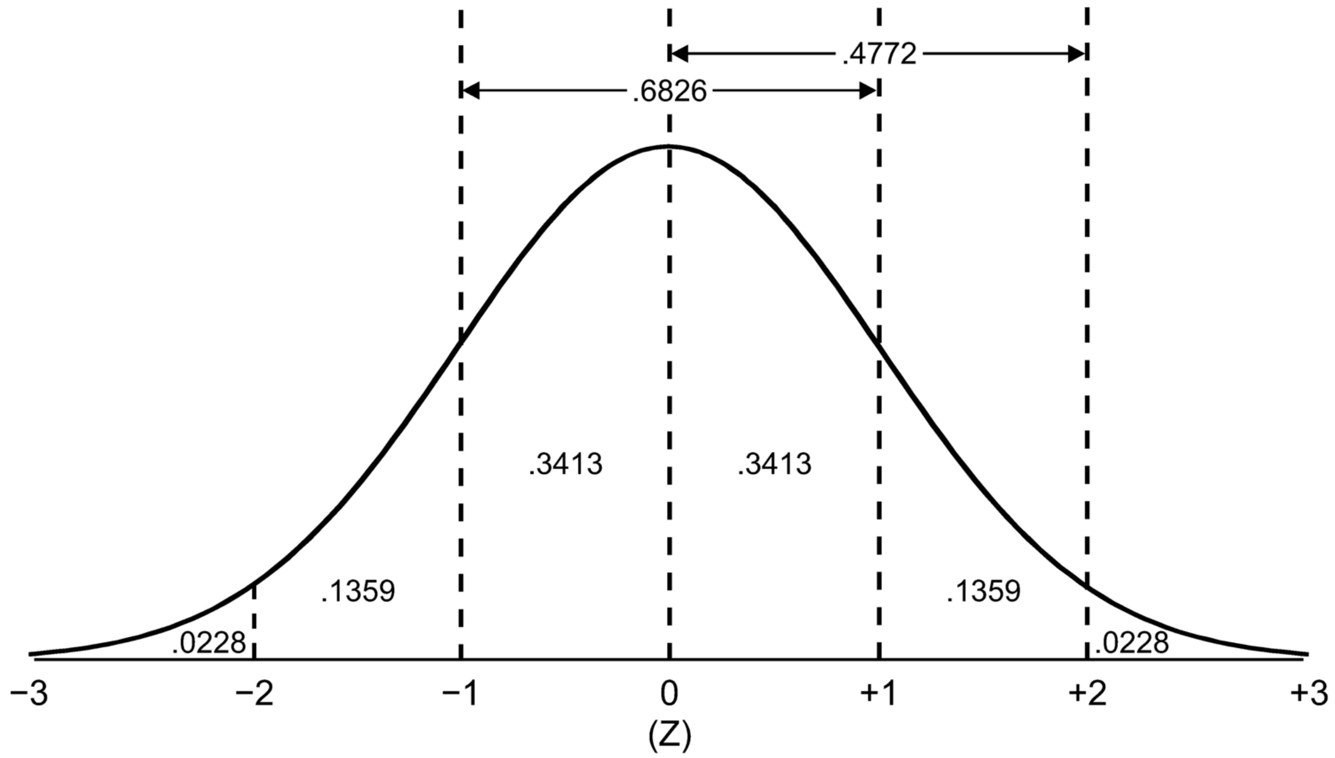
