
Normal Distribution
Z score
Sampling distribution

Probability distributions

- ⇒ Basic prob. rules are used to calculate prob. of an event
- ⇒ How to calculate prob. under more complex conditions
- ⇒ **What is a random variable?**
 - Var. in a study its subjects are randomly selected from a larger population
- ⇒ When all values of the random variable is presented in a frequency table or a histogram, it form a probability distribution
- ⇒ In some distributions, a rule or a formula describe adequately the distribution and you can use the formula to estimate the probability of interest
- ⇒ In other situations a theoretical probability provides a better fit to the distribution of the variables of interest

Common distributions in medicine

⇒ **Binomial distribution**

- ▶▶ For discrete variable
- ▶▶ When the event has only binary outcome (yes, no) (alive, dead)

⇒ **Poisson distribution**

- ▶▶ Named after the French mathematician ***Simeon D. Poisson***
- ▶▶ Discrete variables
- ▶▶ When the outcome is the number of times an event occurs. Used for rare events, (to model no. of cells in a given fluid volume, bacterial colonies in a medium, emission of radioactive particles from radioactive material)

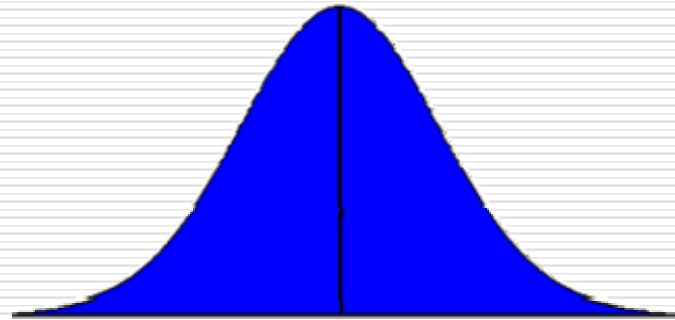
⇒ **Normal distribution**

Normal distribution

Gaussian distribution



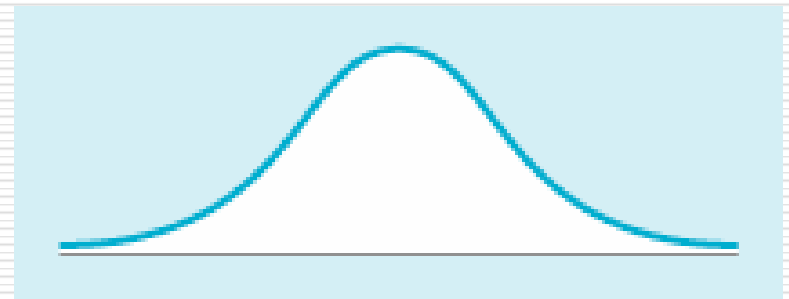
Carl Friedrich Gauss
(1777-1855)



- After German mathematician **Carl Friedrich Gauss**
- Continuous distribution
- Smooth, bell shaped curve
- Symmetric around the mean of the distribution
- Many biological parameters are normally distributed

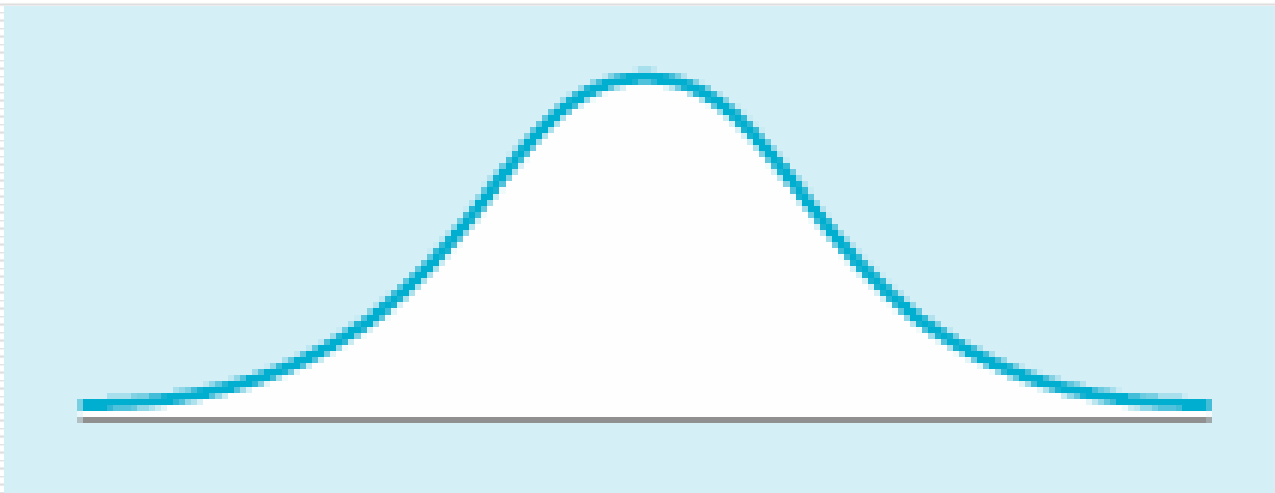
Properties of the Normal Curve

- ⇒ **Bell-shaped**
- ⇒ **Unimodal**
- ⇒ **Symmetrical**
 - ▶▶ Exactly half of the scores above the mean
 - ▶▶ Exactly half of the scores below the mean



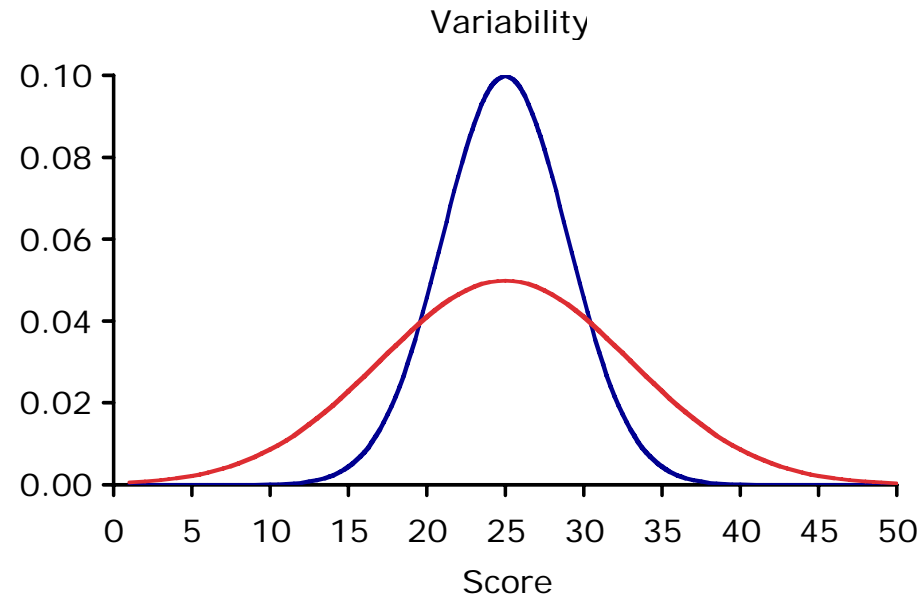
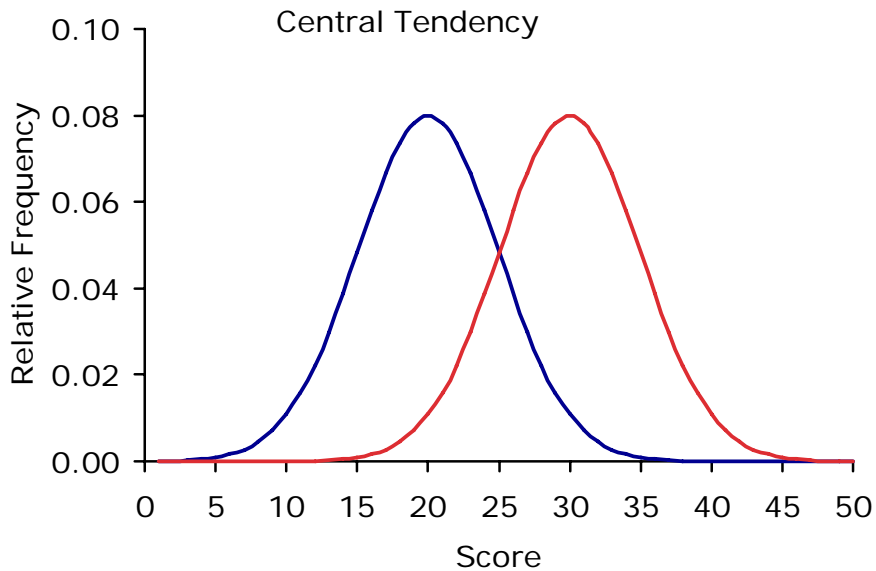
Properties of the Normal Curve

- ⇒ **Bell-shaped**
- ⇒ **Unimodal**
 - ▶ mean = median = mode
- ⇒ **Symmetrical**
- ⇒ **Tails are asymptotic**
 - ▶ i.e., never touch the x axis
- ⇒ **Tails never below x axis**
 - ▶ i.e., probability never negative



