Math 11
Calculus-Based Introductory Probability and Statistics

Eddie Aamari
S.E.W. Assistant Professor

eaamari@ucsd.edu
math.ucsd.edu/~eaamari/
AP&M 5880A

Today:

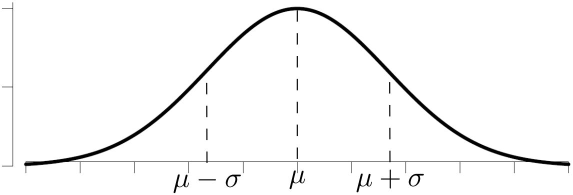
- The Normal distribution
- z-Score, z-Tables

The Normal Distribution

The density function of the **normal distribution** is:

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

where μ is the mean, and σ is the standard deviation.



Notation: $X = N(\mu, \sigma)$.

Other name: Gaussian distribution

Importance of the Normal Distribution

The Normal model is the most important continuous random variable in all of modern statistics.

Roughly speaking, this comes from the fact that any time some quantity is the combination of many independent factors, then this quantity will follow a normal model.

Importance of the Normal Distribution

The Normal model is the most important continuous random variable in all of modern statistics.

Roughly speaking, this comes from the fact that any time some quantity is the combination of many independent factors, then this quantity will follow a normal model.

Examples:

- Human heights in the US (Female: $\mu = 65$ in, $\sigma = 3.5$ in)
- Diastolic blood pressure ($\mu = 77$ mm Hg, $\sigma = 5.5$ mm Hg)
- IQ scores ($\mu = 100$, $\sigma = 15$)

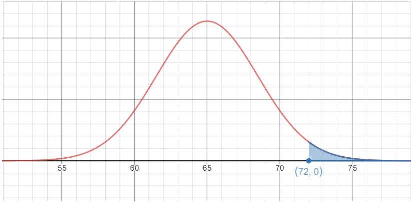
Normal Distribution: Example

Suppose the height of US women is normally distributed with a mean $\mu = 65$ inches and standard deviation $\sigma = 3.5$ inches.
What is the probability the next woman you see has a height over 72 inches?

Normal Distribution: Example

Suppose the height of US women is normally distributed with a mean $\mu = 65$ inches and standard deviation $\sigma = 3.5$ inches.
What is the probability the next woman you see has a height over 72 inches?
Let $X = N(65, 3.5)$. We want

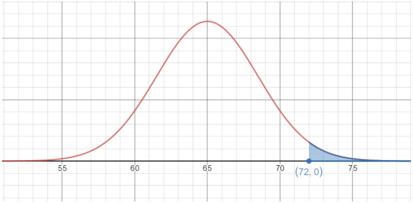
$$P(X \geq 72) = \int_{72}^{\infty} \frac{1}{3.5\sqrt{2\pi}} e^{-\frac{(x-65)^2}{2\times(3.5)^2}} dx$$



Normal Distribution: Example

Suppose the height of US women is normally distributed with a mean $\mu = 65$ inches and standard deviation $\sigma = 3.5$ inches.
What is the probability the next woman you see has a height over 72 inches?
Let $X = N(65, 3.5)$. We want

$$P(X \geq 72) = \int_{72}^{\infty} \frac{1}{3.5\sqrt{2\pi}} e^{-\frac{(x-65)^2}{2\times(3.5)^2}} dx$$
$$= \ldots$$

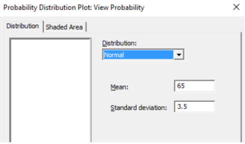


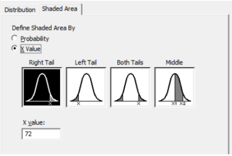
That's a hard integral! This is the first example of a distribution where probabilities cannot be found by hand.

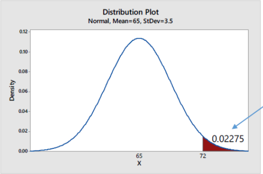
Technology and the z-table

Technology:

Graph >> Probability Distribution Plot >> View Probability





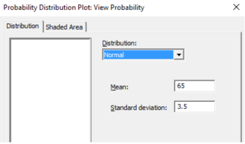


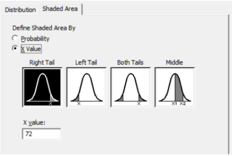
There is a 2.275% chance a random woman is over 6 feet tall.

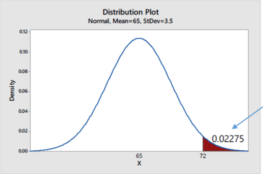
Technology and the z-table

Technology:

Graph >> Probability Distribution Plot >> View Probability







There is a 2.275% chance a random woman is over 6 feet tall.

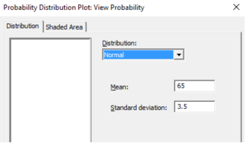
The z-Table:

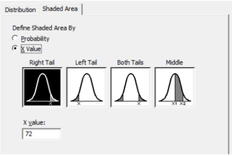
This is a way to look up the areas under part of a normal curve and get the answer 0.02275.

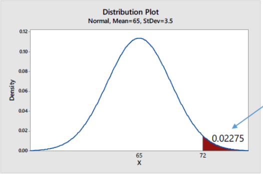
Technology and the z-table

Technology:

Graph >> Probability Distribution Plot >> View Probability







There is a 2.275% chance a random woman is over 6 feet tall.

The z-Table:

This is a way to look up the areas under part of a normal curve and get the answer 0.02275.

Trouble:
We don't want a different table for every possible normal curve (recall that the mean can be any number, and the standard deviation can be any positive number!) We need a way to convert any situation modelled by a normal curve into some standard setup.

z-score

You are in charge of admissions at UCSD. One applicant took the SAT and got a 1775. Another took the ACT and got 27. Which would you admit?

z-score

You are in charge of admissions at UCSD. One applicant took the SAT and got a 1775. Another took the ACT and got 27. Which would you admit?

