

# BIOS 600: Principles of Statistical Inference

## Continuous Probability Distributions

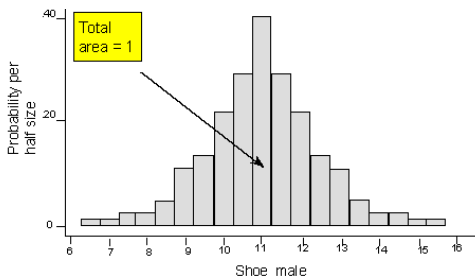
Fall 2012

# Reading

- ▶ Pagano and Gauvreau, Chapter 7, Section 7.4

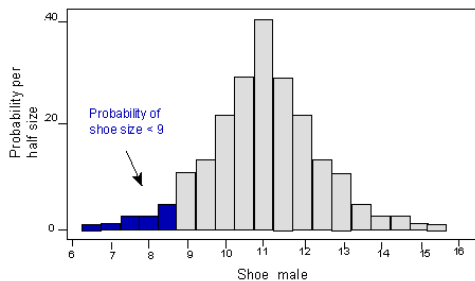
# Probability Distributions

Last time we discussed discrete random variables. As we move to discussion of continuous random variables, we will consider the distribution of shoe size in adult males in the US. Let  $X$  represent the shoe size (whole and half sizes).



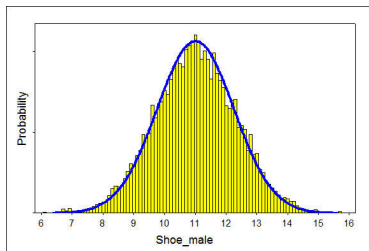
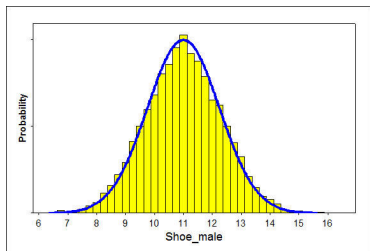
# Probability Distributions

We can use this distribution to determine the probability of shoe size falling in any range.



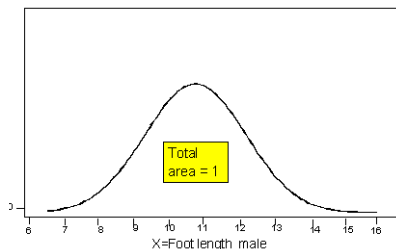
# Probability Distributions

Now suppose we could get *really* well-fitting shoes, using quarter sizes (9, 9.25, 9.5, 9.75, ..., left figure) or even tenth sizes (9, 9.1, 9.2, ..., on right). As the number of intervals increases, the bar width becomes more narrow, and the graph approaches a smooth curve. We will use these smooth curves to describe the probability distributions of continuous random variables, concentrating on the *normal distribution*.



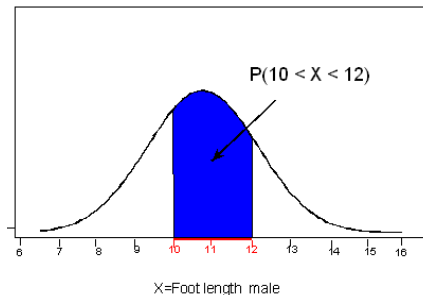
# Continuous Probability Distributions

Now let  $X$  be foot length, which is a continuous random variable which can take values over a very wide range of possibilities (e.g., could be 12.008 inches rather than 12 inches or 12.5 inches). This *probability density curve* shows us the probability of any range of foot lengths we would like to investigate.



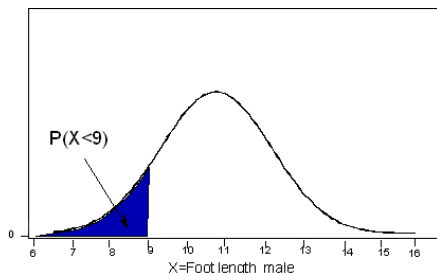
# Continuous Probability Distributions

For example, we may wish to know the probability a man's foot is between 10 and 12 inches,  $Pr(10 < X < 12)$ , which is the shaded area in the plot.



# Continuous Probability Distributions

The probability a man's foot is less than 9 inches is the shaded area in this plot.





