

Statistical Science 1024

Chapter 3

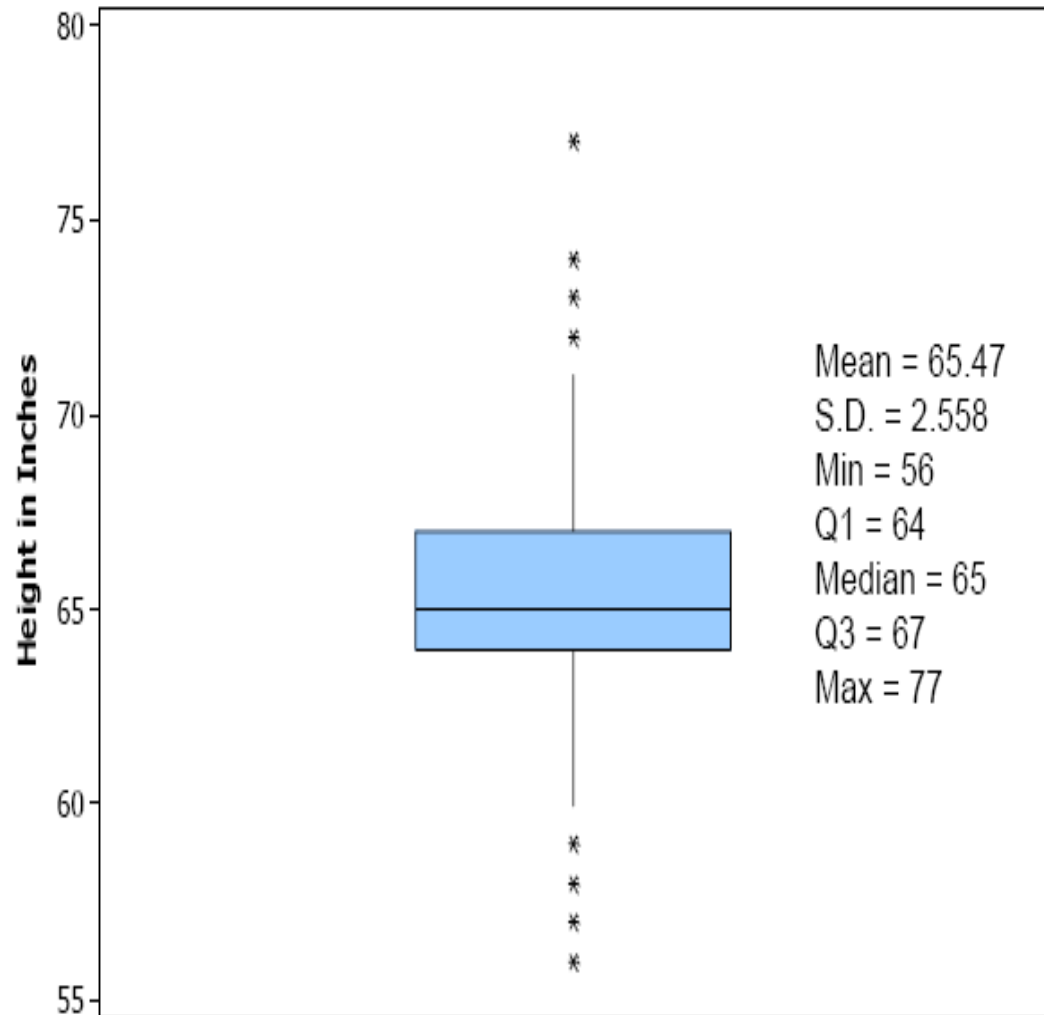
The Normal Distribution

EXPLORING A DISTRIBUTION

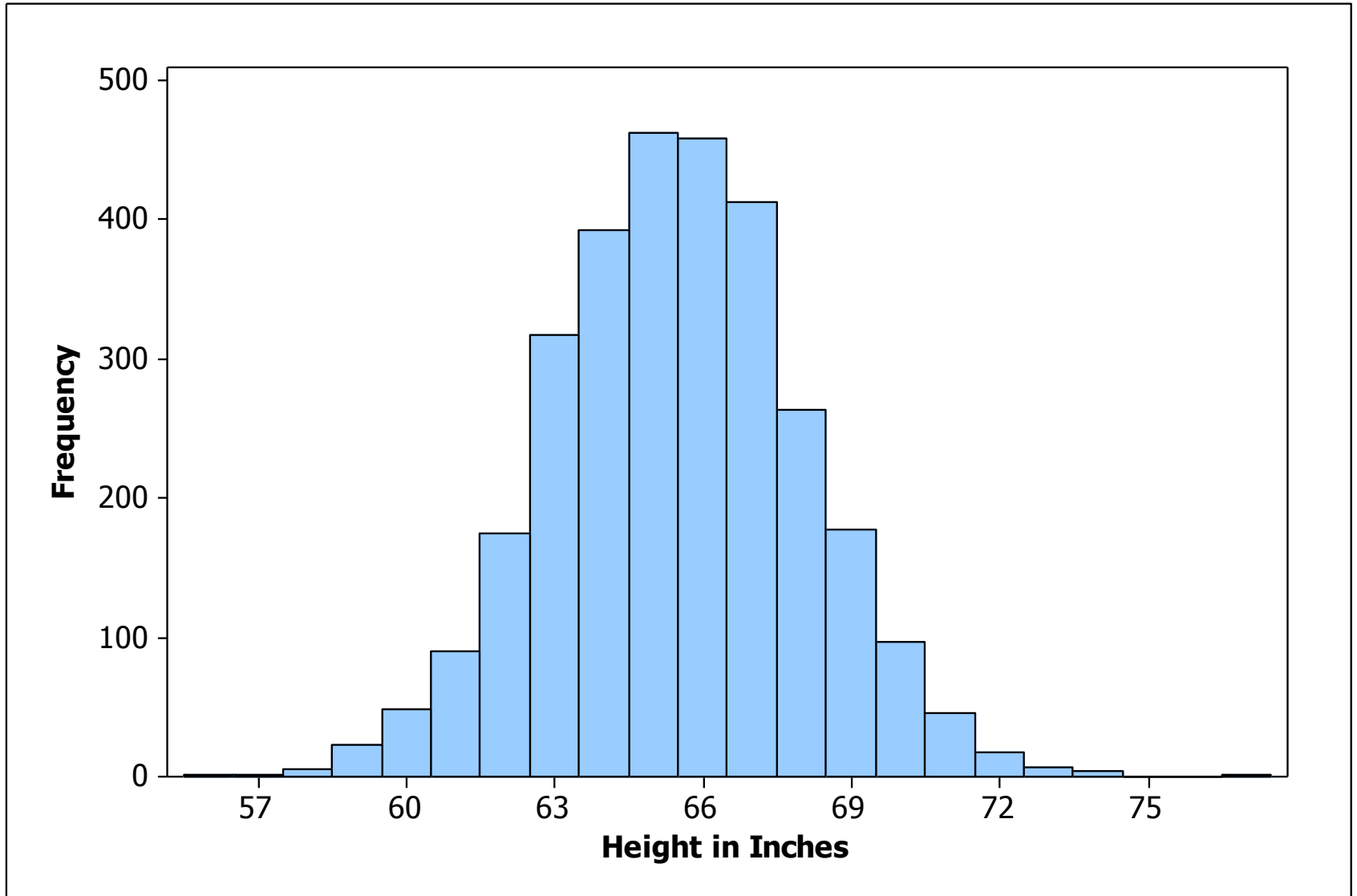
1. Always plot your data: make a graph, usually a histogram or a stemplot.
 2. Look for the overall pattern (shape, center, spread) and for striking deviations such as outliers.
 3. Calculate a numerical summary to briefly describe center and spread.
- Sometimes the overall pattern of a large number of observations is so regular that we can describe it by a smooth curve

