1

Stat 104: Quantitative Methods

Class 13: Continuous Probability Distributions

2

Discrete versus Continuous

The major distinction between a continuous and discrete random variable is the numerical events of interest.

We can list all possible values of a discrete random variable, and it is meaningful to consider the probability that a particular individual value will be assumed.

3

- We cannot list all the values of a continuous random variable- because there is always another possible value between any two of its values.
- Hence the only meaningful events for a continuous random variable are *intervals.*
- The probability that a continuous random variable X will assume any particular value is 0.

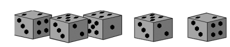
P(X=x) = 0 for any x

There are just too many values out there

4

P(X=x) = 0 for any x if X is a continuous random variable

Huh ?? Let's try to give a basic argument



For a six sided die, what is the probability of rolling a 3 ?

For a 20 sided die, what is the probability of rolling a 3 ?

Several polyhedra in various materials with similar symbols are known from the Roman period. Modern scholarship has not yet established the game for which these dice were used.

For a 100 sided die, what is the probability of rolling a 3 ?

5

Now ponder

•What is happening to the probability of rolling a 3 ? Getting smaller

•Why ? [hint-number of possible outcomes increasing]
The basic axiom of probability says that the sum of all the outcome probabilities must be 1.

If we continue with the die example, it will be impossible to assign positive probability to each outcome so that they all sum to 1.

P(X=x) = 0 for any x if X is a continuous random variable

6

P(X=x) = 0 for any x if X is a continuous random variable

Hence, for a continuous random variable X, it is only meaningful to talk about the probability that the value assumed by X will fall within some *interval of values.*

$$P(a \leq X \leq b)$$

Note that

$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b)[why?]$$

Because the probability of getting something exactly is 0. These are different for discrete random variables though.

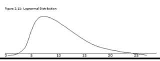
Examples of continuous random variables:

Length of time between arrivals at a hospital clinic, weight of a food item bought at a supermarket, amount of soda in a 12 oz can,

height of sunspots,

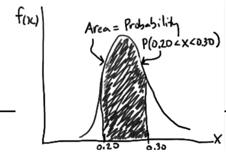
a person's hat size.

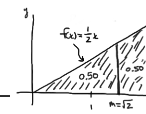
Stock returns



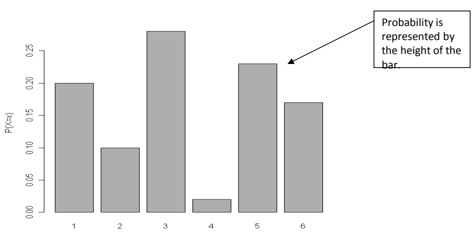
Density Curves

- Graphs are usually the easiest way to describe a continuous random variable. Equivalent to probability mass function for discrete.
- The **probability density function *f(x)*** is a curve that describes the probability associated with the range of values that a continuous random variable can assume.
- The area under the curve of the density function is the probability that any range of values will occur. The total area under the curve has to be one.

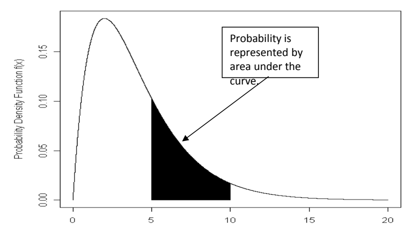




Graph of a Discrete Distribution



Graph of a Continuous Distribution



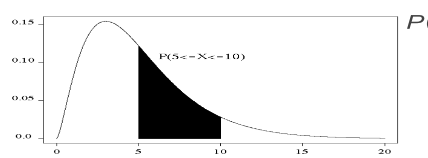
The density curve f(x) is such that f(x)>=0 for all x and the area under the curve=1

The Density Curve f(x)

- Notice that the vertical scale is labeled *f(x)*, not *P(X=x)*.
- This difference in notation emphasizes the fact that the height in a continuous distribution is not a direct measure of probability.
- To **produce probabilities** in the continuous case, we'll need to **measure areas under the curve**. We usually will need a computer or special tables to find the desired area under the curve.
- What is often handy though is to visually inspect the curve to understand what values are likely to occur.
- If we examine the last figure, its clear that values larger than 15 are unlikely (very little area under the curve) and negative values of the random variable are not possible at all (why?).