When data are measured on different scales (= have different units), we need a common way to compare them that is unitless.

z-score

You are in charge of admissions at UCSD. One applicant took the SAT and got a 1775. Another took the ACT and got 27. Which would you admit?

When data are measured on different scales (= have different units), we need a common way to compare them that is unitless.

The z-score of a data point y from a dataset is $\dfrac{y - \bar{y}}{s_y}$.

z-score

You are in charge of admissions at UCSD. One applicant took the SAT and got a 1775. Another took the ACT and got 27. Which would you admit?

When data are measured on different scales (= have different units), we need a common way to compare them that is unitless.

The z-score of a data point y from a dataset is $\dfrac{y - \bar{y}}{s_y}$.

The z-score:

- Is a unitless idea (units in numerator and denominator cancel)
- Tells you how many standard deviations above the mean some piece of data is.

z-score: Example

You see on Google that the SAT has mean 1500 with SD 250, and the ACT has mean 20.8 with SD 4.8. Which do you admit, the 1775 SAT or 27 ACT?

z-score: Example

You see on Google that the SAT has mean 1500 with SD 250, and the ACT has mean 20.8 with SD 4.8. Which do you admit, the 1775 SAT or 27 ACT?

$$z_{SAT} = \frac{1775 - 1500}{250} = 1.1 \qquad z_{ACT} = \frac{27 - 20.8}{4.8} \simeq 1.29.$$

z-score: Example

You see on Google that the SAT has mean 1500 with SD 250, and the ACT has mean 20.8 with SD 4.8. Which do you admit, the 1775 SAT or 27 ACT?

$$z_{SAT} = \frac{1775 - 1500}{250} = 1.1 \qquad z_{ACT} = \frac{27 - 20.8}{4.8} \simeq 1.29.$$

Assuming the two tests are equally difficult, you'd ratther admit the ACT person.

z-score: Example

You see on Google that the SAT has mean 1500 with SD 250, and the ACT has mean 20.8 with SD 4.8. Which do you admit, the 1775 SAT or 27 ACT?

$$z_{SAT} = \frac{1775 - 1500}{250} = 1.1 \qquad z_{ACT} = \frac{27 - 20.8}{4.8} \simeq 1.29.$$

Assuming the two tests are equally difficult, you'd ratther admit the ACT person.

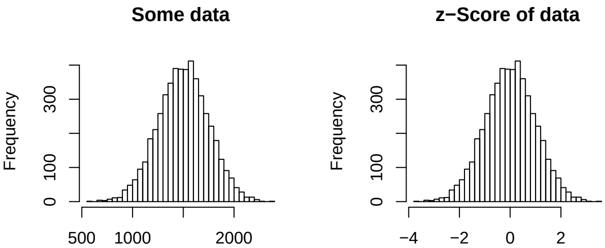
z-scores provide a single "ruler idea" to measure all phenomena, erasing the effect of units.

The z-score says how extreme a data point is relative to its own data set.

z-score

Let's take a data set and find the z-score for every data point.

$$\mu_{SAT} = 1500, \qquad \sigma_{SAT} = 250$$



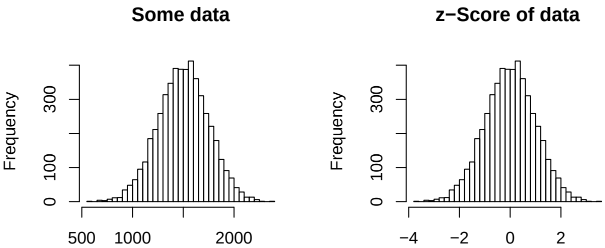
x

$$z_x = \frac{x - \mu_{SAT}}{\sigma_{SAT}}$$

z-score

Let's take a data set and find the z-score for every data point.

$$\mu_{SAT} = 1500, \qquad \sigma_{SAT} = 250$$



x

$$z_x = \frac{x - \mu_{SAT}}{\sigma_{SAT}}$$

It appears:

- The new histogram is similar
- The new mean is 0.
- The new standard deviation is 1.

z-Score Summary

- z-scores allow us to compare two data points from different data sets (with different centers and spreads) and get a sense for which datum is more extreme relative to its own data set.

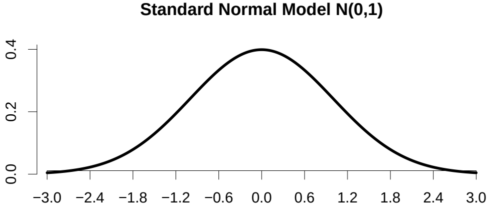
z-Score Summary

- z-scores allow us to compare two data points from different data sets (with different centers and spreads) and get a sense for which datum is more extreme relative to its own data set.

- z-scores allow us to rescale a given data set so it has mean 0 and standard deviation 1.
  In the case of probability models, this allows us to think about whole families of curves using a single "standardized" model.

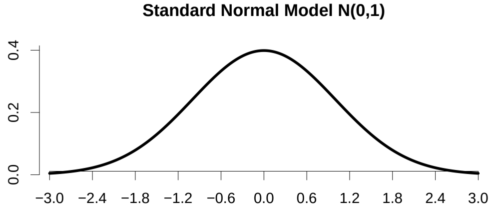
All Roads Lead To … The Standard Normal Model

By rescaling data with the z-score, we turn all Normal models $N(\mu, \sigma)$ into a single one: the **Standard Normal Model** $N(0, 1)$.



All Roads Lead To … The Standard Normal Model

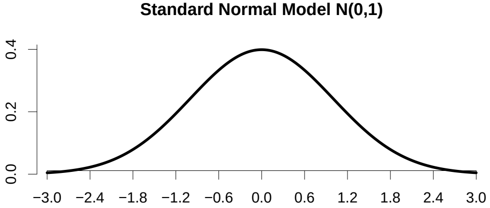
By rescaling data with the z-score, we turn all Normal models $N(\mu, \sigma)$ into a single one: the **Standard Normal Model** $N(0, 1)$.



Any question in the original setting can be reframed as a question on the standard Normal model $N(0, 1)$.

