

Chapter 3. Distribution of Random Variables

Jan 26, 2016

by Huamei Dong

- 1. Normal distribution
- 2. Geometric distribution
- 3. Binomial distribution
- 4. Negative binomial distribution
- 5. Poisson distribution

1. Normal distribution

1. 1. Normal distribution model

Normal distribution is the most common distribution. It is symmetric, unimodal and bell shaped. The normal curve looks like

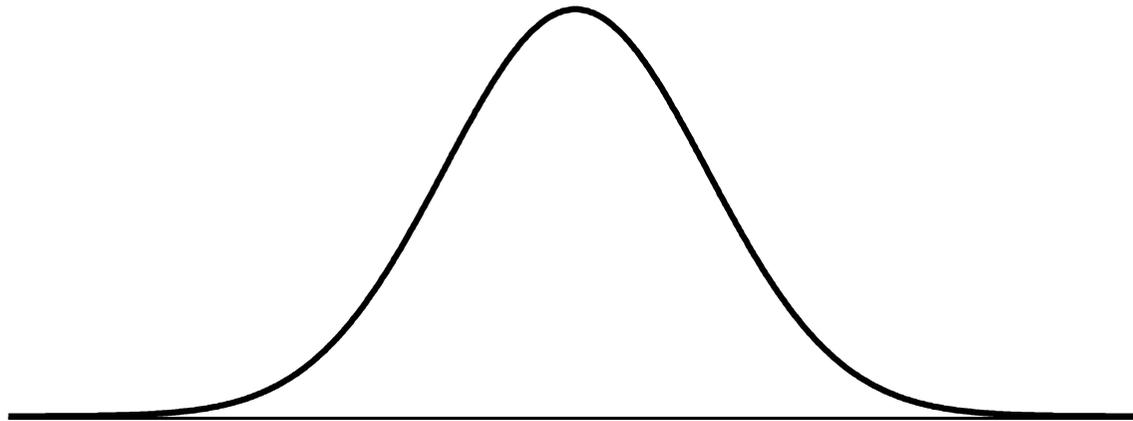


Figure 3.1: A normal curve.

1.2. Standardizing with Z score

The Z score of an observation is the number of standard deviation it falls above or below the mean. We compute the Z score for an observation x that follows a distribution with mean μ and standard deviation σ using

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

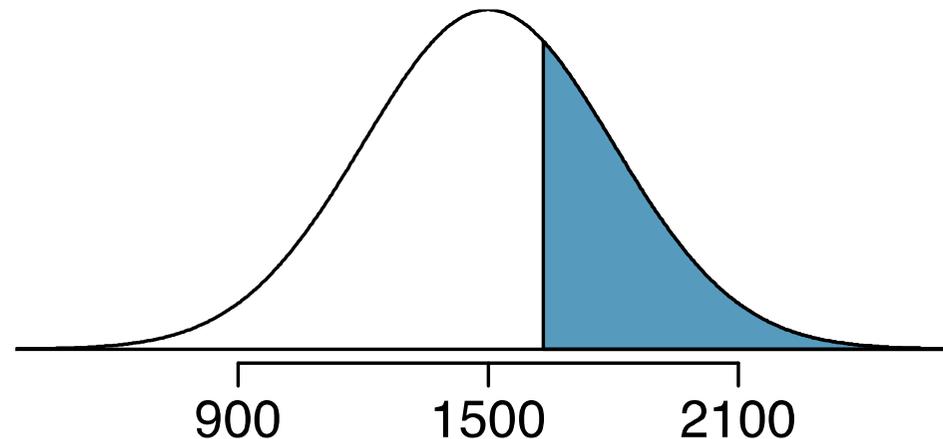
1.3. Normal Probability Table

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

1.4 Normal probability examples

Example 1: Cumulative SAT scores are approximated well by a normal model, $N(\mu=1500, \sigma=300)$. Shannon is a randomly selected SAT aptitude. What is the probability Shannon scores at least 1630 on her SATs?

Answer: First always draw and label a picture of the normal distribution.

