Class 15

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Agenda:

Review Chapters 6-8
(Exam 2 Chapters)

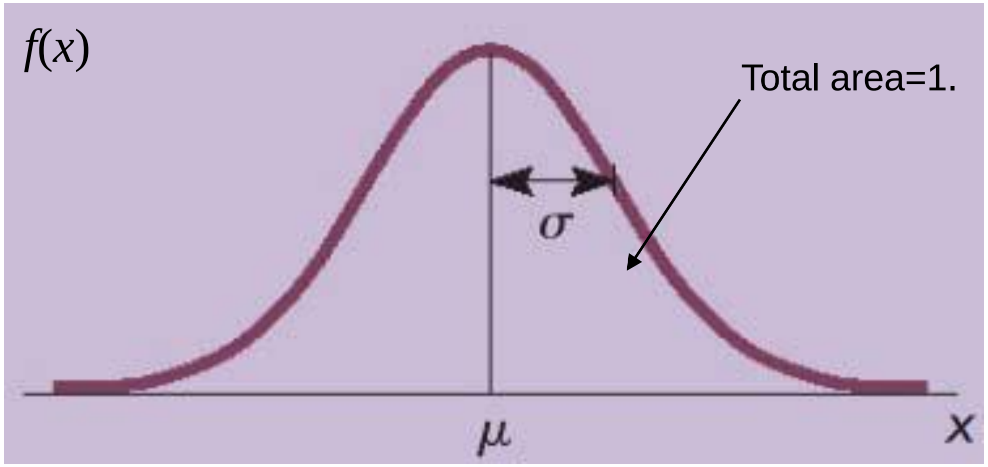
Just the highlights!

6: Normal Probability Distributions

6.1 Normal Probability Distributions

The mathematical formula for the normal distribution is (p 269):

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



where

e = 2.718281828459046…

π = 3.141592653589793…

μ = population mean

σ = population std. deviation

$$-\infty < x, \mu < +\infty$$
$$0 < \sigma$$

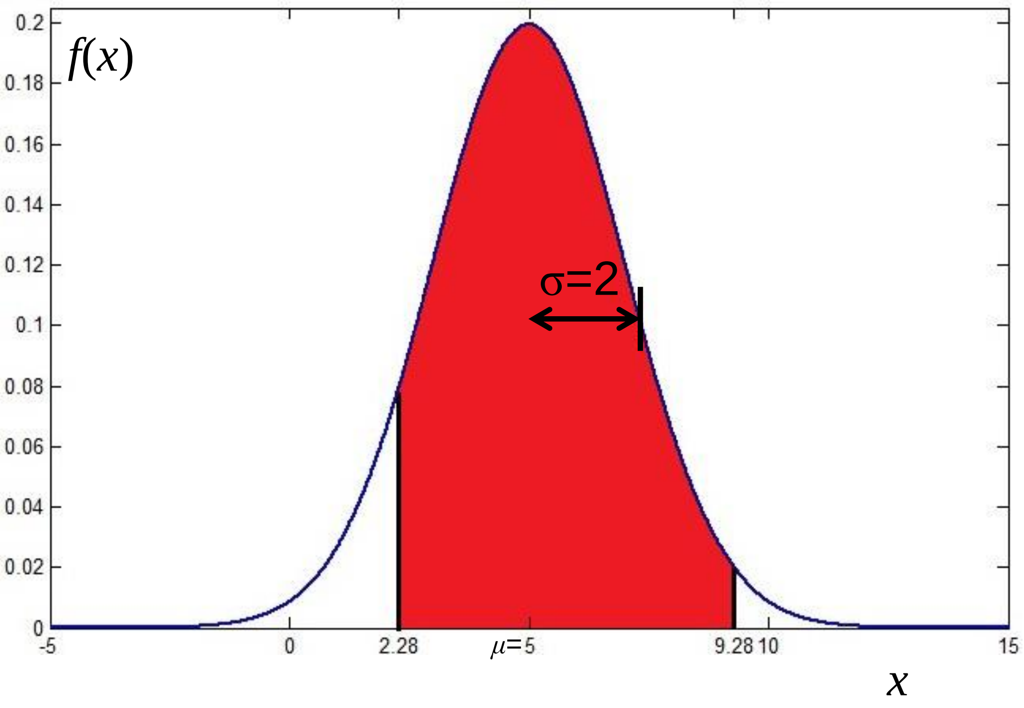
We will not use this formula.

Figure from Johnson & Kuby, 2012.

6: Normal Probability Distributions

6.2 The Standard Normal Probability Distributions

Example: Here is a normal distribution with $\mu = 5$ and $\sigma^2 = 4$.



Let's say we want to know the red area under the normal distribution between $x_1 = 2.28$ and $x_2 = 9.28$.

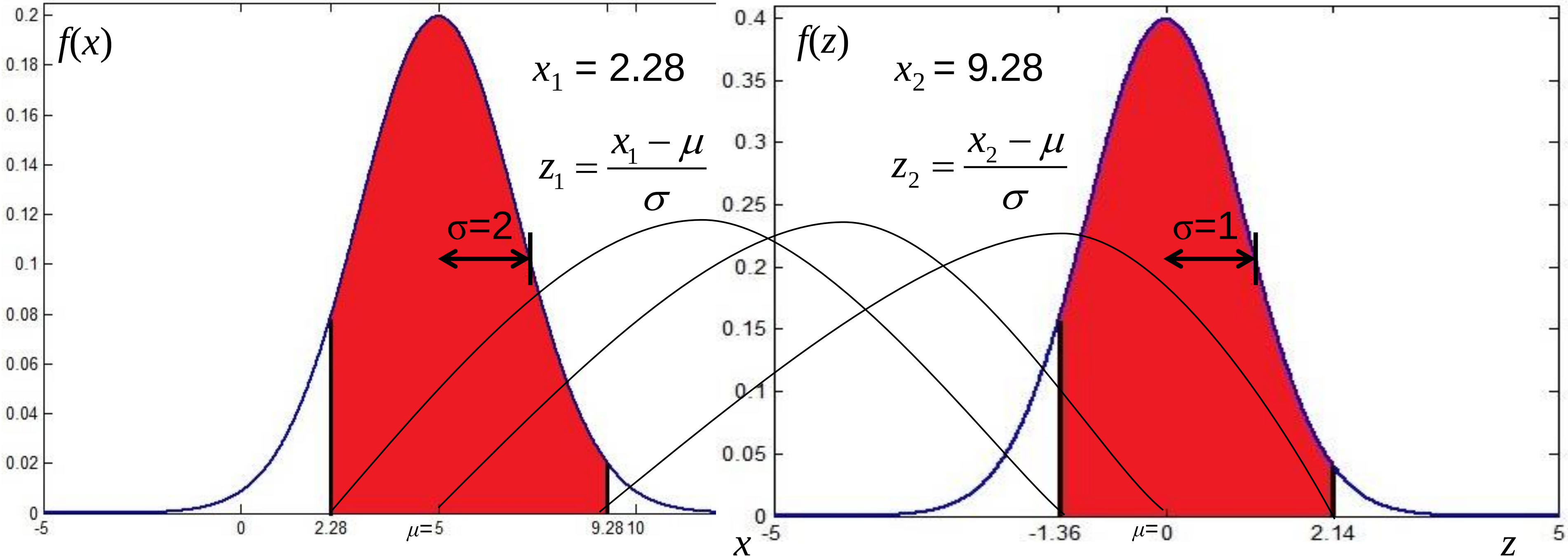
What is the area under the normal distribution between these two values?

6: Normal Probability Distributions

6.2 The Standard Normal Probability Distributions

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

Example: Here is a normal distribution with $\mu = 5$ and $\sigma^2 = 4$.



Area between x_1 and x_2 is the same as area between z_1 and z_2.
We find $z_1 = (x_1 - \mu)/\sigma = -1.36$ and $z_2 = (x_2 - \mu)/\sigma = 2.14$?

↑ ↑ ↑ ↑ ↑ ↑

2.28 5 2 9.28 5 2

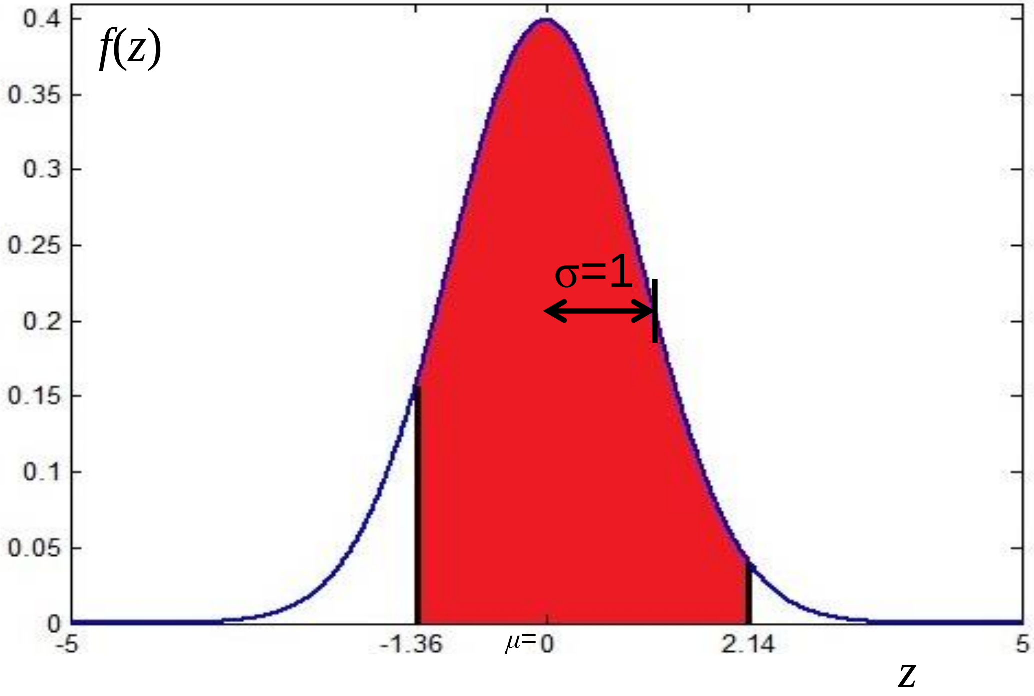
6: Normal Probability Distributions

6.2 The Standard Normal Probability Distributions

Now we can simply look up the z areas in a table.

Appendix B Table 3 Page 716.

Standard normal curve $\mu = 0$ and $\sigma^2 = 1$.



6: Normal Probability Distributions

Appendix B
Table 3
Page 716

$z_1 = -1.36$
$z_2 = 2.14$

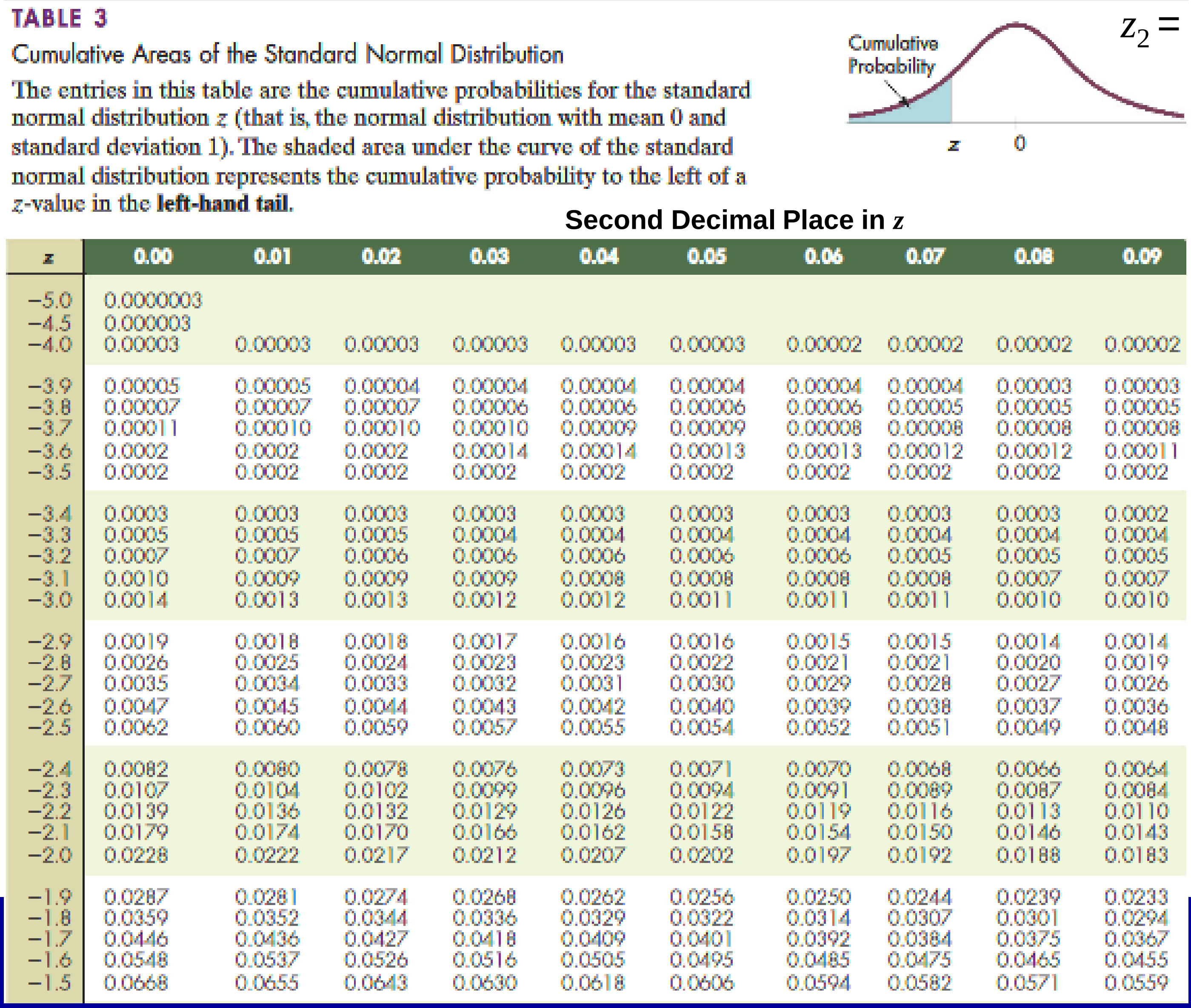


TABLE 3

Cumulative Areas of the Standard Normal Distribution

The entries in this table are the cumulative probabilities for the standard normal distribution z (that is, the normal distribution with mean 0 and standard deviation 1). The shaded area under the curve of the standard normal distribution represents the cumulative probability to the left of a z-value in the **left-hand tail.**