All Roads Lead To … The Standard Normal Model

By rescaling data with the z-score, we turn all Normal models $N(\mu, \sigma)$ into a single one: the **Standard Normal Model** $N(0, 1)$.



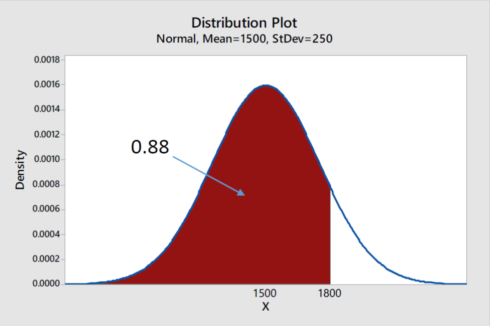
Any question in the original setting can be reframed as a question on the standard Normal model $N(0, 1)$.

If we understand $N(0, 1)$, we understand all the Normal models.

One Question, Many Ways to Solve

Find the percentage of students that has an SAT score below 1800.

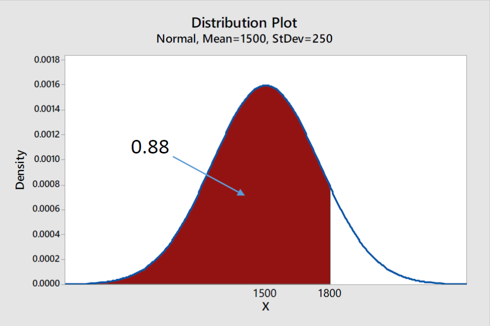
A priori, we could find this probability by finding the area under the density curve using an integral. For normal distributions, this is too difficult to do by hand.



One Question, Many Ways to Solve

Find the percentage of students that has an SAT score below 1800.

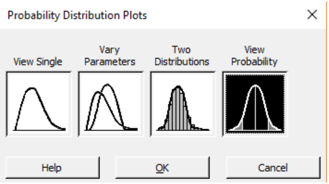
A priori, we could find this probability by finding the area under the density curve using an integral. For normal distributions, this is too difficult to do by hand.



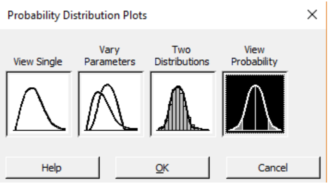
Option 1: Original setting + technology. $X = N(1500, 250)$, and we want $P(X \leq 1800)$.

Doing This in Minitab

Go to Graph » Probability Distribution Plot » View Probability

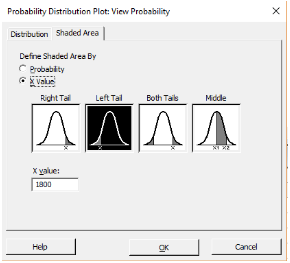


Doing This in Minitab

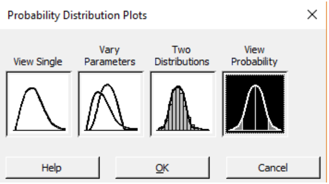


Go to Graph » Probability Distribution Plot » View Probability

Set up the distribution you want (here: Normal with mean 1500 and SD 250), then click on "Shaded Area" tab. Specify the area you are trying to find.

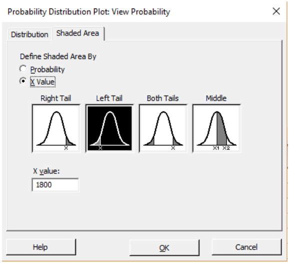


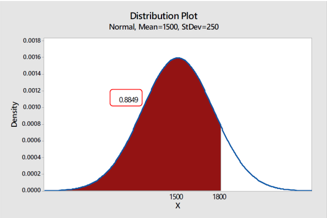
Doing This in Minitab



Go to Graph » Probability Distribution Plot » View Probability

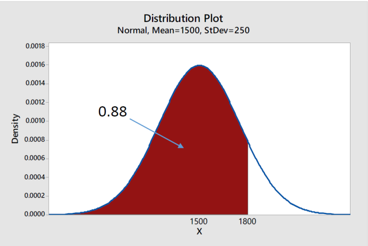
Set up the distribution you want (here: Normal with mean 1500 and SD 250), then click on “Shaded Area” tab. Specify the area you are trying to find.

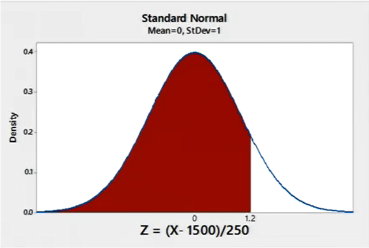




You get a nice plot with the answer displayed. Note that this does NOT require converting 1800 to a z-score.

One Question, Many Ways to Solve

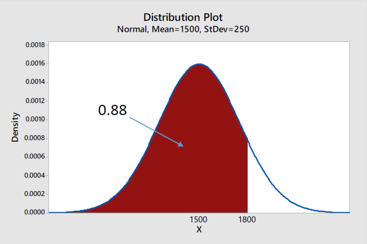


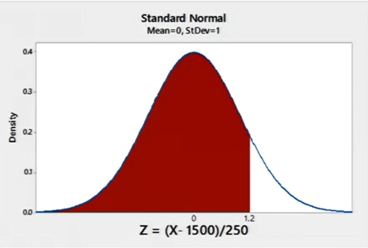


Option 2: Standardized setting + tables.
We notice than

$$P(X \leq 1800) = P\left(\frac{X - 1500}{250} \leq \frac{1800 - 1500}{250}\right) = P(Z \leq 1.2).$$

One Question, Many Ways to Solve



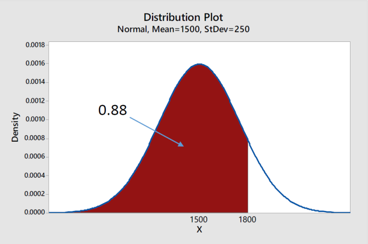


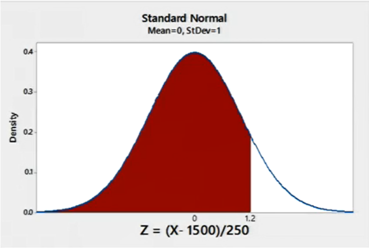
Option 2: Standardized setting + tables.
We notice than

$$P(X \leq 1800) = P\left(\frac{X - 1500}{250} \leq \frac{1800 - 1500}{250}\right) = P(Z \leq 1.2).$$

We want the area less than $z = 1.20$ under the Standard Normal.

