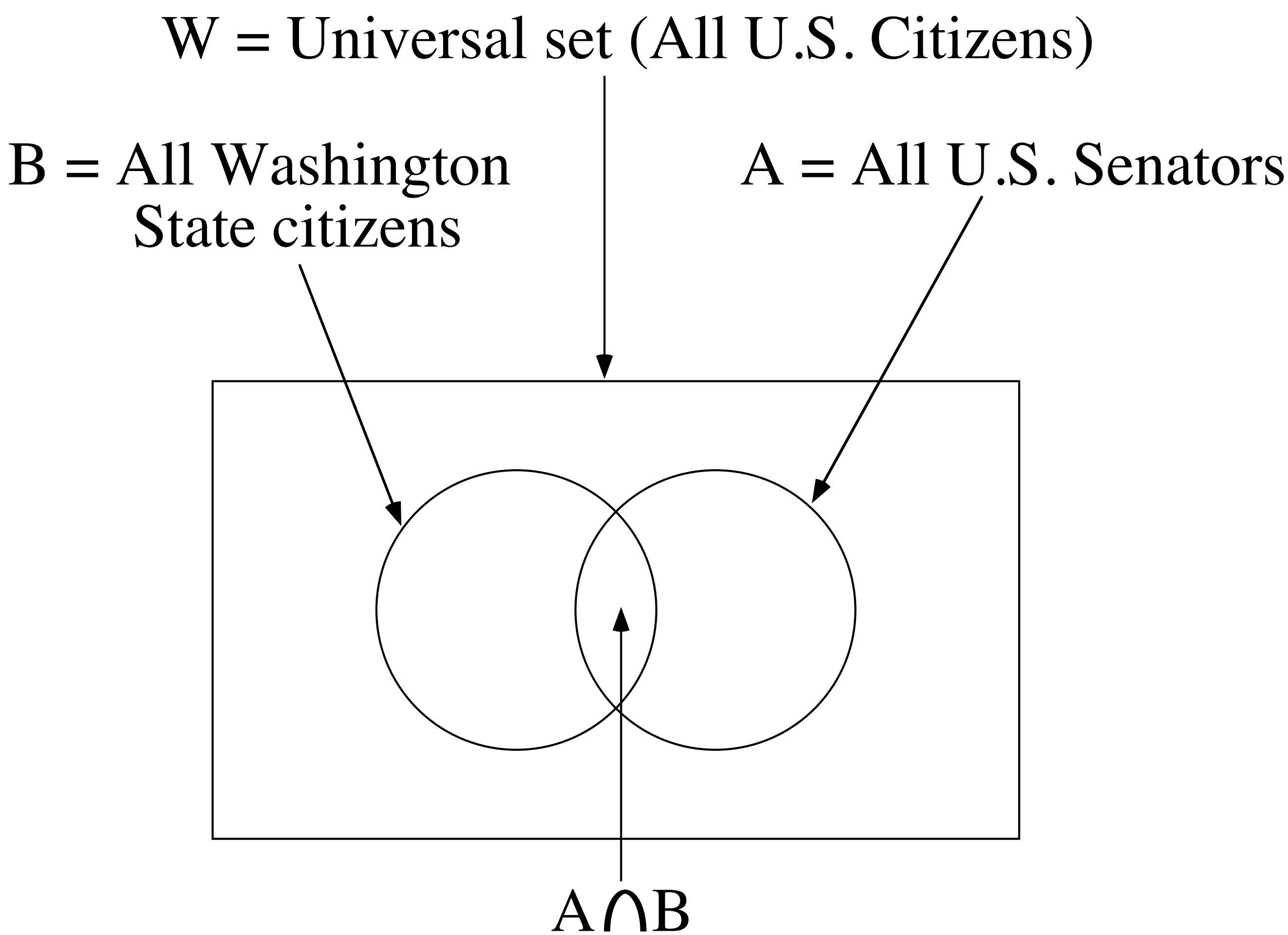
P317: Puyallup Man Skull Sizes

Volumes of 5 Skulls (cubic inches)

Case I	Case II	Case III	Case IV	Case V
96	66	96	100	96
105	75	96	100	105
90	60	96	100	90
100	70	96	100	100
89	59	96	100	89
				95
				85
				108
				⋮

P317: Venn Diagrams



P317: Baby Girls and Boys

<u>Question</u>

How many sequences of girls and boys are there in a four-child family?