Second Decimal Place in z

First Decimal Place in z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−5.0	0.0000003									
−4.5	0.000003									
−4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
−3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
−3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
−3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
−3.6	0.0002	0.0002	0.0002	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
−3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0014	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0042	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0076	0.0073	0.0071	0.0070	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0126	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

6: Normal Probability Distributions

Appendix B, Table 3, Page 716

z_1 = -1.36
z_2 = 2.14

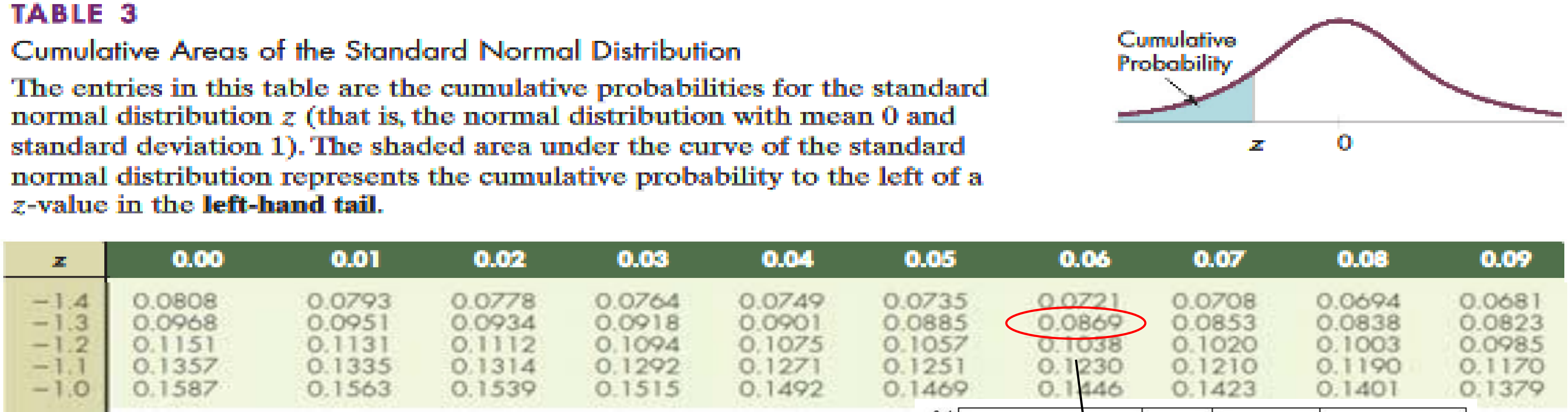


TABLE 3

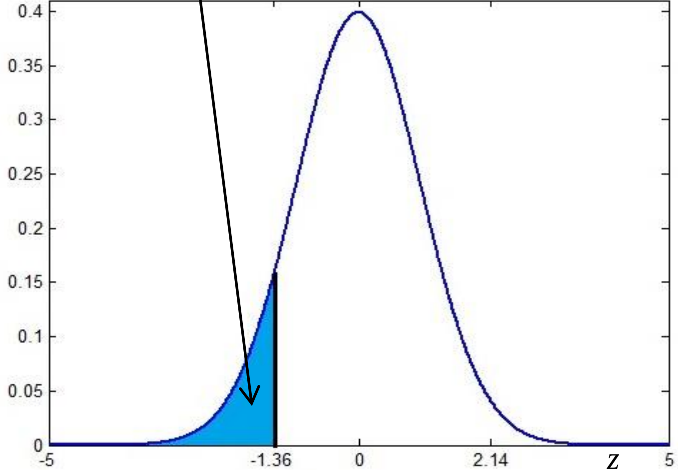
Cumulative Areas of the Standard Normal Distribution

The entries in this table are the cumulative probabilities for the standard normal distribution z (that is, the normal distribution with mean 0 and standard deviation 1). The shaded area under the curve of the standard normal distribution represents the cumulative probability to the left of a z-value in the **left-hand tail.**

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1094	0.1075	0.1057	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1563	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379

$P(z<-1.36)$=Area less than -1.36.

We get this from Table 3.
Row labeled -1.3 over to column
Labeled .06.



6: Normal Probability Distributions

Appendix B, Table 3, Page 717

z_1 = -1.36
z_2 = 2.14

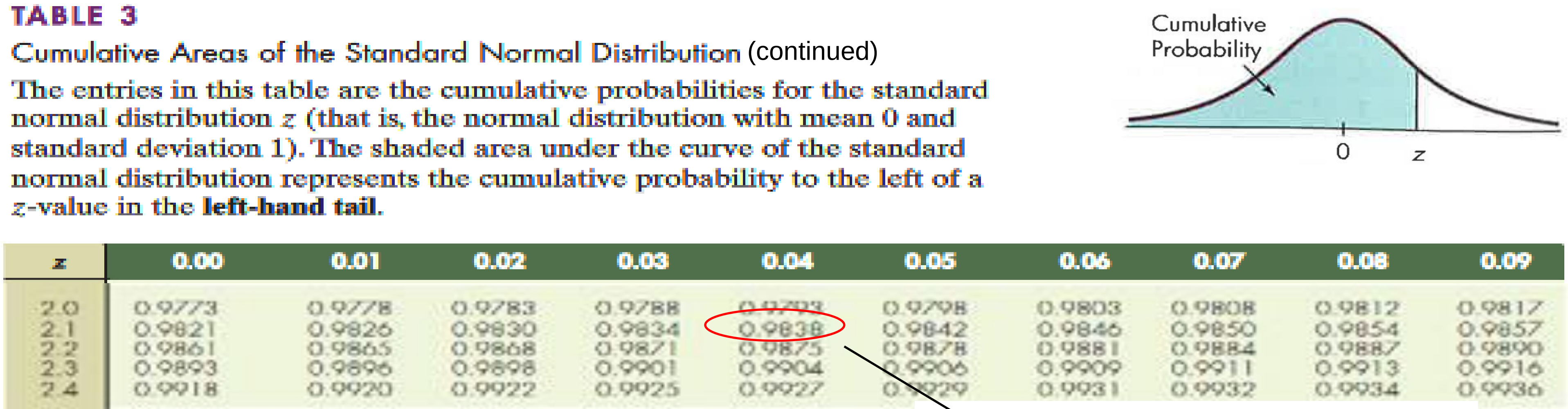


TABLE 3

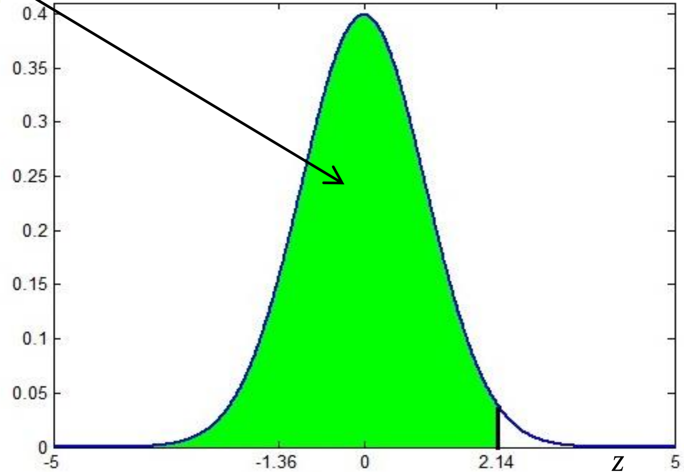
Cumulative Areas of the Standard Normal Distribution (continued)

The entries in this table are the cumulative probabilities for the standard normal distribution z (that is, the normal distribution with mean 0 and standard deviation 1). The shaded area under the curve of the standard normal distribution represents the cumulative probability to the left of a z-value in the **left-hand tail.**

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936

$P(z<2.14)$=Area less than 2.14.

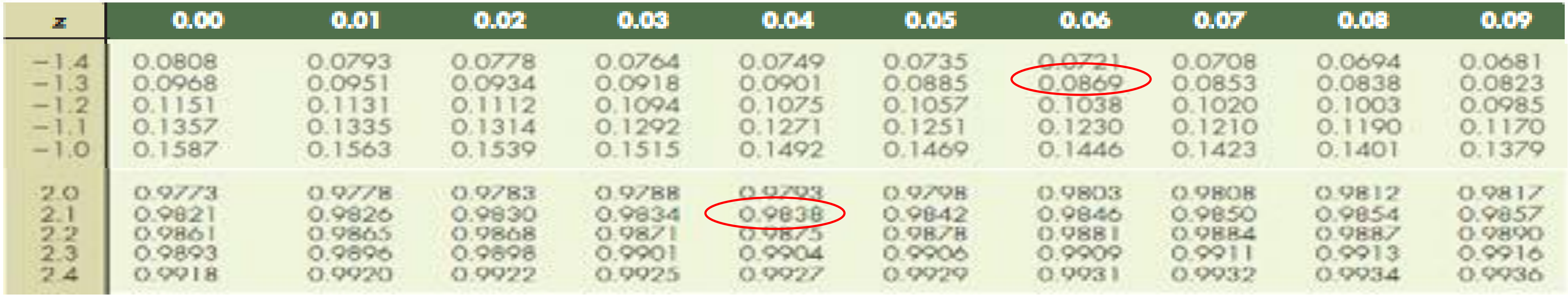
We get this from Table 3.
Row labeled 2.1 over to column
Labeled .04.



6: Normal Probability Distributions

Appendix B, Table 3, Page 716-717

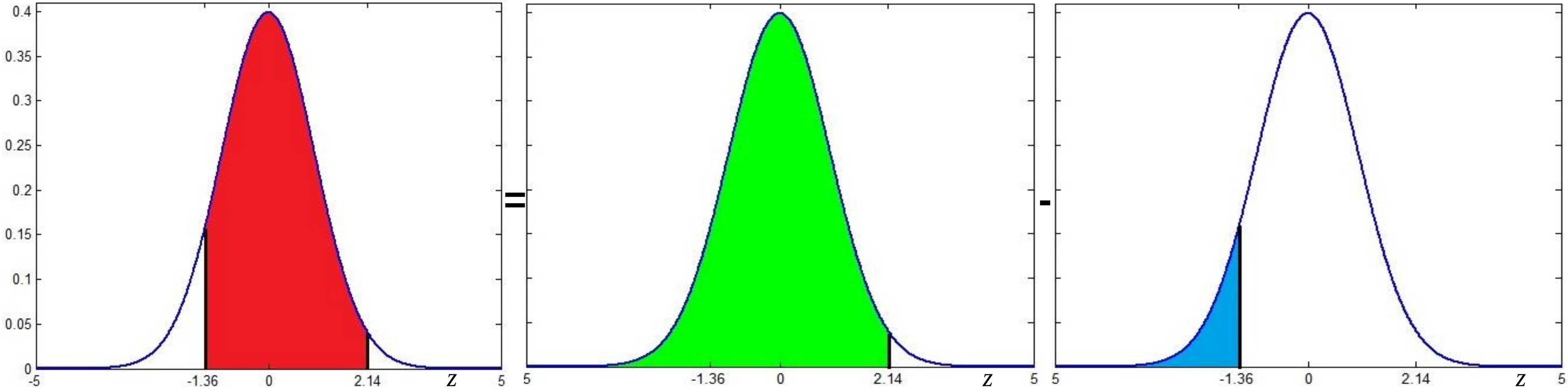
z_1 = -1.36
z_2 = 2.14



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1094	0.1075	0.1057	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1563	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936