One Question, Many Ways to Solve





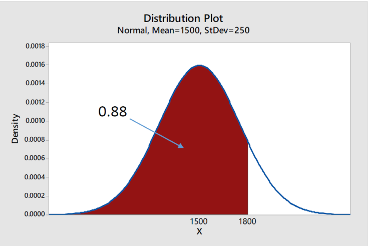
Option 2: Standardized setting + tables.
We notice than

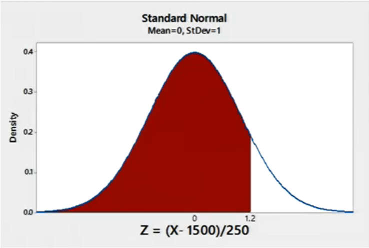
$$P(X \leq 1800) = P\left(\frac{X - 1500}{250} \leq \frac{1800 - 1500}{250}\right) = P(Z \leq 1.2).$$

We want the area less than $z = 1.20$ under the Standard Normal.

Any question about Normal curves can be converted to an equivalent question on the standard normal curve. We just need a lookup table of "areas under the curve" for the standard normal!

One Question, Many Ways to Solve





Option 2: Standardized setting + tables.
We notice than

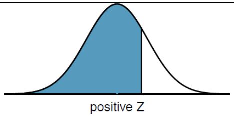
$$P(X \leq 1800) = P\left(\frac{X - 1500}{250} \leq \frac{1800 - 1500}{250}\right) = P(Z \leq 1.2).$$

We want the area less than $z = 1.20$ under the Standard Normal.

Any question about Normal curves can be converted to an equivalent question on the standard normal curve. We just need a lookup table of "areas under the curve" for the standard normal!

On tests, you won't have access to Minitab, so you **will have to** use this approach.

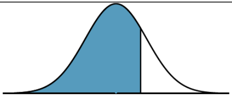
z-Table (see at the end of your book)



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
	Second decimal place of Z									
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

The values in the table are all areas (probabilities) The number along the top and left side are the $z - value$ broken into its two parts.

z-Table (see at the end of your book)

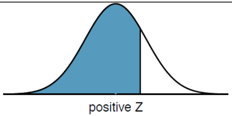


positive Z

Z	Second decimal place of 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

The values in the table are all areas (probabilities) The number along the top and left side are the $z - value$ broken into its two parts.

z-Table (see at the end of your book)



Z	Second decimal place of 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

The values in the table are all areas (probabilities) The number along the top and left side are the $z - value$ broken into its two parts.

Here, $z = 1.20 = 1.2 + 0.00$, so we find $P(Z \leq 1.20) \simeq 0.8849$.

Disneyland

10-year-olds, regardless of gender, have heights (in inches) well-modeled by $N(55, 6)$. What percentage of 10-year-olds can ride Disneyland's Space Mountain, which has a height requirement of 44"?

Disneyland

10-year-olds, regardless of gender, have heights (in inches) well-modeled by $N(55, 6)$. What percentage of 10-year-olds can ride Disneyland's Space Mountain, which has a height requirement of 44"?
Remark: Always define your random variable and draw a picture!

Disneyland

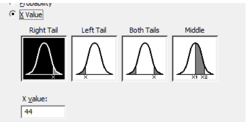
10-year-olds, regardless of gender, have heights (in inches) well-modeled by $N(55, 6)$. What percentage of 10-year-olds can ride Disneyland's Space Mountain, which has a height requirement of 44"?
Remark: Always define your random variable and draw a picture!

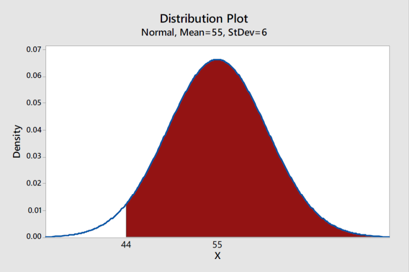
$$\text{Let } X = N(55, 6). \text{ We want } P(X \geq 44).$$

Disneyland

10-year-olds, regardless of gender, have heights (in inches) well-modeled by $N(55, 6)$. What percentage of 10-year-olds can ride Disneyland's Space Mountain, which has a height requirement of 44"?
Remark: Always define your random variable and draw a picture!

$$\text{Let } X = N(55, 6). \text{ We want } P(X \geq 44).$$

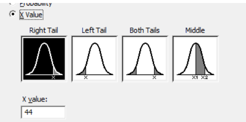


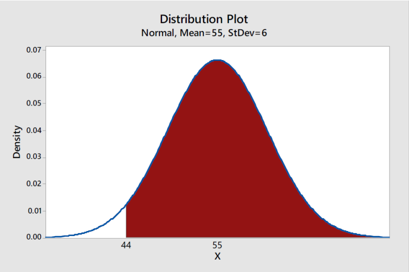


Disneyland

10-year-olds, regardless of gender, have heights (in inches) well-modeled by $N(55, 6)$. What percentage of 10-year-olds can ride Disneyland's Space Mountain, which has a height requirement of 44"?
Remark: Always define your random variable and draw a picture!