Properties of the Normal Curve

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



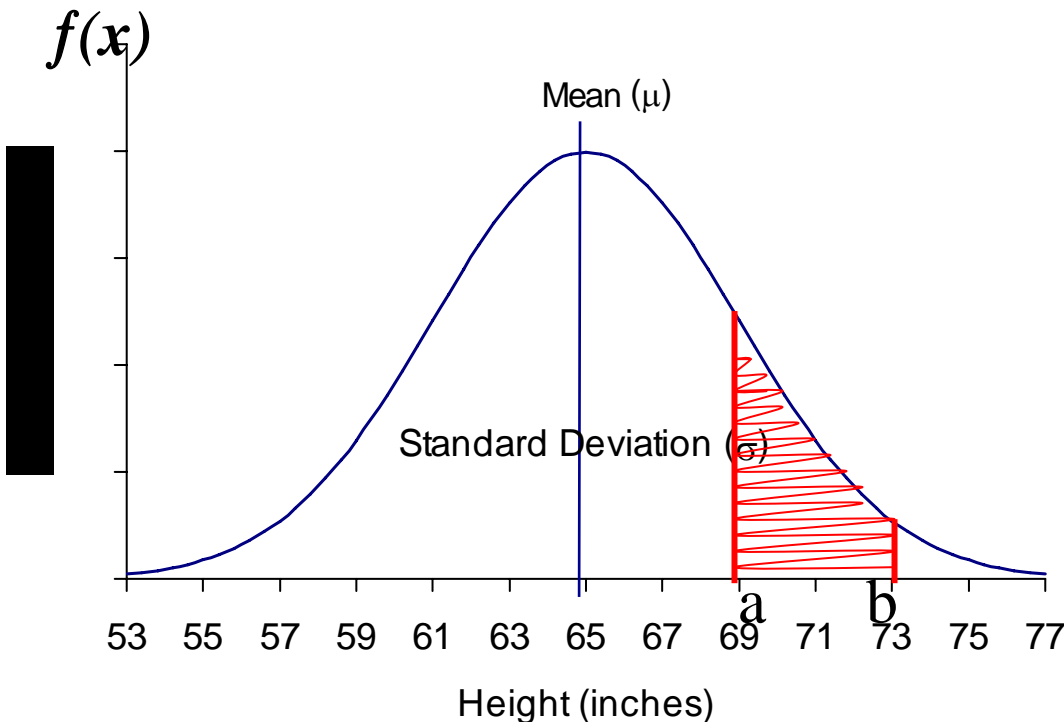
The Normal Curve

⇒ The Theoretical Normal Distribution

▶▶ Area under the curve = 1

▶ no matter what the mean and standard deviation of the curve

▶▶ The Y-axis is labelled **Relative Frequency, $f(x)$**



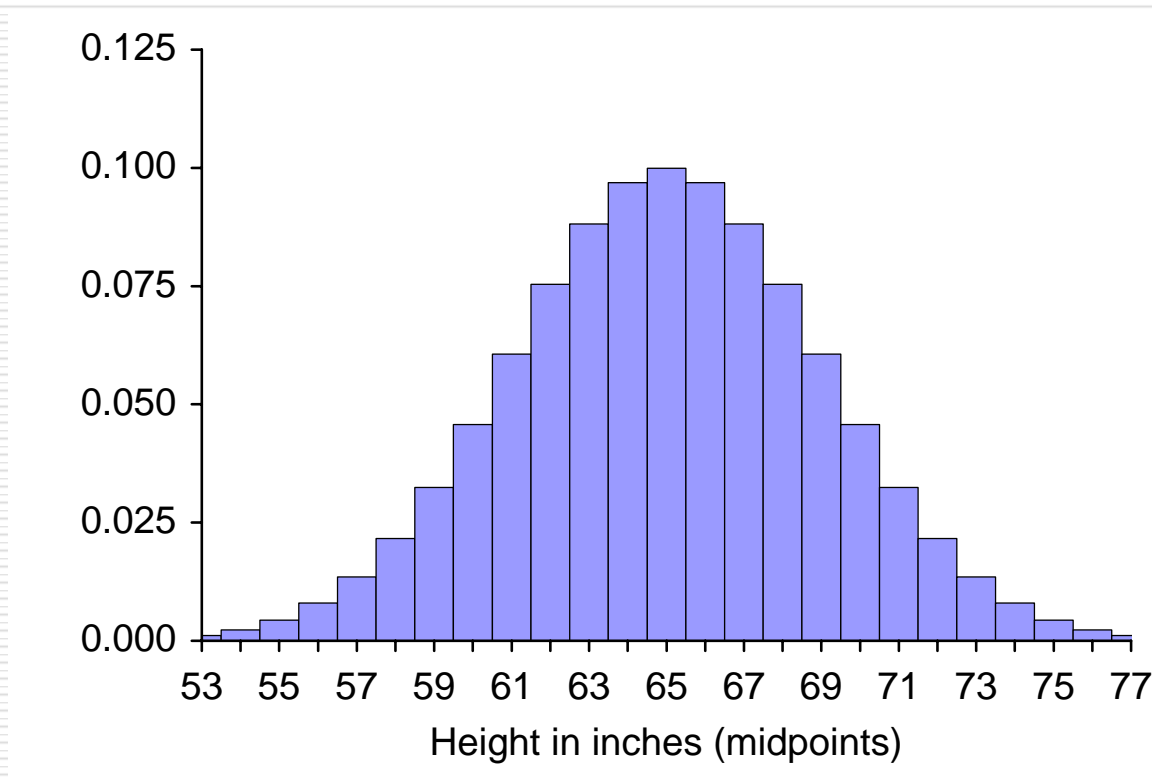
$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

Theoretical probability function called probability density function (pdf)

The Normal Curve

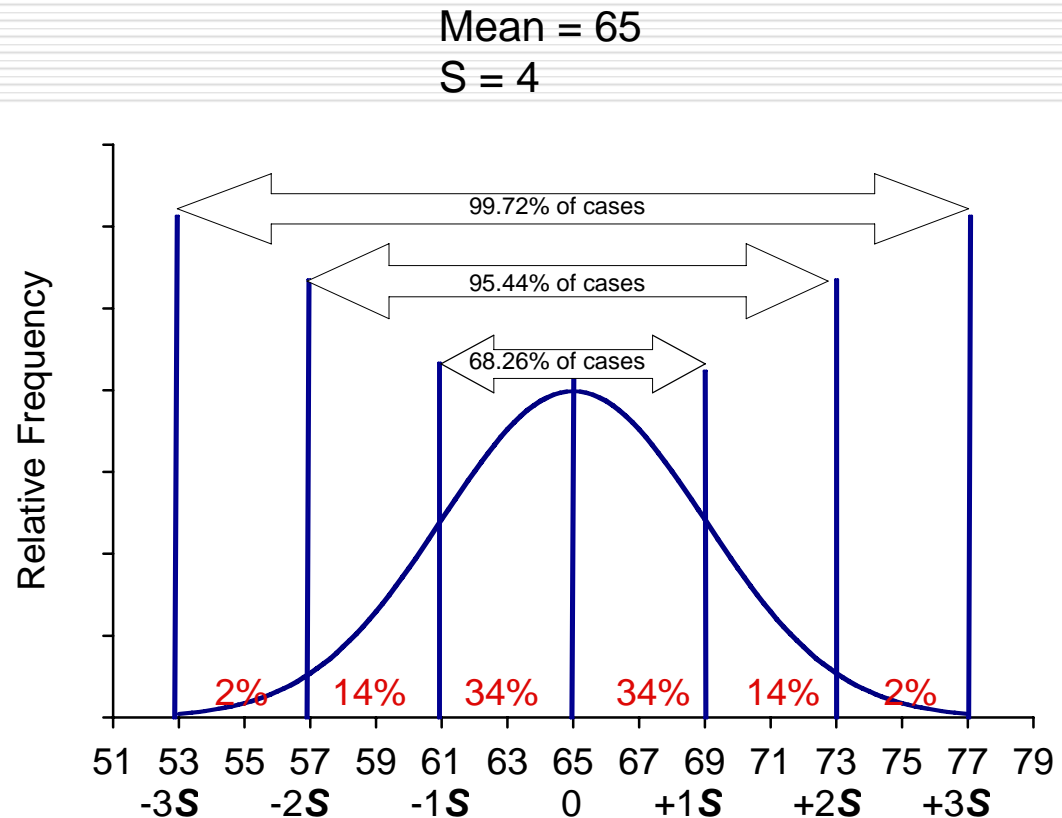
⇒ The Empirical Normal Distribution

- ▶ The empirical distribution is an approximation to the theoretical distribution