$P(-1.36<z<2.14) = P(z<2.14) - P(z<-1.36)$

0.8969 = 0.9838 - 0.0869



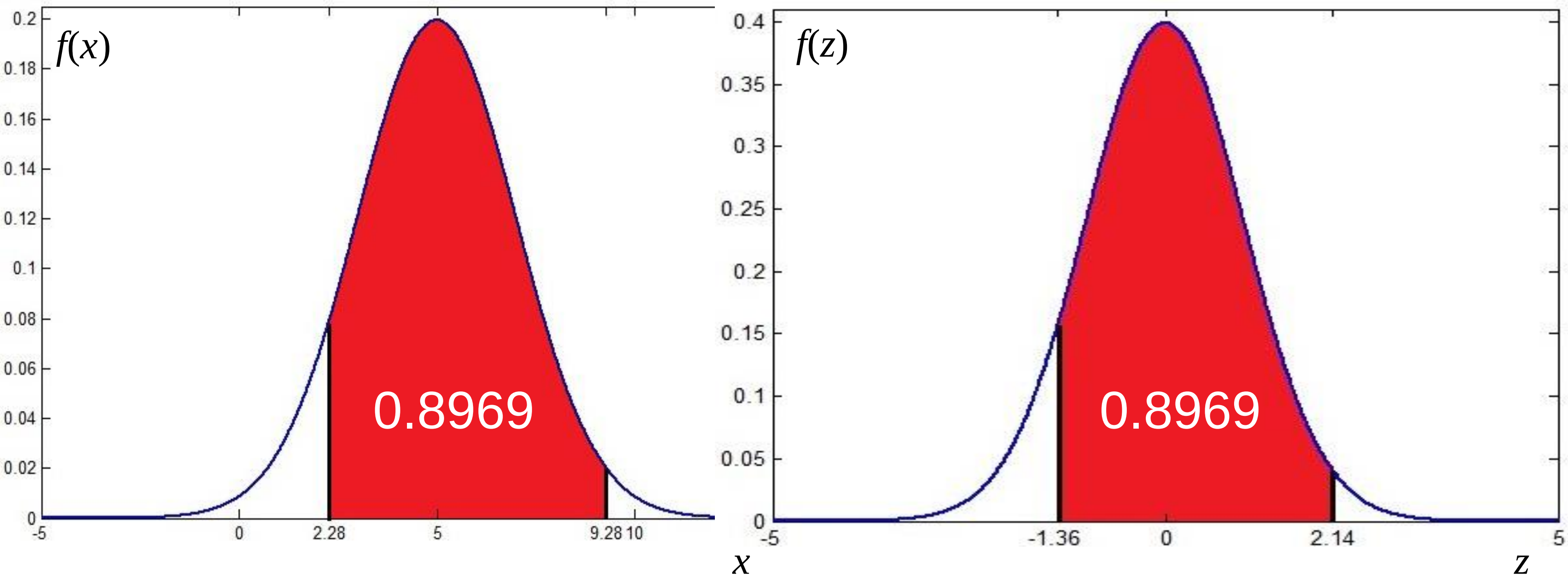
Red Area=0.8969

6: Normal Probability Distributions

6.2 The Standard Normal Probability Distributions

$z_1 = -1.36$
$z_2 = 2.14$

Example: Here is a normal distribution with $\mu = 5$ and $\sigma^2 = 4$.



Area between x_1 and x_2 is same as the area between z_1 and z_2.

6: Normal Probability Distributions

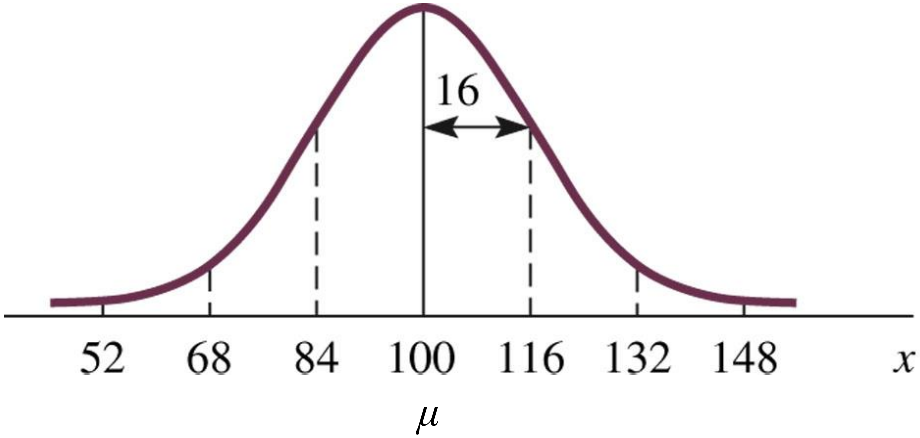
6.3 Applications of Normal Distributions

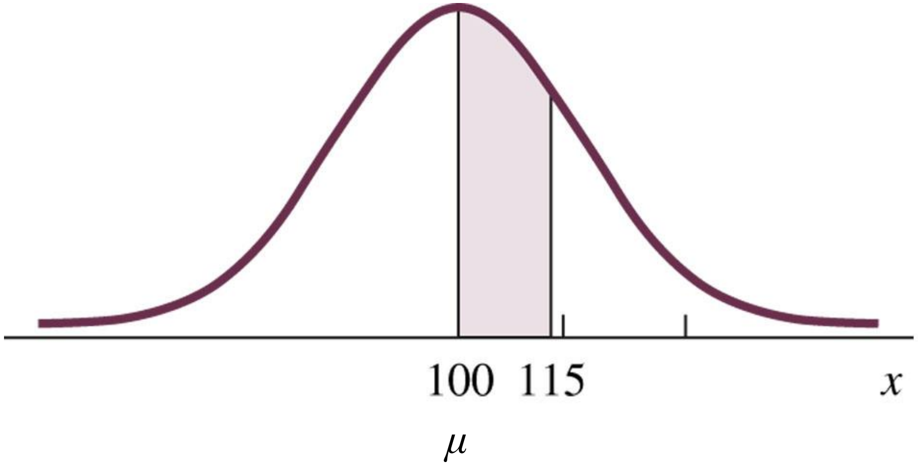
Example:

Assume that IQ scores are normally distributed with a mean μ of 100 and a standard deviation σ of 16.

If a person is picked at random, what is the probability that his or her IQ is between 100 and 115?

i.e. $P(100 < x < 115)$?





Figures from Johnson & Kuby, 2012.

6: Normal Probability Distributions

6.3 Applications of Normal Distributions

IQ scores normally distributed μ=100 and σ=16.

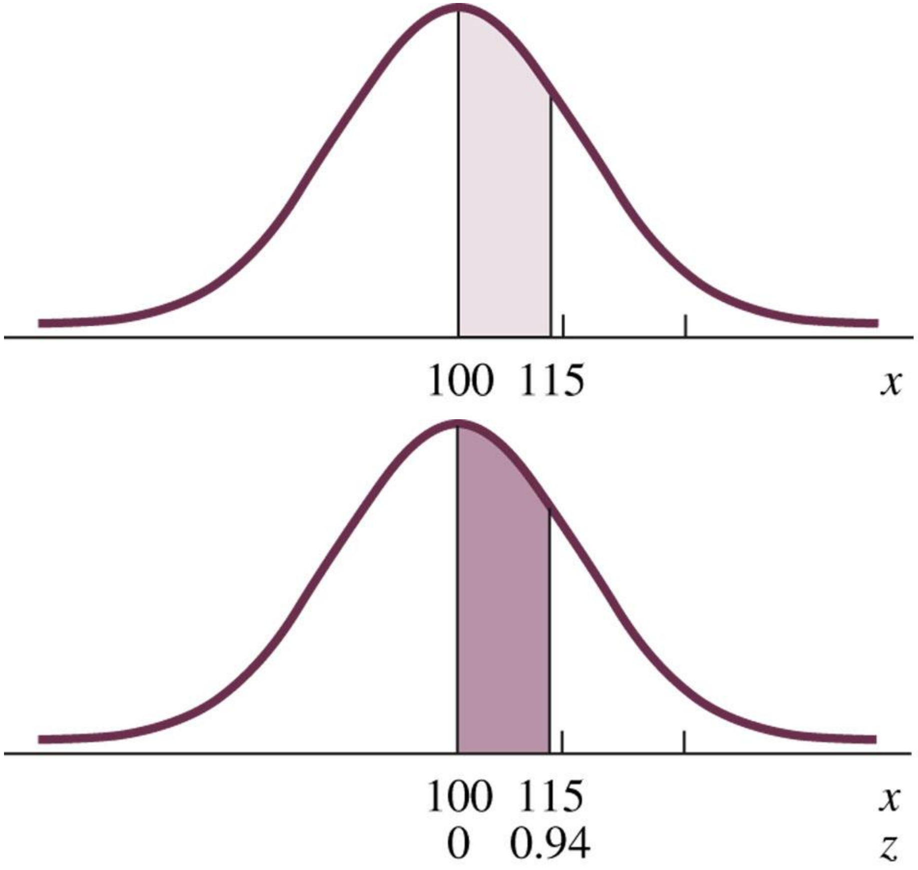
$$P(100 < x < 115)$$

$$z = \frac{x-\mu}{\sigma} \qquad x_1 = 100$$

$$x_2 = 115$$

$$z_1 = \frac{x_1-\mu}{\sigma} = \frac{100-100}{16} = 0$$

$$z_2 = \frac{x_2-\mu}{\sigma} = \frac{115-100}{16} = 0.94$$

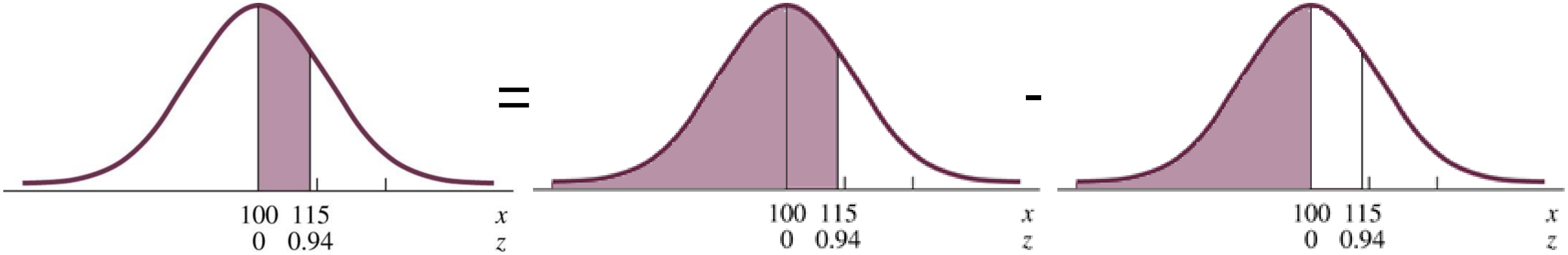


Figures from Johnson & Kuby, 2012.

6: Normal Probability Distributions

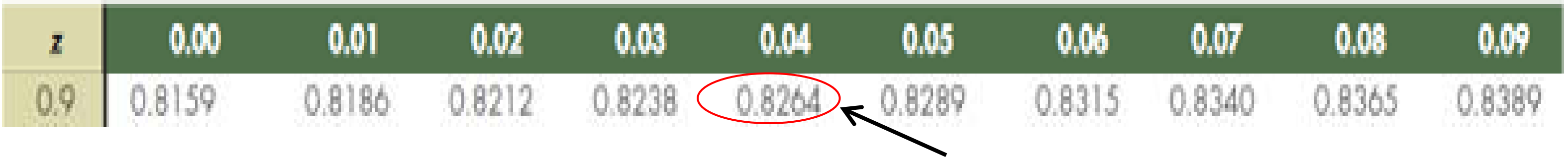
6.3 Applications of Normal Distributions

Now we can use the table.



$$
\begin{aligned}
P(0 < z < 0.94) &= P(z < 0.94) - P(z < 0) \\
&= 0.8264 - .5 \\
&= 0.3264
\end{aligned}
$$

Figures from Johnson & Kuby, 2012.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389

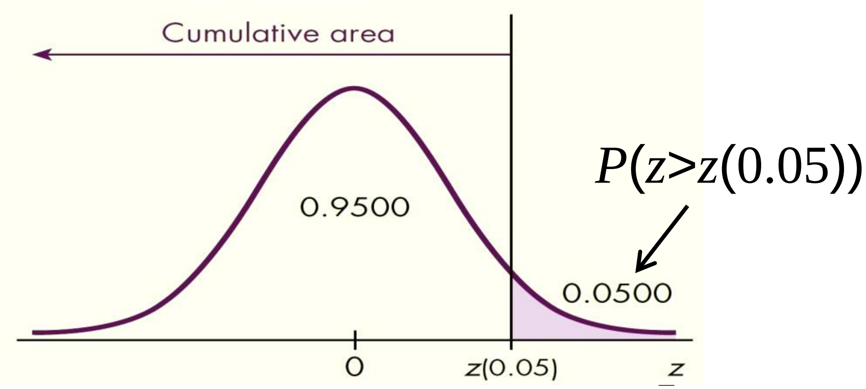
6: Normal Probability Distributions

6.4 Notation

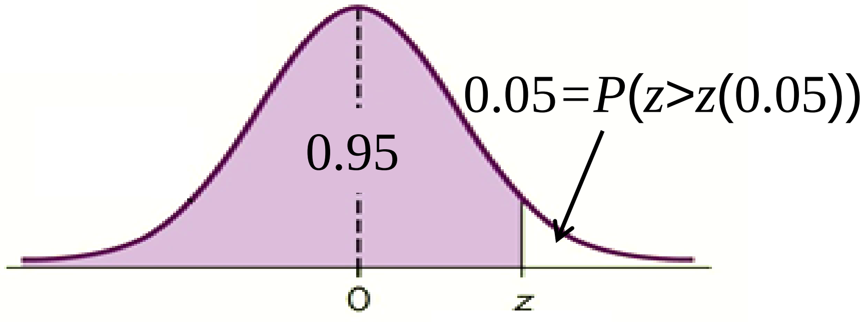
Example:

Let α=0.05. Let's find $z(0.05)$.