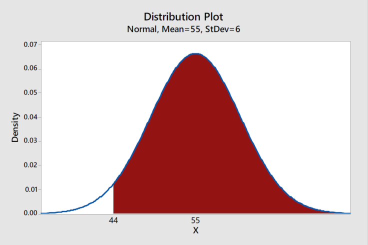
$$\text{Let } X = N(55, 6). \text{ We want } P(X \geq 44).$$

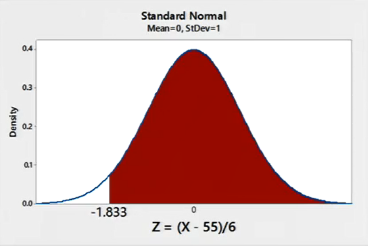




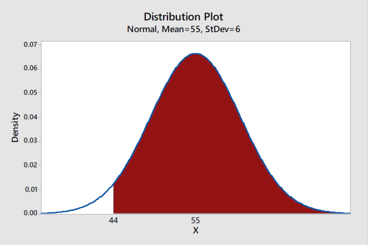
Minitab gives use $p = 0.9666$.
So, 96.7% of 10-yearolds can ride the Space Mountain ride at Disneyland.

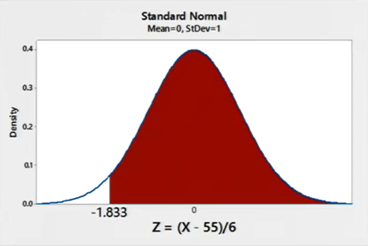
Practice With Tables





Practice With Tables





Writing $Z = \dfrac{X - 55}{6}$, we see that we want

$$
\begin{aligned}
P(X \geq 44) &= P\left(Z \geq \frac{44 - 55}{6}\right) \\
&= P(Z \geq -1.83) \\
&= 1 - P(Z < -1.83)
\end{aligned}
$$

Practice With Tables

Second decimal place of Z										
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Z
0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	−3.4
0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005	−3.3
0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	−3.2
0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010	−3.1
0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013	−3.0
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	−2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	−2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	−2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	−2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	−2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	−2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	−2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	−2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	−2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	−2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	−1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	−1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	−1.7

We read $P(Z < -1.83) \simeq 0.0336$, so

$$P(X \geq 44) \simeq 1 - P(Z < -1.83) \simeq 1 - 0.0336 = 0.9664.$$

Practice With Tables

Second decimal place of Z										
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Z
0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	−3.4
0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005	−3.3
0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	−3.2
0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010	−3.1
0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013	−3.0
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	−2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	−2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	−2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	−2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	−2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	−2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	−2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	−2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	−2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	−2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	−1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	−1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	−1.7

We read $P(Z < -1.83) \simeq 0.0336$, so

$$P(X \geq 44) \simeq 1 - P(Z < -1.83) \simeq 1 - 0.0336 = 0.9664.$$

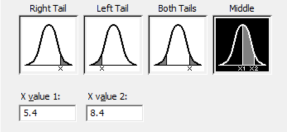
Remark: On these types of problems, never worry about rounding issues or slight difference in answers, we just want approximations.

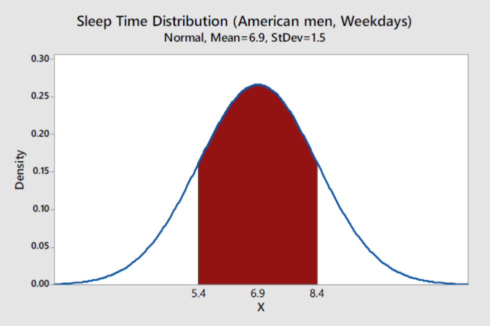
Sleep Time

The sleep times (in hours) of American men on a weekday are well-modeled by $N(6.9, 1.5)$. What percentage of American men are within 1 standard deviation of the mean sleep time?

Sleep Time

The sleep times (in hours) of American men on a weekday are well-modeled by $N(6.9, 1.5)$. What percentage of American men are within 1 standard deviation of the mean sleep time?

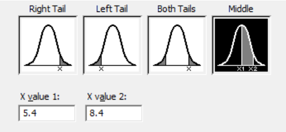


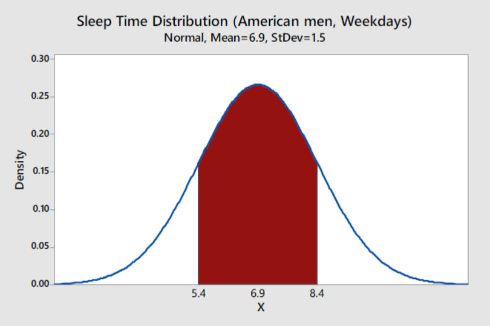


Let $X = N(6.9, 1.5)$. We want $P(5.4 \leq X \leq 8.4)$.

Sleep Time

The sleep times (in hours) of American men on a weekday are well-modeled by $N(6.9, 1.5)$. What percentage of American men are within 1 standard deviation of the mean sleep time?



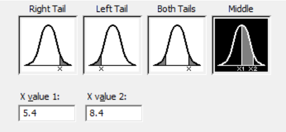


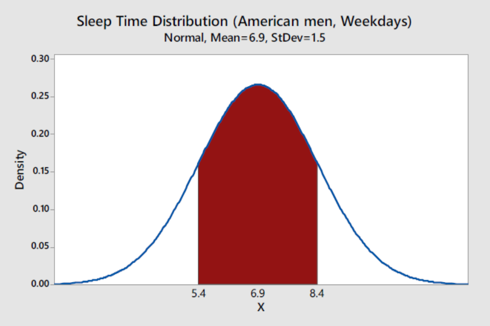
Let $X = N(6.9, 1.5)$. We want $P(5.4 \le X \le 8.4)$.
To find this, we solve

$$P(5.4 \le X \le 8.4) = P(X \le 8.4) - P(X < 5.4).$$

Sleep Time

The sleep times (in hours) of American men on a weekday are well-modeled by $N(6.9, 1.5)$. What percentage of American men are within 1 standard deviation of the mean sleep time?