<u>Answer:</u>

S = { BBBB, BBBG, BBGB, BBGG,
BGBB, BGBG, BGGB, BGGG,
GBBB, GBBG, GBGB, GBGG,
GGBB, GGBG, GGGB, GGGG }

P317: Contingency Tables

Example 1: Employment Status/Collar

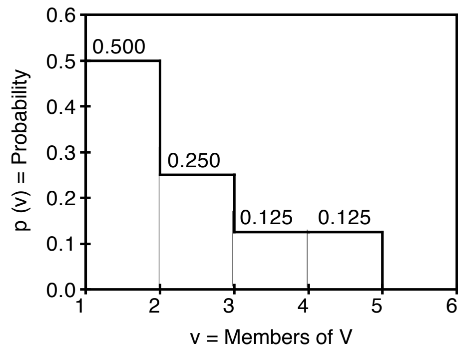
	B : Blue Collar:	$\overline{B}$: White Collar	
E: Employed:	$f(E \cap B) =$ $p(E \cap B) =$	$f(E \cap \overline{B}) =$ $p(E \cap \overline{B}) =$	$f(E) =$ $p(E) =$
$\overline{E}$: Unemployed:	$f(\overline{E} \cap B) =$ $p(\overline{E} \cap B) =$	$f(\overline{E} \cap \overline{B}) =$ $p(\overline{E} \cap \overline{B}) =$	$f(\overline{E}) =$ $p(\overline{E}) =$
	$f(B) =$ $p(B) =$	$f(\overline{B}) =$ $p(\overline{B}) =$	$f(S) =$ $p(S) =$

Example 2: Employment Status/Gender

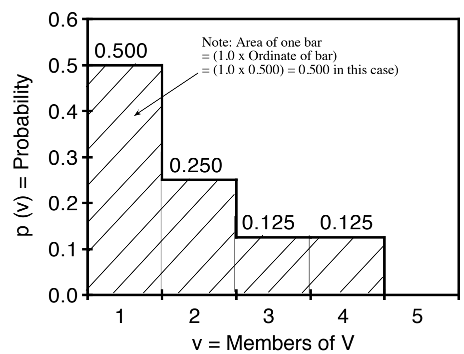
	F : Female:	$\overline{F}$: Male	
E: Employed:	$f(E \cap F) =$ $p(E \cap F) =$	$f(E \cap \overline{F}) =$ $p(E \cap \overline{F}) =$	$f(E) =$ $p(E) =$
$\overline{E}$: Unemployed:	$f(\overline{E} \cap F) =$ $p(\overline{E} \cap F) =$	$f(\overline{E} \cap \overline{F}) =$ $p(\overline{E} \cap \overline{F}) =$	$f(\overline{E}) =$ $p(\overline{E}) =$
	$f(F) =$ $p(F) =$	$f(\overline{F}) =$ $p(\overline{F}) =$	$f(S) =$ $p(S) =$