$P(z>z(0.05))$=0.05.



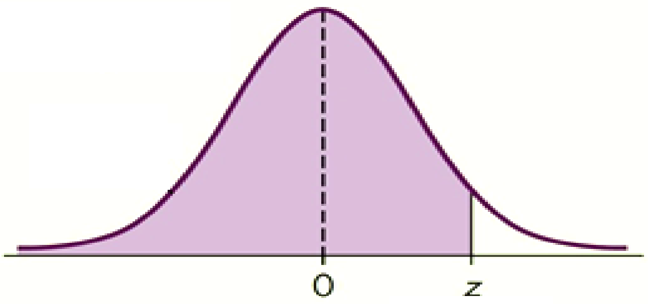
Same as finding $P(z<z(0.05))$=1-0.05.



Figures from Johnson & Kuby, 2012.

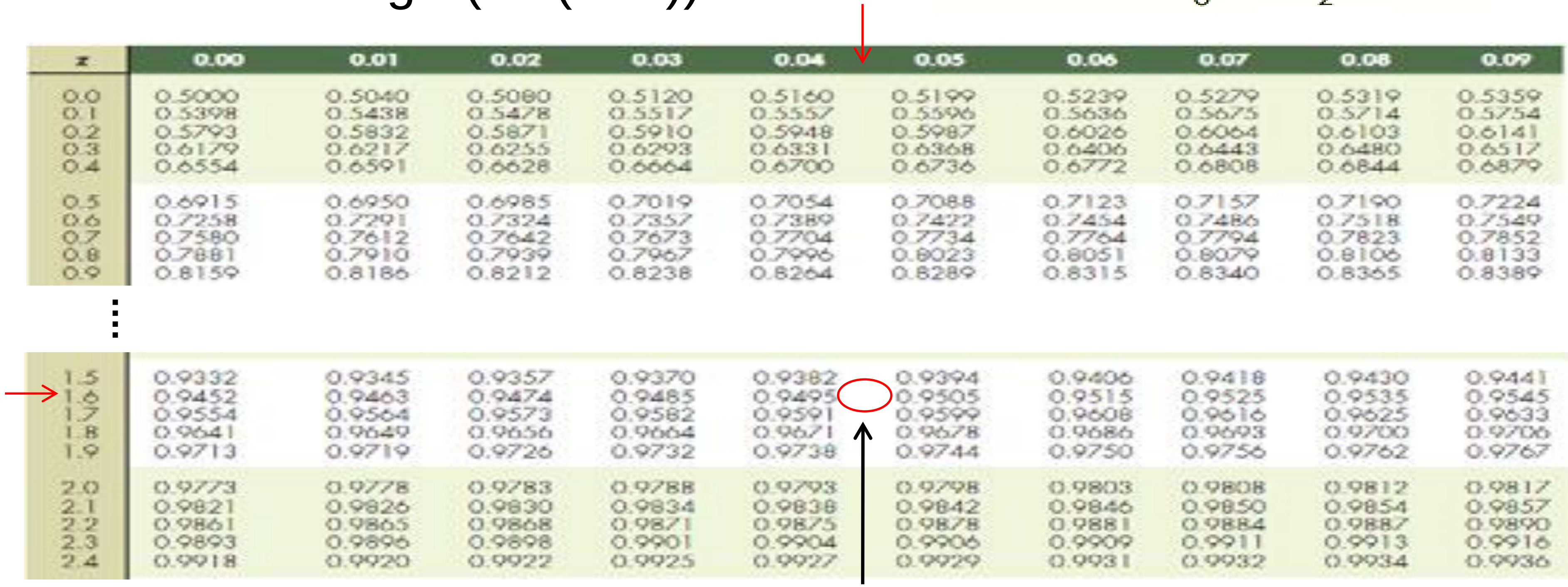
6: Normal Probability Distributions

6.4 Notation



Example:

Same as finding $P(z<z(0.05))=0.95$.



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	0.5754
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	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	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	0.7996	0.8023	0.8051	0.8079	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
⋮										
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	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
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936

1.645

Figures from Johnson & Kuby, 2012.

7: Sample Variability

7.2 The Sampling Distribution of Sample Means

When we take a random sample $x_1,\ldots, x_n$ from a population,

one of the things that we do is compute the sample mean $\bar{x}$.

The value of $\bar{x}$ is not μ. Each time we take a random sample

of size n (with replacement), we get a different set of values

$x_1,\ldots, x_n$ and a different value for $\bar{x}$.

There is a distribution of possible $\bar{x}$'s.

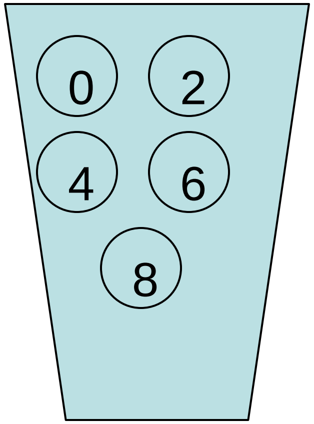
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

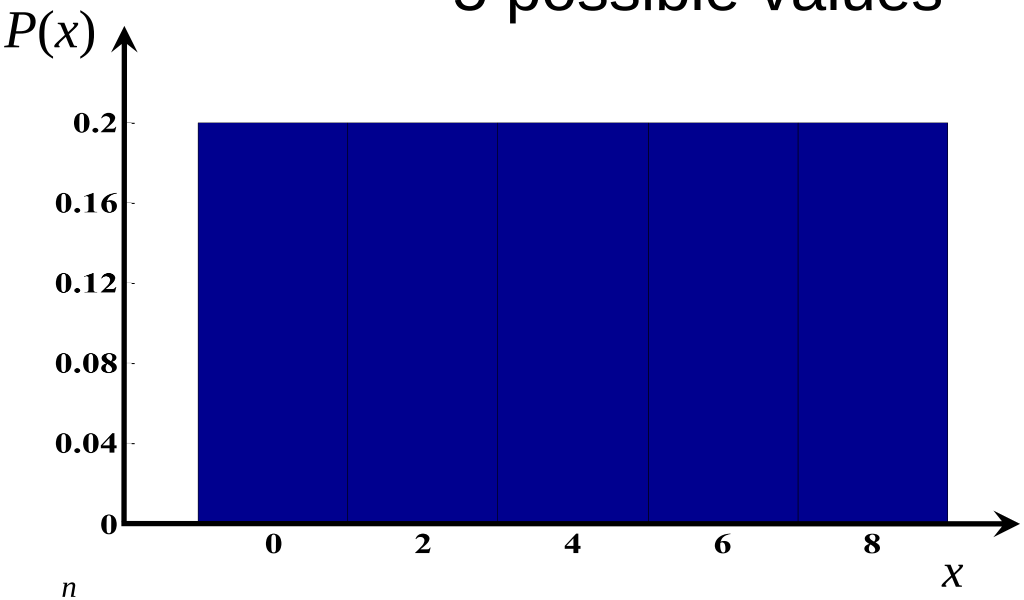
N=5 balls in bucket, select n=1 *with* replacement.

Population data values: 0, 2, 4, 6, 8.

5 possible values



x	$P(x)$
0	1/5
2	1/5
4	1/5
6	1/5
8	1/5



$$\mu = \sum_{i=1}^{n} [x_i P(x_i)] = 4$$

$$\sigma^2 = \sum_{i=1}^{n} [(x_i - \mu)^2 P(x_i)] = 8 \qquad \sigma = \sqrt{8} = 2\sqrt{2}$$

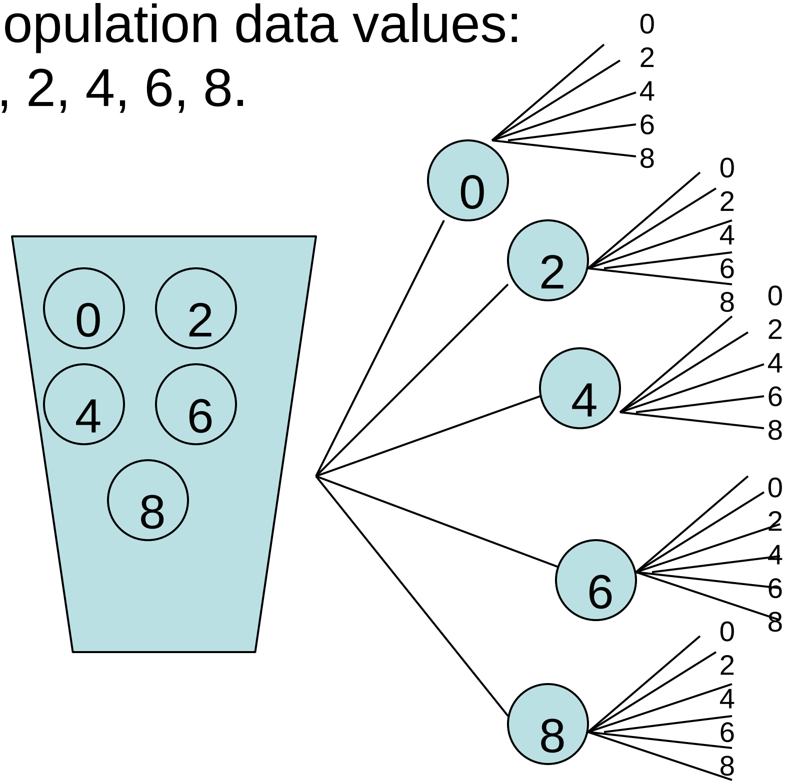
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

Example:

N=5 balls in bucket, select n=2 *with* replacement.

Population data values:
0, 2, 4, 6, 8.



$(0,0)$	$(2,0)$	$(4,0)$	$(6,0)$	$(8,0)$
$(0,2)$	$(2,2)$	$(4,2)$	$(6,2)$	$(8,2)$
$(0,4)$	$(2,4)$	$(4,4)$	$(6,4)$	$(8,4)$
$(0,6)$	$(2,6)$	$(4,6)$	$(6,6)$	$(8,6)$
$(0,8)$	$(2,8)$	$(4,8)$	$(6,8)$	$(8,8)$

25 possible samples

7: Sample Variability

7.2 The Sampling Distribution of Sample Means

25 possible samples.

$(0,0)$	$(2,0)$	$(4,0)$	$(6,0)$	$(8,0)$
$(0,2)$	$(2,2)$	$(4,2)$	$(6,2)$	$(8,2)$
$(0,4)$	$(2,4)$	$(4,4)$	$(6,4)$	$(8,4)$
$(0,6)$	$(2,6)$	$(4,6)$	$(6,6)$	$(8,6)$
$(0,8)$	$(2,8)$	$(4,8)$	$(6,8)$	$(8,8)$

Example: N=5, values: 0, 2, 4, 6, 8, n=2 (with replacement).

Prob. of each samples mean = 1/25 = 0.04

0	1	2	3	4
1	2	3	4	5
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8

$\bar{x} = 0,$ one time

$\bar{x} = 1,$ two times

$\bar{x} = 2,$ three times

$\bar{x} = 3,$ four times

$\bar{x} = 4,$ five times

$\bar{x} = 5,$ four times

$\bar{x} = 6,$ three times

$\bar{x} = 7,$ two times

$\bar{x} = 8,$ one time

$P(\bar{x} = 0) = 1/25$

$P(\bar{x} = 1) = 2/25$

$P(\bar{x} = 2) = 3/25$

$P(\bar{x} = 3) = 4/25$

$P(\bar{x} = 4) = 5/25$

$P(\bar{x} = 5) = 4/25$

$P(\bar{x} = 6) = 3/25$

$P(\bar{x} = 7) = 2/25$

$P(\bar{x} = 8) = 1/25$

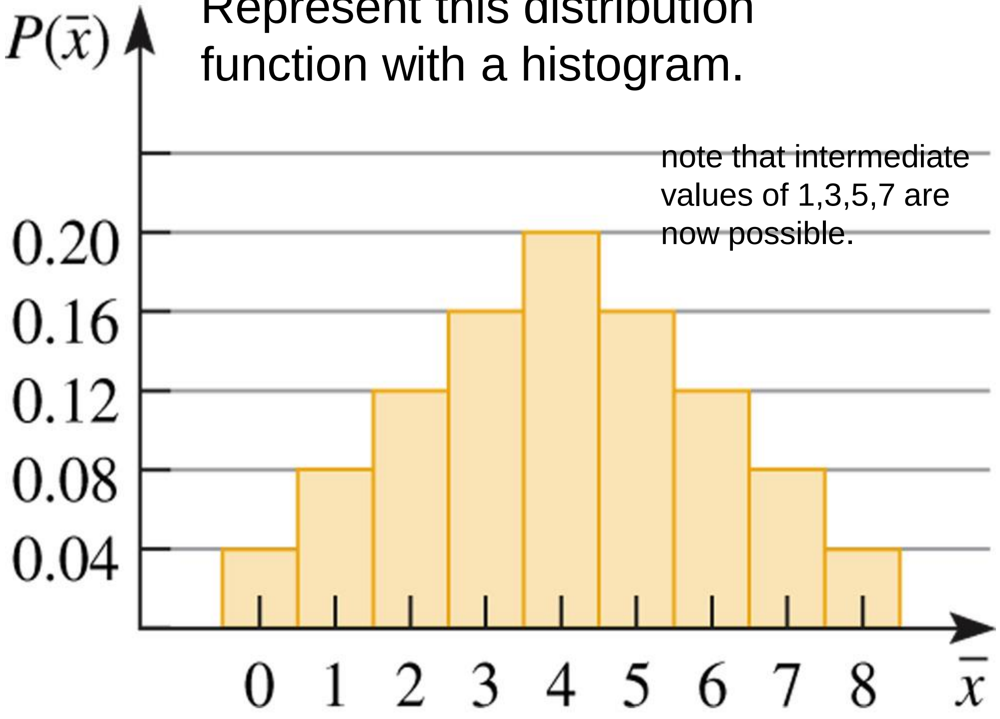
7: Sample Variability

7.2 The Sampling Distribution of Sample Means

Example: N=5, values: 0, 2, 4, 6, 8, n=2 (with replacement).

$$P(\bar{x}=0)=1/25$$
$$P(\bar{x}=1)=2/25$$
$$P(\bar{x}=2)=3/25$$
$$P(\bar{x}=3)=4/25$$
$$P(\bar{x}=4)=5/25$$
$$P(\bar{x}=5)=4/25$$
$$P(\bar{x}=6)=3/25$$
$$P(\bar{x}=7)=2/25$$
$$P(\bar{x}=8)=1/25$$

Represent this distribution function with a histogram.



note that intermediate values of 1,3,5,7 are now possible.