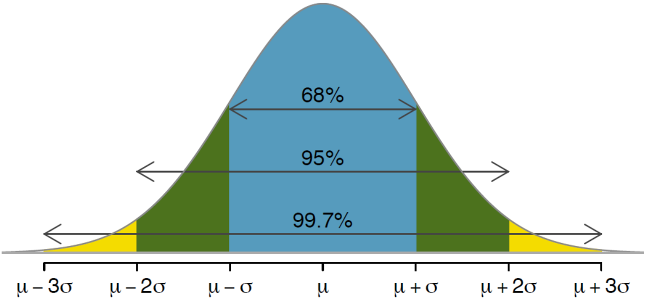


Let $X = N(6.9, 1.5)$. We want $P(5.4 \leq X \leq 8.4)$.
To find this, we solve

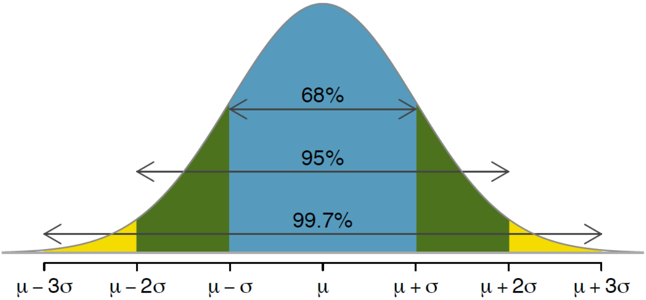
$$P(5.4 \leq X \leq 8.4) = P(X \leq 8.4) - P(X < 5.4).$$

About 68% American men are within one σ of the mean.

The 68 – 95 – 99.7% (Approximate) Rule



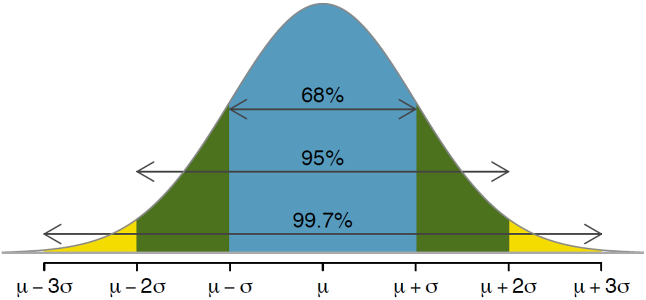
The 68 – 95 – 99.7% (Approximate) Rule



This holds for <u>any</u> data set that is normally distributed.

- About 68% of the data values are within 1 SD of the mean (blue).
- About 95% are within 2 SDs (blue and green).
- About 99.7% are within 3 SDs (blue, green, and yellow).

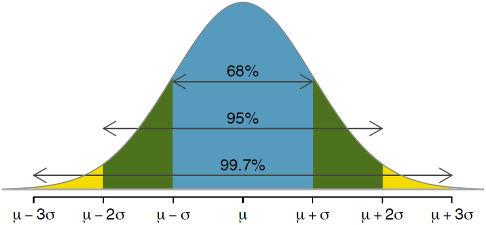
The 68 – 95 – 99.7% (Approximate) Rule



This holds for <u>any</u> data set that is normally distributed.

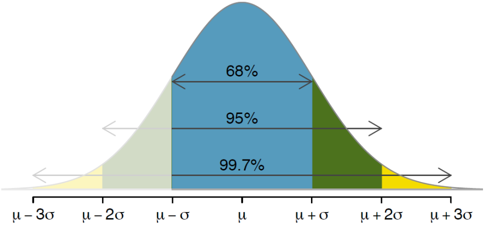
- About 68% of the data values are within 1 SD of the mean (blue).
- About 95% are within 2 SDs (blue and green).
- About 99.7% are within 3 SDs (blue, green, and yellow).

A Question using the 68 – 95 – 99.7% Rule



What percentage of students score above a 1250 on the SAT? ($\mu = 1500$, $\sigma = 250$)

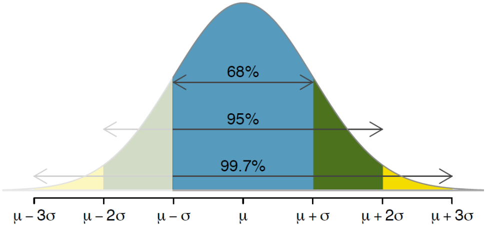
A Question using the $68 - 95 - 99.7\%$ Rule



What percentage of students score above a 1250 on the SAT? ($\mu = 1500$, $\sigma = 250$)

Notice that 1250 is one SD below the mean.

A Question using the $68 - 95 - 99.7\%$ Rule

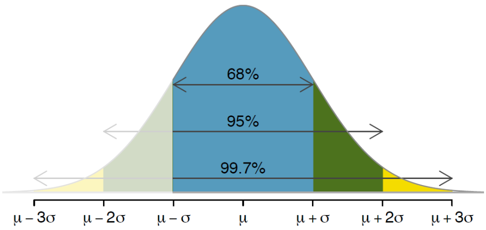


What percentage of students score above a 1250 on the SAT? ($\mu = 1500$, $\sigma = 250$)

Notice that 1250 is one SD below the mean.

The area outside the 1 SD windows is $100 - 68 = 32\%$.

A Question using the 68 – 95 – 99.7% Rule



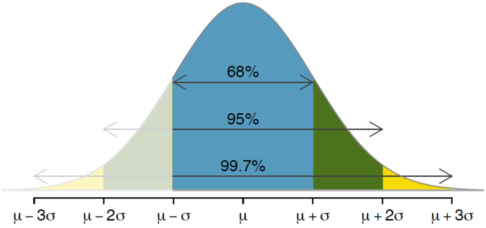
What percentage of students score above a 1250 on the SAT? ($\mu = 1500$, $\sigma = 250$)

Notice that 1250 is one SD below the mean.

The area outside the 1 SD windows is $100 - 68 = 32\%$.

So, the area to the left of -1 SD is $32/2 = 16\%$ (by symmetry)

A Question using the 68 – 95 – 99.7% Rule



What percentage of students score above a 1250 on the SAT? ($\mu = 1500$, $\sigma = 250$)

Notice that 1250 is one SD below the mean.

The area outside the 1 SD windows is $100 - 68 = 32\%$.

So, the area to the left of -1 SD is $32/2 = 16\%$ (by symmetry)

The desired percentage is $100 - 16 = \boxed{84\%}$.