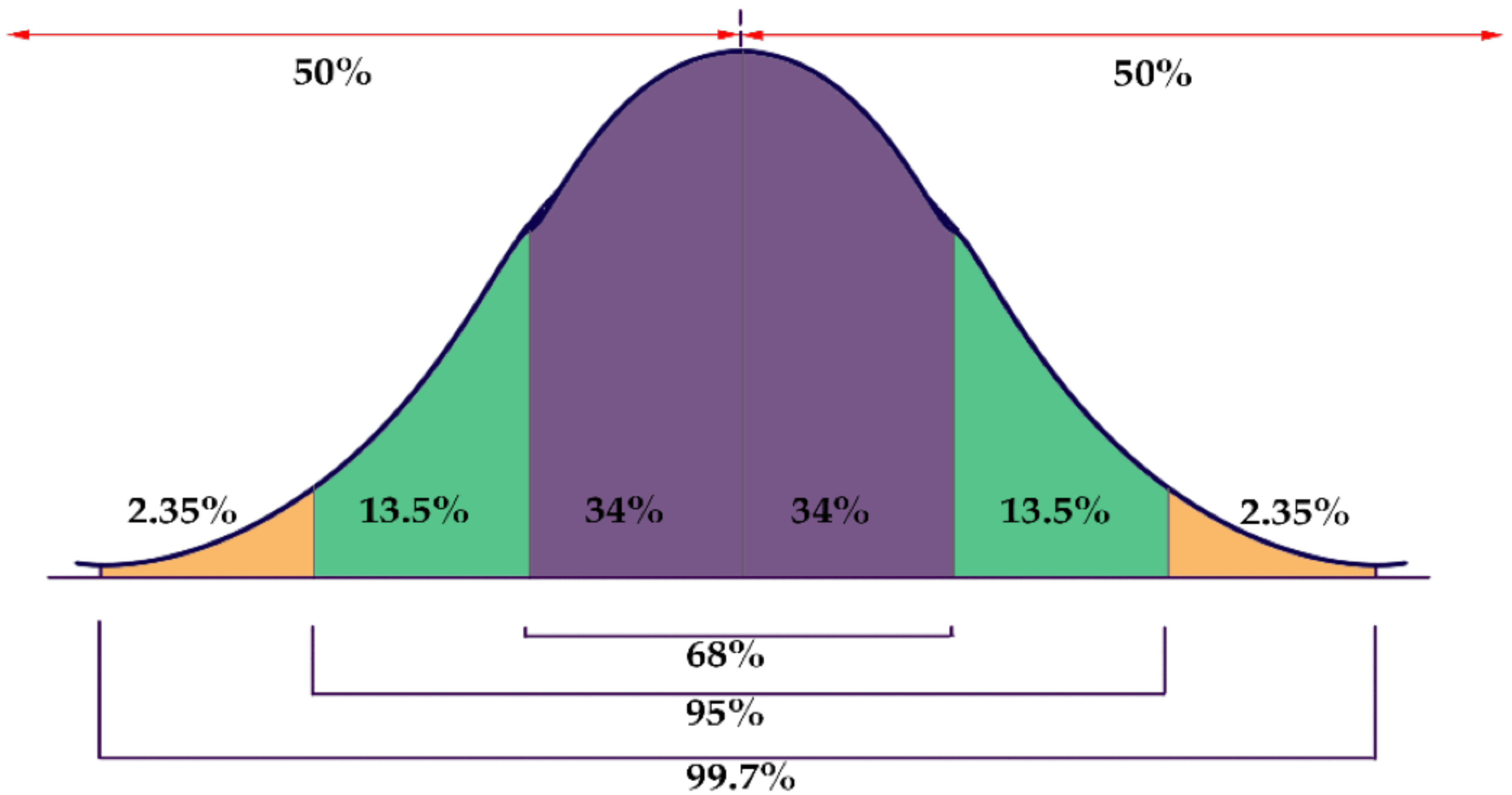