# Continuous Probability Distributions

- ▶ Density curves, like histograms, can have any shape. The area under a density curve is always 1.
- ▶ How do you find the area of interest in the plots?
  - ▶ Calculus!
  - ▶  $Pr(a \leq X \leq b) = \text{area between } a \text{ and } b \text{ below the curve} = \int_a^b f(x)dx$  where  $f(x)$  represents the density curve
  - ▶ While you will need to know calculus for courses like BIOS 660, for BIOS 600, we can rely on tables or computers to provide the probabilities we need.
  - ▶ We will see density curves for several important distributions – normal,  $t$ ,  $\chi^2$ , and  $F$  random variables.

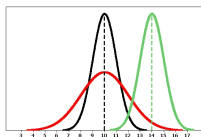
# Normal Distribution

For the normal distribution,

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

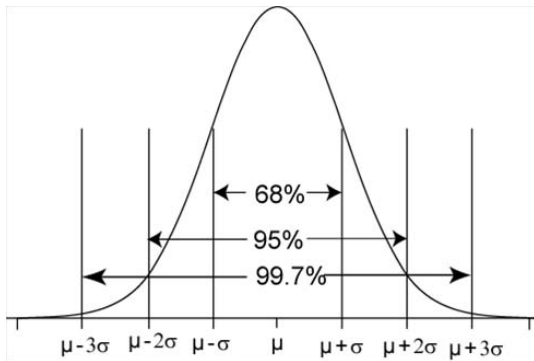
where the mean is given by  $\mu$ , the variance by  $\sigma^2$ , and the standard deviation by  $\sigma$ . The notation  $N(\mu, \sigma^2)$  is often used.

Also called Gaussian distribution or bell curve. This distribution is symmetric.



Which of these normal distributions has the biggest mean?  
standard deviation?

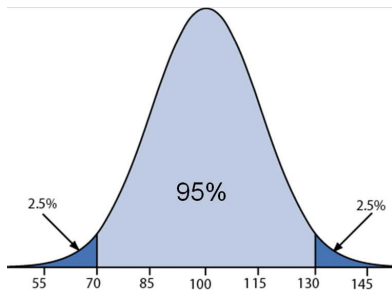
## Normal Distribution: Standard Deviation Rule



A useful rule for normal distributions is that roughly 68% of the area under the curve is within one standard deviation ( $\sigma$ ) of the mean, 95% is within two standard deviations ( $2\sigma$ ), and 99.7% is within  $3\sigma$ .

# Normal Distribution

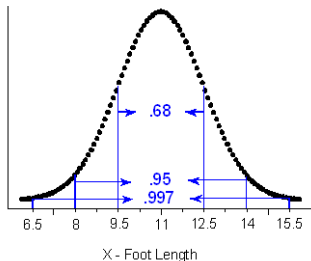
The symmetry of the normal distribution also allows us to calculate the probability of values falling in the tails:



5% of the data are further than two standard deviations ( $2\sigma$ ) from the mean, 2.5% in each tail.

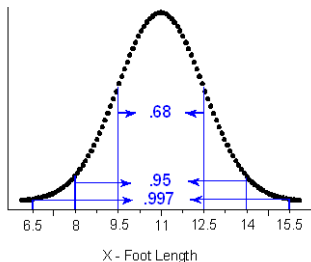
## Normal Distribution: Example

Suppose US male foot length follows a normal distribution with  $\mu = 11$  and  $\sigma = 1.5$ . Then the probability density of foot length is given by



What is the probability that a randomly chosen US male will have a foot length between 8 and 14 inches?

## Normal Distribution: Example

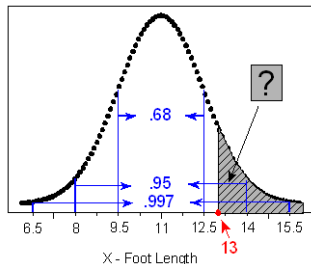


What is the probability that a randomly chosen US male will have a foot length greater than 12.5 inches?

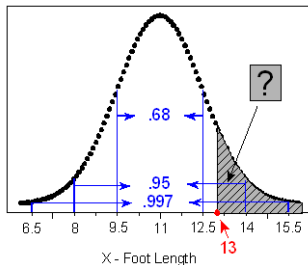
## Normal Distribution: Example

What is the probability that a randomly chosen US male will have a foot length greater than 13 inches?

- ▶  $\Pr(\text{foot length} > 12.5) = \frac{1}{2}(1 - 0.68) = 0.16$
- ▶  $\Pr(\text{foot length} > 14) = 0.025$
- ▶  $0.025 < \Pr(\text{foot length} > 13) < 0.16$
- ▶ While the standard deviation rule gives us some idea, it is pretty limited in the questions we can answer. We need a better approach!
- ▶ Solution: standard normal table and computerized standard normal probabilities



## Normal Distribution: Example



How many standard deviations above the mean is a 13 inch foot?