Example Data: Heights of 3000 Criminals

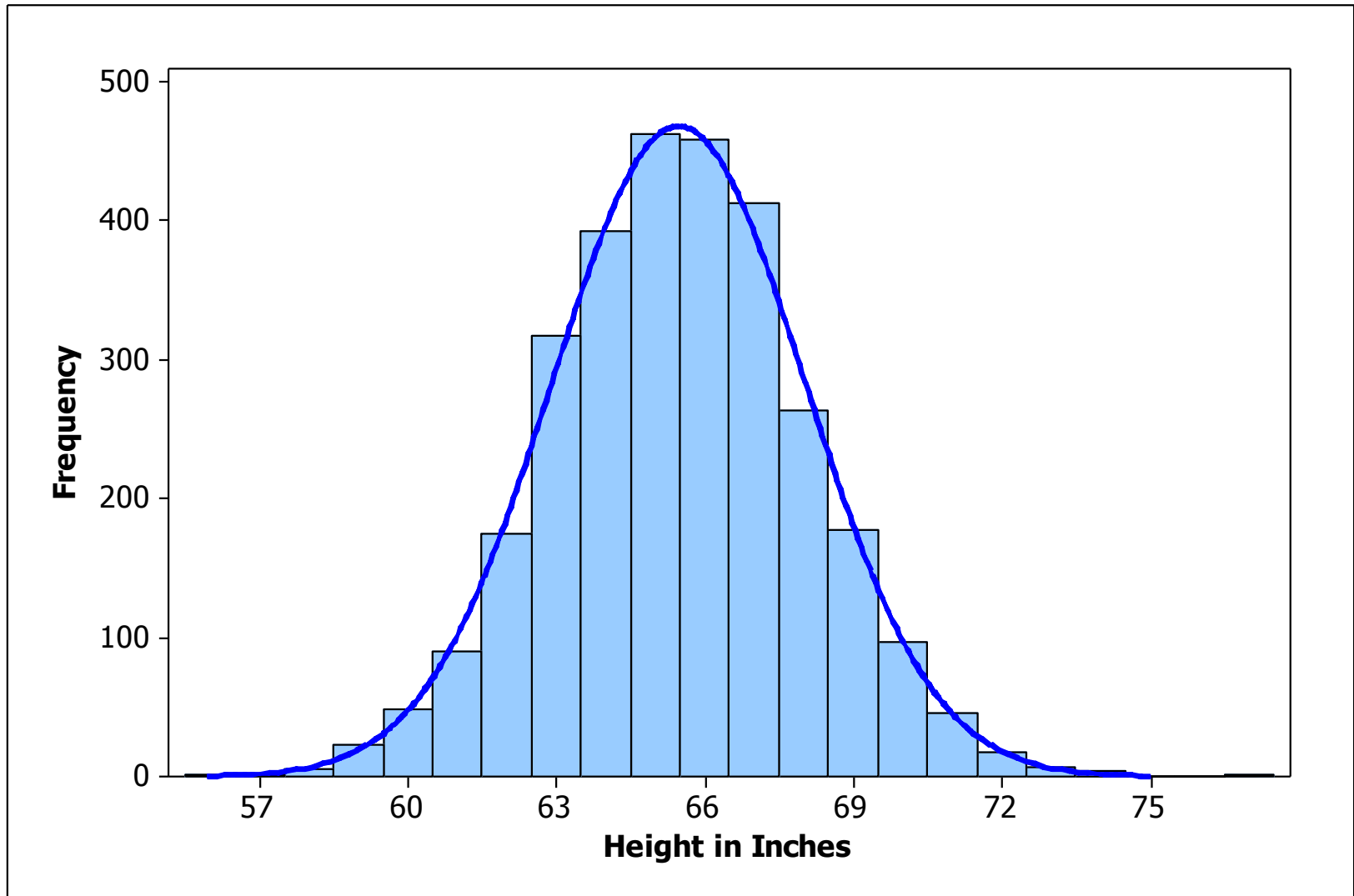
- distribution of heights of 3000 criminals in 1902
- distribution is nearly symmetric



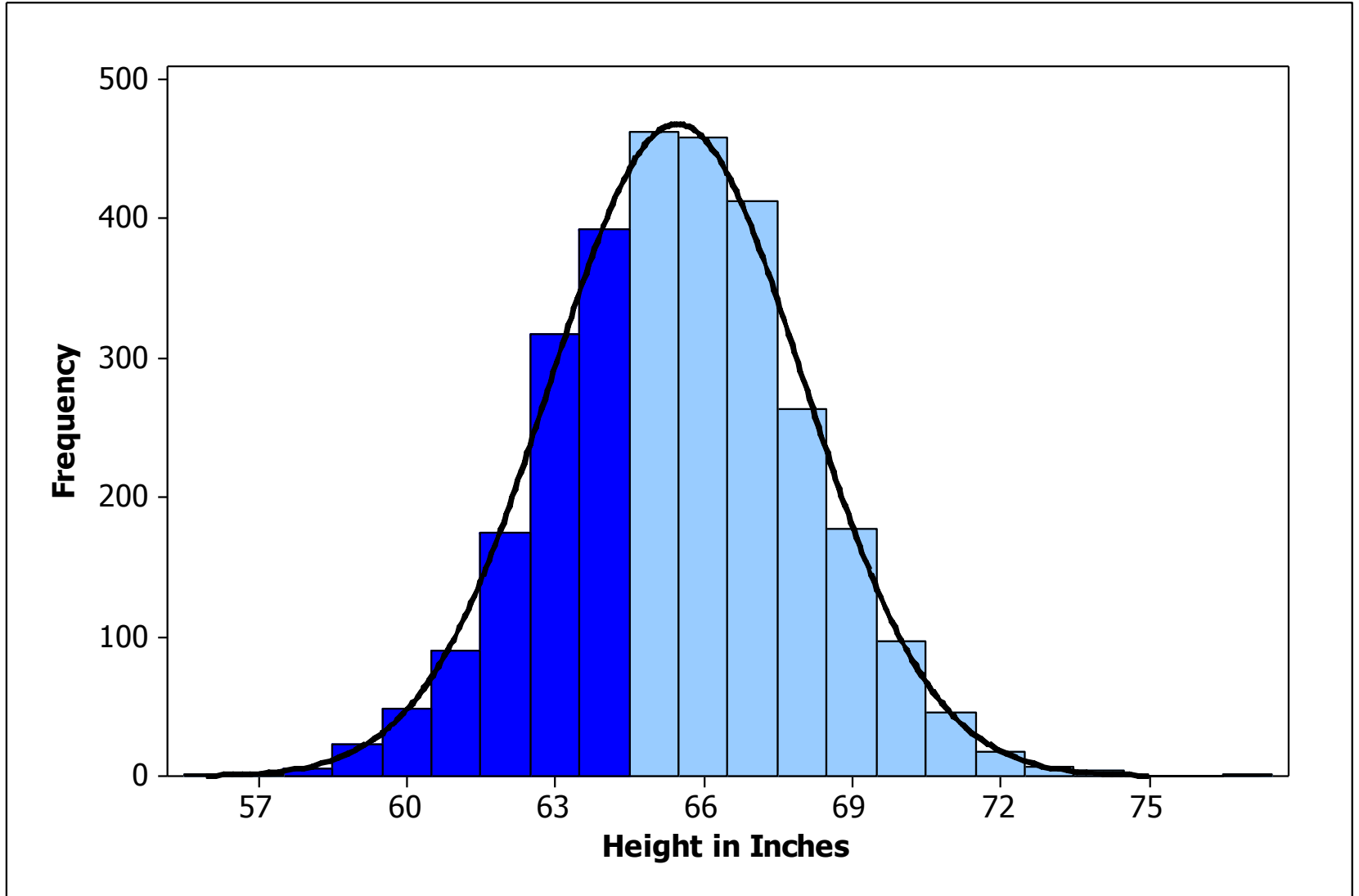
Histogram of Heights



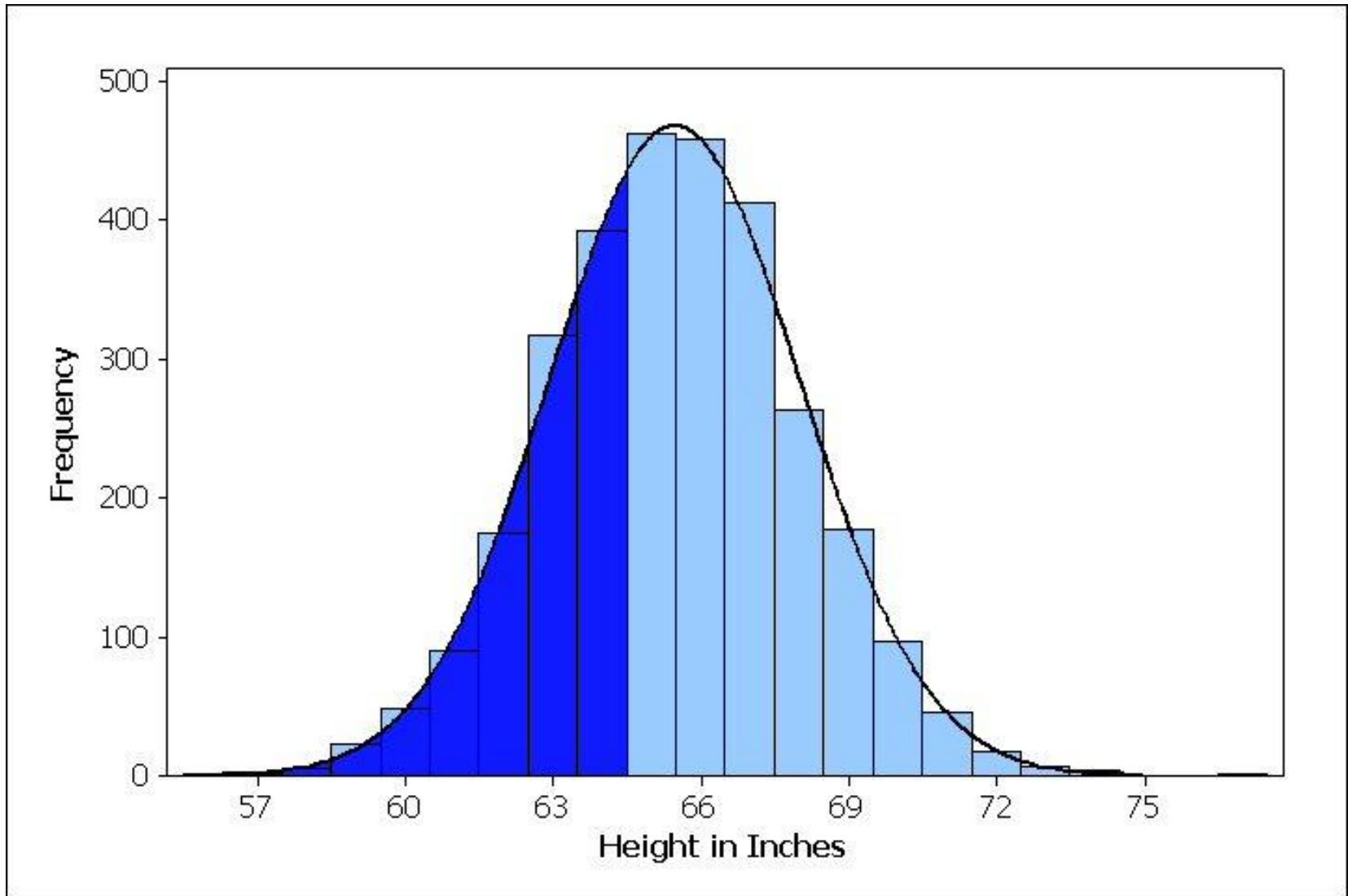
Histogram to Density Curve



Histogram Areas



Density Curve Area



DENSITY CURVE

A **density curve** is a curve that

- is always on or above the horizontal axis, and
- has area exactly 1 underneath it.