More on the Density Curve

- As previously mentioned, the probability that X will take any specific value is zero. Given a pdf f(x), the area under the graph of f(x) between the two values a and b is the probability that X will take a value between a and b. In calculus notation, we have that



$$P(a \le X \le b) = \int_a^b f(x)dx$$

=area under curve between a and b

CDF: Cumulative Distribution Function

- The cumulative distribution function (cdf) is useful because it gives us the probability of being less than some value:

$$F_X(X \le x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx$$

- We will (usually) never calculate this directly; we (usually) obtain these values from the computer or a table.

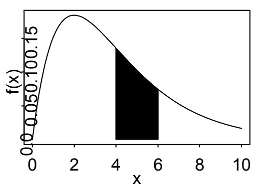
Probability of an Interval

- Using the CDF, we can calculate the probability of any interval:

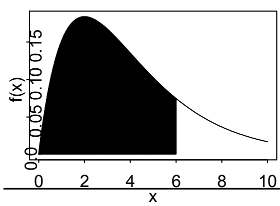
$$P(a \le X \le b) = F_X(b) - F_X(a)$$

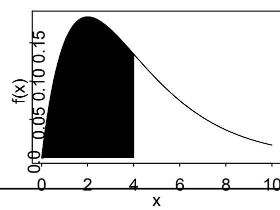
Everything upto a
Everyting upto b
Subtracting the two: area between a and b: Probability of interval

To find this shaded area, subtract the two shaded areas below.



=





Mean of a continuous RV

- The expected value of a continuous random variable *X* is defined to be

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

- This is just the continuous version of

$$\mu_X = E(X) = \sum_{all\ x_i} x_i P_X(x_i)$$

This formula is given for completeness-we don't assume calculus knowledge.

Variance of a continuous RV

- The variance of a continuous random variable *X* is defined to be

$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

- This is just the continuous version of

$$\sigma_X^2 = \sum_{all\ x_i} (x_i - \mu)^2 P(X = x_i)$$

This formula is given for completeness-we don't assume calculus knowledge.

The Uniform Distribution

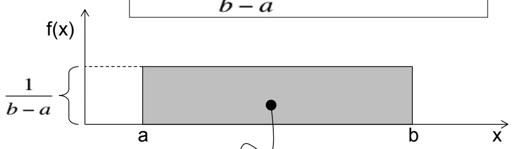
- One of the simplest examples of a continuous probability distribution is the <u>uniform distribution.</u>
- We'll use it here to illustrate typical continuous probability distribution characteristics.

19

The Uniform Distribution

- Consider the ***uniform probability distribution***
- It is described by the function:

$$f(x) = \frac{1}{b-a}, \quad where \quad a \le x \le b$$



area = width x height = (b – a) x $\frac{1}{b-a}$ = 1

20

General Characteristics

- In general the uniform density is given by

$$f(x) = \begin{cases} \frac{1}{(b-a)} & \text{for x values between a and b} \\ 0 & \text{everywhere else} \end{cases}$$

- The mean is then $E(X) = \frac{(a+b)}{2}$

Don't forget on cheat sheet

- The variance is $Var(X) = \frac{(b-a)^2}{12}$

21

Example

- The amount of gasoline sold daily at a service station is uniformly distributed with a minimum of 2,000 gallons and a maximum of 5,000 gallons.

b a

  - ***Find the probability that daily sales will fall between 2,500 and 3,000 gallons.***
  - ***What is the probability that the service station will sell <u>at least</u> 4,000 gallons?***
  - ***What is the probability that the station will sell <u>exactly</u> 2,500 gallons?*** 0

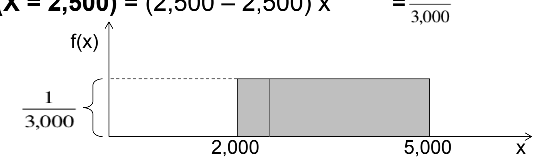
22

Example

- We have X~U(2000,5000) so that the density function is

23

- ***What is the probability that the station will sell <u>exactly</u> 2,500 gallons?***
- **P(X = 2,500)** = (2,500 – 2,500) x $= \frac{1}{3,000}$



- ***"The probability that the gas station will sell <u>exactly</u> 2,500 gallons is zero"***

24

Finding Probabilities

- ***Find the probability that daily sales will fall between 2,500 and 3,000 gallons.*** length * height (area under the curve, here, rectangle)
- **P(2,500 ≤ X ≤ 3,000)** = (3,000 – 2,500) x $\frac{1}{3,000}$ = .1667
- ***What is the probability that the service station will sell <u>at least</u> 4,000 gallons?***
- **P(X ≥ 4,000)** = (5,000 – 4,000) x $\frac{1}{3,000}$ = .3333

length * height (area under the curve, here, rectangle)

The Normal Distribution

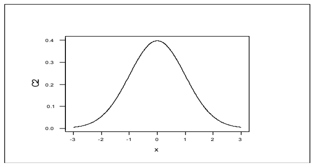
The normal distribution is the most fundamental continuous distribution used in statistics so I guess we better learn it. Many methods that are widely used in economics, finance, and marketing are based on an assumption of a normal distribution.