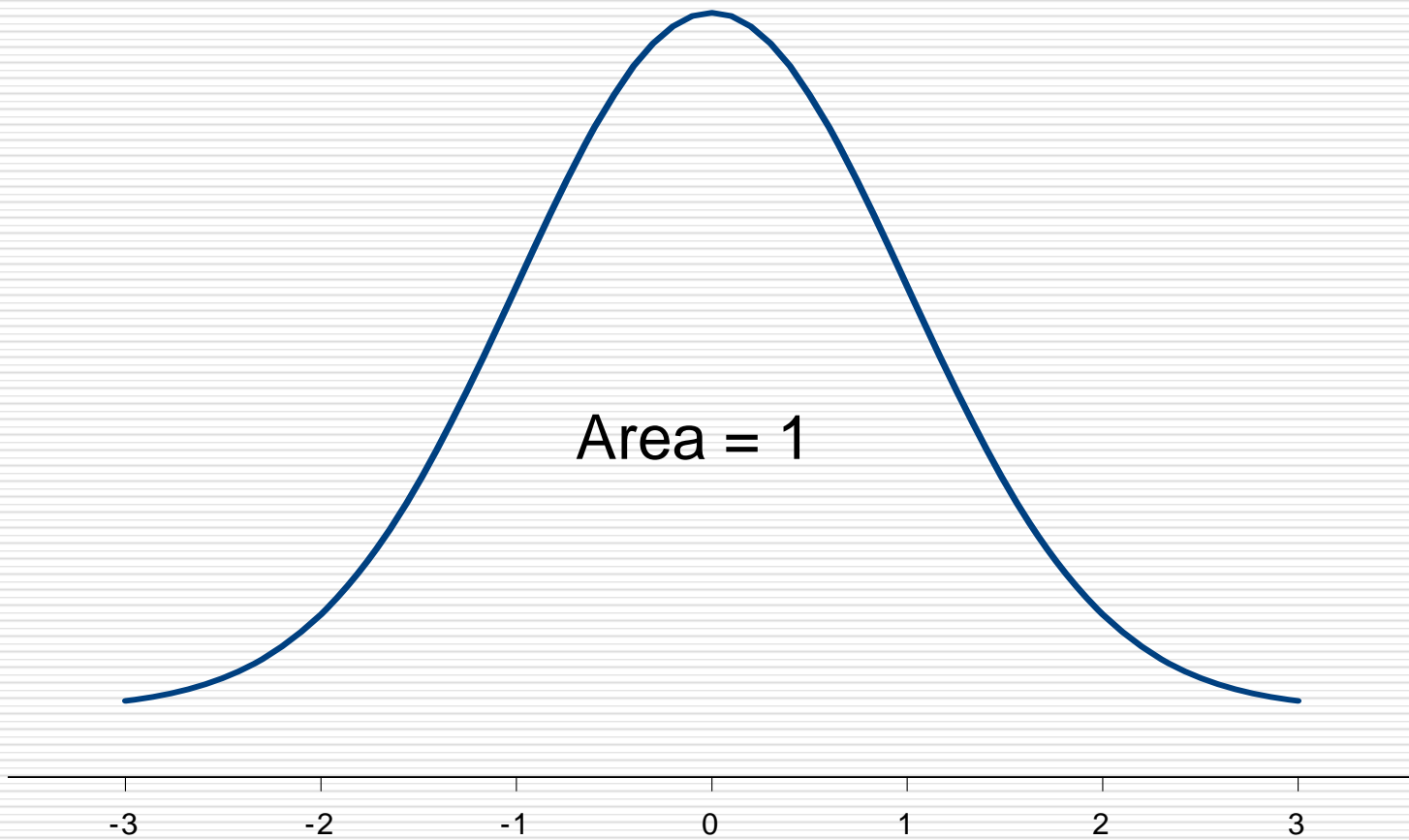


The Normal Curve

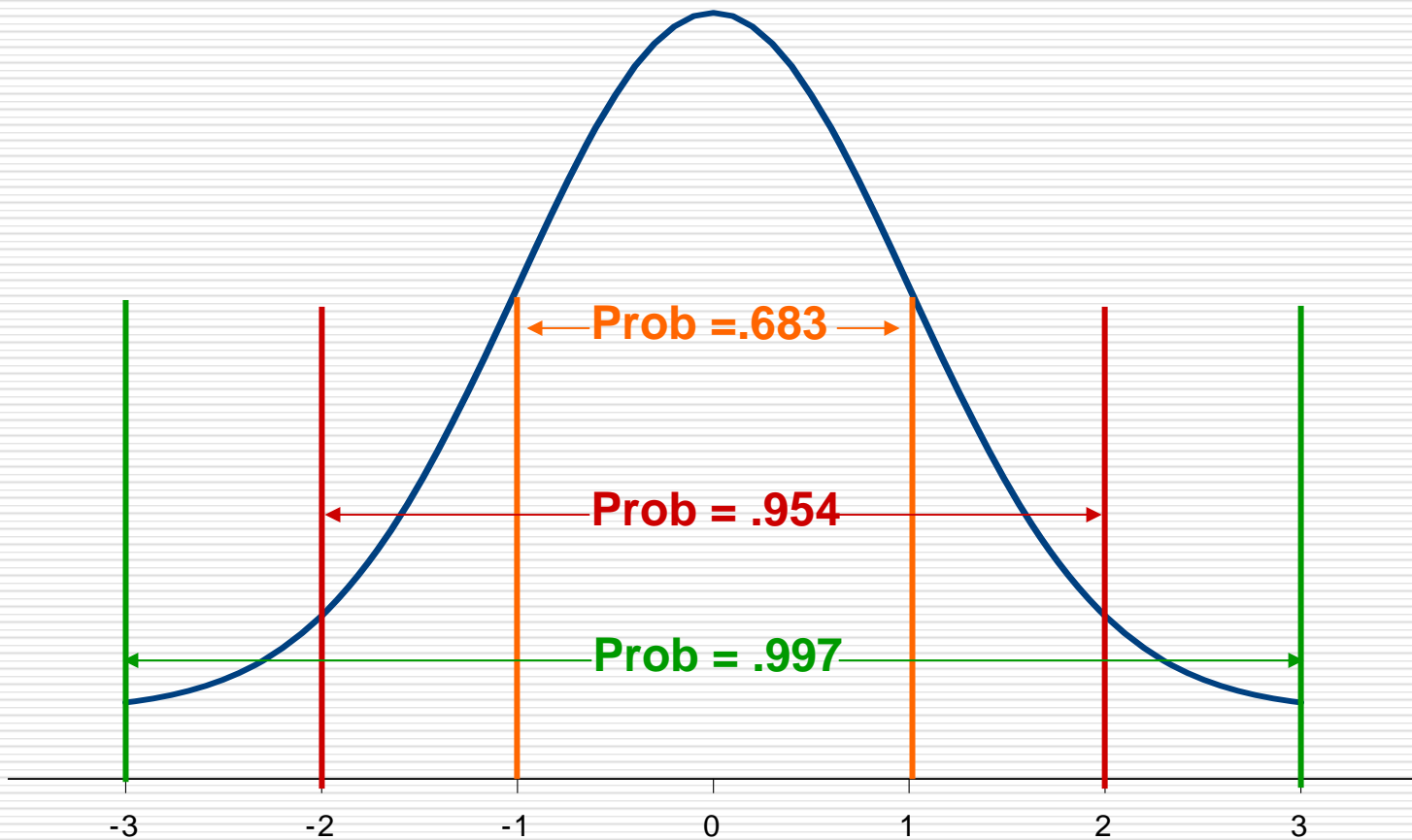
- ⇒ There are known percentages of scores above or below any given point on a normal curve
- ▶▶ 34% of scores between the mean and 1 SD above or below the mean
 - ▶▶ An additional 14% of scores between 1 and 2 SDs above or below the mean
 - ▶▶ Thus, about 96% of all scores are within 2 SDs of the mean (34% + 34% + 14% + 14% = 96%)
- ⇒ Note: 34% and 14% figures can be useful to remember



Normal Probability Distribution

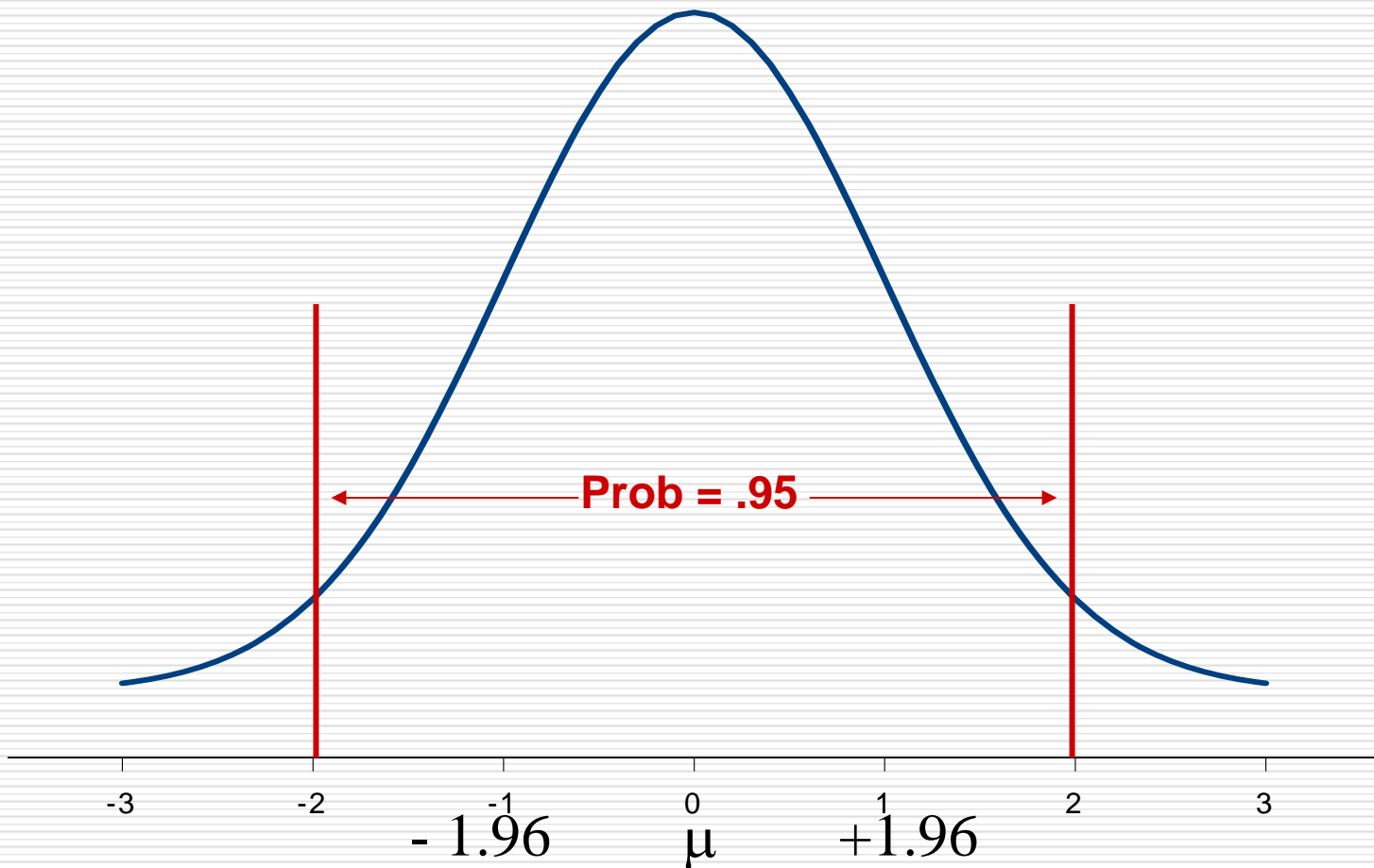


Normal Probability Distribution



Normal Probability Distribution

95% of the population = normal range

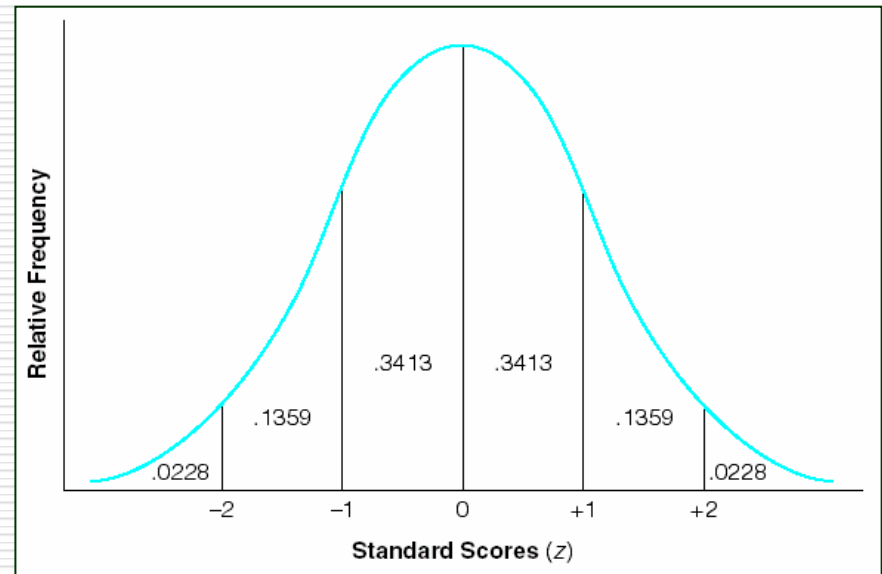


The Standard Normal Curve

- ⇒ “A standard score expresses a score’s position in relation to the mean of the distribution, using the standard deviation as the unit of measurement”
- ⇒ “A z-score states the number of standard deviations by which the original score lines above or below the mean”
- ⇒ Any score can be converted to a z-score as follows

$$\Rightarrow Z = (x - \mu) / \sigma$$

- ⇒ The **standard normal distribution** has a mean of 0 and a standard deviation of 1.



$$z = \frac{X - \bar{X}}{S}$$

The Normal Curve Table

- ⇒ **Normal curve table gives the precise percentage of scores between the mean (z score of 0) and any other z score.**
- ⇒ **Can be used to determine**
 - ▶▶ Proportion of scores above or below a particular z score
 - ▶▶ Proportion of scores between the mean and a particular z score
 - ▶▶ Proportion of scores between two z scores
- ⇒ **Assumption:** Using a z score table assumes that we are dealing with a normal distribution
 - ▶▶ If scores are drawn from a non-normal distribution (e.g., a rectangular distribution) converting these to z scores does not produce a normal distribution.