A density curve describes the overall pattern of a distribution. The area under the curve and above any range of values is the proportion of all observations that fall in that range.

More on Density Curves

- a density curve describes the theoretical pattern or distribution of the data that can be obtained. The description is in terms of a mathematical function
- for the theory to coincide with the observations, the histogram and the density curve should be similar

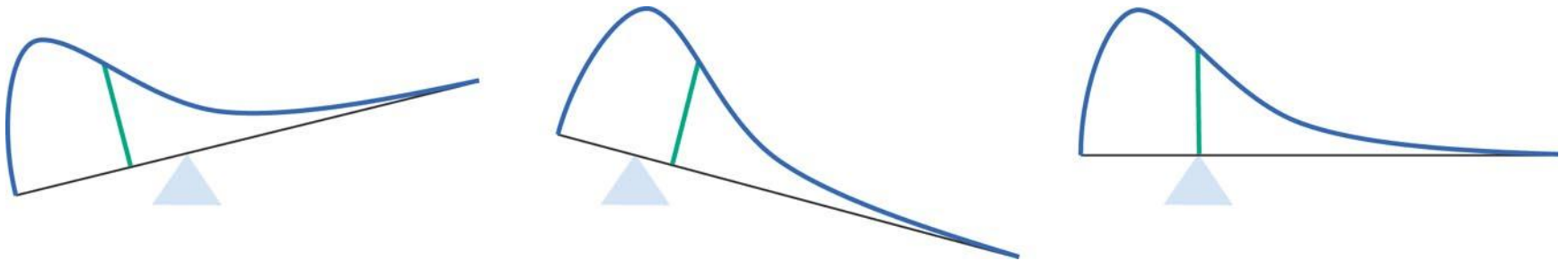
MEDIAN AND MEAN OF A DENSITY CURVE

The **median** of a density curve is the equal-areas point, the point that divides the area under the curve in half.

The **mean** of a density curve is the balance point, at which the curve would balance if made of solid material.

The median and mean are the same for a symmetric density curve. They both lie at the center of the curve. The mean of a skewed curve is pulled away from the median in the direction of the long tail.

The Mean as the Centre of Gravity of the Density Curve



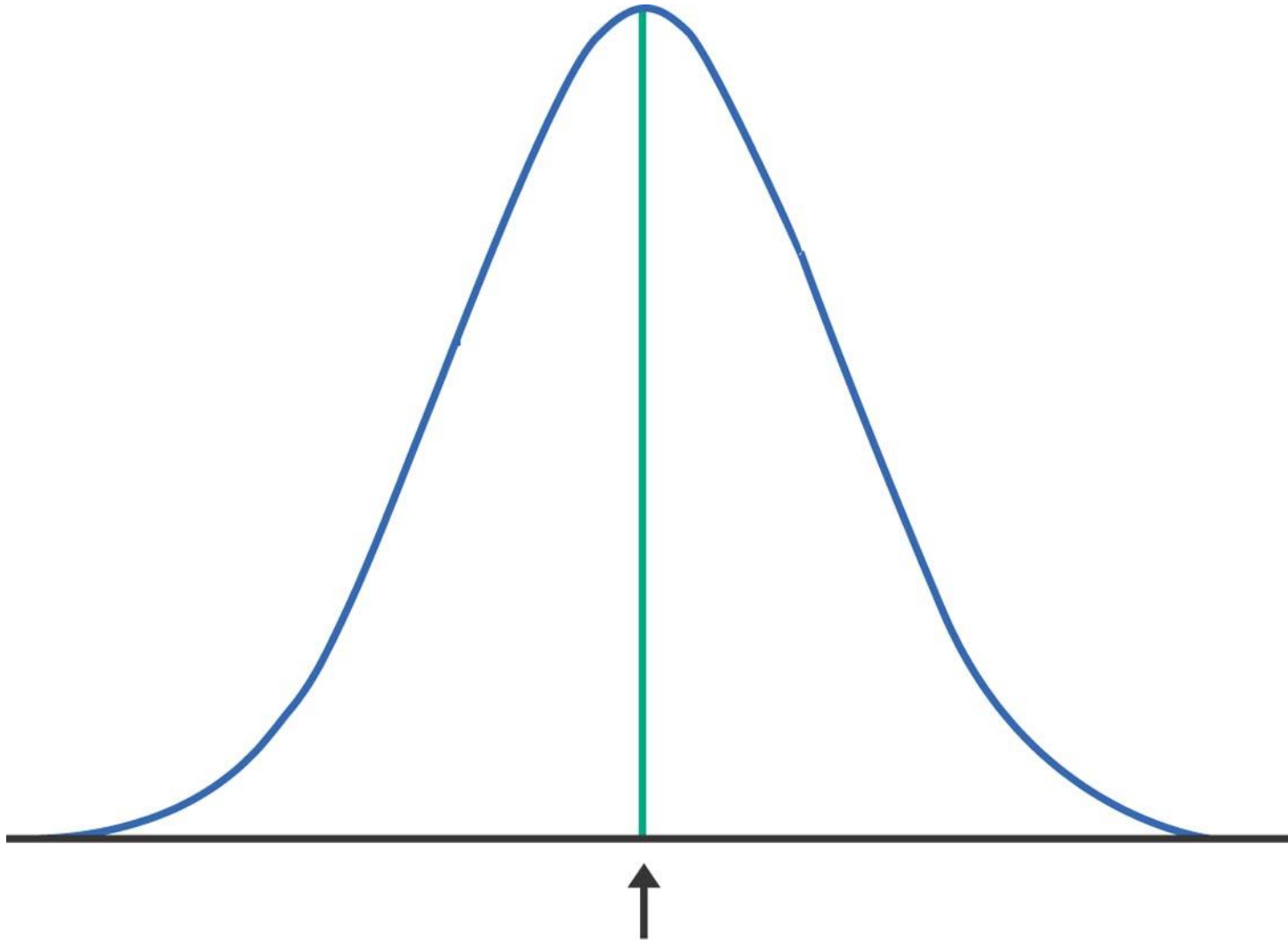
Properties of Means and Medians of Density Curves