Normal Dist.

Gaussian Dist.

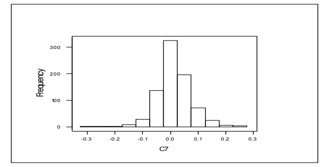
Bell-Shaped Curve

equivalent names

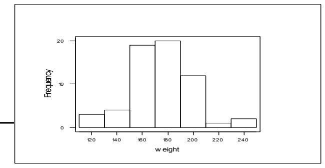


We kinda have a graphical feel for the normal distribution already. It is the familiar *bell shaped* curve we hear people talk about.

Here are monthly returns for coca-cola

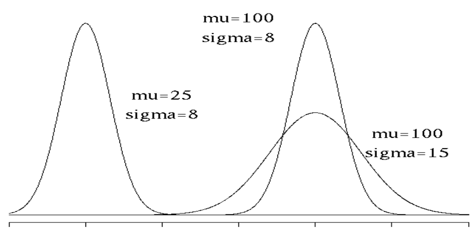


Here is the weight from my class last year (just male weights)



What controls the shape of the curve ?

The normal distribution is governed by the **two parameters :** μ (the *mean*) and σ (the *standard devia*



Realize that sigma governs height and width.
Area has to be 1 so as you get higher, you must get skinnier.

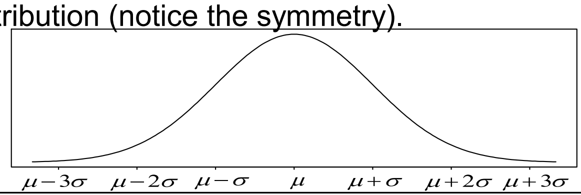
Notation for the normal curve

We use the notation

$$X \sim N(\mu, \sigma^2).$$

Notation uses squared even though formulae us single

Here is a general picture of a normal distribution (notice the symmetry).



Using R

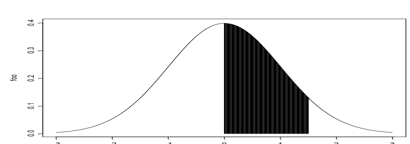
P(X<=x) =

- P(X<x) =pnorm(x,m,s)
- f(x) = dnorm(x,m,s) [this is NOT P(X=x)]

For discrete we cared about dbinom because P(X=x) is useful.
Not so much here

Finding normal probabilities

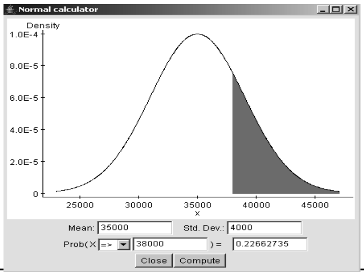
- $P(a \leq X \leq b)$ = area under the curve between *a* and *b*.



We use the computer or tables to do this

Example: The tread life of a particular brand of tire has a normal distribution with mean 35,000 miles and standard deviation 4,000 miles.

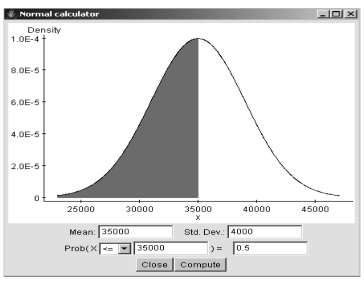
What proportion of these tires has tread lives of more than 38,000 miles ?



Using R

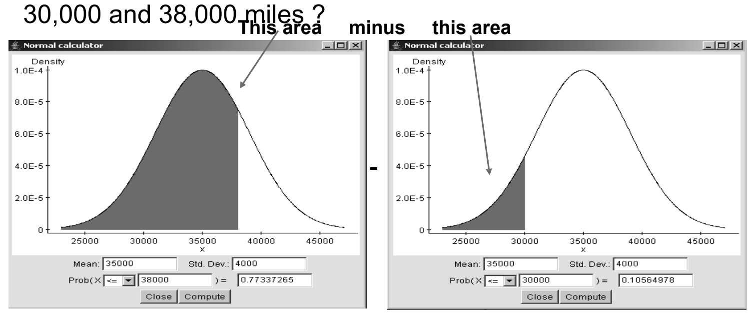
```
> 1-pnorm(38000,35000,4000)
[1] 0.2266274
```


What proportion of these tires has tread lives less than 35,000 miles ?