Figure from Johnson & Kuby, 2012.

7: Sample Variability

7.2 The Sampling Distribution of Sample Means

Example: N=5, values: 0, 2, 4, 6, 8, n=1 or 2 (with replacement).

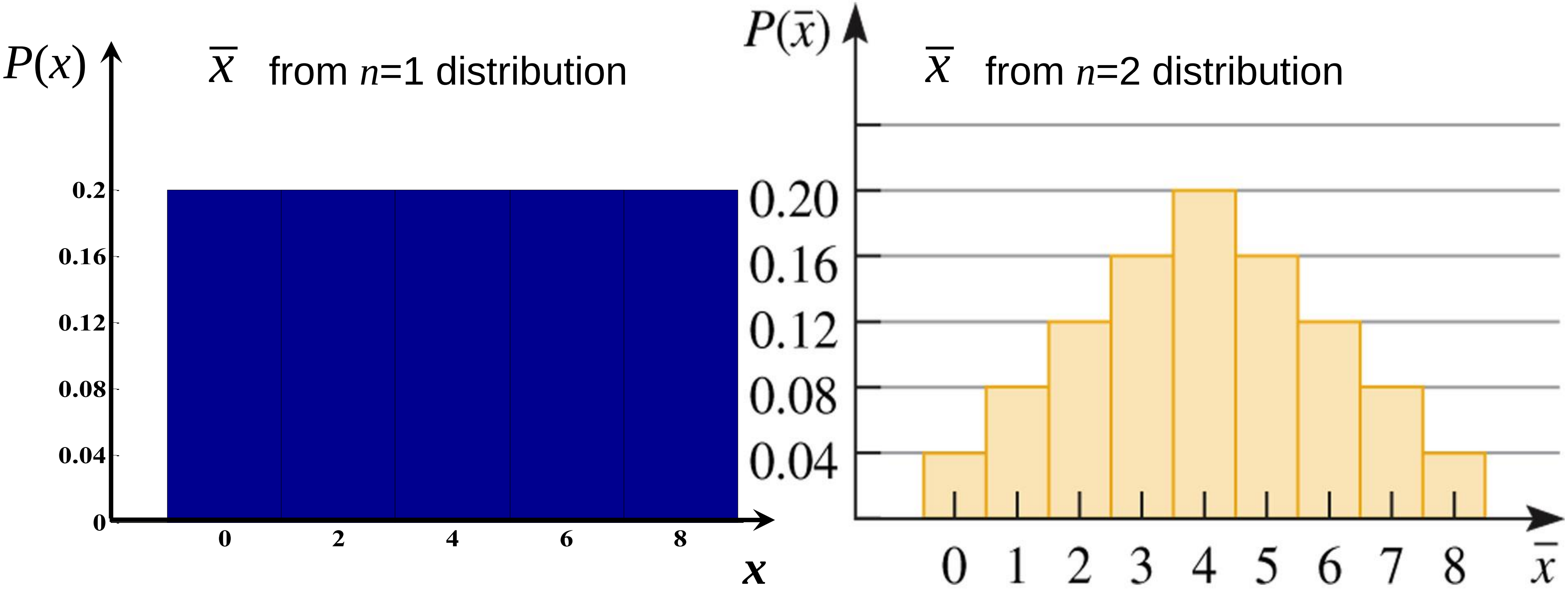


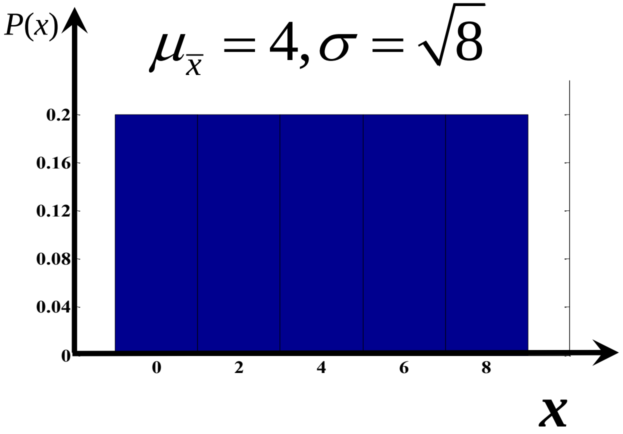
Figure from Johnson & Kuby, 2012.

1. A mean $\mu_{\bar{x}}$ equal to μ
2. A standard deviation $\sigma_{\bar{x}}$ equal to $\frac{\sigma}{\sqrt{n}}$

7: Sample Variability

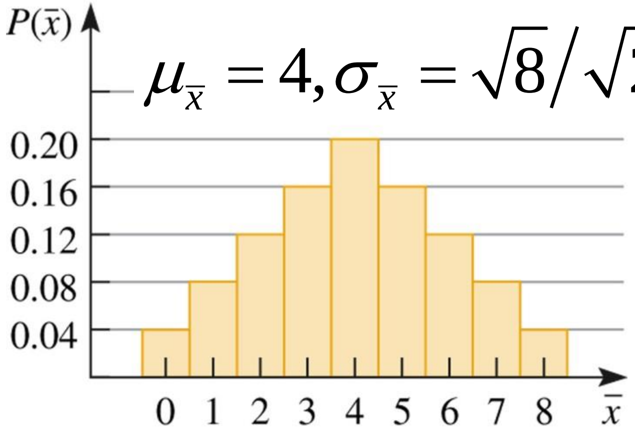
7.2 The Sampling Distribution of Sample Means

$\bar{x}$ from n=1 distribution



$$\mu_{\bar{x}} = 4, \sigma = \sqrt{8}$$

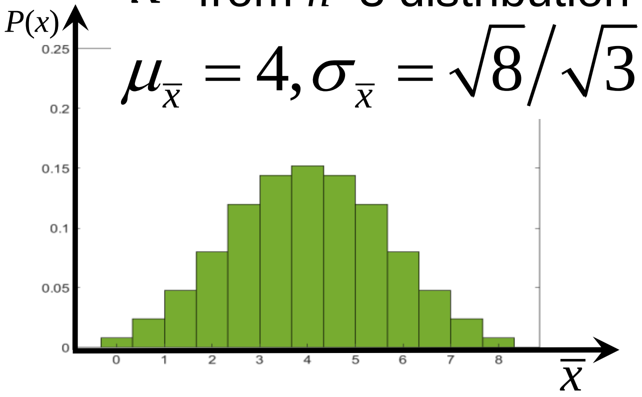
$\bar{x}$ from n=2 distribution



$$\mu_{\bar{x}} = 4, \sigma_{\bar{x}} = \sqrt{8}/\sqrt{2}$$

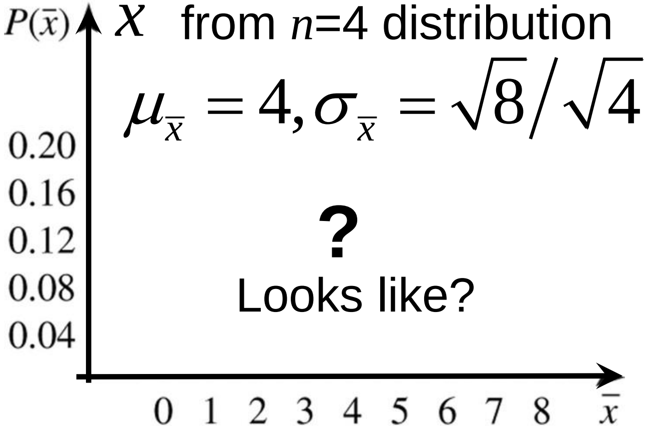
Figure from Johnson & Kuby, 2012.

$\bar{x}$ from n=3 distribution



$$\mu_{\bar{x}} = 4, \sigma_{\bar{x}} = \sqrt{8}/\sqrt{3}$$

$\bar{x}$ from n=4 distribution



$$\mu_{\bar{x}} = 4, \sigma_{\bar{x}} = \sqrt{8}/\sqrt{4}$$

?

Looks like?

$$\xrightarrow{n \text{ large?}}$$

$$\mu_{\bar{x}} = 4$$

$$\sigma_{\bar{x}} = \sqrt{8}/\sqrt{n}$$

Looks like?

See end of slides for details

7: Sample Variability

7.2 The Sampling Distribution of Sample Means

Sample distribution of sample means (SDSM):
If random samples of size n, are taken from ANY population with mean μ and standard deviation σ, then the SDSM has:

1. A mean $\mu_{\bar{x}}$ equal to μ
2. A standard deviation $\sigma_{\bar{x}}$ equal to $\dfrac{\sigma}{\sqrt{n}}$

Central Limit Theorem (CLT): The sampling distribution of sample means will more closely resemble the normal distribution as the sample size n increases.

7: Sample Variability

7.2 The Sampling Distribution of Sample Means

The CLT: Assume that we have a population (arbitrary distribution) with mean μ and standard deviation σ.

If we take random samples of size n (with replacement), then for “large” n, the distribution of the sample means the $\bar{x}$‘s is approximately normally distributed with

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where in general $n \geq 30$ is sufficiently “large,” but can be as small as15 or as big as 50 depending upon the shape of distribution!

7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

Example:

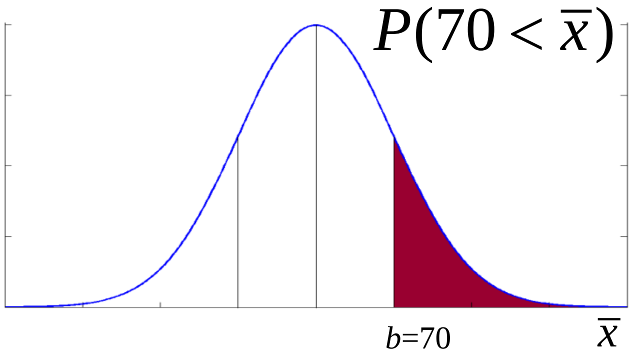
What is probability that sample mean $\bar{x}$ from a random sample of n=15 heights is greater than 70” when $\mu = 67.5$ and $\sigma = 3.7$?

7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

Example:

What is probability that sample mean $\bar{x}$ from a random sample of n=15 heights is greater than 70” when $\mu = 67.5$ and $\sigma = 3.7$?