For a symmetric density curve

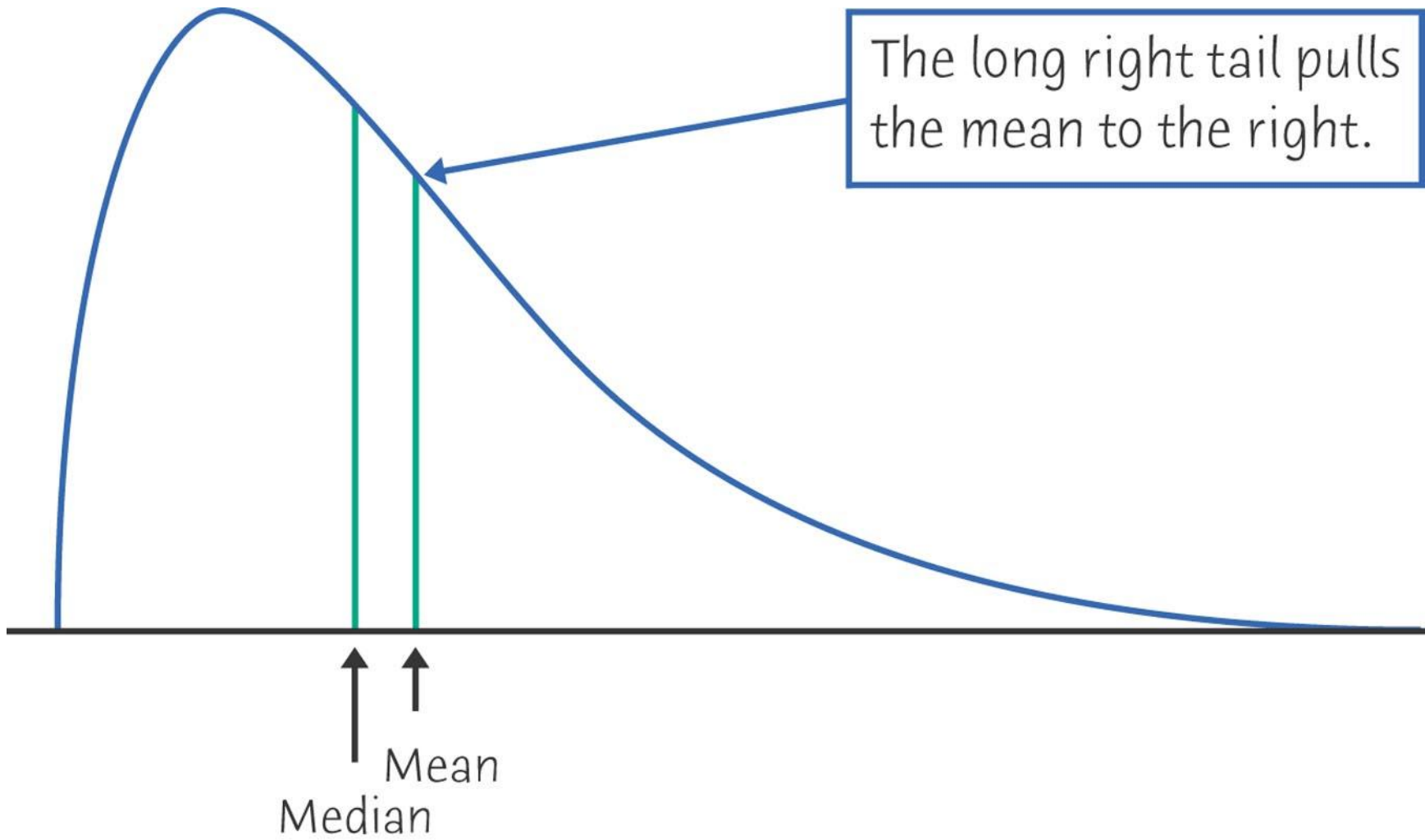
- mean = median

For an asymmetric or skewed density curve

- mean $>$ median if the long tail of the curve is in the right of the distribution (skewed to the right)
- mean $<$ median if the long tail of the curve is in the left of the distribution (skewed to the left)



Median and mean



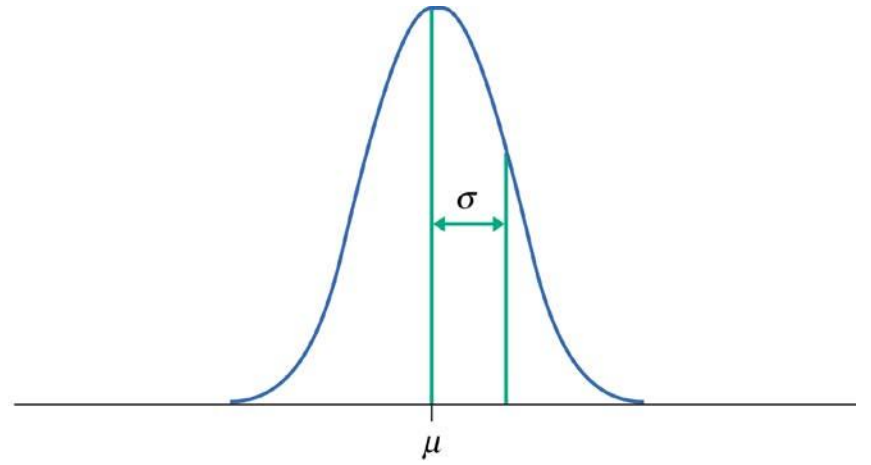
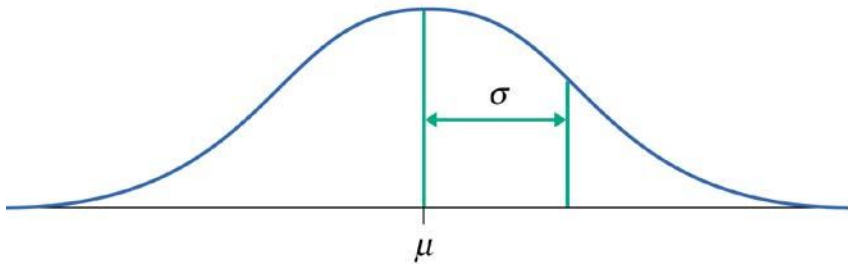
NORMAL DISTRIBUTIONS

A **Normal distribution** is described by a Normal density curve. Any particular Normal distribution is completely specified by two numbers, its mean and standard deviation.

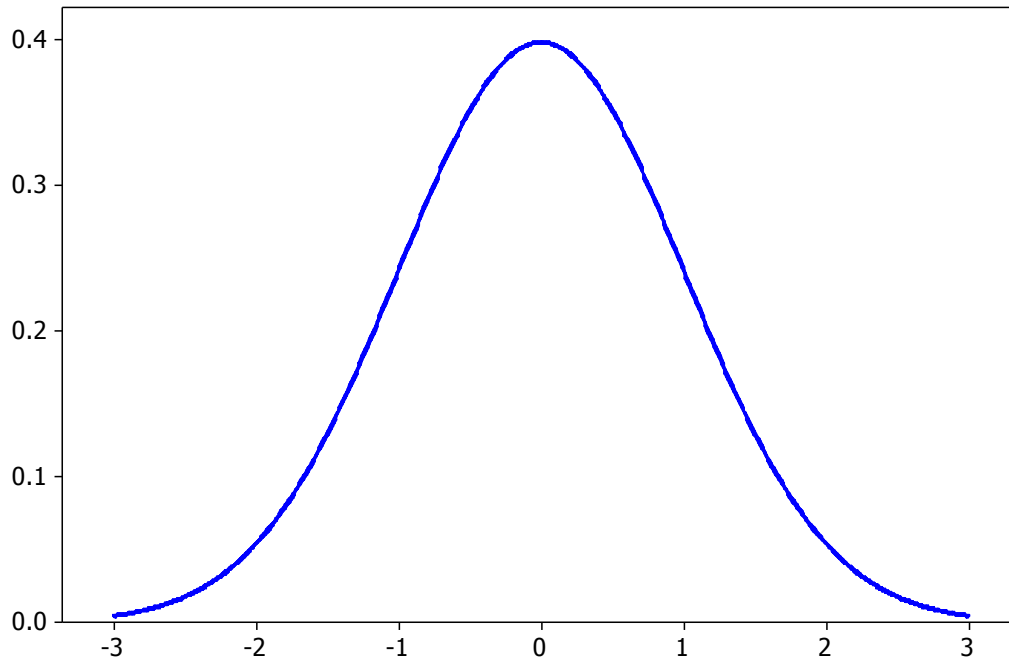
The mean of a Normal distribution is at the center of the symmetric Normal curve. The standard deviation is the distance from the center to the change-of-curvature points on either side.

More on Normal Distributions

- the mean μ of the distribution determines the location of the distribution or the centre of the bell-shaped curve
- the standard deviation (SD) σ determines the shape or how spread out the bell shape is.
- a normal distribution is completely described by the mean and SD.



Standard Normal Distribution

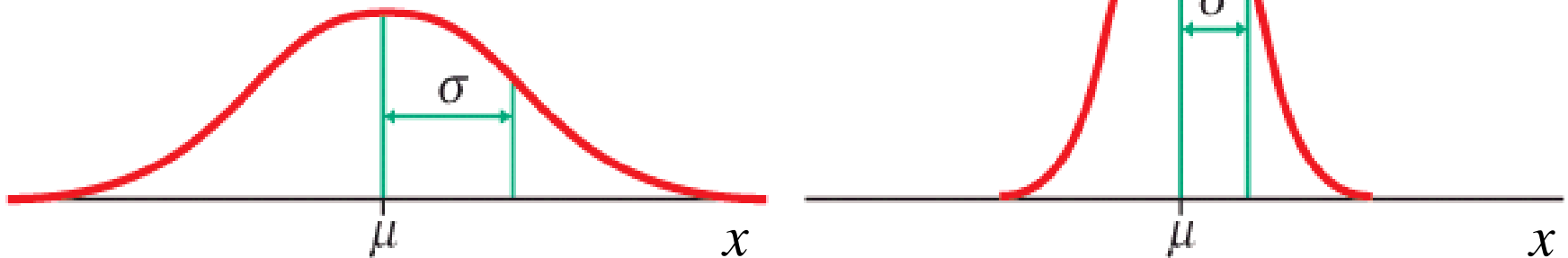


- bell-shaped curve centered around 0 with most of the curve between -3 and +3
- symmetric around 0 so that the mean and median are both 0
- has standard deviation equal to 1

Normal Distributions: a Little Math

Normal—or Gaussian—distributions are a family of symmetrical, bell-shaped density curves

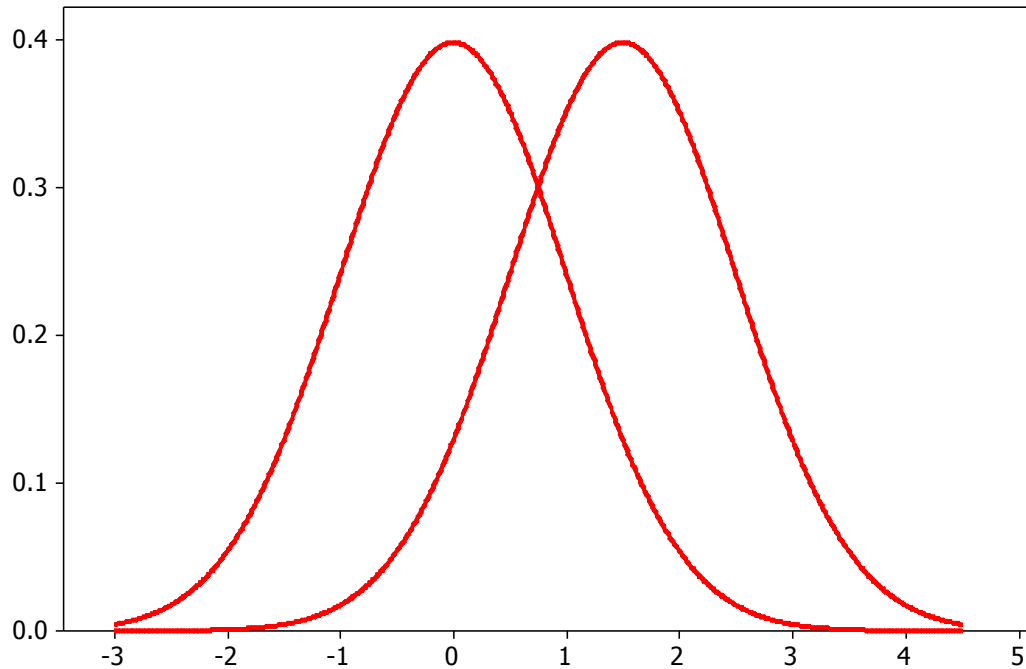
$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