The Normal Curve Table

Table A-2. Areas under the standard normal curve.¹

z	Area Between $-z$ & $+z$	Area in Two Tails ($< -z$ & $> +z$)	Area in One Tail ($< -z$ or $> +z$)
0.00	0.000	1.000	0.500
0.05	0.040	0.960	0.480
0.10	0.080	0.920	0.460
0.15	0.119	0.881	0.440
0.20	0.159	0.841	0.421
0.25	0.197	0.803	0.401
0.30	0.236	0.764	0.382
0.35	0.274	0.726	0.363
0.40	0.311	0.689	0.345
0.45	0.347	0.653	0.326
0.50	0.383	0.617	0.309
0.55	0.418	0.582	0.291
0.60	0.451	0.549	0.274
0.65	0.484	0.516	0.258
0.70	0.516	0.484	0.242
0.75	0.547	0.453	0.227
0.80	0.576	0.424	0.212
0.85	0.605	0.395	0.198
0.90	0.632	0.368	0.184
0.95	0.658	0.342	0.171
1.00 ²	0.683	0.317	0.159
1.05	0.706	0.294	0.147
1.10	0.729	0.271	0.136
1.15	0.750	0.250	0.125
1.20	0.770	0.230	0.115
1.25	0.789	0.211	0.106
1.28 ²	0.800	0.200	0.100

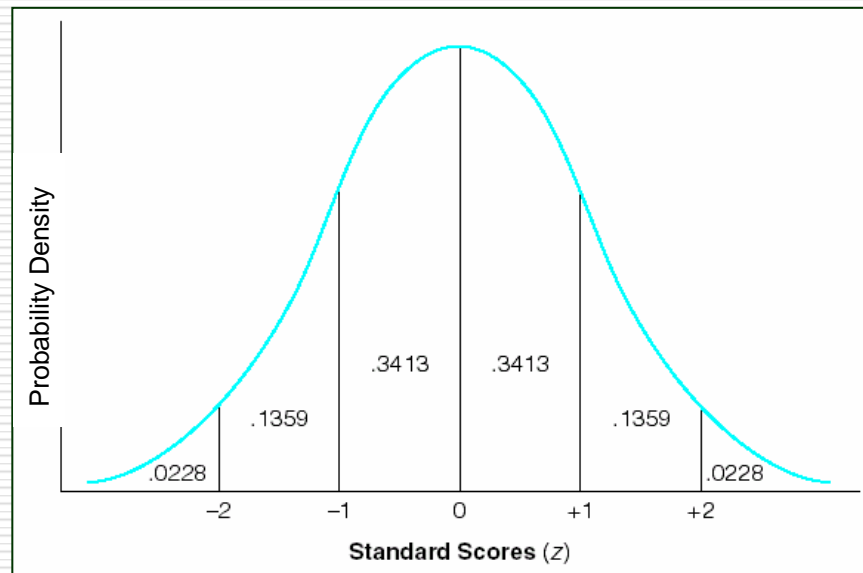
The Normal Curve Table

1.30	0.806	0.194	0.097
1.35	0.823	0.177	0.089
1.40	0.838	0.162	0.081
1.45	0.853	0.147	0.074
1.50	0.866	0.134	0.067
1.55	0.879	0.121	0.061
1.60	0.890	0.110	0.055
1.645 ²	0.900	0.100	0.050
1.65	0.901	0.099	0.049
1.70	0.911	0.089	0.045
1.75	0.920	0.080	0.040
1.80	0.928	0.072	0.036
1.85	0.936	0.064	0.032
1.90	0.943	0.057	0.029
1.95	0.949	0.051	0.026
1.96 ²	0.950	0.050	0.025
2.00	0.954	0.046	0.023
2.05	0.960	0.040	0.020
2.10	0.964	0.036	0.018
2.15	0.968	0.032	0.016
2.20	0.972	0.028	0.014
2.25	0.976	0.024	0.012
2.30	0.979	0.021	0.011
2.326 ²	0.980	0.020	0.010
2.35	0.981	0.019	0.009
2.40	0.984	0.016	0.008
2.45	0.986	0.014	0.007
2.50	0.988	0.012	0.006
2.55	0.989	0.011	0.005
2.575 ²	0.990	0.010	0.005
2.60	0.991	0.009	0.005
2.65	0.992	0.008	0.004
2.70	0.993	0.007	0.003
2.75	0.994	0.006	0.003

Normal Curve Table Continued

⇒ **The Z table can also be used to.**

- ▶ determine a z score for a particular proportion of scores under the normal curve, and
- ▶ Determine the proportion of scores below (or above) a z score



Find the probability

μ SBP = 120, σ = 10, proportion of those with SBP > 125

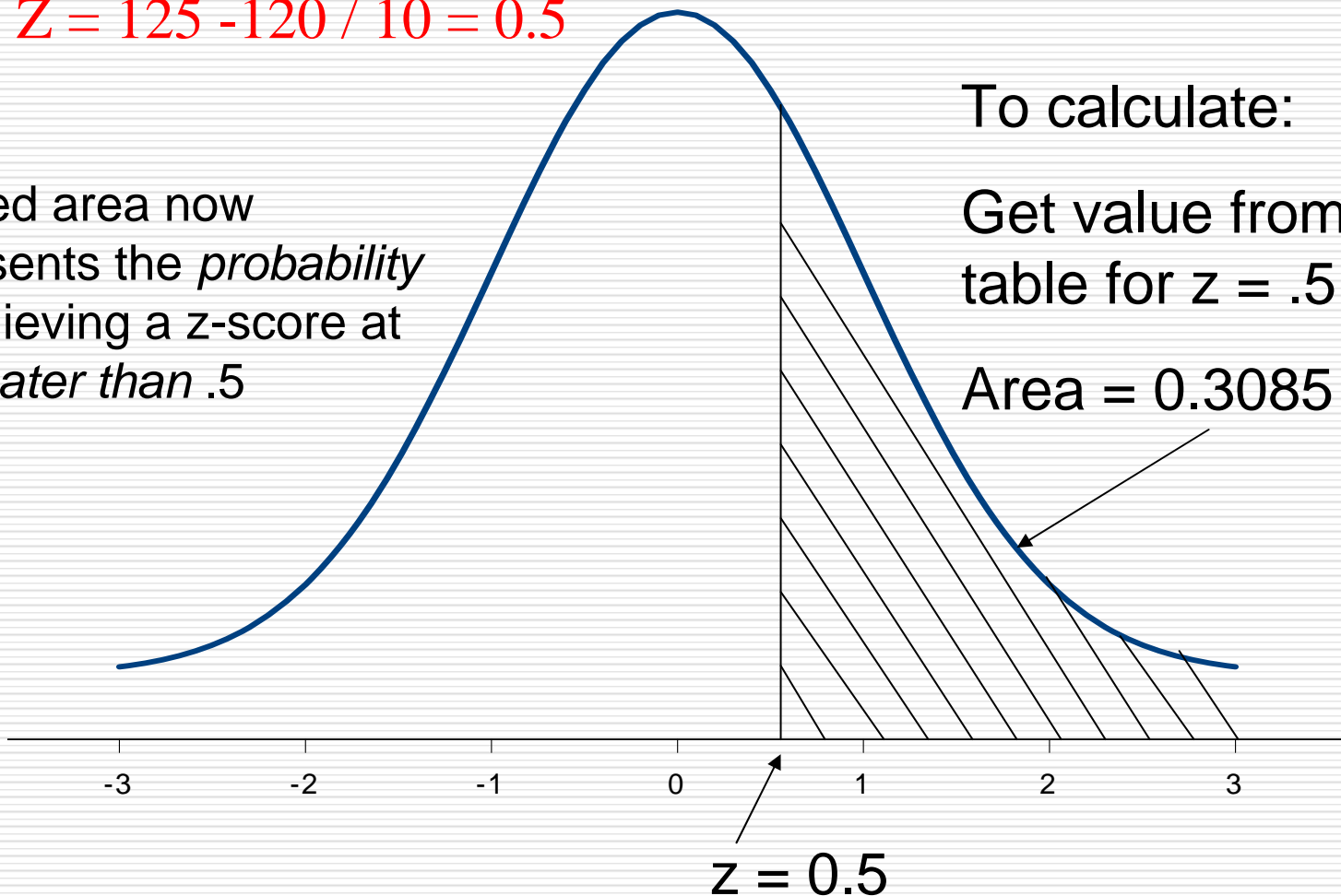
$$Z = 125 - 120 / 10 = 0.5$$

Shaded area now represents the *probability* of achieving a z-score at or greater than .5

To calculate:

Get value from the table for $z = .5$

Area = 0.3085



Find the probability

proportion of those with SBP < 125

To calculate:

Shaded area represents the probability of achieving a z-score at or less than .5

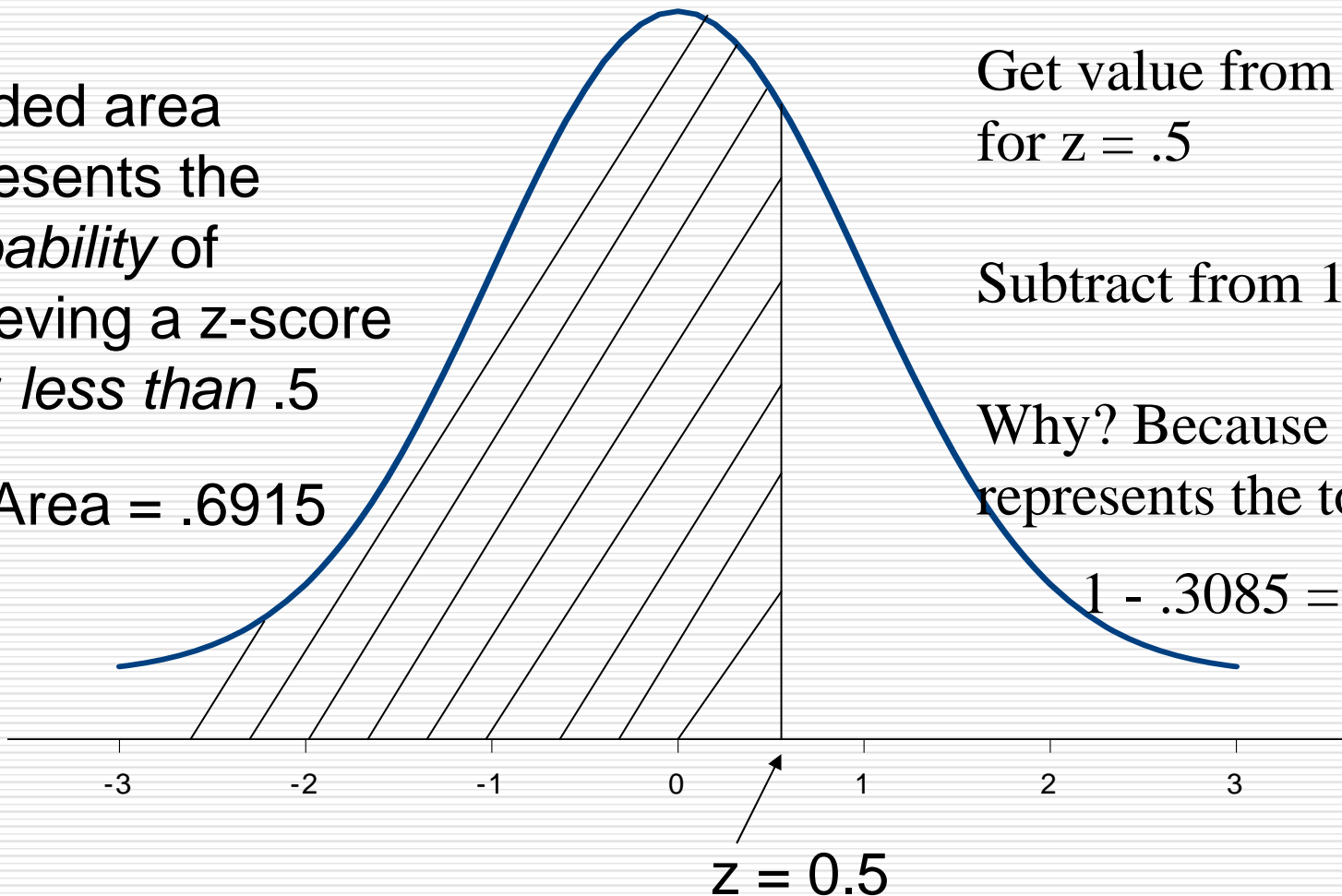
Area = .6915

Get value from the table for $z = .5$

Subtract from 1

Why? Because 1 represents the total area

$$1 - .3085 = .6915$$



Find the probability

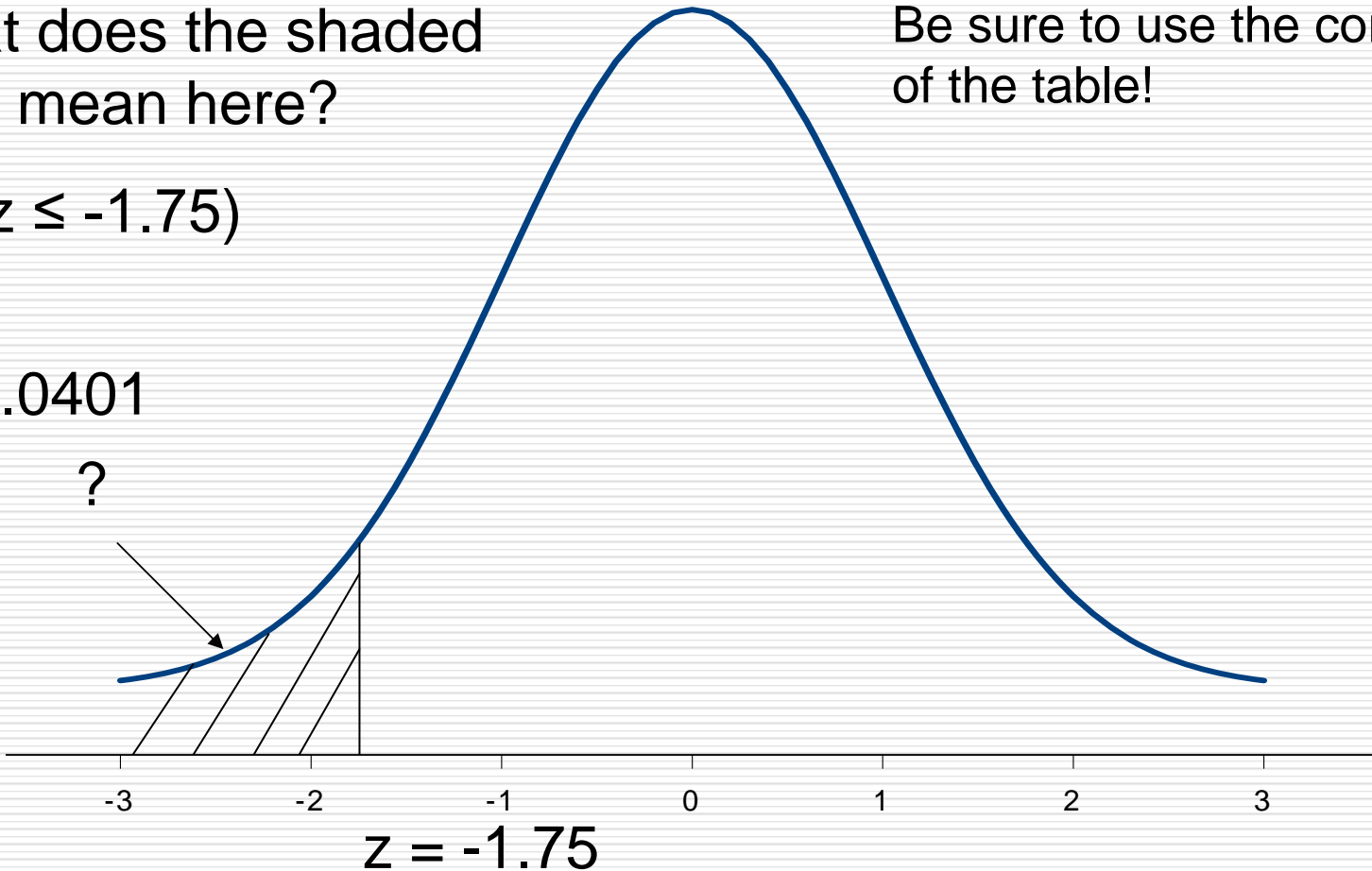
What does the shaded area mean here?

Be sure to use the correct side of the table!

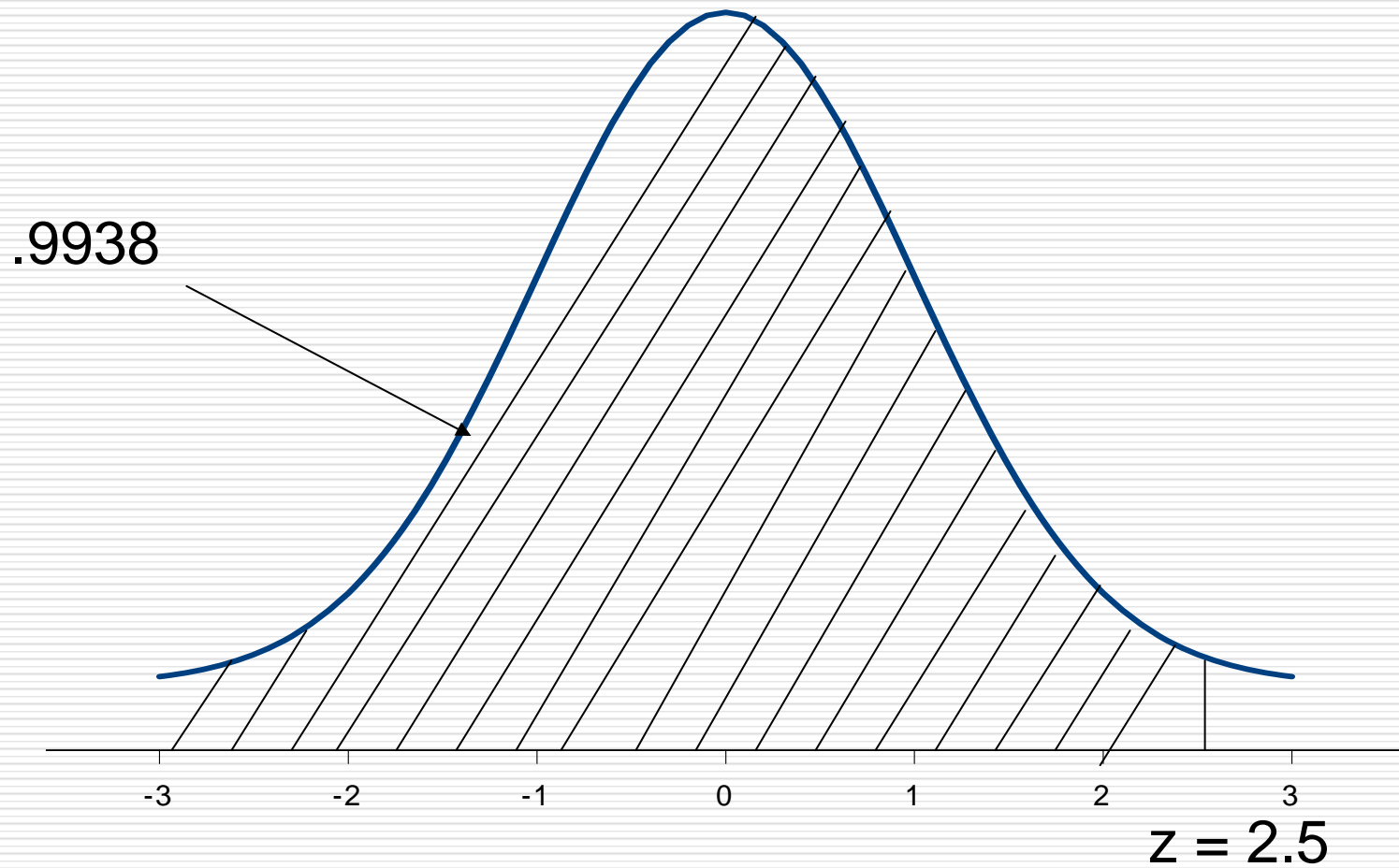
$$P(z \leq -1.75)$$

.0401

?