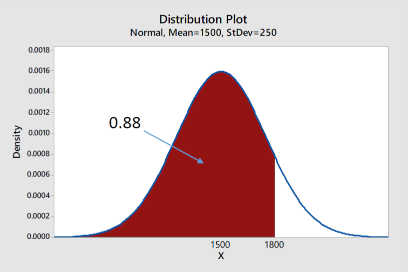
Percentiles and “Going Backwards”

For any x value (or z-score, if you convert to a standard normal), the **percentile** is simply the area to the left of this value.

Percentiles and “Going Backwards”

For any x value (or z-score, if you convert to a standard normal), the **percentile** is simply the area to the left of this value.

Example: the value $x = 1800$ on the SAT is about the 88th percentile.



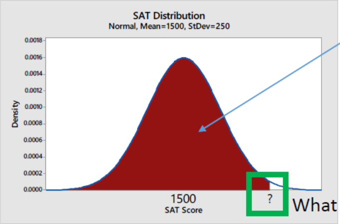
This means you scored higher than 88% of people who took the SAT.

Percentiles and "Going Backwards"

Suppose a college only takes students who reach the 99th percentile (or better) on the SAT. What cutoff must you attain?

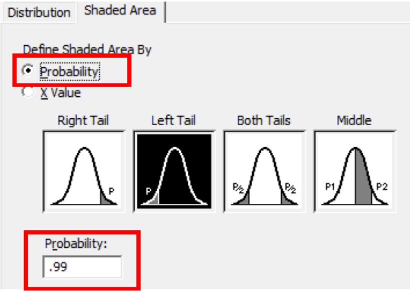
Percentiles and “Going Backwards”

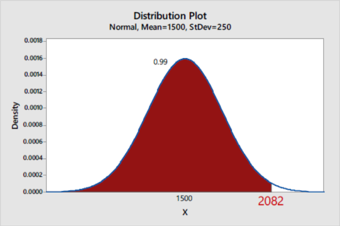
Suppose a college only takes students who reach the 99th percentile (or better) on the SAT. What cutoff must you attain?



We need this area to be 0.99.

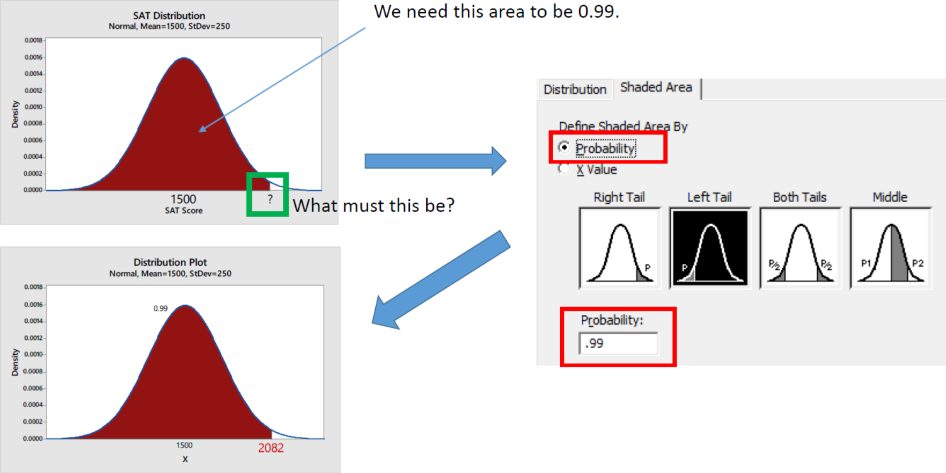
What must this be?





Percentiles and "Going Backwards"

Suppose a college only takes students who reach the 99th percentile (or better) on the SAT. What cutoff must you attain?



You must score at or above a 2082 to get into this college!

"Going Backwards" with Tables

As a New Year's resolution, you decide to get more sleep than 93% of American men. What is the least amount you can sleep per night? We remember that sleep is modelled by $N(6.9, 1.5)$.

	Second decimal place of Z									
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

We find the area in the table that gives 93%:

“Going Backwards” with Tables

As a New Year’s resolution, you decide to get more sleep than 93% of American men. What is the least amount you can sleep per night? We remember that sleep is modelled by $N(6.9, 1.5)$.

Z	Second decimal place of 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

We find the area in the table that gives 93%:

"Going Backwards" with Tables

As a New Year's resolution, you decide to get more sleep than 93% of American men. What is the least amount you can sleep per night? We remember that sleep is modelled by $N(6.9, 1.5)$.

	Second decimal place of Z									
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

We find the area in the table that gives 93%: $z_{93\%} = 1.48$

"Going Backwards" with Tables

As a New Year's resolution, you decide to get more sleep than 93% of American men. What is the least amount you can sleep per night? We remember that sleep is modelled by $N(6.9, 1.5)$.

Z	Second decimal place of 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

We find the area in the table that gives 93%: $z_{93\%} = 1.48$

Since the initial model is $X = N(6.9, 1.5)$, we get

$$z_{93\%} = \frac{x_{93\%} - 6.9}{1.5} \text{ which we solve to get}$$

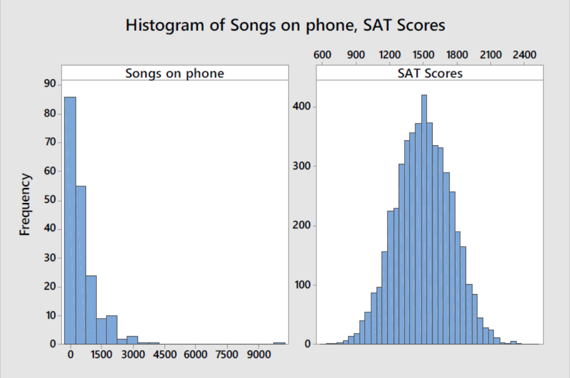
$$x_{93\%} = 1.5 \times z_{93\%} + 6.9 = \boxed{9.12} \text{ hours/night.}$$

But Wait! Are My Data Really Normal?

But Wait! Are My Data Really Normal?