$e = 2.71828\dots$ The base of the natural logarithm

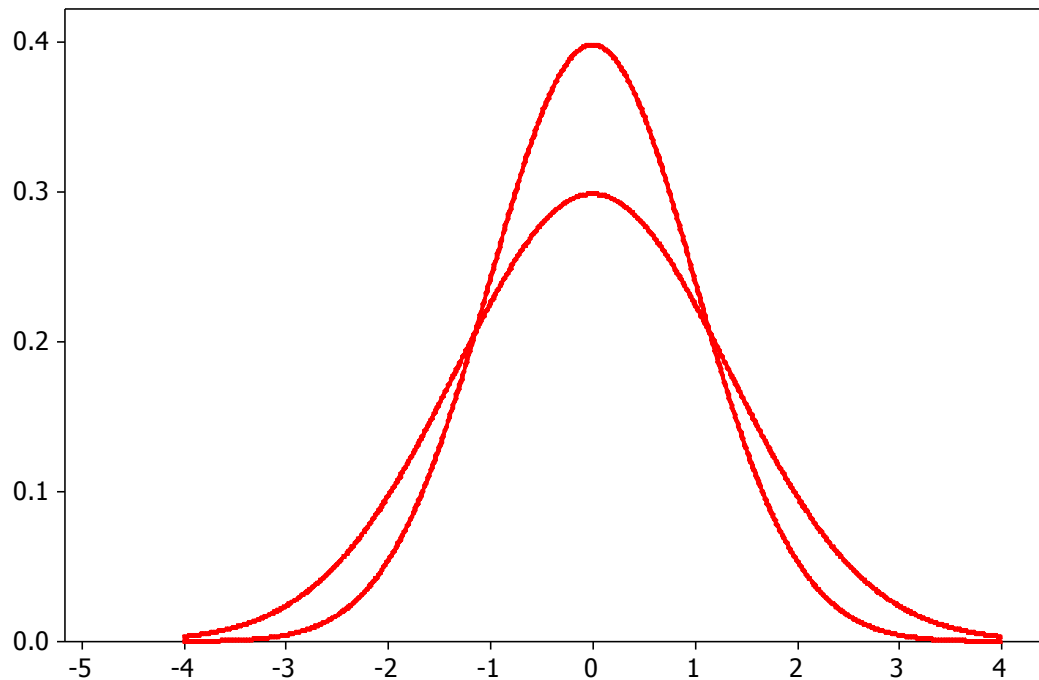
$\pi = \text{pi} = 3.14159\dots$

Same Standard Deviations, Different Means



- the curve on the right has a larger mean than the curve on the left
- the amount of the shift is equal to the difference in the means

Same Means, Different Standard Deviations



- the lower curve has a larger standard deviation
- the spread of the curve increases with the standard deviation

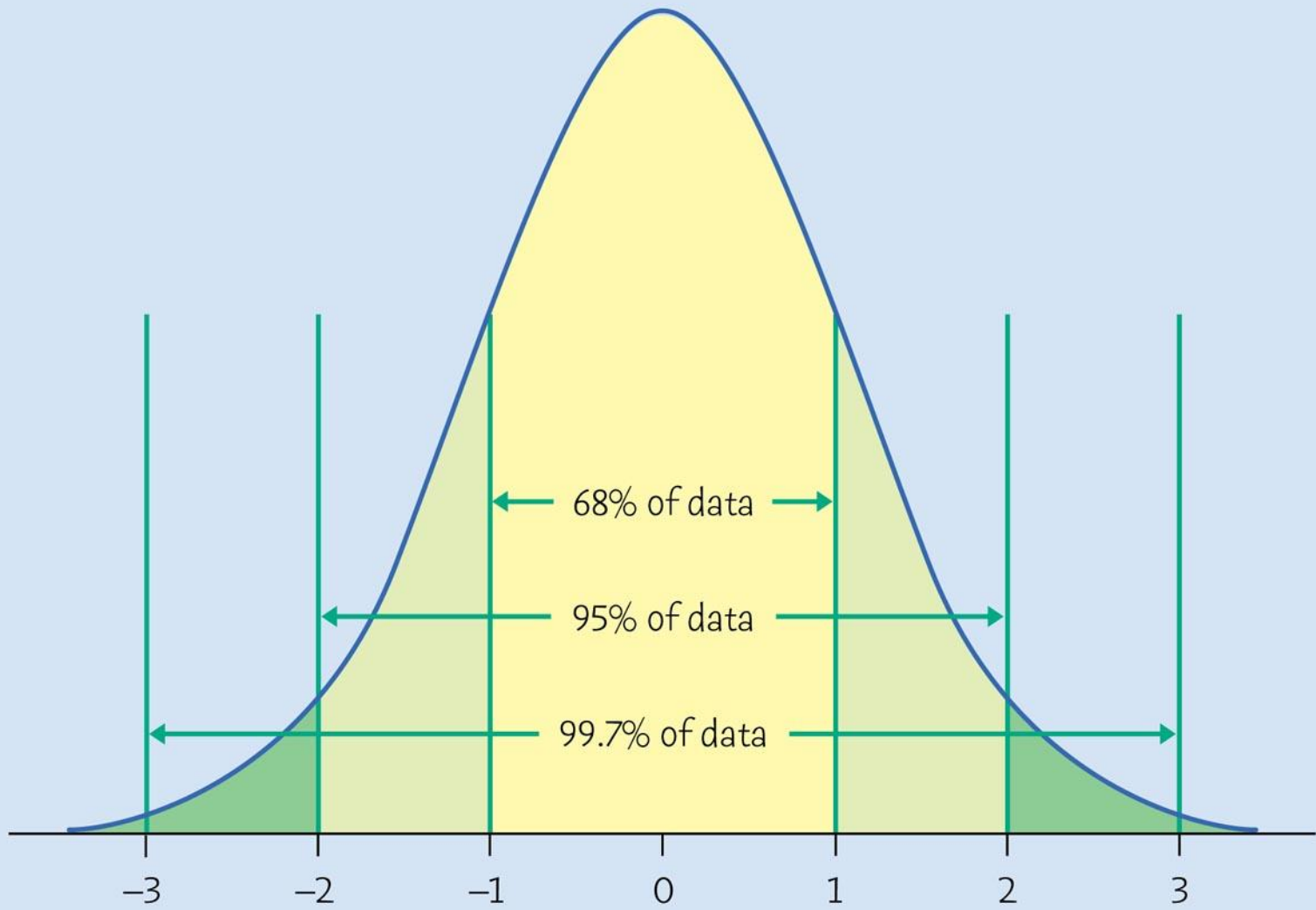
THE 68-95-99.7 RULE

In the Normal distribution with mean μ and standard deviation σ :

- Approximately **68%** of the observations fall within σ of the mean μ .
- Approximately **95%** of the observations fall within 2σ of μ .
- Approximately **99.7%** of the observations fall within 3σ of μ .

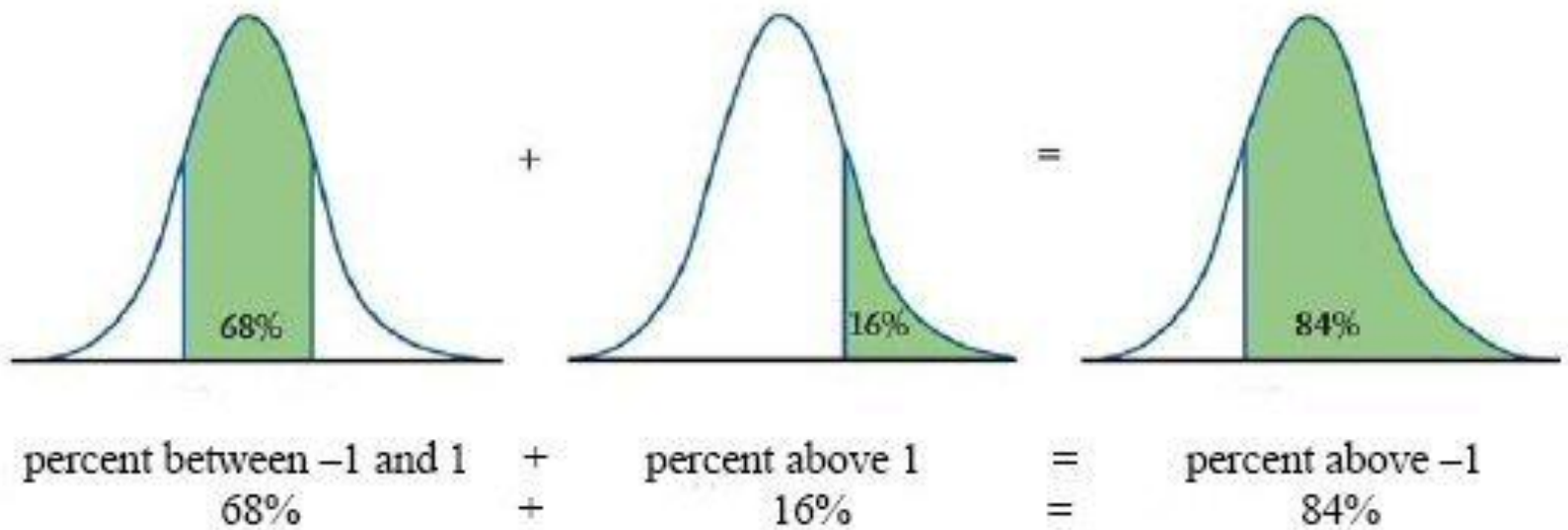
Same Rule, Different Expression

- 68% of the area under the curve is between $\mu - \sigma$ and $\mu + \sigma$
- 95% of the area under the curve is between $\mu - 2\sigma$ and $\mu + 2\sigma$
- 99.7% of the area under the curve is between $\mu - 3\sigma$ and $\mu + 3\sigma$