P317: Properties of Probability Distributions

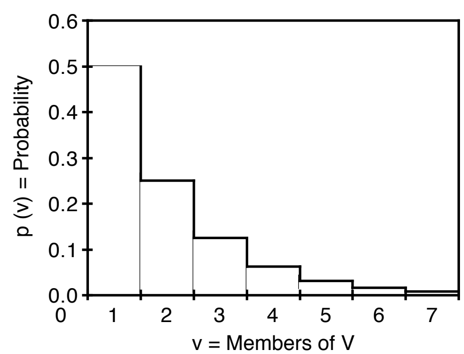
1. Ordinates add to 1.0



2. Areas add to 1.0

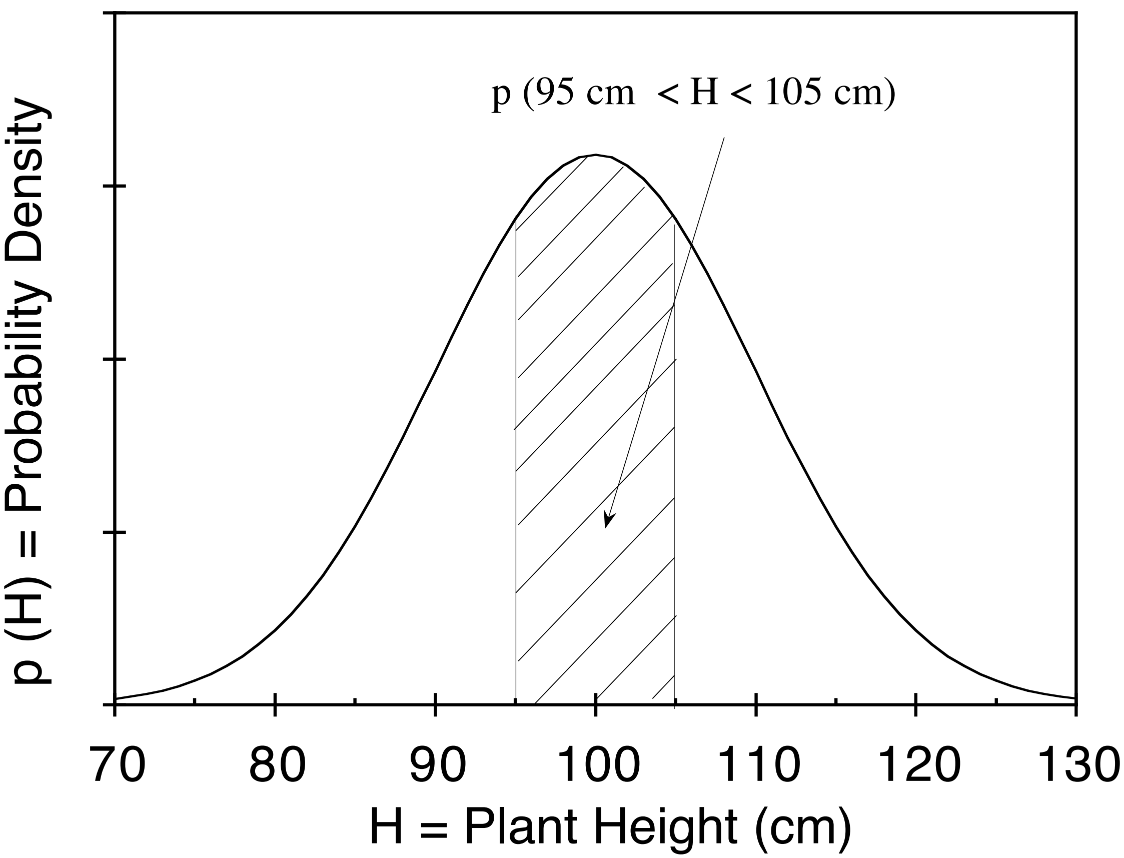


3. In probability distributions with infinite values, areas and ordinates still sum to 1.0



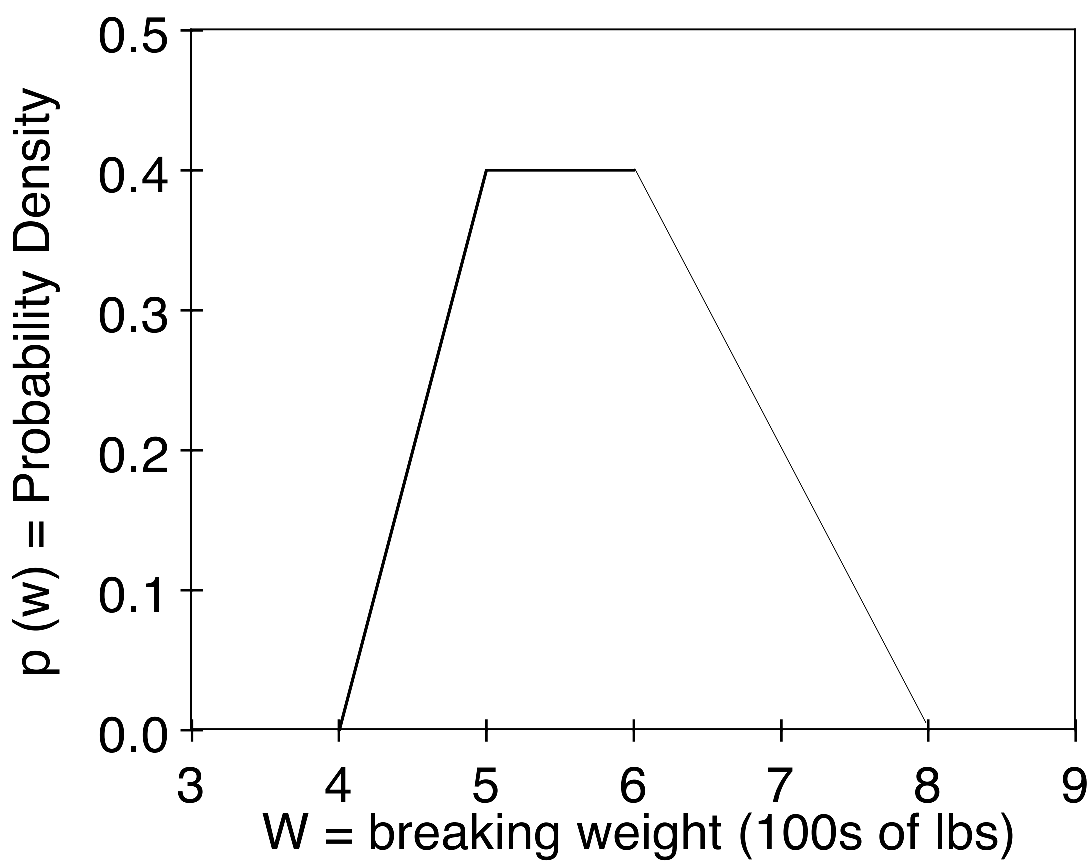
P317: Areas Under Curves

In a continuous probability distribution, probability of some interval is represented by the area under the curve within that interval.



P317: Area Demonstration

The Acme Rope Co. makes climbing ropes which have the breaking weight (strength) distribution depicted below.



P317: Inventing a Measure of Variability

Large Deviations:

Original Scores (X_i)	Deviation Scores (X_i - M)	Absolute Deviations $|X_i - M|$	Squared Deviations $(X_i - M)^2$
0			
4			
8			
12			
16			

Small Deviations:

Original Scores (X_i)	Deviation Scores (X_i - M)	Absolute Deviations $|X_i - M|$	Squared Deviations $(X_i - M)^2$
6			
7			
8			
9			
10			

P317: Demonstration of z-Scores