Two Tests for Normality:

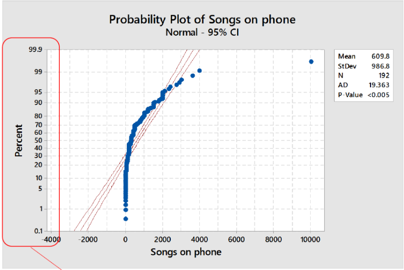
1) The human eye:
Does the histogram look unimodal and symmetric?

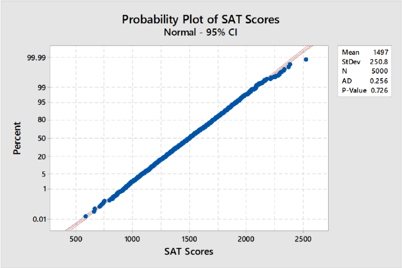


But Wait! Are My Data Really Normal?

Two Tests for Normality:

2) Use a “probability plot”. Minitab: Graph » Probability Plot
If the data fall on a straight line, that implies normality. If not, your data are non-normal.





This axis is scaled strangely so that normally-distributed data fall on a line!

Practice

Joe Bob is totally average in every way. What will his z-score be when he takes the SAT?

1. 1500
2. 0
3. 1
4. -1

Practice

Joe Bob is totally average in every way. What will his z-score be when he takes the SAT?

1. 1500
2. 0
3. 1
4. -1

Answer: 2., by definition of the z-score!

Practice

A recent stat test had a mean of 80% with a SD of 3%. What test score has a z-score -3?

1. 89%
2. 89
3. 71%
4. 71

Practice

A recent stat test had a mean of 80% with a SD of 3%. What test score has a z-score -3?

1. 89%
2. 89
3. 71%
4. 71

Answer: 3. Notice that $(71\% - 80\%)/3\% = -3$ (no unit)
The 89% has a z-score of $+3$.

Practice

Who was better in his sport?

- Michael Phelps, swimming, 27 olympic medals
  (mean olympic medals won by swimmers: 0.3, SD: 0.1)
- Barry Bonds, baseball, 762 home runs
  (mean home runs by baseball player: 4, SD: 3)

1. Phelps
2. Bonds
3. Equally good

Practice

Who was better in his sport?

- Michael Phelps, swimming, 27 olympic medals
  (mean olympic medals won by swimmers: 0.3, SD: 0.1)
- Barry Bonds, baseball, 762 home runs
  (mean home runs by baseball player: 4, SD: 3)

1. Phelps
2. Bonds
3. Equally good

Answer: 1., since Phelps z-score is (27-0.3)/0.1 = 267, and Bonds' is (762-4)/3 = 252.7, and 267 > 252.7.

Practice

The median on a test is 60. If the teacher halves everyone's score and adds 50, what is the new median?

1. 55
2. 110
3. 30
4. 80

Practice

The median on a test is 60. If the teacher halves everyone's score and adds 50, what is the new median?

1. 55
2. 110
3. 30
4. 80

Answer: 4., since the median undergoes all the transformation the data does. So $60/2 + 50 = 80$ is the new median.

(This reasoning would also apply to the mean)

Practice

The SD of temperature in a city is 10 degrees Celsius (C).
What is the SD if the data are measured in Fahrenheit (F)?
Recall that $F = (9/5)C + 32$

1. 50° F
2. 18° F
3. 10° F
4. -3° F

Practice

The SD of temperature in a city is 10 degrees Celsius (C).
What is the SD if the data are measured in Fahrenheit (F)?
Recall that $F = (9/5)C + 32$

1. 50° F
2. 18° F
3. 10° F
4. -3° F

Answer: 2., since $(9/5) \times 10 = 18$.
(SD's are only affected by scaling, not shifting)

Practice

Billy takes the weight of everyone in his class and converts them to z-scores. What is the mean of the data when written as z-scores?

1. Cannot determine with given info
2. 0
3. 1
4. -1

Practice

Billy takes the weight of everyone in his class and converts them to z-scores. What is the mean of the data when written as z-scores?

1. Cannot determine with given info
2. 0
3. 1
4. -1

Answer: 2., by definition of z-scores!!

Practice

Billy takes the weight of everyone in his class and converts them to z-scores. What is the SD of the data when written as z-scores?

1. Cannot determine with given info
2. 0
3. 1
4. -1

Practice

Billy takes the weight of everyone in his class and converts them to z-scores. What is the SD of the data when written as z-scores?

1. Cannot determine with given info
2. 0
3. 1
4. -1

Answer: 3.
We have $New\ Data = \dfrac{Old\ data - mean}{SD}$.
The SD ignores shifts like substracting the mean. To find the new SD, we just apply the scaling in the formula (division by the old SD). So the new SD is $SD/SD = 1$.

Are Those Claims Really True?

Suppose you data set collects random variables X with mean μ and standard deviation σ.

Are Those Claims Really True?

Suppose you data set collects random variables X with mean μ and standard deviation σ.

By moving to z-scores, we create a new random variable $Z = \dfrac{X - \mu}{\sigma}$.

Are Those Claims Really True?

Suppose you data set collects random variables X with mean μ and standard deviation σ.

By moving to z-scores, we create a new random variable $Z = \dfrac{X-\mu}{\sigma}$.

Note that

$$\begin{aligned} E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) &= \frac{1}{\sigma}(E(X)-\mu) \\ &= \frac{1}{\sigma}(\mu-\mu) = 0. \end{aligned}$$

Are Those Claims Really True?

Suppose you data set collects random variables X with mean μ and standard deviation σ.

By moving to z-scores, we create a new random variable $Z = \frac{X - \mu}{\sigma}$.

Note that

$$
\begin{aligned}
E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) &= \frac{1}{\sigma}(E(X) - \mu) \\
&= \frac{1}{\sigma}(\mu - \mu) = 0.
\end{aligned}
$$

Also,

$$
\begin{aligned}
SD(Z) = SD\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}SD(X-\mu) &= \frac{1}{\sigma}SD(X) \\
&= \frac{1}{\sigma}\sigma = 1.
\end{aligned}
$$