Standard deviations

Example Calculation Using the 68-95-99.7 Rule



STANDARDIZING AND z-SCORES

If x is an observation from a distribution that has mean μ and standard deviation σ , the **standardized value** of x is

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

A standardized value is often called a **z-score**.

STANDARD NORMAL DISTRIBUTION

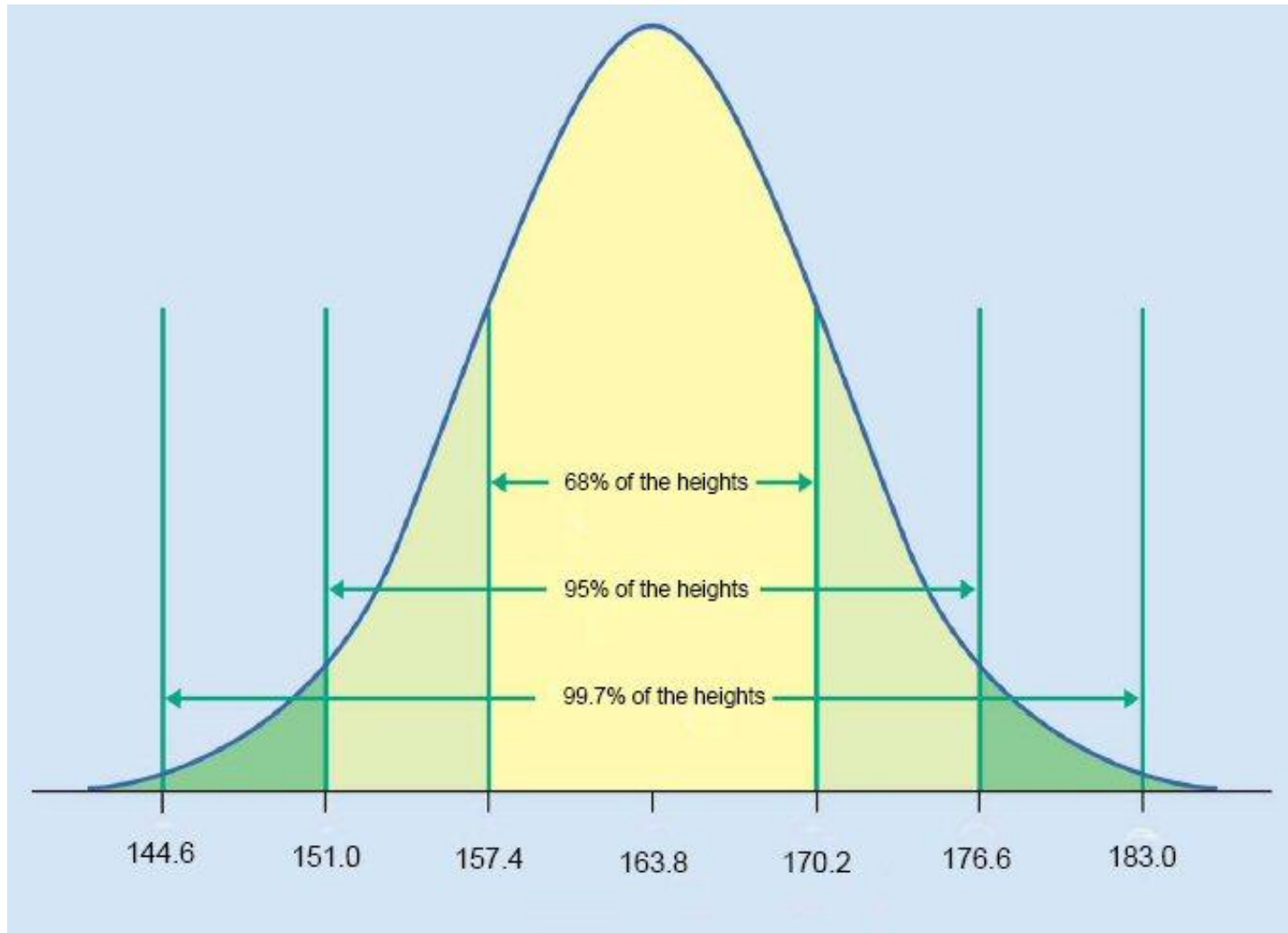
The **standard Normal distribution** is the Normal distribution $N(0, 1)$ with mean 0 and standard deviation 1.

If a variable x has any Normal distribution $N(\mu, \sigma)$ with mean μ and standard deviation σ , then the standardized variable

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

has the standard Normal distribution.

Example: Heights of Women



Assume women's heights have a normal distribution with mean 163.8 cm and a standard deviation of 6.4 cm

Then

$$\frac{\text{height} - 163.8}{6.4}$$

has a standard normal distribution

USING TABLE A TO FIND NORMAL PROPORTIONS

1. State the problem in terms of the observed variable x . **Draw a picture** that shows the proportion you want in terms of cumulative proportions.
2. **Standardize** x to restate the problem in terms of a standard Normal variable z .
3. Use **Table A** and the fact that the total area under the curve is 1 to find the required area under the standard Normal curve.

Table A

Table A gives the area under the standard Normal curve to the left of any z-value.

TABLE A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-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

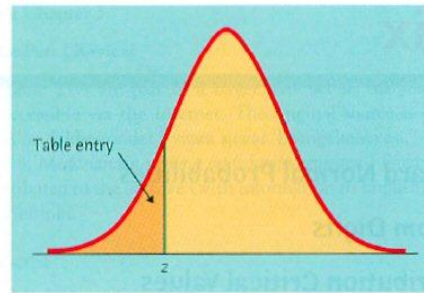


Table entry for z is the area under the standard Normal curve to the left of z .

Table for
z-values < 0

TABLE A Standard Normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-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
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

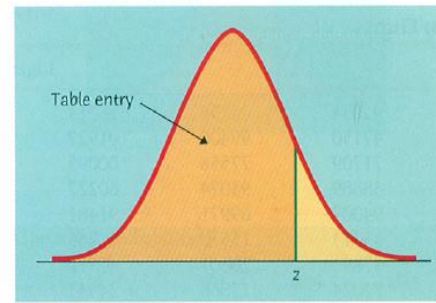


Table entry for z is the area under the standard Normal curve to the left of z .

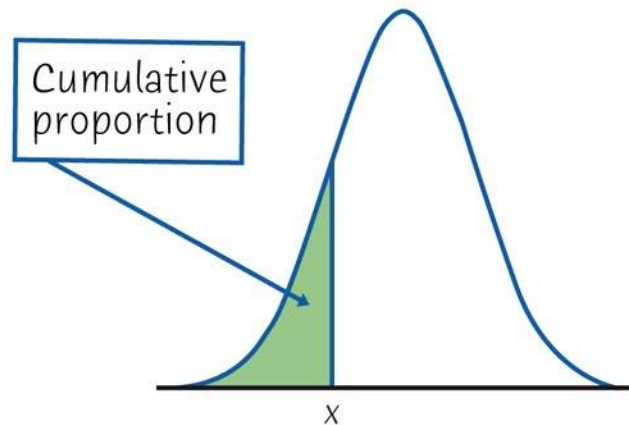
TABLE A Standard Normal probabilities (continued)

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
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9988	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