- ▶ Recall  $\mu = 11$  and  $\sigma = 1.5$
- ▶ 13 inch foot is 2 inches above the mean
- ▶ 2 inches is  $\frac{2}{1.5} = 1.33$  standard deviations
- ▶  $\frac{13-11}{1.5} = 1.33$  is known as a *z-score*



## Z-scores

Z-scores are used in many public health settings, including

- ▶ Evaluating health effects related to body mass index (BMI)
- ▶ Measuring child growth (height-for-age, weight-for-age, weight-for-height, BMI-for-age) and identifying malnourished children
- ▶ Testing safety of food, water, and environmental samples
- ▶ Reporting bone density scan results

## z-scores

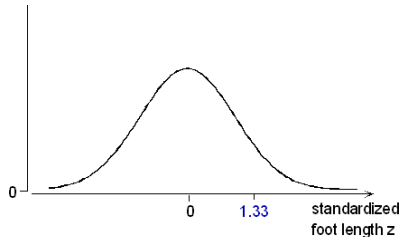
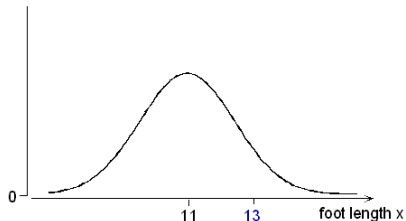
The *z-score* is a standardized normal variable that tells us how many standard deviations above (positive z-scores) or below (negative z-scores) the mean our original value (13 inch foot) is. That is,

$$z = \frac{x - \mu}{\sigma} = \frac{\text{value} - \text{mean}}{\text{standard deviation}}.$$

The z-score for that 12.5 inch foot is  $\frac{12.5-11}{1.5} = 1$ .

## z-scores

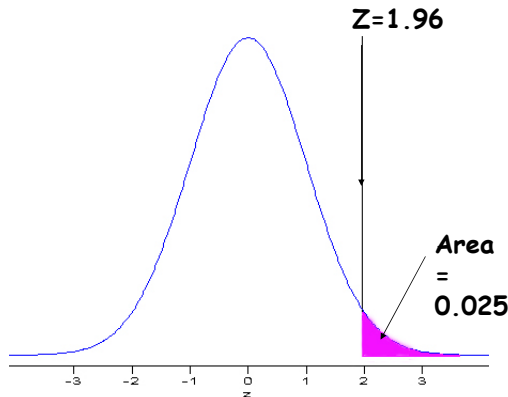
When we standardize by finding z-scores, we change the normal distribution by moving the location (mean moves to zero) and changing the scale (so the standard deviation is one), but the *relative* position of our 13 inch (or 1.33 standard deviation) foot remains the same relative to the rest of the distribution.



# Standard Normal Distribution

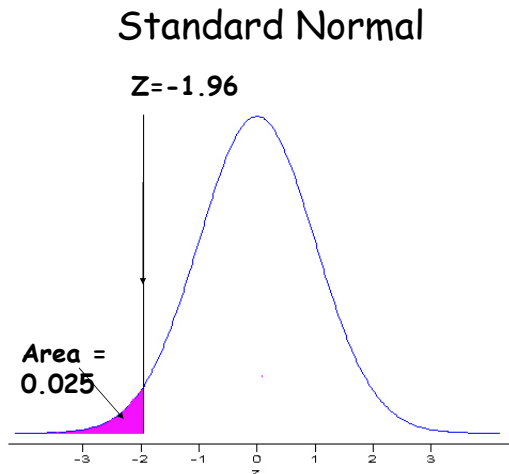
The distribution of the z-scores is known as the *standard normal distribution*. Here are a few important reference points for this distribution.

## Standard Normal



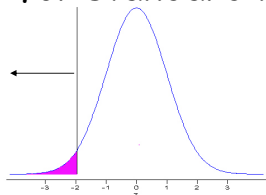
# Standard Normal Distribution

The distribution of the z-scores is known as the *standard normal distribution*.