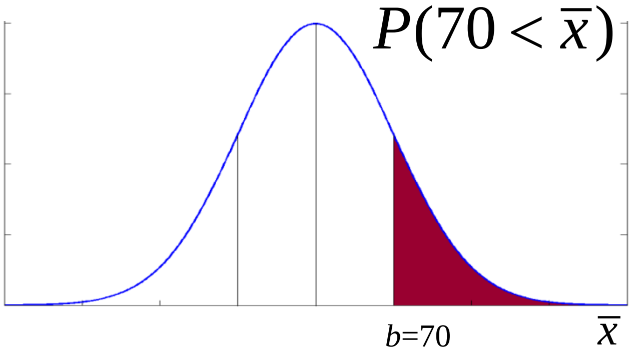


7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

Example:

What is probability that sample mean $\bar{x}$ from a random sample of n=15 heights is greater than 70” when $\mu = 67.5$ and $\sigma = 3.7$?



we first convert to z scores

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

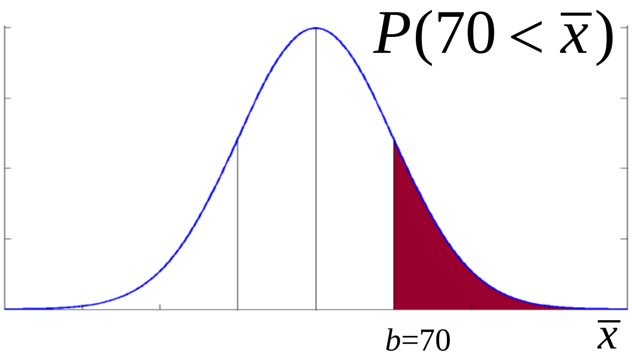
$$d = \frac{b - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

7: Sample Variability

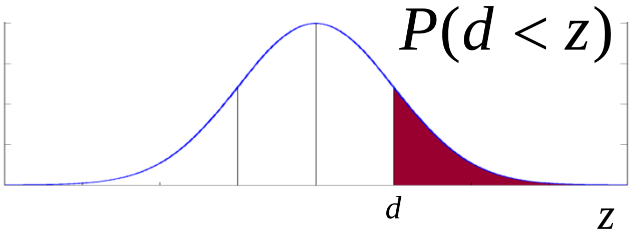
7.3 Application of the Sampling Distribution of Sample Means

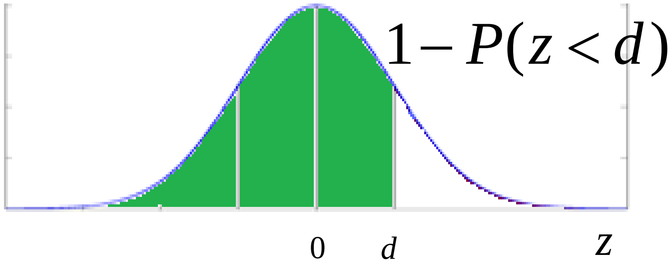
Example:

What is probability that sample mean $\bar{x}$ from a random sample of n=15 heights is greater than 70” when $\mu = 67.5$ and $\sigma = 3.7$?



we first convert to z scores





$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

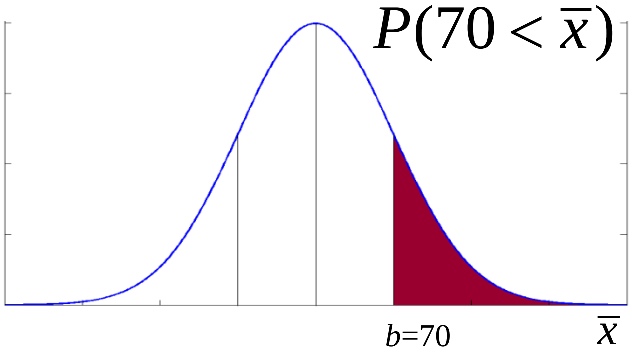
$$d = \frac{b - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

7: Sample Variability

7.3 Application of the Sampling Distribution of Sample Means

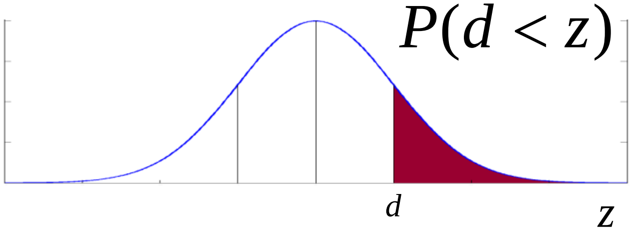
Example:

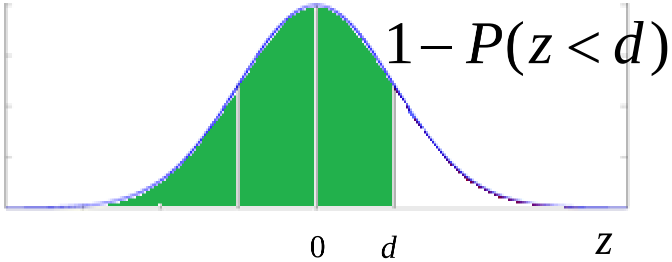
What is probability that sample mean $\bar{x}$ from a random sample of n=15 heights is greater than 70” when $\mu = 67.0$ and $\sigma = 4.3$?



we first convert to z scores

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$





where $d = \dfrac{b - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \dfrac{70 - 67.0}{4.3 / \sqrt{15}} = 2.70$, then use the table in book.

$$1 - P(z < 2.70) = 1 - .9965 = 0.0035$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974

8: Introduction to Statistical Inference

8.1 The Nature of Estimation

Point estimate for a parameter: A single number …, to estimate a parameter … usually the .. **sample statistic**.

i.e. $\bar{x}$ is a point estimate for μ

Interval estimate: An interval bounded by two values and used to estimate the value of a population parameter. ….

i.e. $\bar{x} \pm \text{(some amount)}$ is an interval estimate for μ.

$\text{point estimate} \pm \text{some amount}$

8: Introduction to Statistical Inference

8.1 The Nature of Estimation

Significance Level: Probability parameter outside interval, α.

$$P(\mu \text{ not in } \bar{x} \pm \text{some amount}) = \alpha$$

Level of Confidence 1-α:

$$P(\bar{x} - \text{some amount} < \mu < \bar{x} + \text{some amount}) = 1 - \alpha$$

Confidence Interval:

point estimator ± some amount that depends on confidence level

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

By SDSM

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

What this implies is that $z = \dfrac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

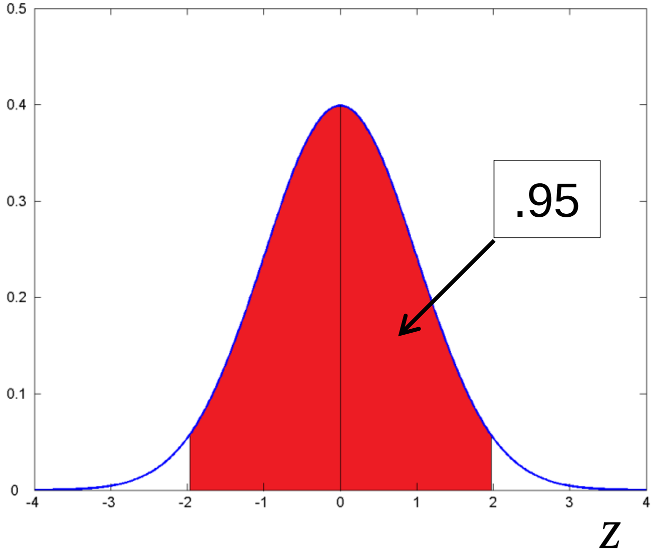
has an approximate standard normal distribution!

$$P(-1.96 < z < 1.96) = 0.95 \qquad \alpha = .05$$

Or more generally,

$$P(-z(\alpha/2) < z < z(\alpha/2)) = 1 - \alpha \quad .$$

$z(\alpha/2)$ called the confidence coefficient.



$-z(.025)=-1.96$ $z(.025))=1.96$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

$$P(-z(\alpha/2) < z < z(\alpha/2)) = 1-\alpha$$

With some algebra, we saw that

$$\begin{aligned}
-z(\alpha/2) &< z \\
-z(\alpha/2) &< \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} \\
-z(\alpha/2)\frac{\sigma}{\sqrt{n}} &< \bar{x}-\mu \\
-z(\alpha/2)\frac{\sigma}{\sqrt{n}}-\bar{x} &< -\mu \\
\bar{x}+z(\alpha/2)\frac{\sigma}{\sqrt{n}} &> \mu
\end{aligned}$$

and

$$\begin{aligned}
z(\alpha/2) &> z \\
z(\alpha/2) &> \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} \\
z(\alpha/2)\frac{\sigma}{\sqrt{n}} &> \bar{x}-\mu \\
z(\alpha/2)\frac{\sigma}{\sqrt{n}}-\bar{x} &> -\mu \\
\bar{x}-z(\alpha/2)\frac{\sigma}{\sqrt{n}} &< \mu
\end{aligned}$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Thus, a (1-α)×100% confidence interval for μ is

$$\bar{x} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

which if α=0.05, a 95% confidence interval for μ is

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} \quad .$$

z(.025))=1.96

Confidence Interval for Mean:

$$\bar{x} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} \quad \textbf{to} \quad \bar{x} + z(\alpha/2)\frac{\sigma}{\sqrt{n}} \qquad \textbf{(8.1)}$$

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

Philosophically, μ is fixed and the interval varies.

If we take a sample of data, $x_1,\ldots,x_n$ and determine a confidence interval from it, we get.

$$\bar{x} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

If we had a different sample of data, $y_1,\ldots,y_n$ we would have determined a different confidence interval.

$$\bar{y} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

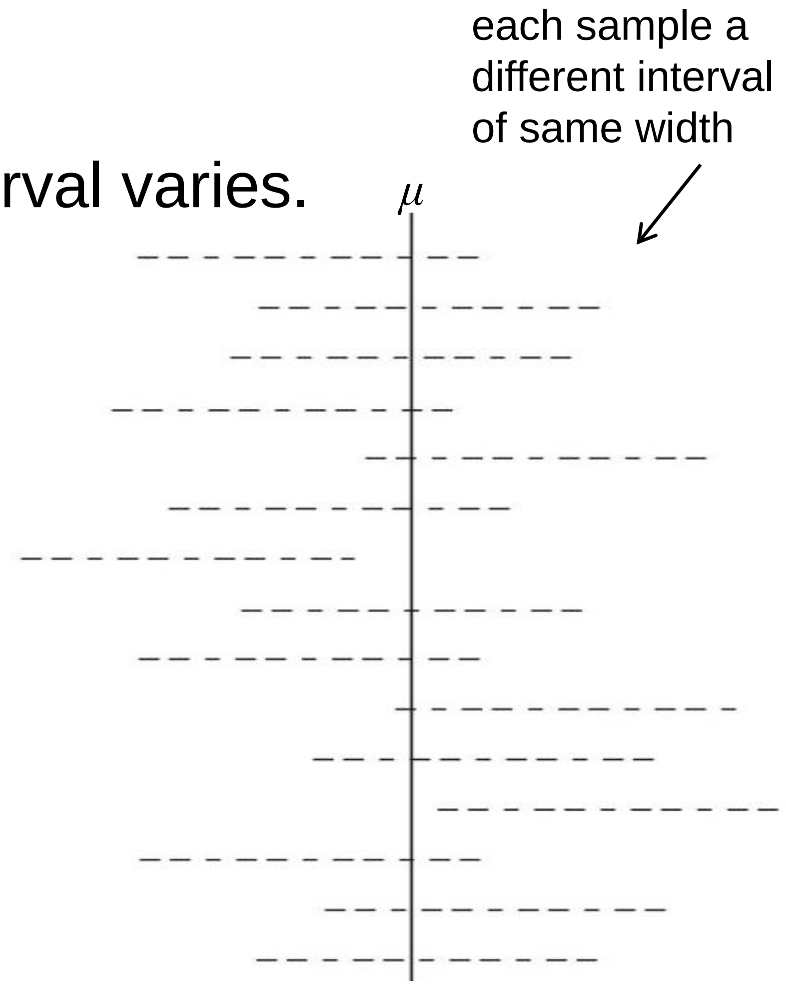


Figure from Johnson & Kuby, 2012.

8: Introduction to Statistical Inference

8.2 Estimation of Mean μ (σ Known)

We never truly know if our CI from our sample of data will contain the true population mean μ.

But we do know that there is a (1-α)×100% chance that a confidence interval from a sample of data will contain μ.

8: Introduction to Statistical Inference

8.4 The Nature of Hypothesis Testing

Example 1: Friend's Party.

H_0: "The party will be a dud"

vs.

H_a: "The Party will be a great time"

Example 2: Math 1700 Students Height

H_0: The mean height of Math 1700 students is 69", $\mu = 69$".

vs.

H_a: The mean height of Math 1700 students is not 69", $\mu \neq 69$".

8: Introduction to Statistical Inference

8.4 The Nature of Hypothesis Testing

Example 1: Friend’s Party

H_0: “The party will be a dud”

vs.

H_a: “The Party will be a great time”

Four outcomes from a hypothesis test.

	Party Great	Party a dud.
We go.	Correct Decision	Type II Error
We do not go.	Type I Error	Correct Decision

If do not go to party and it’s great, we made an error in judgment.

If go to party and it’s a dud, we made in error in judgment.

8: Introduction to Statistical Inference

8.4 The Nature of Hypothesis Testing

Example 2: Math 1700 Height

H_0: μ = 69”

vs.

H_a: $\mu \neq$ 69”

Four outcomes from a hypothesis test.

	μ = 69	$\mu \neq$ 69
Fail to reject H_0.	Correct Decision	Type II Error
Reject H_0.	Type I Error	Correct Decision

If we reject H_0 and it is true, we made in error in judgment.

If we do not reject H_0 and it is false, we have made an error in judgment.

8: Introduction to Statistical Inference

8.4 The Nature of Hypothesis Testing

Type I Error: …true null hypothesis H_0 is rejected.

Level of Significance (α): The probability of committing a type I error. (Sometimes α is called the false positive rate.)

Type II Error: … favor … null hypothesis that is actually false.

Type II Probability (β): The probability of committing a type II error.

	H_0 True	H_0 False
Do Not Reject H_0	Type A Correct Decision (1-α)	Type II Error (β)
Reject H_0	Type I Error (α)	Type B Correct Decision (1-β)

8: Introduction to Statistical Inference

8.4 The Nature of Hypothesis Testing

We need to determine a measure that will quantify what we should believe.

Test Statistic: A random variable whose value is calculated from the sample data and is used in making the decision "reject H_0: or "fail to reject H_0."