FIGURE 6.2
Selected Areas of the Normal Curve

Graphical (Properties of the normal distribution)- Percentages



Normal distribution

A curve that follows this equation

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

Equation (just for reference)

There are 2 parameters:

Mean: μ

Standard deviation: σ

You can *model* the probability density of a given variable if:

1) it is normally distributed

2) you know the mean

3) you know the standard deviation

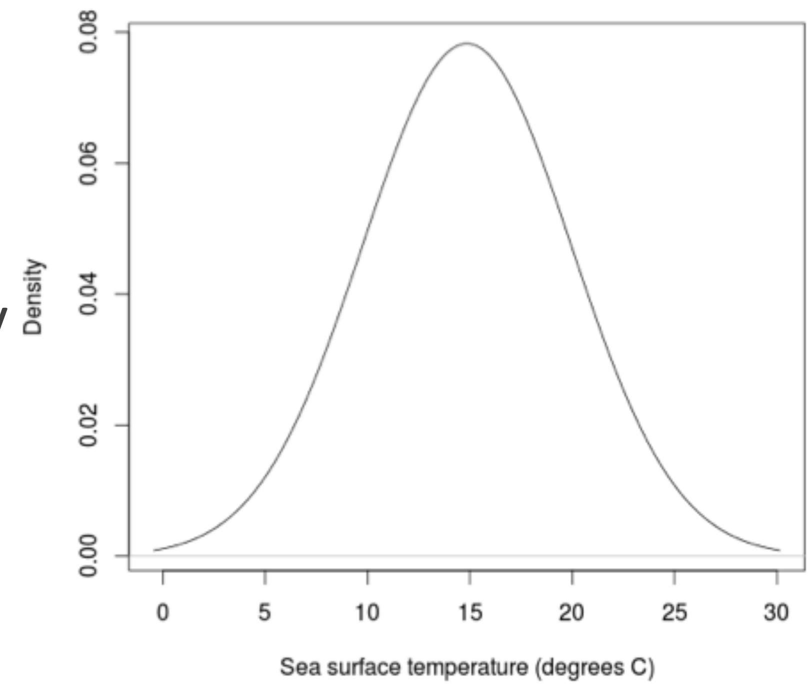


Figure x. A normal curve example.

Is it normal?

- Since data must be normally distributed (or approximately so) to do many advanced statistics
- A very basic question is:
  - “Is this variable normally distributed??”
- **Answer:**
  - Make a histogram
  - Are the mean, median & modal class near each other?
  - Check the shape of the distribution:
    - Skew
    - Kurtosis

Standard scores

- It is often a pain to determine how many standard deviations a given value is from the mean
- So, we standardize the data by:
  1. subtracting the mean – to get deviation
  2. dividing by the standard deviation
- This is called a standard score (or z-score)
- Result is a variable with:
  - mean = 0
  - standard deviation = 1

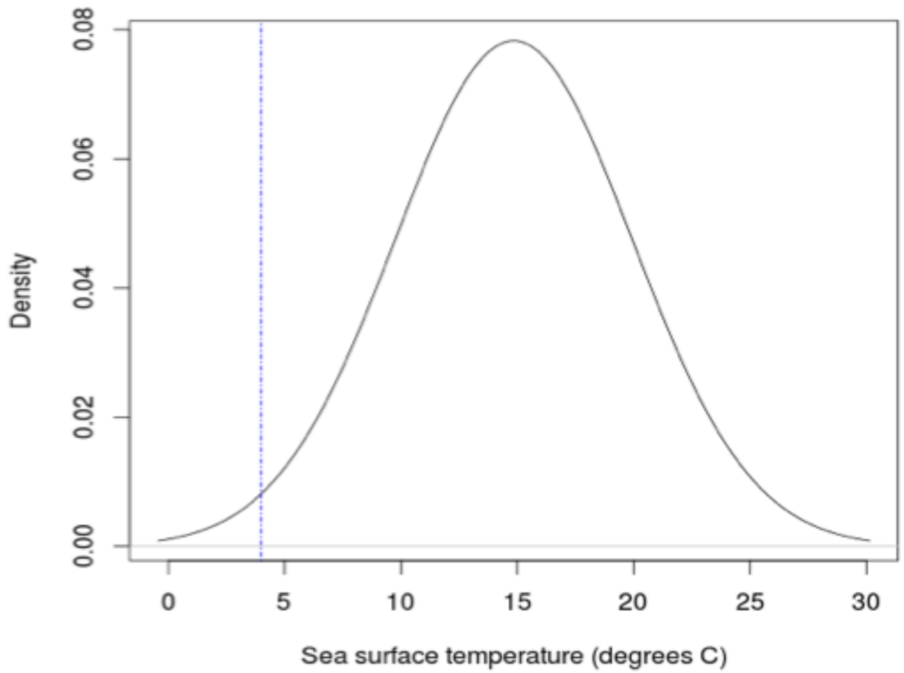
$$\boldsymbol{z = \frac{(x_i - \bar{x})}{s}}$$