Don't need calculator. Left of mean is 0.5.

What proportion of these tires has tread lives between 30,000 and 38,000 miles ?



Using R

```
> pnorm(38000,35000,4000)-pnorm(30000,35000,4000)
[1] 0.6677229
```


This is Easy

- Yes it is…….
- Finding normal probabilities aren't that hard-heck the computer does all the work.
- However, will you have a computer on the exam ?????

The Z score

Many years ago someone realized that if

$$X \sim N(\mu, \sigma^2).$$

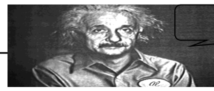
These values are not tabulated for every possible normal distribution.

Then [using the a+bX rule]

$$Z = \frac{X - E(X)}{SD(X)} = \frac{X - \mu}{\sigma} \sim N(0,1).$$

These values are tabulated.

So what ?

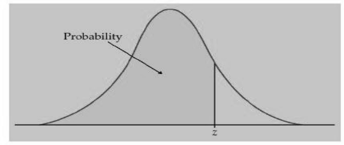


I got a C in high school math

A normal after Z transformation, remains a normal.

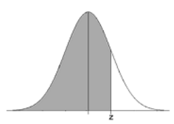
The Z~N(0,1) is called the standard normal and areas under the curve for this distribution are tabulated and found on the course web site:

F(z)=P(Z<z)



Standard Normal Cumulative Probability Table

Example P(z ≤ 0.65)= 0.7422



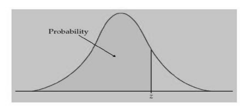
Cumulative probabilities for POSITIVE z-values are shown in the following table:

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

Standard Normal Table

TABLE A Standard normal probabilities

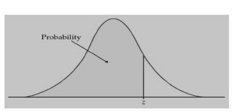
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−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



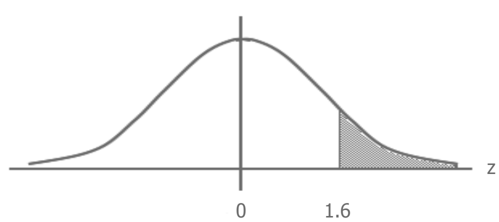
Standard Normal Table

TABLE A Standard normal probabilities (*continued*)

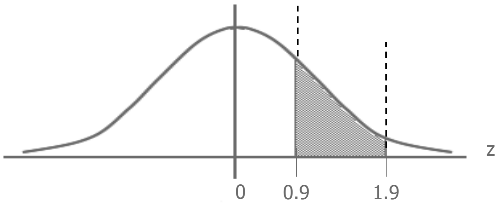
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.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	.9989	.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



Practice: Find P(Z>1.6)



Practice: Find P(0.9<Z<1.9)



Standardization

- The values Z takes are called z-scores.
- Z-score is literally the number of standard deviations the related X is from its mean.
- For any normal distribution X it holds that:

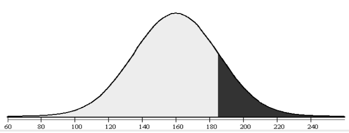
$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$$

$$\text{and since } Z = \frac{X-\mu}{\sigma}, \text{ we have :}$$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

Note-we interchangeably use ≤ and <. No biggie.

Example: A plumbing company has found that the length of time, in minutes, for installing a bathtub is normally distributed with mean *160 minutes and standard deviation 25 minutes.* What percent of the bathtubs take longer than 185 minutes to be installed?



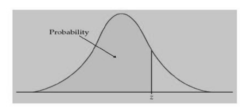
$$\begin{aligned} P[X \geq 185] &= P\left[(X-160) \geq (185-160)\right] \\ &= P\left[\left(\frac{X-160}{25}\right) \geq \left(\frac{185-160}{25}\right)\right] \\ &= P[Z \geq 1] \end{aligned}$$

We want P(Z>1).

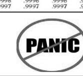
The table gives us P(Z<1):

TABLE A Standard normal probabilities (*continued*)

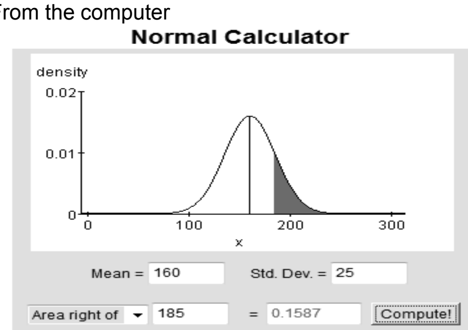
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.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	.9989	.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



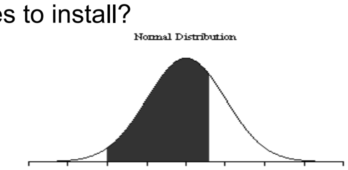
What do we do ?