Find the probability

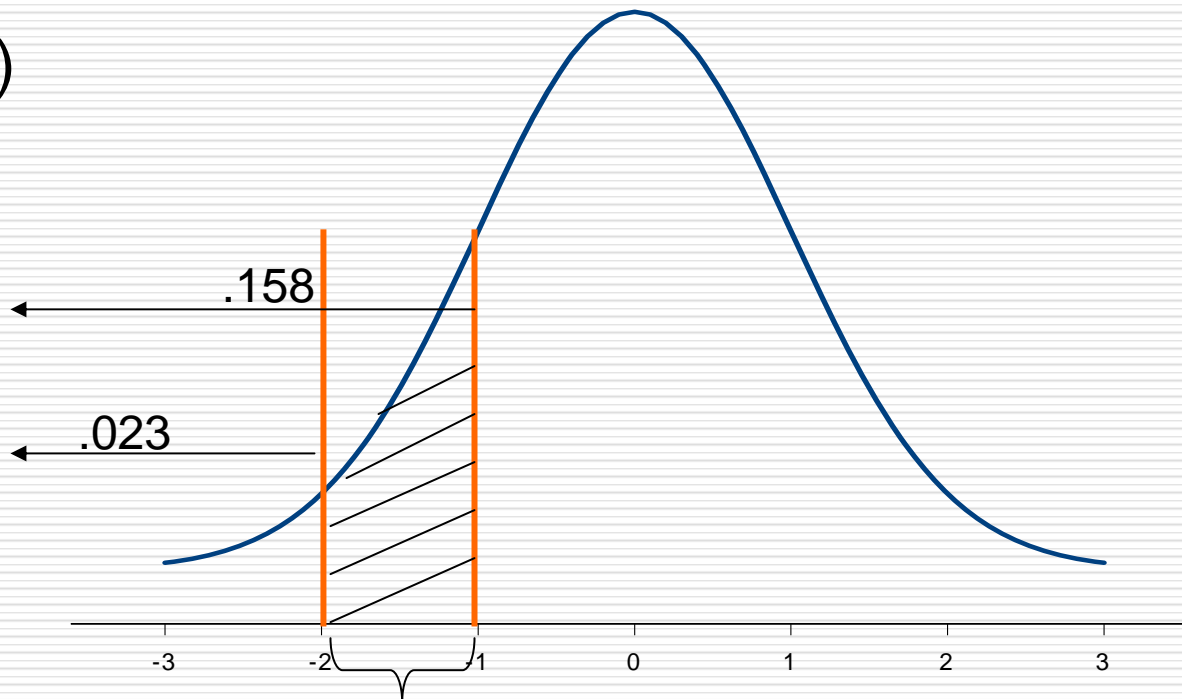


Find the probability

Proportion with SBP < 110 and > 100

What does the shaded area mean?

$$P(-2 < z < -1)$$

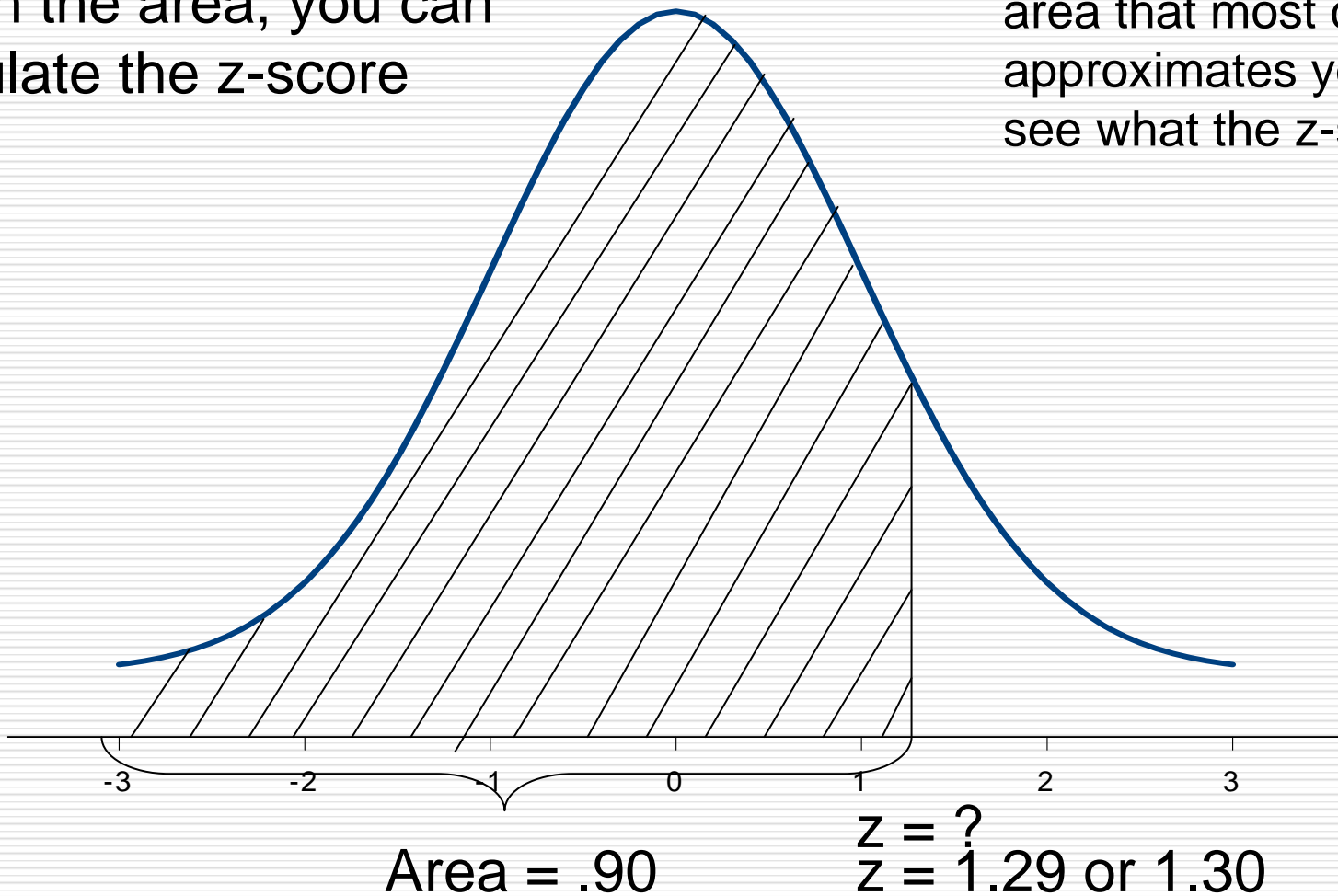


Find the probability your z-score is between -1 and -2

Find the z-score

Given the area, you can calculate the z-score

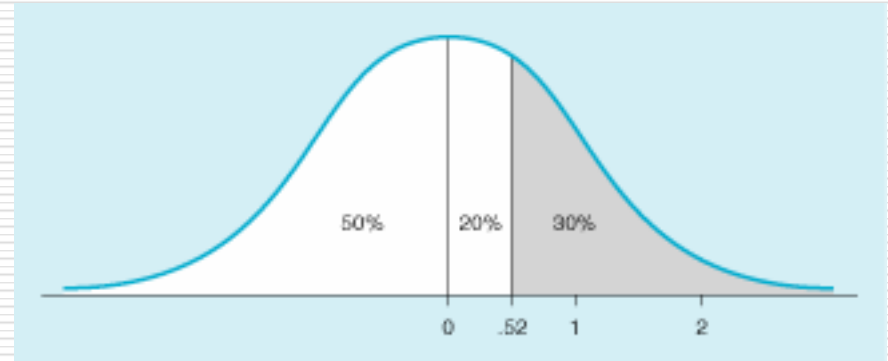
Look in table A-2 for the area that most closely approximates yours and see what the z-score is



Exercise

⇒ **To find the proportion of the curve that lies above or below a particular score**

- ▶▶ Convert raw score to z score, if necessary
- ▶▶ Draw a normal curve
 - ▶ Indicate where z score falls
 - ▶ Shade area you're trying to find
- ▶▶ Make rough estimate of shaded area's percentage
- ▶▶ Find exact percentage with normal curve table
- ▶▶ Check to verify that it's close to your estimate



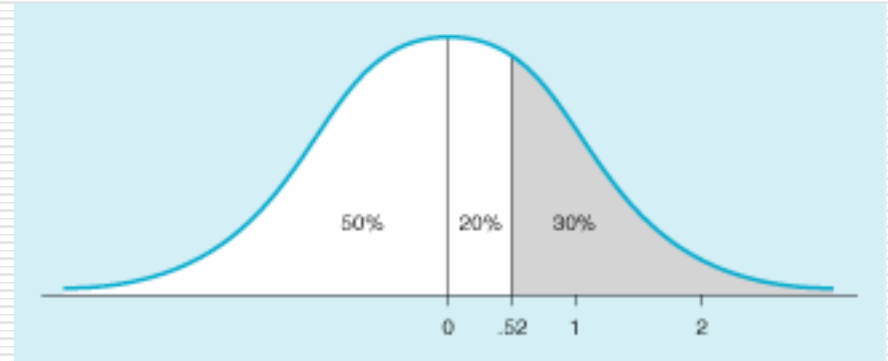
For a normal distribution with a mean of 10 and a standard deviation of 2

- ▶▶ Find the percentage of the distribution that
 - ▶ falls above 12
 - ▶ falls below 12
 - ▶ falls above 8
 - ▶ falls below 8
 - ▶ fall above 9
 - ▶ falls below 7

Exercise

⇒ **To find the proportion of the curve that lies between two scores**

- ▶▶ Convert the raw scores to z scores
- ▶▶ Draw a normal curve
 - ▶ Indicate where the two z scores fall
 - ▶ Shade area you're trying to find
- ▶▶ Make rough estimate of shaded area's percentage
- ▶▶ Find exact percentage with normal curve table
- ▶▶ Check to verify that it's close to your estimate

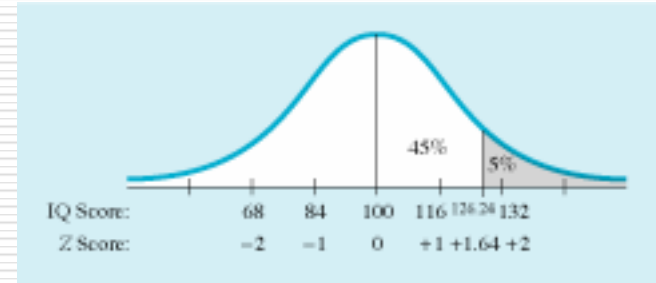


⇒ **For a normal distribution with a mean of 10 and a standard deviation of 2**

- ▶▶ Find the percentage of the distribution that
 - ▶ falls between 10 and 12
 - ▶ falls between 8 and 10
 - ▶ falls between 6 and 10
 - ▶ falls between 6 and 8
 - ▶ falls between 10.5 and 11
 - ▶ falls between 8.5 and 11

Exercise

- ⇒ Draw normal curve, shading approximate area for the percentage desired
- ⇒ Make a rough estimate of the **Z** score where the shaded area starts
- ⇒ Find the exact **Z** score using normal curve table
- ⇒ Check to verify that it's close to your estimate
- ⇒ Convert **Z** score to raw score, if desired



- ⇒ For a normal distribution with a mean of 10 and a standard deviation of 2
 - ▶▶ Find the raw score for which
 - ▶ 50% of the distribution falls above it
 - ▶ 84% of the distribution falls below it
 - ▶ 98% falls above it
 - ▶ 62% falls below it
 - ▶ 30% falls above it

!!! Remember !!!