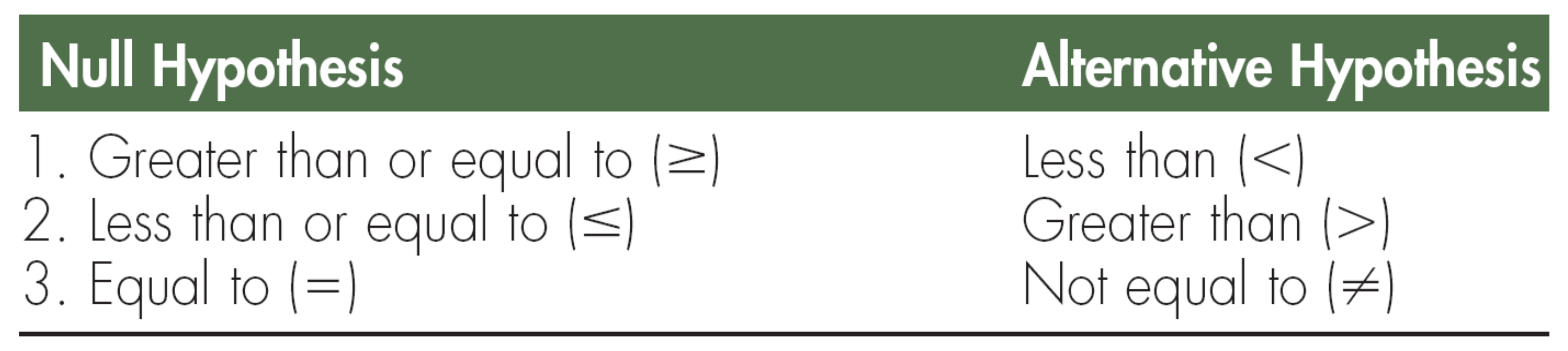
Example: Friend's Party
Fraction of parties that were good.

Example: Math 1700 Heights
Sample mean height.

8: Introduction to Statistical Inference

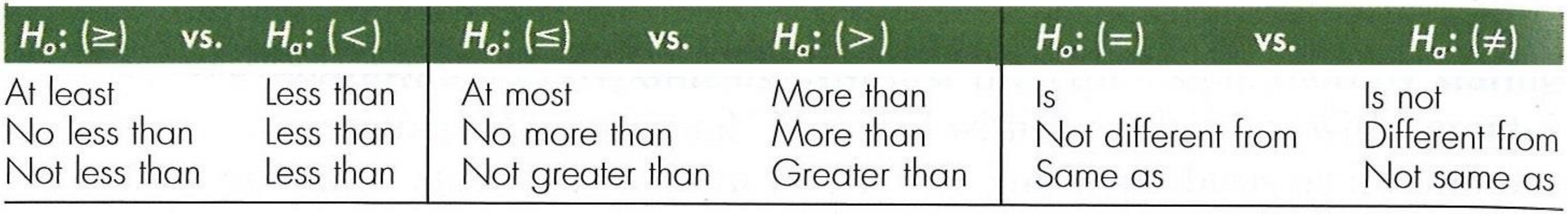
8.4 The Nature of Hypothesis Testing

HYPOTHESIS TESTING PAIRS



Null Hypothesis	Alternative Hypothesis
1. Greater than or equal to (≥)	Less than (<)
2. Less than or equal to (≤)	Greater than (>)
3. Equal to (=)	Not equal to (≠)

TABLE 8.6 Common Phrases and Their Negations



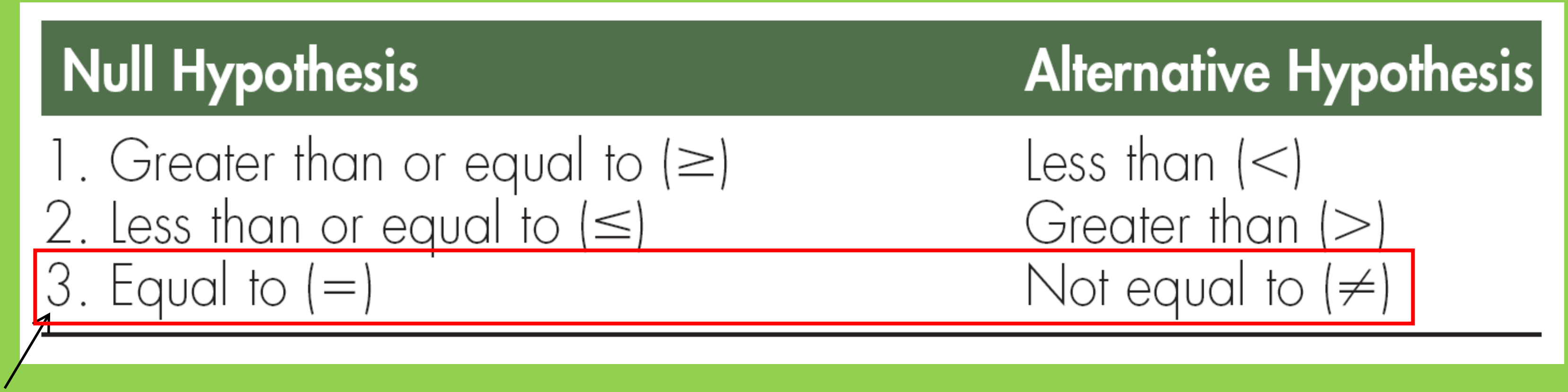
H_o: (≥) vs.	H_a: (<)	H_o: (≤) vs.	H_a: (>)	H_o: (=) vs.	H_a: (≠)
At least	Less than	At most	More than	Is	Is not
No less than	Less than	No more than	More than	Not different from	Different from
Not less than	Less than	Not greater than	Greater than	Same as	Not same as

Figure from Johnson & Kuby, 2012.

8: Introduction to Statistical Inference

8.4 Hypothesis Test of Mean (σ Known): Probability Approach

HYPOTHESIS TESTING PAIRS



Null Hypothesis	Alternative Hypothesis
1. Greater than or equal to (≥)	Less than (<)
2. Less than or equal to (≤)	Greater than (>)
3. Equal to (=)	Not equal to (≠)

$H_0: \mu = 69"$ vs. $H_a: \mu \neq 69"$

Figure from Johnson & Kuby, 2012.

8: Introduction to Statistical Inference

8.4 Hypothesis Test of Mean (σ Known): Probability Approach

Step 1 The Set-Up: Null (H_0) and alternative (H_a) hypotheses

H_0: μ = 69" vs. H_a: $\mu \neq$ 69"

Step 2 The Hypothesis Test Criteria: Test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

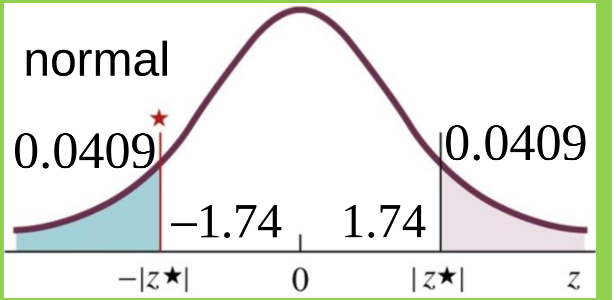
σ known, n is "large" so by CLT $\bar{x}$ is normal

z^* is normal

Step 3 The Sample Evidence: Calculate test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{67.2 - 69}{4 / \sqrt{15}} = -1.74$$

n=15, $\bar{x}$=67.2, $\sigma = 4$



Step 4 The Probability Distribution:

$P(z > | z^* |) = p - \text{value} \longrightarrow$ 0.0819

Step 5 The Results:

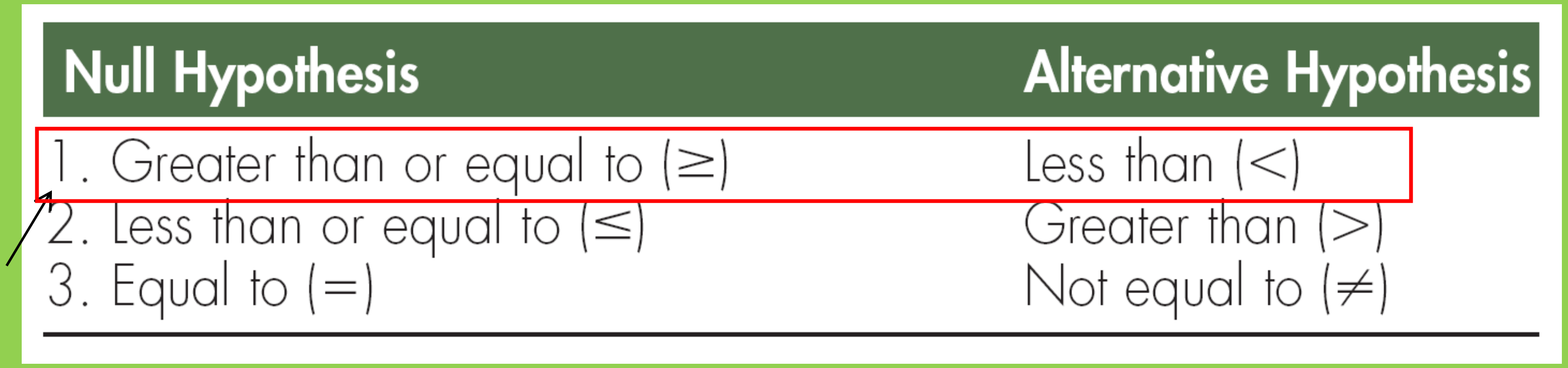
$p - \text{value} \leq \alpha$, **reject H_0,** $p - \text{value} > \alpha$ **fail to reject H_0**

α = 0.05

8: Introduction to Statistical Inference

8.4 Hypothesis Test of Mean (σ Known): Probability Approach

HYPOTHESIS TESTING PAIRS



Null Hypothesis	Alternative Hypothesis
1. Greater than or equal to (≥)	Less than (<)
2. Less than or equal to (≤)	Greater than (>)
3. Equal to (=)	Not equal to (≠)

H_0: $\mu \geq 69”$ vs. H_a: $\mu < 69”$

Figure from Johnson & Kuby, 2012.

8: Introduction to Statistical Inference

8.4 Hypothesis Test of Mean (σ Known): Probability Approach

Step 1 The Set-Up: Null (H_0) and alternative (H_a) hypotheses

$H_0: \mu \geq 69$” vs. $H_a: \mu < 69$”

Step 2 The Hypothesis Test Criteria: Test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

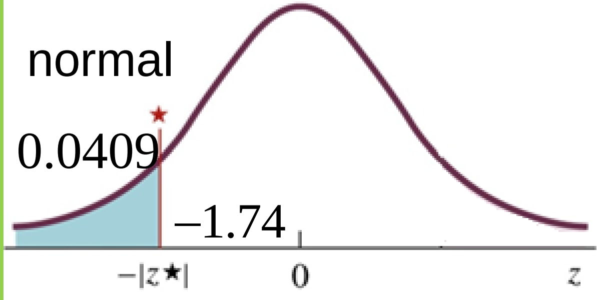
σ known, n is “large” so by CLT $\bar{x}$ is normal

z^* is normal

Step 3 The Sample Evidence: Calculate test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{67.2 - 69}{4 / \sqrt{15}} = -1.74$$

n=15, $\bar{x}$ =67.2, $\sigma = 4$



Step 4 The Probability Distribution:

$P(z < z^*) = p - \text{value} \rightarrow$ 0.0409

Step 5 The Results:

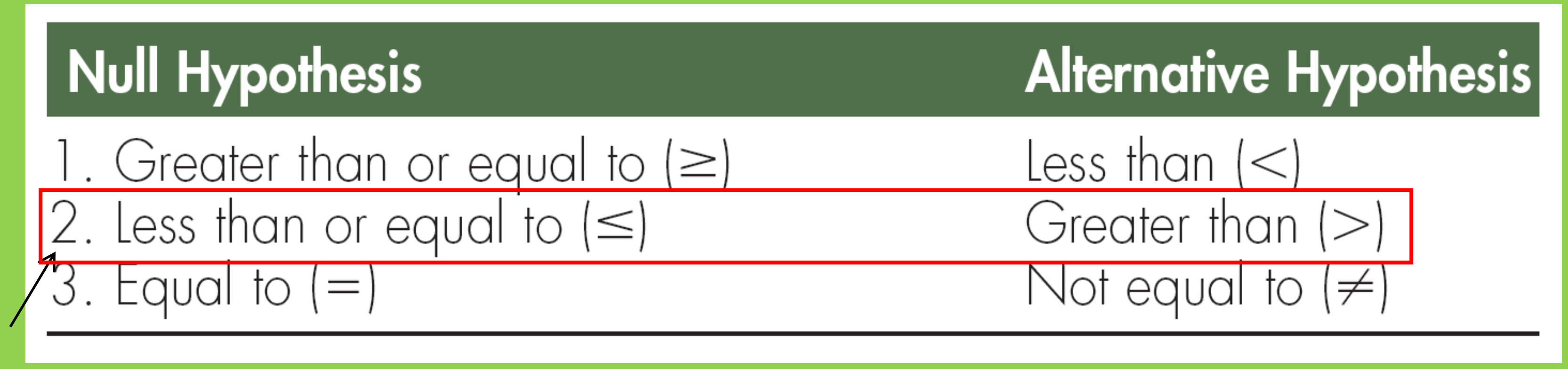
$p - \text{value} \leq \alpha$, **reject H_0,** $p - \text{value} > \alpha$ **fail to reject H_0**

α = 0.05

8: Introduction to Statistical Inference

8.4 Hypothesis Test of Mean (σ Known): Probability Approach

HYPOTHESIS TESTING PAIRS



Null Hypothesis	Alternative Hypothesis
1. Greater than or equal to (≥)	Less than (<)
2. Less than or equal to (≤)	Greater than (>)
3. Equal to (=)	Not equal to (≠)

$H_0: \mu \le 69”$ vs. $H_a: \mu > 69”$

Figure from Johnson & Kuby, 2012.