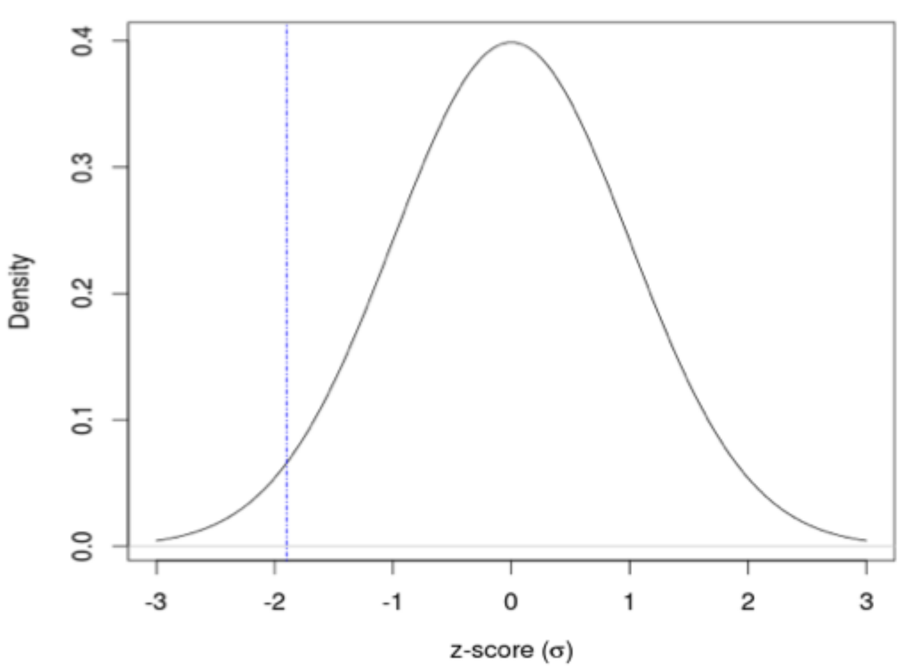
Example

If temperature of the ocean near Newfoundland is 4.2°C

Mean: 15.6°C, Standard deviation: 6.1°C

z-score = (4.2 – 15.6) / 6.1 = -1.87





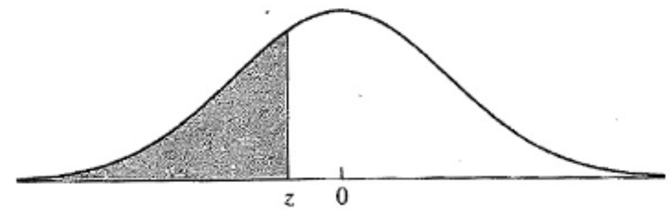
Finding the probability using z-scores

- Now that you have a z-score, what is the probability of getting a *smaller value*?

- Use a table
- Use software

- Z-score = -1.87
- P(z < -1.87) = 0.0307

TABLE A.2 Cumulative normal distribution (z table)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681

Finding the probability using z-scores

What is the probability of getting a *smaller value*?

$P(z < -1.87) = 0.0307$

What is the probability of getting a *larger value*?

$P(z > -1.87) = 1 - P(z < -1.87)$ → It's the *complement!*

$= 1 – 0.0307 = 0.9693$

What is the probability of getting *between two* z-scores?

$P(-1.87 < z < +1.23)$

$= P(z < +1.23) - P(z < -1.87)$

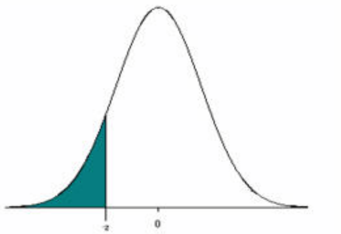
$= 0.8907 – 0.0307$

$= 0.8600$

If you know the probability...