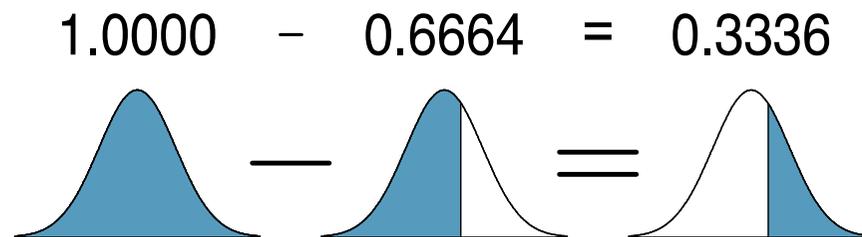


Here we have $x=1630$, $\mu=1500$, $\sigma=300$

So the Z score is
$$Z = \frac{x - \mu}{\sigma} = \frac{1630 - 1500}{300} = \frac{130}{300} = 0.43$$

We look up $Z=0.43$ in the normal probability table which yields 0.6664. However, this percentile describes those who had a Z score lower than 0.43. That means this percentile describes those who had a SAT less than 1630.

What we need is the percentile for having SAT at least 1630.
So the answer $1-0.6664=0.3336$



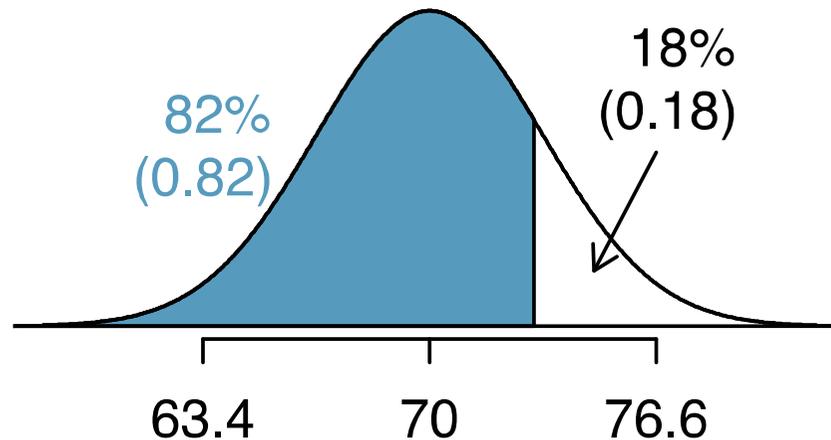
Comments: (1) Always draw the figure first and mark the area you want.

(2) Guess the area is less than 0.5 or big than 0.5

(3) Z table only gives you the area to the left of the Z score (i.e. less than Z score)

Example 2. Based on a sample of 100 men, the height of male adults between the age 20 and 62 in the US is nearly normal with mean 70.0" and standard deviation 3.3". What is the adult male height at the 82nd percentile?

Answer: We draw the figure first



Assume the height at the 82nd percentile is x .

We know the area below the Z score is 0.82. Look at the Z table, we find $Z=0.92$.

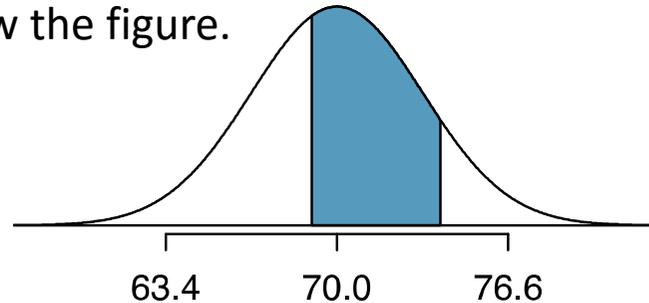
Now we need find x using the Z score 0.82.

$$0.92 = Z = \frac{x - \mu}{\sigma} = \frac{x - 70}{3.3}$$

Solving $x=73.04$ "

Example 3. What is the probability that a random adult male is between 5'9" and 6'2"? Assume it is normal distribution and mean is 70" and standard deviation is 3.3".

Answer: First we draw the figure.



If we can find the area of the two tails, we can find the shaded area.
Find the area for the left tail using

$$Z = \frac{69-70}{3.3} = -0.3030$$

The left tail area **0.3821**.