# Standard Normal Distribution

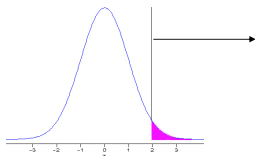
Areas for standard normal



z	area
0	0.5
-1.65	0.049
-1.96	0.025
-2.58	0.005
-3	0.001

# Standard Normal Distribution

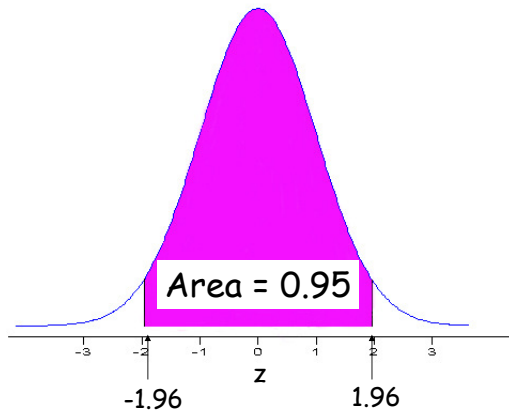
## Areas for standard normal



z	area
0	0.5
1.65	0.049
1.96	0.025
2.58	0.005
3	0.001

# Standard Normal Distribution

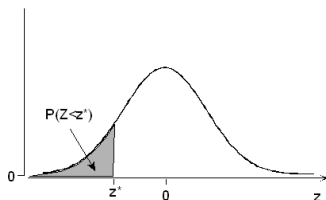
## Standard Normal





# Calculating Normal Probabilities

A *standard normal table* allows you to calculate values based on the standard normal distribution.



It tells you how much area is under the normal curve to the *left* of the specified value (lower tail area). Because the standard normal distribution is symmetric with mean zero,  $Pr(Z \leq 0) = 0.5$ .

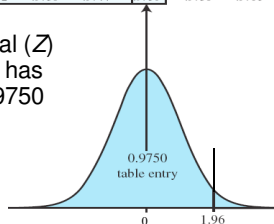
# Standard Normal Table for $Pr(Z \leq z)$

z tenths	hundredths									
	.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

**Example:** A Standard Normal ( $Z$ ) variable with a value of 1.96 has a cumulative probability of .9750

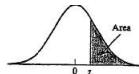
- That is,

$$Pr(Z \leq 1.96) = 0.9750$$



# Standard Normal Table for $Pr(Z \geq z)$ (Table A.3)

Table 4. Normal curve areas  
Standard normal probability in right-hand tail  
(for negative values of  $z$  areas are found by symmetry)

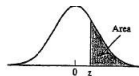


		Second decimal place of $z$									
$z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641	
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247	
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859	
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483	
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121	
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776	
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451	
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148	
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867	
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611	
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379	
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170	
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985	
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823	
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681	
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559	
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455	
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367	
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294	
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233	
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183	
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143	
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110	

# Back to Feet...

Using the normal table, what is the probability of a foot bigger than 13 inches (z-score 1.33)?

Table 4. Normal curve areas  
Standard normal probability in right-hand tail  
(for negative values of z areas are found by symmetry)



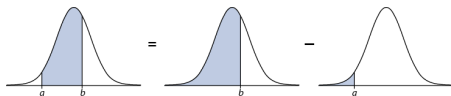
		Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641	
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247	
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859	
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483	
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121	
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776	
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451	
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148	
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867	
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611	
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379	
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170	
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985	
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823	
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681	
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559	
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455	
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367	
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294	
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233	
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183	
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143	
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110	

# Standard Normal Distribution: Probabilities Between Two Values

## Probabilities Between Two Points

Let  $a$  represent the lower boundary and  $b$  represent the upper boundary of a range:

$$\Pr(a \leq Z \leq b) = \Pr(Z \leq b) - \Pr(Z \leq a)$$



## Case Study: Body Mass Index

Suppose the distribution of body mass index (BMI) in young adults has a mean  $\mu = 26$  and standard deviation  $\sigma = 6$  (based on figures from a 2009 Australian study). NHLBI classifies BMI as underweight ( $<18.5$ ), normal weight ( $18.5 \leq \text{BMI} < 25$ ), overweight ( $25 \leq \text{BMI} < 30$ ), or obese ( $\text{BMI} \geq 30$ ). Suppose we plan to randomly sample a woman from a similar population.

- ▶ What is the probability she will be underweight?
- ▶ What is the probability she will be obese?
- ▶ What is the probability she will be normal weight?
- ▶ What is the probability she will be overweight but not obese?

Seriously. This is 2012! Isn't there an alternative to a table?

Online calculators for normal probabilities abound. For example, try this [normal probability calculator](#).

**Stata:** `display normal(z)` will calculate  $Pr(Z \leq z)$  for the standard normal ( $N(0, 1)$ ) distribution.

# Reading for Next Time

- ▶ Pagano and Gauvreau, Chapter 8