8: Introduction to Statistical Inference

8.4 Hypothesis Test of Mean (σ Known): Probability Approach

Step 1 The Set-Up: Null (H_0) and alternative (H_a) hypotheses

H_0: $\mu \leq$ 69” vs. H_a: $\mu >$ 69”

Step 2 The Hypothesis Test Criteria: Test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

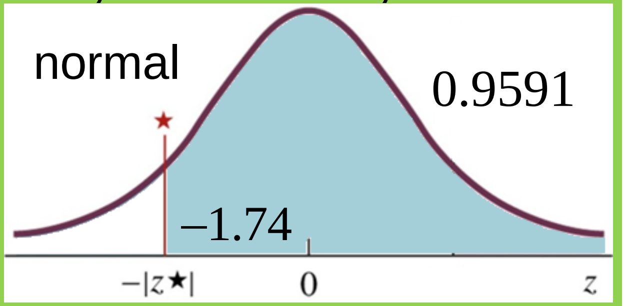
σ known, n is “large” so by CLT $\bar{x}$ is normal

z^* is normal

Step 3 The Sample Evidence: Calculate test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{67.2 - 69}{4 / \sqrt{15}} = -1.74$$

n=15, $\bar{x}$ =67.2, $\sigma = 4$



Step 4 The Probability Distribution:

$P(z > z^*) = p - \text{value}$ ⟶ 0.9691

Step 5 The Results:

$p - \text{value} \leq \alpha$, **reject H_0**, $p - \text{value} > \alpha$ **fail to reject H_0**

α = 0.05

8: Introduction to Statistical Inference

8.4 Hypothesis Test of Mean (σ Known): Probability Approach

Step 1 The Set-Up: Null (H_0) and alternative (H_a) hypotheses

H_0: μ = 69” vs. H_a: $\mu \neq$ 69”

Step 2 The Hypothesis Test Criteria: Test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

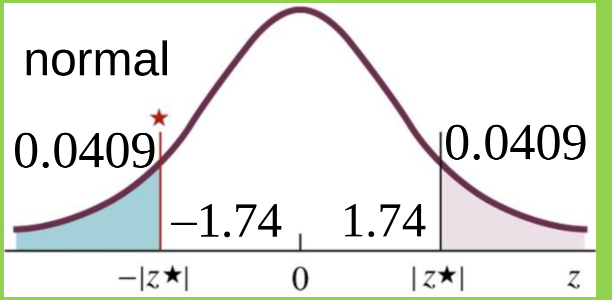
σ known, n is “large” so by CLT $\bar{x}$ is normal

z^* is normal

Step 3 The Sample Evidence: Calculate test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{67.2 - 69}{4 / \sqrt{15}} = -1.74$$

n=15, $\bar{x}$=67.2, $\sigma = 4$



Step 4 The Probability Distribution:

$P(z >| z^* |) = p - \text{value} \longrightarrow$ 0.0819

Step 5 The Results:

$p - \text{value} \leq \alpha$, **reject H_0**, $p - \text{value} > \alpha$ **fail to reject H_0**

α = 0.05

8: Introduction to Statistical Inference

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

Step 1 The Set-Up: Null (H_0) and alternative (H_a) hypotheses

H_0: μ = 69” vs. H_a: $\mu \neq$ 69”

Step 2 The Hypothesis Test Criteria: Test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

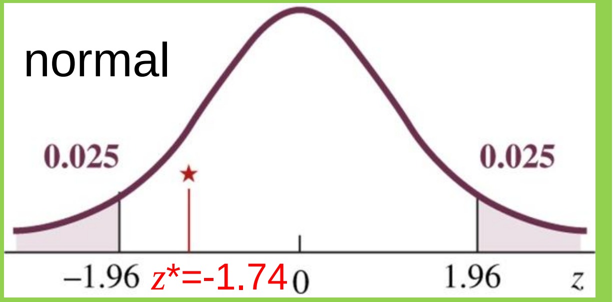
σ known, n is “large” so by CLT $\bar{x}$ is normal

z^* is normal

Step 3 The Sample Evidence: Calculate test statistic.

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{67.2 - 69}{4 / \sqrt{15}} = -1.74$$

n=15, $\bar{x}$ =67.2, $\sigma = 4$



Step 4 The Probability Distribution:

α = 0.05, $z(\alpha/2)$=1.96

Step 5 The Results:

$|z^*| > z(\alpha/2)$, **reject H_0**, $|z^*| \leq z(\alpha/2)$ **fail to reject H_0**

8: Introduction to Statistical Inference

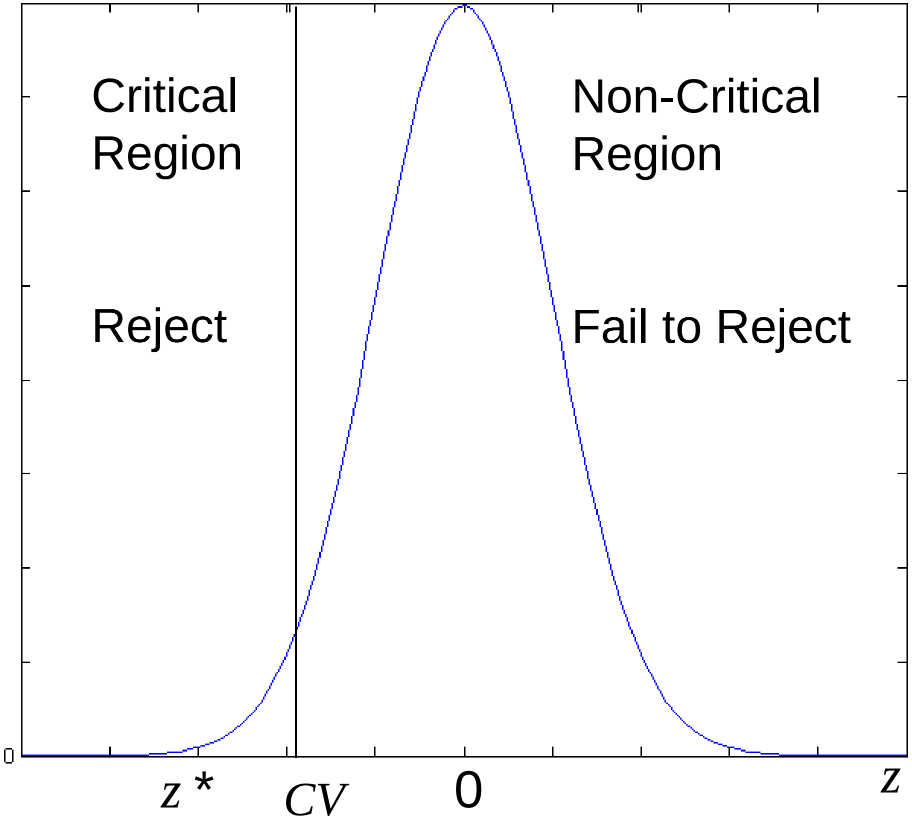
8.5 Hypothesis Test of Mean (σ Known): Classical Approach

There are three possible hypothesis pairs for the mean.

H_0: $\mu \geq \mu_0$ vs. H_a: $\mu < \mu_0$

Reject H_0 if $$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$ less than $$-z(\alpha)$$

data indicates $\mu < \mu_0$ because $\bar{x}$ is "a lot" smaller than μ_0



8: Introduction to Statistical Inference

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

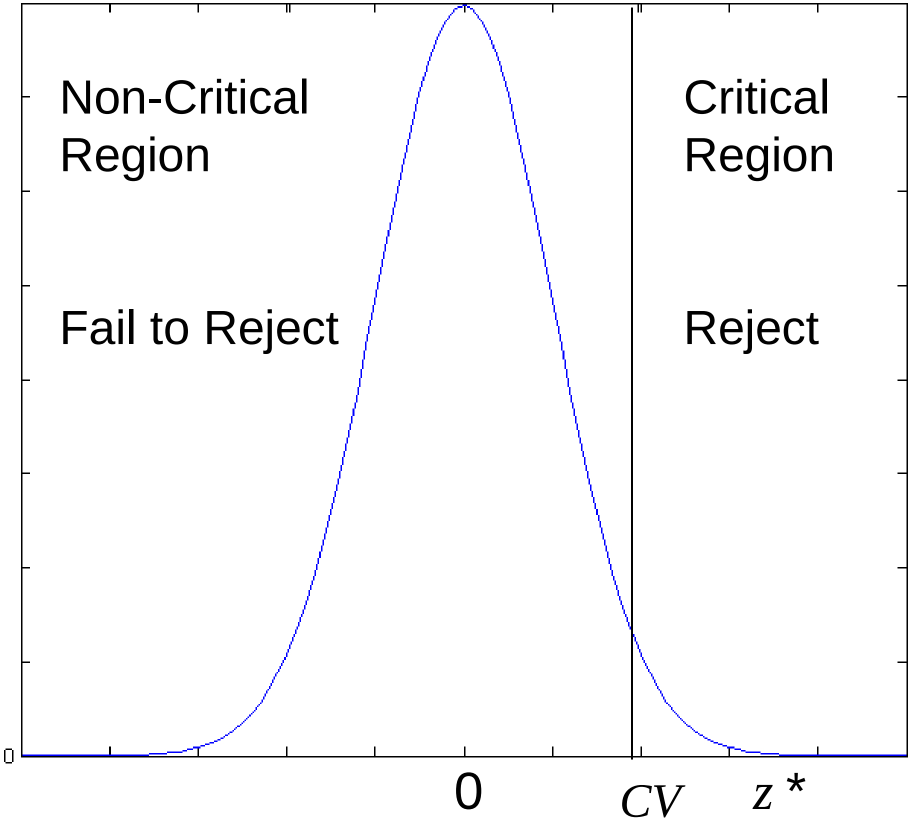
There are three possible hypothesis pairs for the mean.

H_0: $\mu \le \mu_0$ vs. H_a: $\mu > \mu_0$

Reject H_0 if greater then

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \qquad z(\alpha)$$

data indicates $\mu > \mu_0$
because $\bar{x}$ is “a lot”
larger than μ_0



8: Introduction to Statistical Inference

8.5 Hypothesis Test of Mean (σ Known): Classical Approach

There are three possible hypothesis pairs for the mean.

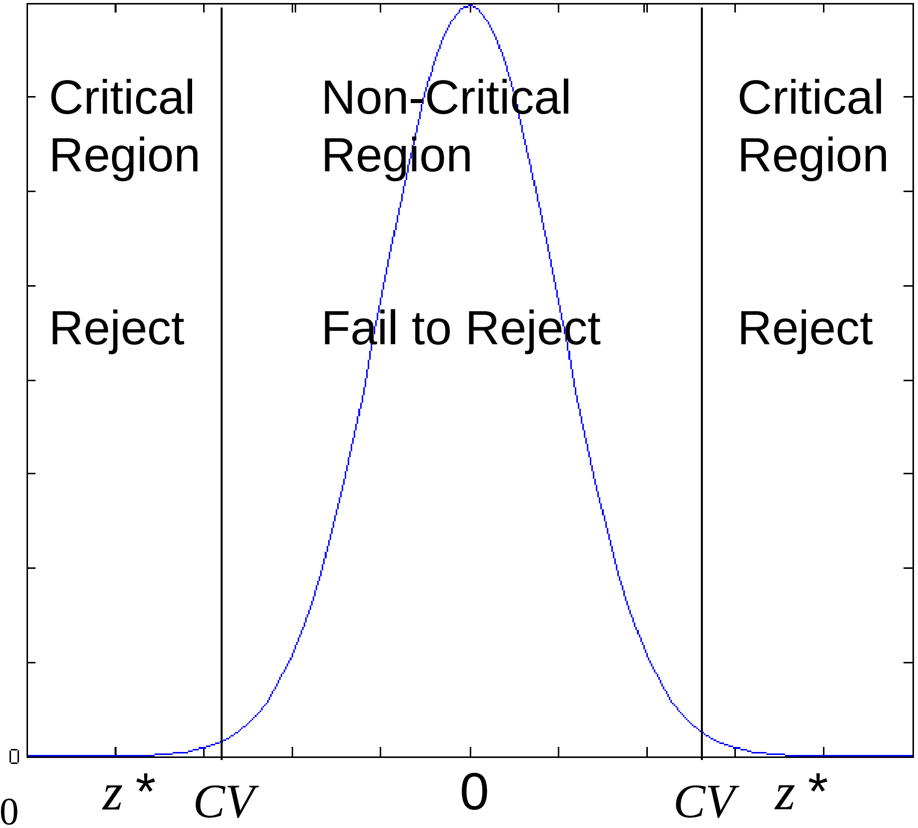
H_0: $\mu = \mu_0$ vs. H_a: $\mu \neq \mu_0$

Reject H_0 if less than

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \qquad -z(\alpha / 2)$$

or if is greater than

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \qquad z(\alpha / 2)$$



data indicates $\mu \neq \mu_0$, $\bar{x}$ far from μ_0

Exam 2 Next Class