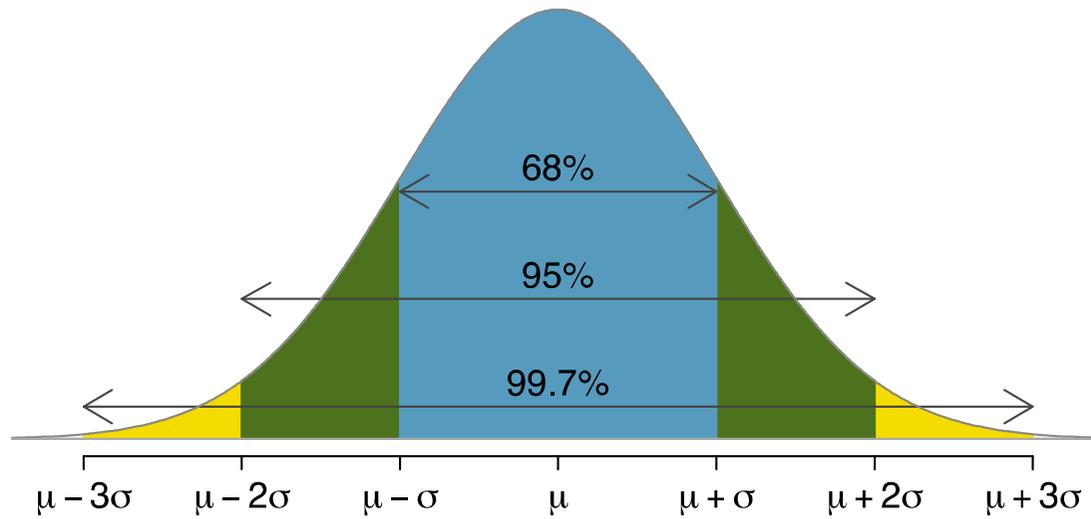
Find the area for the right tail using

$$Z = \frac{74-70}{3.3} = 1.2121$$

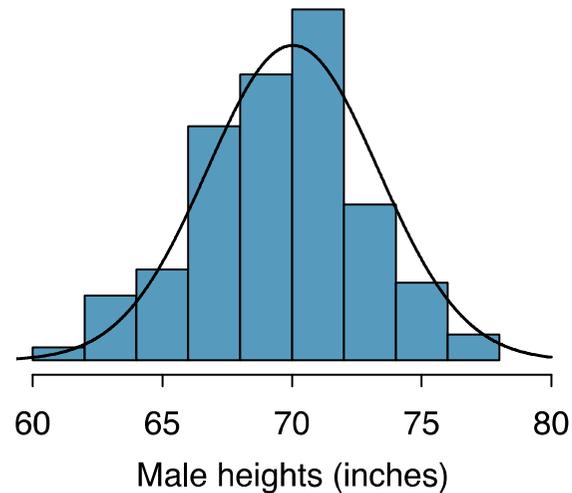
The right tail area is **1-0.8869=0.1131**

So the answer is $1-0.3821-0.1131=0.5048$

1.5 68-95-99.7 rule



1.6 Checking Normality Using Histogram



Sometimes you would like to know if your data fits normal model. To check the normality, you can use histogram. The closer the curve fits the histogram, the more reasonable the normal model is.

Example 4. Using R to find the probability that a random adult male is between 5'9"

And 6'2" ?

Answer:

```
> x1<-69
```

```
> x2<-74
```

```
> mean<-70
```

```
> SD<-3.3
```

```
> z1<-(x1-mean)/SD
```

```
> z2<-(x2-mean)/SD
```

```
> z1
```

```
[1] -0.3030303
```

```
> z2
```

```
[1] 1.212121
```

```
> tail1<-pnorm(z1)
```

```
> tail2<-1-pnorm(z2)
```

```
> tail1
```

```
[1] 0.3809334
```

```
> tail2
```

```
[1] 0.112733
```

```
> 1-tail1-tail2
```

```
[1] 0.5063336
```

Chapter 3 Homework#3 (due 02/02/2016) : Page 130 3.4, Page 134 3.14, Page 135 3.17, Page 136 3.20.