⇒ **We can use the standard normal distribution table **ONLY** when our distribution of scores is normal.**

▶▶ Using the standard normal table is not appropriate if our distribution differs markedly from normality

▶▶ e.g.,

- ▶ rectangular
- ▶ skewed
- ▶ leptokurtic
- ▶ bimodal

Comparing Scores from Different Distributions

⇒ **Again: The standard normal distribution has a mean of 0 and standard deviation of 1**

⇒ **Consider two sections of statistics**

▶▶ Class A has a mean of 80 and S of 5

▶▶ Class B has a mean of 70 and S of 5

▶ Student 1 gets 80 in Class A

▶ Student 2 gets 75 in Class A

▶▶ Which student did better?

Percentile Ranks and the Normal Distribution

- ⇒ **When we ask what proportion of a distribution lies below a particular z score, we are actually asking what is the percentile rank of the score**
- ⇒ **e.g., in a distribution with a mean of 100 and standard deviation of 15, 84% of the distribution falls below a score of 115 [$z = (115-100)/15 = 1$].**
 - ▶▶ Therefore, the percentile rank of 115 is 84%

Sampling distribution

The Concept of Sampling Distributions

Parameter – numerical descriptive measure of a population. It is usually unknown

Sample Statistic - numerical descriptive measure of a sample. It is usually known

Sampling distribution – the probability distribution of a sample statistic, calculated from a very large number of samples of size n

The Concept of Sampling Distributions

19, 19, 20, 21, 20, 25, 22, 18, 18, 17

We can take 45 samples of size 2 from this group of 10 observations

$\mu = 19.9$

If we take one random sample and get (19, 20),

Another random sample may yield (22, 25), with

$$\bar{x} = 19.5$$

$$\bar{x} = 23.5$$

The Concept of Sampling Distributions

Taking all possible samples of size 2, we can graph them and come up with a sampling distribution of the sample statistic \bar{x}

Sampling distributions can be derived for any statistic

Knowing the properties of the underlying sampling distributions allows us to judge how accurate the statistics are as estimates of parameters

The Concept of Sampling Distributions

Decisions about which sample statistic to use must take into account the sampling distribution of the statistics you will be choosing from.

The Concept of Sampling Distributions

Given the probability distribution

X	0	6	9
p(x)	1/3	1/3	1/3

Find the sampling distribution of mean and median of x

Taking all possible samples of size 3

Sampling distribution

Possible Samples	\bar{x}	m	Probability
0,0,0	0	0	1/27
0,0,6	2	0	1/27
0,0,9	3	0	1/27
0,6,0	2	0	1/27
0,6,6	4	6	1/27
0,6,9	5	6	1/27
0,9,0	3	0	1/27
0,9,6	5	6	1/27
0,9,9	6	9	1/27
6,0,0	2	0	1/27
6,0,6	4	6	1/27
6,0,9	5	6	1/27
6,6,0	4	6	1/27
6,6,6	6	6	1/27
6,6,9	7	6	1/27
6,9,0	5	6	1/27
6,9,6	7	6	1/27
6,9,9	8	9	1/27
9,0,0	3	0	1/27
9,0,6	5	6	1/27
9,0,9	6	9	1/27
9,6,0	5	6	1/27
9,6,6	7	6	1/27
9,6,9	8	9	1/27
9,9,0	6	9	1/27
9,9,6	8	9	1/27
9,9,9	9	9	1/27

of x

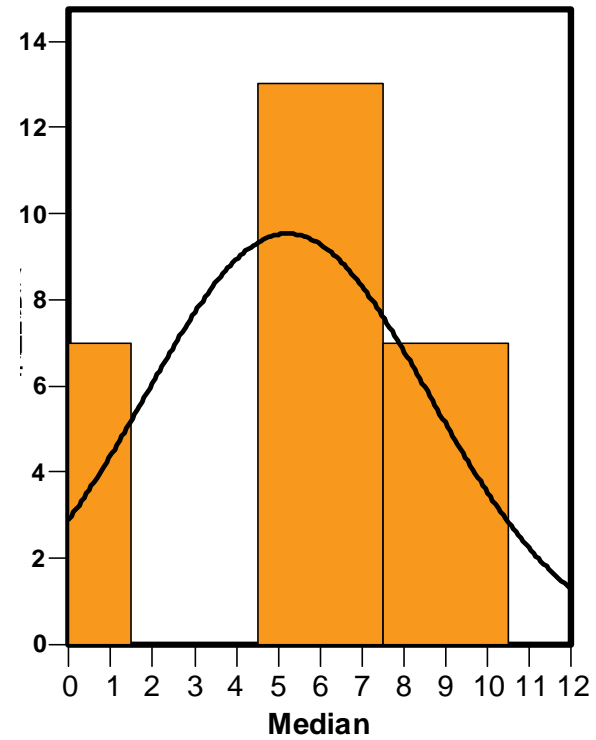
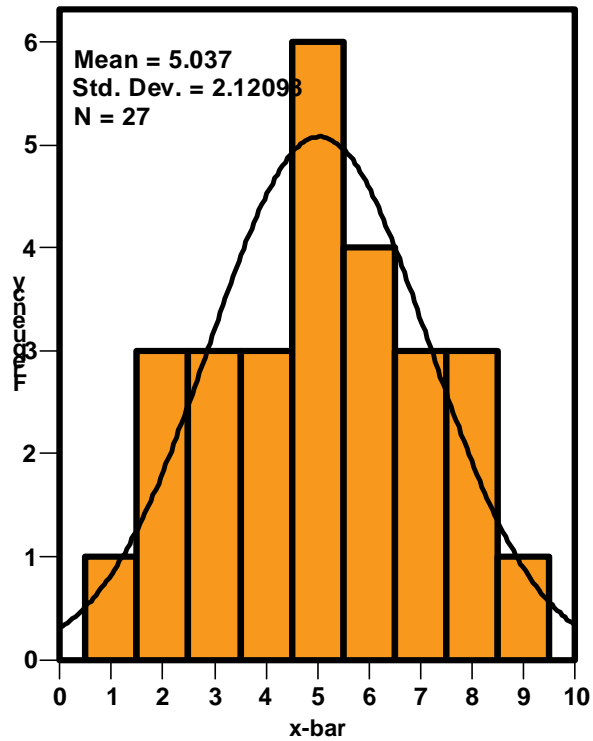
\bar{x}	$P(\bar{x})$
0	1/27
2	3/27
3	3/27
4	3/27
5	6/27
6	4/27
7	3/27
8	3/27
9	1/27

Sampling distribution
of median

m	$P(m)$
0	7/27
6	13/27
9	7/27

Sampling distribution of mean and median

Sample size = 3



The Concept of Sampling Distributions

Simulating a Sampling Distribution

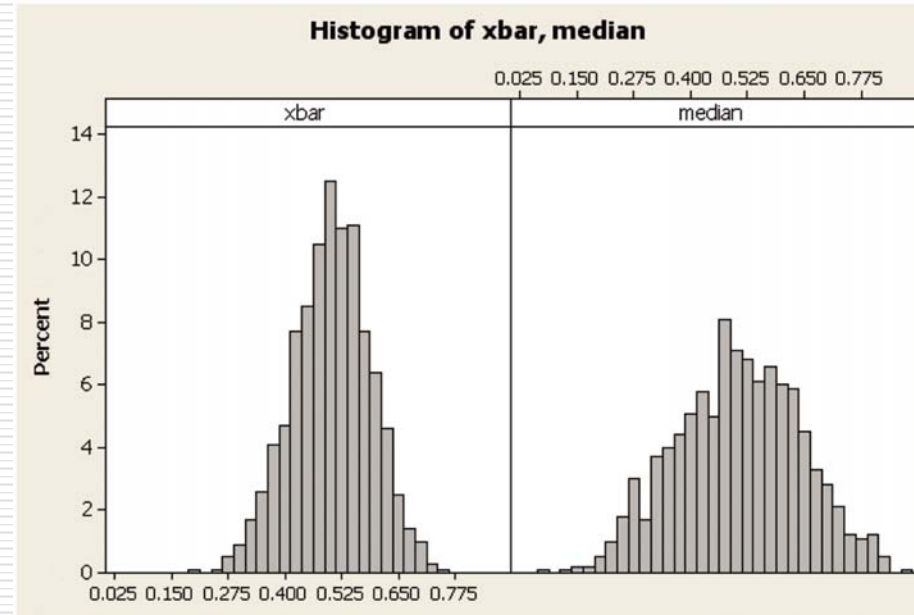
Use a software package to generate samples of size $n = 11$ from a population with a known $\mu = .5$

Calculate the mean and median for each sample

Generate histograms for the means and medians of the samples

Note the greater clustering of the values of \bar{x} around μ

These histograms are approximations of the sampling distributions of \bar{x} and m



Properties of Sampling Distributions: Unbiasedness and Minimum Variance

Point Estimator – formula or rule for using sample data to calculate an estimate of a population parameter

Point estimators have sampling distributions

These sampling distributions tell us how accurate an estimate the point estimator is likely to be

Sampling distributions can also indicate whether an estimator is likely to under/over estimate a parameter

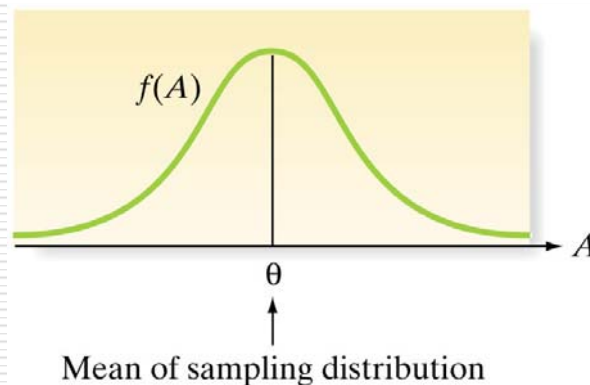
Properties of Sampling Distributions: Unbiasedness and Minimum Variance