Check : From the computer



What is the likelihood that it will take between 110 and 175 minutes to install?



$$\begin{aligned} P(110 \leq X \leq 175) &= P\left(\frac{110-160}{25} \leq \frac{X-160}{25} \leq \frac{175-160}{25}\right) \\ &= P(-2 \leq Z \leq 0.6) \end{aligned}$$

How do we find this using our table ?

$P(-2 \le Z \le 0.6)$

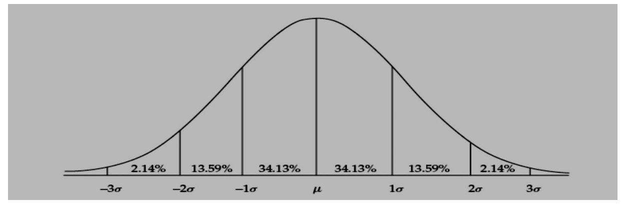
Normal Coverage Rule

For $X \sim N(\mu, \sigma^2)$,

$P(\mu - \sigma \le X \le \mu + \sigma) = 0.68$

$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 0.95$

$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0.997$



Review coverage rule

- Chebyshev's theorem says for *any* random variable
  $P(\mu - 2\sigma < X < \mu + 2\sigma) \ge 75\%$
- However, if we are told *X* is normally distributed
  $P(\mu - 2\sigma < X < \mu + 2\sigma) = 95\%$

Reverse Look Up

- Sometimes we use the N(0,1) table in reverse, looking up probability to find a Z score.
- This is best explained with an example.
- In 2000, Verbal SAT ~~SAT~~ scores were normally distributed with μ of 505 ~~1019~~ and σ of ~~209~~ 110
- What Verbal SAT score will place a student in the top 10% of the population?

Reverse Look Up

- From the table find z such that P(Z<z)=.90

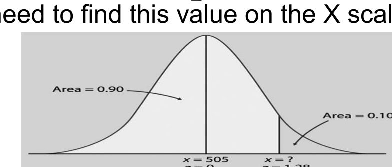
TABLE A Standard normal probabilities (*continued*)

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

- So P(Z<1.28) = 0.90 (approximately)

Reverse Look Up

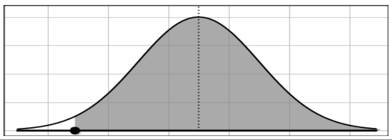
- We need to find this value on the X scale



- Easily done by reversing the standardization
- X = σ·Z + μ = 110·1.28 + 505 = 646
- So, a student needs a Verbal SAT score of 646 in order to be in the top 10% of all students

Reverse Look Up Practice

- Lifetimes of light bulbs that are advertised to last 5000 hours are normally distributed with **μ=**5100, **σ=**200.
  - What figure should be advertised if the company wishes to be sure that 98% of all bulbs last longer than the advertised figure?



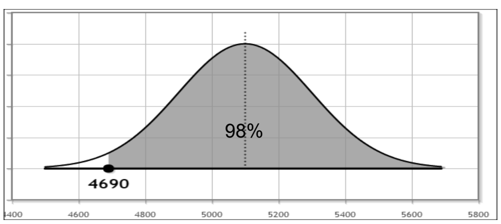
Practice

- What figure should be advertised if the company wishes to be sure that 98% of all bulbs last longer than the advertised figure?
- From our Normal table we find P(Z> -2.05)=0.98

−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

- $X = \sigma \cdot Z + \mu = -200 \cdot 2.05 + 5100 = 4690$
- So P(X>4690) = 98%.

Pictorially



Using R

- Reverse look-up in R can be done as follows.
- Lifetimes of light bulbs that are advertised to last 5000 hours are normally distributed with **μ=**5100, **σ=**200.
  - What figure should be advertised if the company wishes to be sure that 98% of all bulbs last longer than the advertised figure?

```
> qnorm(.02,5100,200)
[1] 4689.25
```

qnorm(q,mu,s) finds x so that P(X<x)=q

<u>Things you should know</u>

- ❑Discrete versus Continuous random variables
- ❑Uniform distribution
- ❑Normal distribution
- ❑Z scores
- ❑Using the N(0,1) table
- ❑Reverse Look-Up