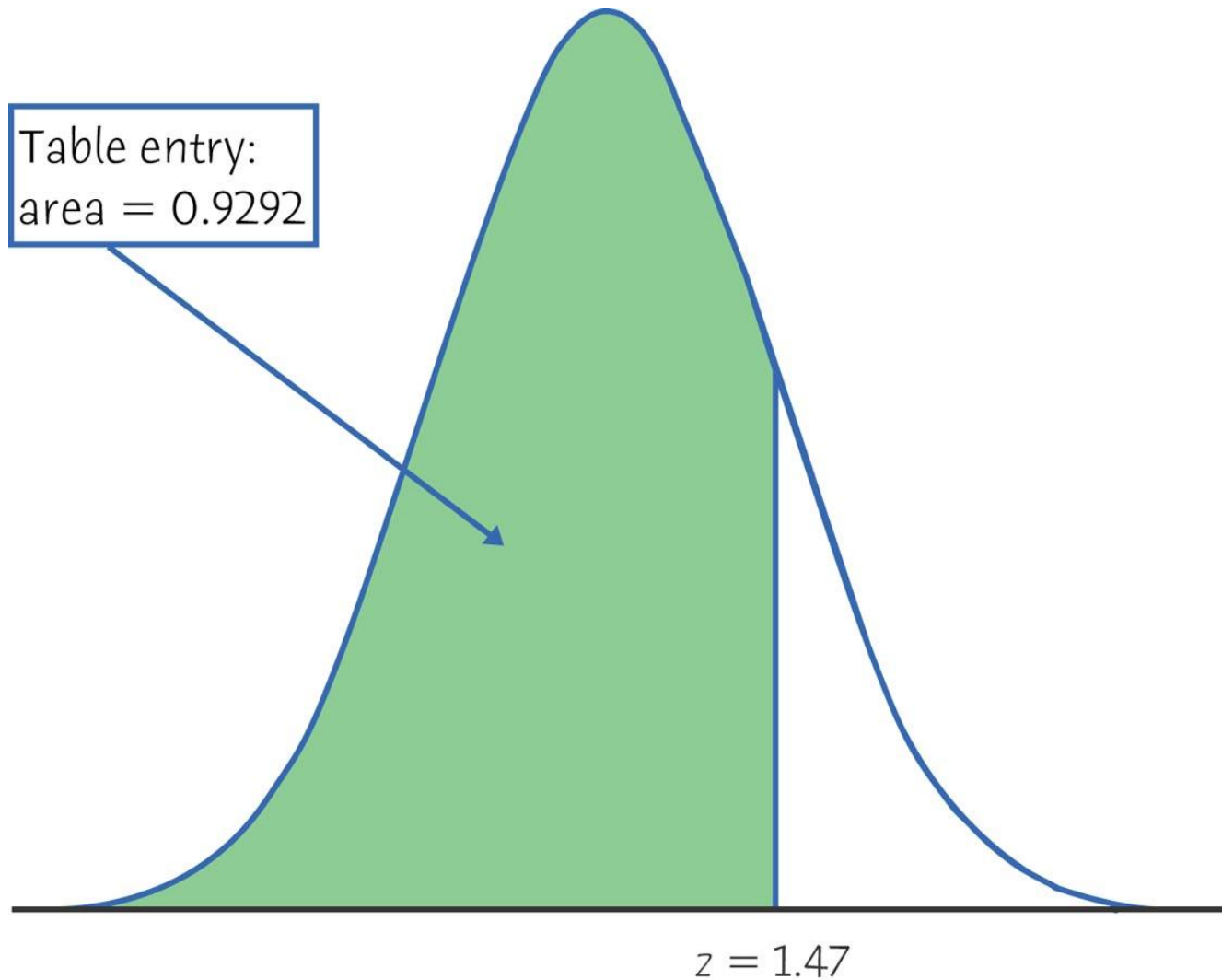
Table for
z-values > 0

CUMULATIVE PROPORTIONS

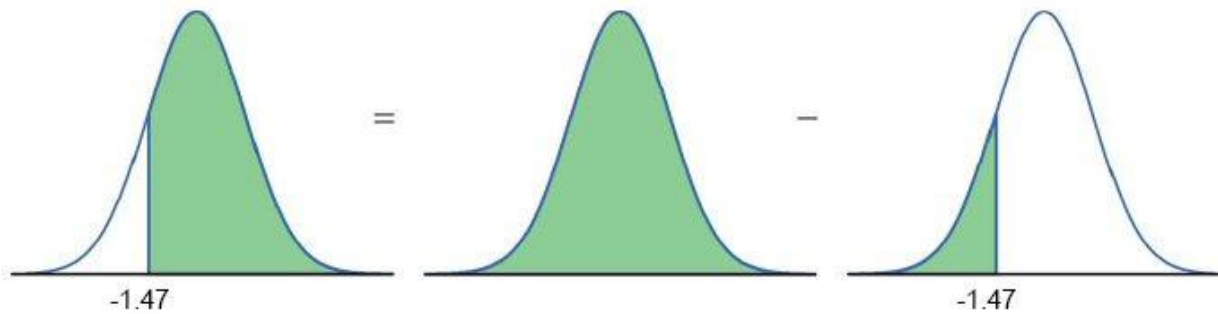
The **cumulative proportion** for a value x in a distribution is the proportion of observations in the distribution that lie at or below x .



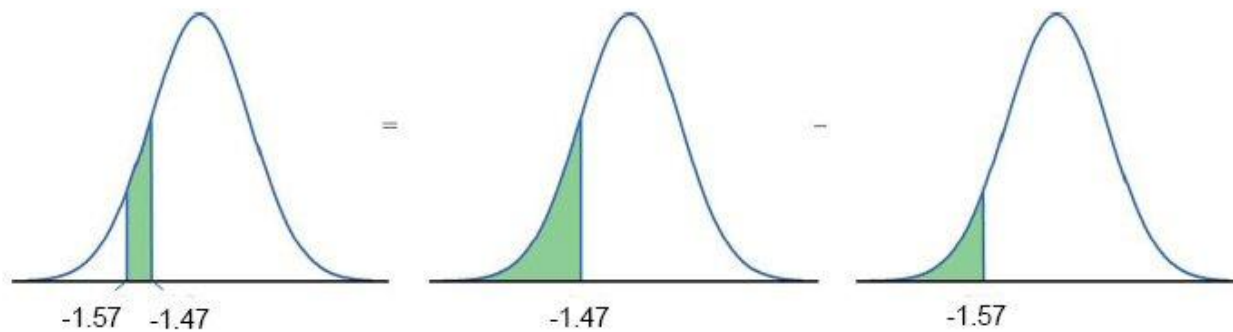
Tables for the Normal Distribution are Based on $N(0,1)$



Any area can be obtained from the basic table calculation



$$1 - 0.0708 \\ = 0.9292$$



$$0.0708 - 0.0582 \\ = 0.0126$$

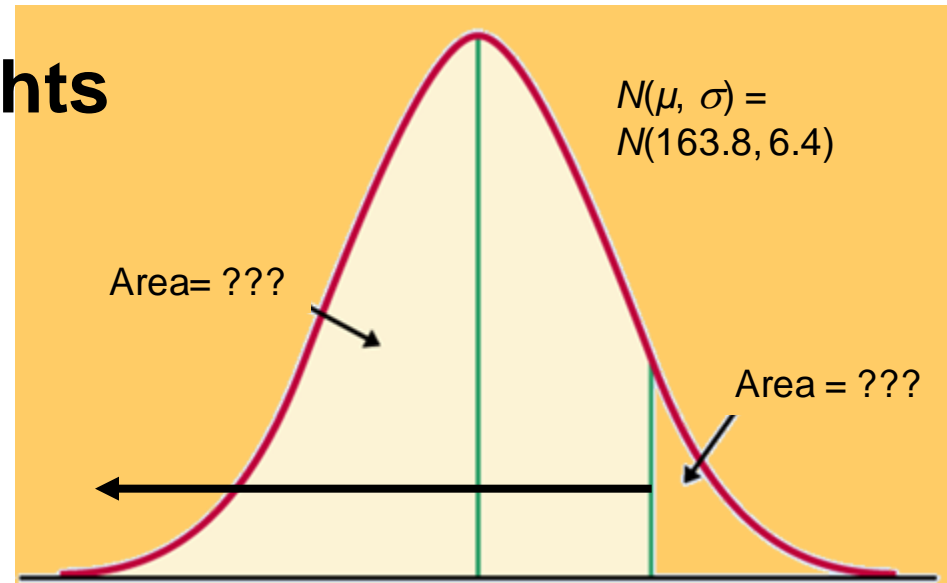
Example: Women heights

Women's heights follow the $N(163.8, 6.4)$ distribution. What percent of women are shorter than 170.2 cm tall?

mean $\mu = 163.8$ cm

standard deviation $\sigma = 6.4$ cm

x (height) = 170.2 cm



$$\begin{aligned} \mu &= 163.8 & x &= 170.2 \\ z &= 0 & z &= 1 \end{aligned}$$

We calculate z , the standardized value of x :

$$z = \frac{(x - \mu)}{\sigma}, \quad z = \frac{(170.2 - 163.8)}{6.4} = \frac{6.4}{6.4} = 1 \Rightarrow 1 \text{ stand.dev. from mean}$$

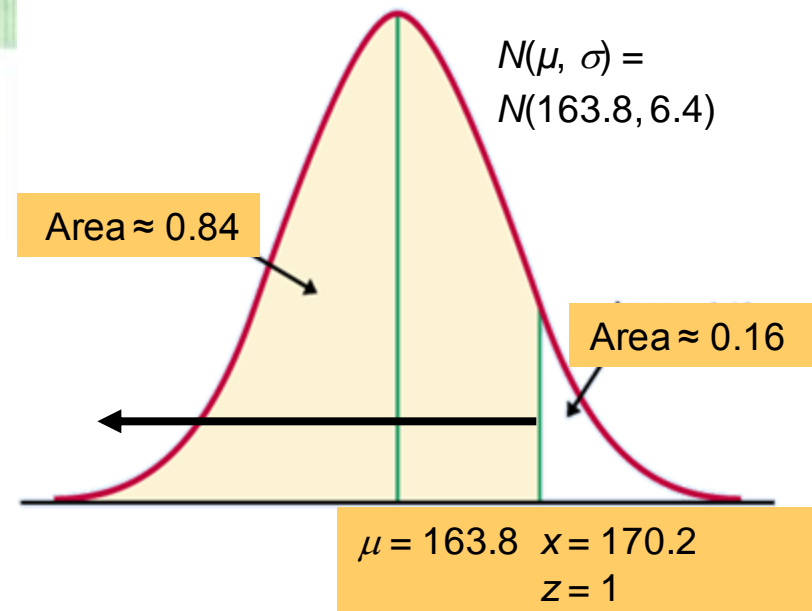
Because of the 68-95-99.7 rule, we can conclude that the percent of women shorter than 170.2 cm should be, approximately, $.68 + \text{half of } (1 - .68) = .84$, or 84%.