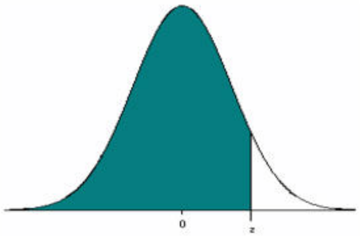
- You want to know the temperature associated with the 40^{th} percentile
  - 40^{th} percentile → 40% of observations will be at or below this level (or 0.4 is the area under the normal curve)
- So look up the z-score associated with this probability
- =NORM.S.INV(0.4)
- z-score = -0.253
- Convert to temperature →
  - z-score = (X – 15.6) / 6.1 = -0.253 = 14.1°C

Table of Standard Normal Probabilities for Negative Z-scores



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

Table of Standard Normal Probabilities for Positive Z-scores

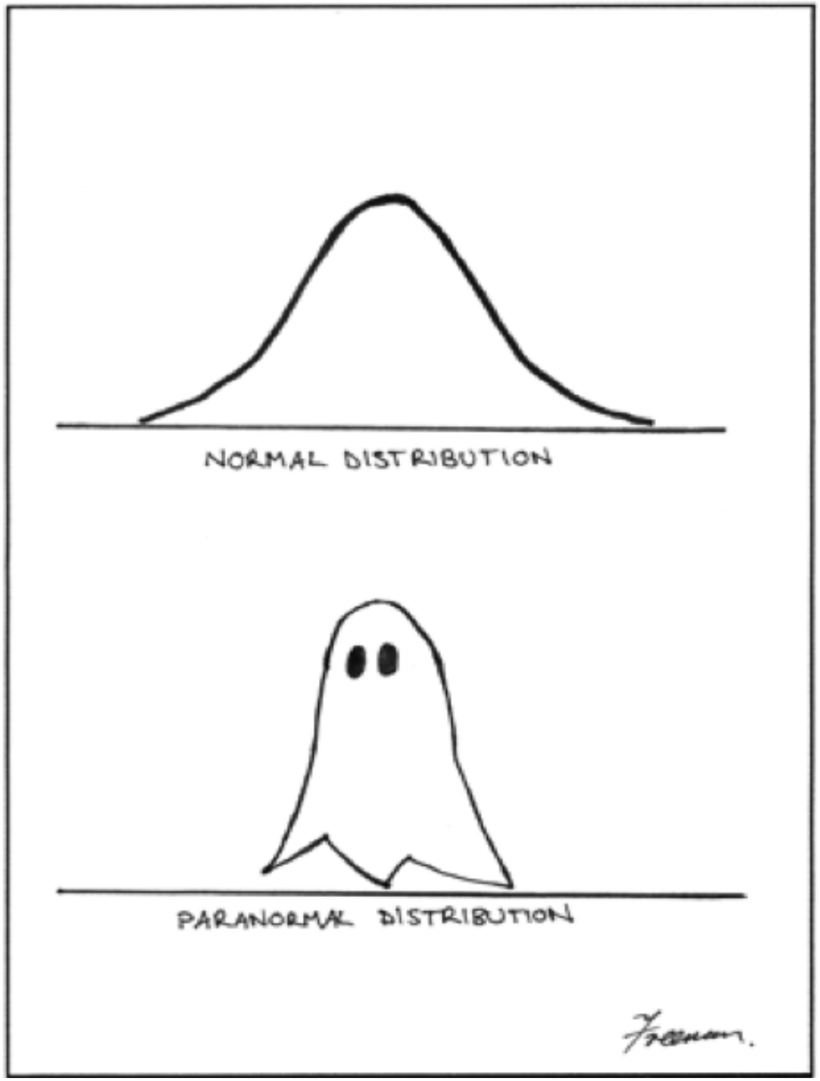


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Note that the probabilities given in this table represent the area to the LEFT of the z-score.
The area to the RIGHT of a z-score = 1 – the area to the LEFT of the z-score

Other probability distributions

- The normal curve is for continuous data
- There are other distributions:
  - Discrete data:
    - Binomial (Binary Categories)
    - Poisson (Count)
- F-distribution
- T-distribution
- X^2 distribution



From: A visual comparison of normal and paranormal distributions Matthew Freeman J Epidemiol Community Health 2006;60:6.

Summary

- Probability theory deals with random events
- In statistics, random chance/probability is a benchmark
  - Is the sample drawn randomly and independently from a given population?
  - Is the probability of some occurrence/event greater than random chance?

- Often data follow the normal distribution and so it is a convenient model
  - You can standardize data and determine probabilities of getting more or less than that value using the normal distribution

- Other resources: https://www.khanacademy.org/math/precalculus/prob-comb/basic-prob-precalc/v/basic-probability