Two point estimators, A and B, of parameter θ

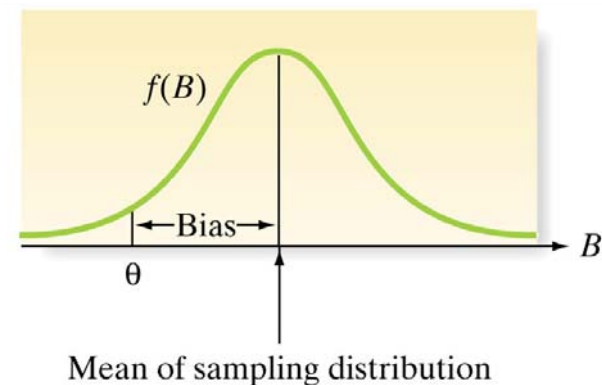
After generating the sampling distributions of A and B, we can see that

A is an **unbiased estimator** of θ

B is a **biased estimator** of θ , with a bias toward overstatement



a. Unbiased sample statistic for the parameter θ



b. Biased sample statistic for the parameter θ

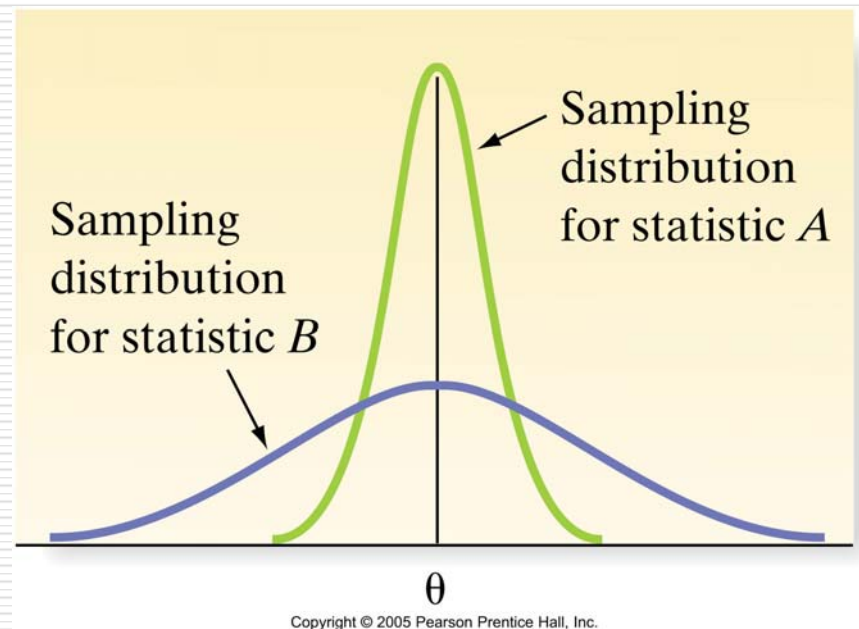
Properties of Sampling Distributions: Unbiasedness and Minimum Variance

What if A and B are both unbiased estimators of θ ?

Look at the sampling distributions and compare their standard deviations

A has a smaller standard deviation than B

Which would you use as your estimator?



The Sampling Distribution of \bar{X} and the Central Limit Theorem

Assume 1000 samples of size n taken from a population, with \bar{x} calculated for each sample. What are the Properties of the Sampling Distribution of \bar{x} ?

Mean of sampling distribution equals mean of sampled population

$$\mu_{\bar{x}} = \mu$$

Standard deviation of sampling distribution equals

Standard deviation of sampled population

Square root of sample size

or,
$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$

$\sigma_{\bar{x}}$ is referred to as the standard error of the mean

The Sampling Distribution of \bar{X} and the Central Limit Theorem

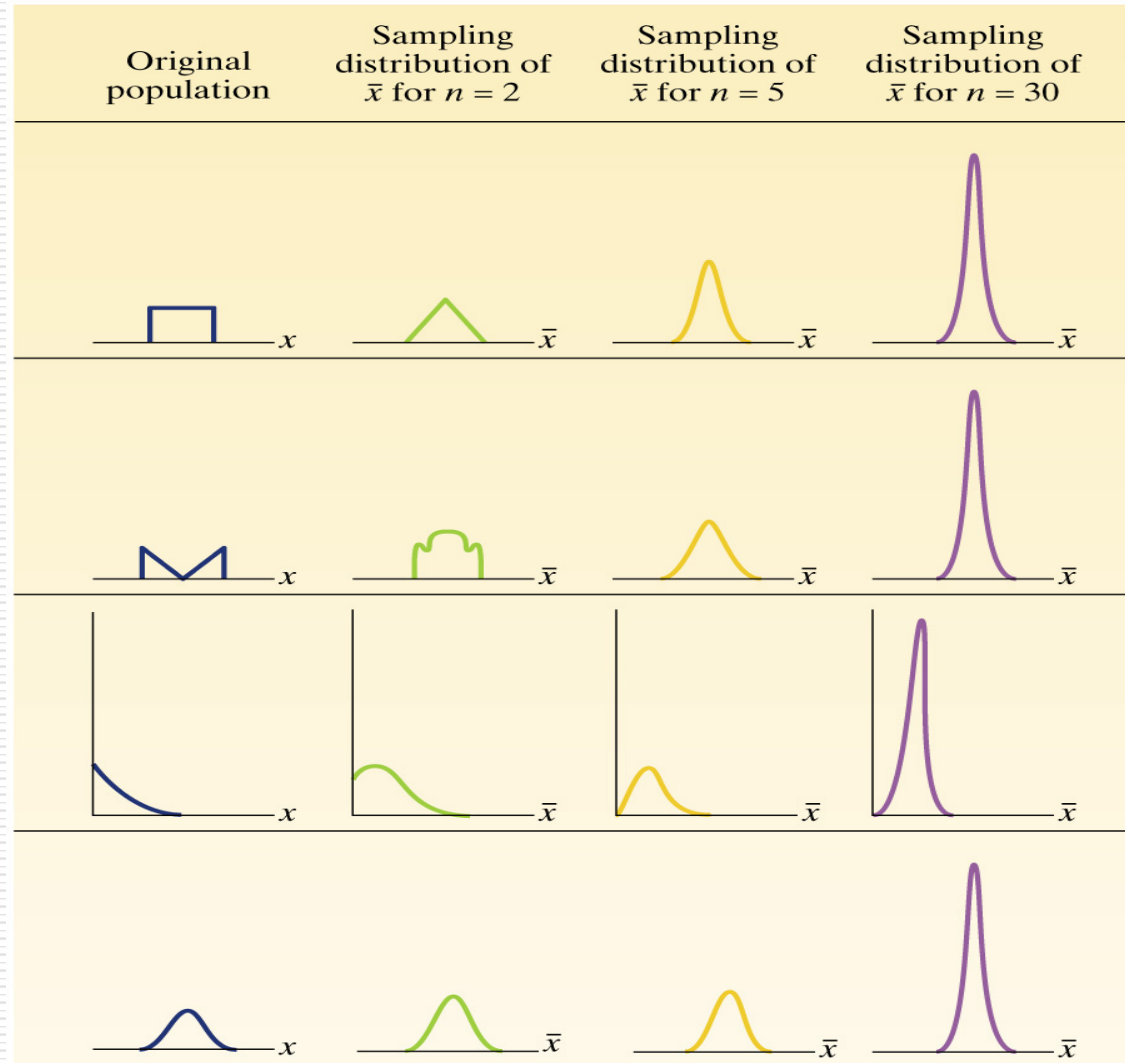
If we sample n observations from a normally distributed population, the sampling distribution of \bar{x} will be a normal distribution

Central Limit Theorem

In a population with standard deviation σ and mean μ , the distribution of sample means from samples of n observations will approach a normal distribution with standard deviation of $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ and mean of $\mu_{\bar{x}} = \mu$ as n gets larger. The larger the n , the closer the sampling distribution of \bar{x} to a normal distribution.

The Sampling Distribution of \bar{X} and the Central Limit Theorem

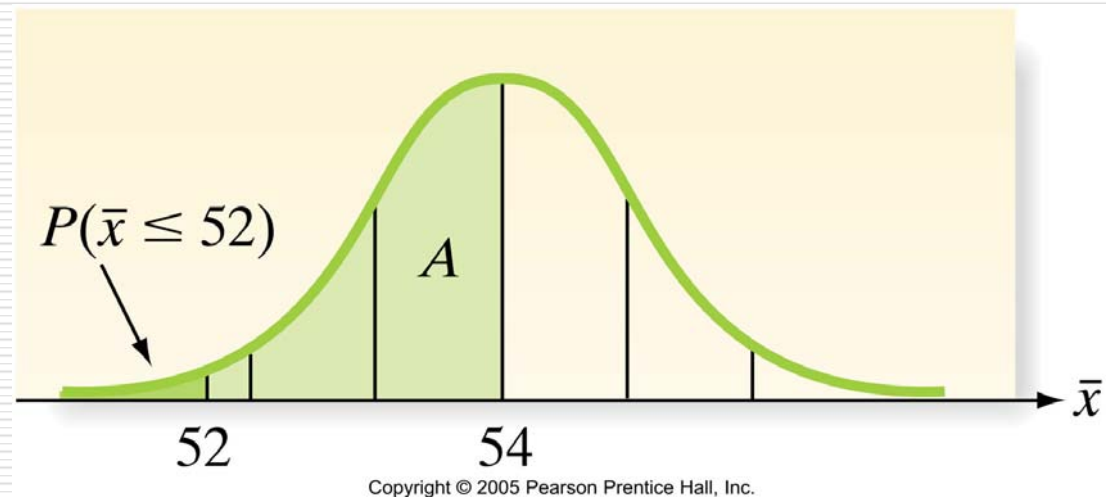
Note how the sampling distribution approaches the normal distribution as n increases, whatever the shape of the distribution of the original population



The Sampling Distribution of \bar{X} and the Central Limit Theorem

Assume a population with $\mu = 54$, $\sigma = 6$. If a sample of 50 is taken from this population, what is the probability that the sample mean is less than or equal to 52?

Sketch the curve of \bar{x} and identify area of interest



The Sampling Distribution of \bar{X} and the Central Limit Theorem

Convert 52 to z value

First, calculate the standard deviation of the sampling distribution

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{50}} = .85$$

Then calculate the z value

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{52 - 54}{.85} = -2.35$$

Use the tables to find probability of interest

$$P(\bar{x} \leq 52) = .5 - .4906 = .0094$$

Standard error of the mean SEM and standard variation

- ⇒ In real life we usually do not have the luxury or time to do repeated sampling, only one sample to estimate a population parameter.
- ⇒ With relatively large sample sizes and with the use of central limit theorem, we have a reliable estimate of population parameter
- ⇒ Standard deviation expresses the amount of variability among individual observations of the sample
- ⇒ $SEM = \sigma / \sqrt{n}$ it measures the amount of variability in the sample mean, it indicates how closely the population mean is likely to be estimated by sample mean
- ⇒ SEM is a function of both population variability and sample size
- ⇒ Because standard deviation and SEM are often confused it is very important that they are clearly labelled when presented in tables of results