Percent of women shorter than 170.2 cm: Exact Calculation

TABLE A Standard normal probabilities (continued)

<i>z</i>	.00	.01	.02	.03	.04
0.0	.5000	.5040	.5080	.5120	.5160
0.1	.5398	.5438	.5478	.5517	.5557
0.2	.5793	.5832	.5871	.5910	.5948
0.3	.6179	.6217	.6255	.6293	.6331
0.4	.6554	.6591	.6628	.6664	.6700
0.5	.6915	.6950	.6985	.7019	.7054
0.6	.7257	.7291	.7324	.7357	.7389
0.7	.7580	.7611	.7642	.7673	.7704
0.8	.7881	.7910	.7939	.7967	.7995
0.9	.8159	.8186	.8212	.8238	.8264
1.0	.8413	.8438	.8461	.8485	.8508
1.1	.8643	.8665	.8686	.8708	.8729
1.2	.8849	.8869	.8888	.8907	.8925
1.3	.9032	.9049	.9066	.9082	.9099

For $z = 1.00$, the area under the standard Normal curve to the left of z is 0.8413.



Conclusion:

84.13% of women are shorter than 170.2 cm.

By subtraction, $1 - 0.8413$, or 15.87%, of women are taller than 170.2 cm.

Finding a value given a proportion

When you know the proportion, but you don't know the x -value that represents the cut-off, you need to use Table A backward.

1. State the problem and draw a picture.
2. Use Table A backward, from the inside out to the margins, to find the corresponding z .
3. Unstandardize to transform z back to the original x scale by using the formula:

$$X = \mu + z\sigma$$

Example: Women's heights

Women's heights follow the $N(163.8, 6.4)$ distribution. What is the 25th percentile for women's heights?

mean $\mu = 163.8$ cm

standard deviation $\sigma = 6.4$ cm

proportion = area under curve = 0.25

We use Table A backward to get the z .

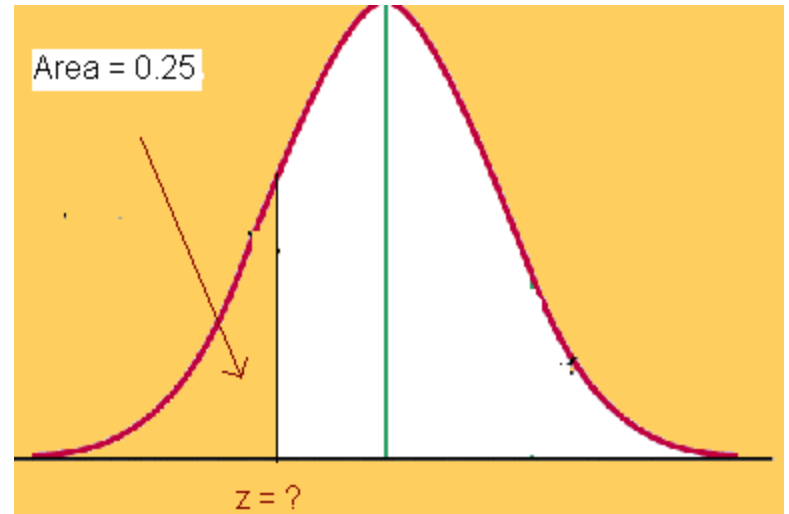
On the left half of Table A (with proportions ≤ 0.5), we find that a proportion of 0.25 is between $z = -0.67$ and -0.68 .

We'll use $z = -0.67$.

Now convert back to x :

$$x = \mu + z\sigma = 163.8 + (-0.67)(6.4) = 159.512 \text{ cm}$$

The 25th percentile for women's heights is about 159.5 cm



Applying the Concepts

A bank's loan officer rates applicants for credit. The ratings are normally distributed with a mean of 200 and a standard deviation of 50. If an applicant is randomly selected, find the probability of a rating that is between 170 and 220.

$$\begin{aligned} P(170 \leq Y \leq 220) &= P\left(\frac{170 - 200}{50} \leq z \leq \frac{220 - 200}{50}\right) \\ &= P(-0.6 \leq Z \leq 0.4) \\ &= 0.6554 - 0.2743 = 0.3811 \end{aligned}$$

Applying the Concepts

As reported in *Runner's World* magazine, the times of the finishers in the New York City 10 km run are normally distributed with a mean of 61 minutes and a standard deviation of 9 minutes. Let Y be the time of a randomly selected finisher. Find $P(Y > 75)$.

$$\begin{aligned} P(Y > 75) &= P\left(Z > \frac{75 - 61}{9}\right) = P(Z > 1.56) \\ &= 1 - 0.9406 = 0.0594 \end{aligned}$$

Applying the Concepts

The life of a calculator manufactured by Intal corporation has a normal distribution with a mean of 54 months and a standard deviation of one year. What should the warranty be to replace a malfunctioning calculator if the company does not want to replace more than 1% of all the calculators sold?

Want k such that $P(Y > k) = 0.01$. Then

$$P\left(Z > \frac{k-54}{1}\right) = 0.01 \text{ or } k - 54 = 2.33 \text{ or } k = 56.33$$

Applying the Concepts

A soft-drink machine can be regulated so that it discharges an average of μ oz. per cup. If the ounces of fill are normally distributed, with a standard deviation of 0.4 oz., what value should μ be set at so that 6-oz. cups will overflow only 2% of the time?

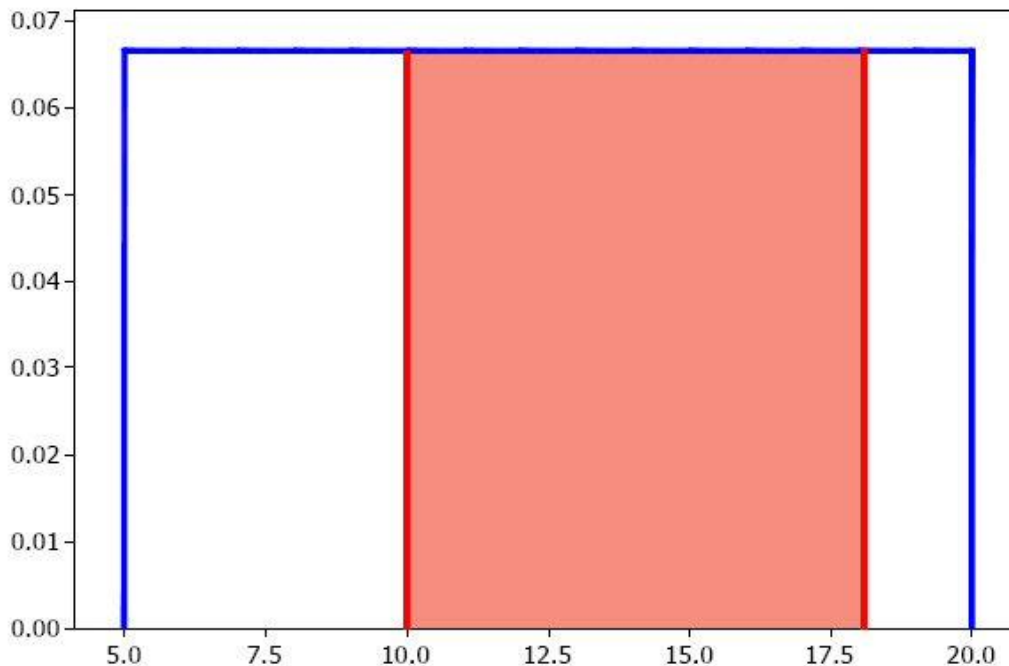
$$\text{Want } P(Y > 6) = 0.02 = P\left(Z > \frac{6-\mu}{0.4}\right) = 0.02$$

$$\text{Therefore } \frac{6-\mu}{0.4} = 2.05 \text{ or } \mu = 5.18.$$

Other Densities: Example

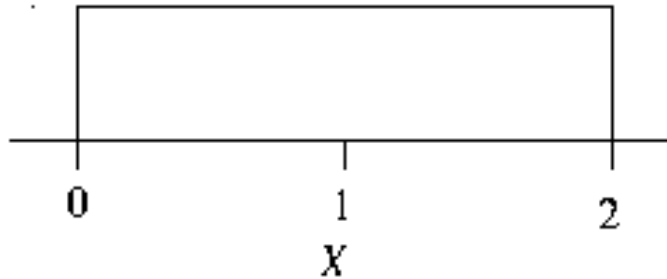
Uniform Distribution

Suppose that the time it takes a student to eat their lunch is uniformly distributed between 5 and 20 minutes. What percentage of the time does it take a student more than 10 but less than 18 minutes to eat lunch?



Another Uniform Density

The probability density of a random variable X is given in the figure below.



What is the probability that X is ***at least*** 1.5?