Two cases for Charlie Cutler:
Scores represent sales for 7 salespeople
Charlie's score: X = 25

Case I	Case II
0	5
5	15
10	15
15	15
20	15
25	15
30	25
$M = \Sigma X / 7 = 105 / 7 = 15$	$M = \Sigma X / 7 = 105 / 7 = 15$
$S^2 = \Sigma X^2 / 7 - M^2 = 2275 / 7 - 15^2$	$S^2 = \Sigma X^2 / 7 - M^2 = 1775 / 7 - 15^2$
$S^2 = 325 - 225 = 100$	$S^2 = 253.571 - 225 = 28.571$
$S \sqrt{100} = 10$	$S = \sqrt{28.571} = 5.345$
$z = (X - M) / S = (25 - 15) / 10$	$z = (X - M) / S = (25 - 15) / 5.345$
$10 / 10 = 1.000$	$10 / 5.345 = 1.871$

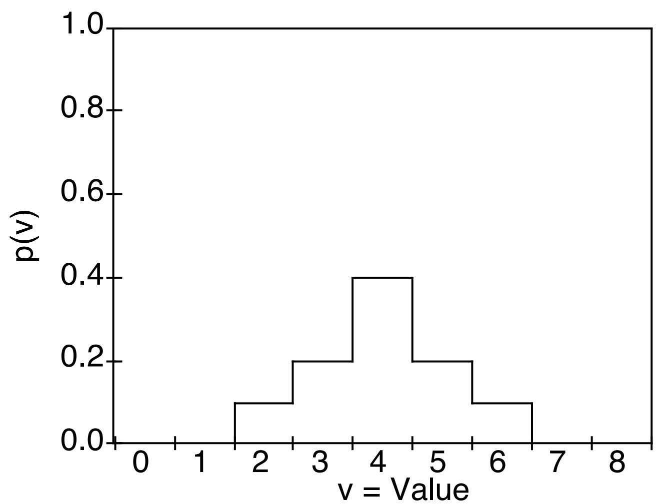
P317: Means and Variances of Probability Distributions: Large Variability

v	p(v)	vp(v)	v^2	$v^2p(v)$
1	0.0	0.0	1	0.0
2	0.1	0.2	4	0.4
3	0.2	0.6	9	1.8
4	0.4	1.6	16	6.4
5	0.2	1.0	25	5.0
6	0.1	0.6	36	3.6
7	0.0	0.0	49	0.0
SUMS:	1.0	4.0	140.0	17.2

$E(x) = \Sigma v p(v) = 4.0$

$\text{Variance} = \Sigma v^2 p(v) - [E(x)]^2 = 17.200 - 16.000 = 1.200$

$\text{Standard Deviation} = \sqrt{\text{Variance}} = \sqrt{1.200} = 1.095$



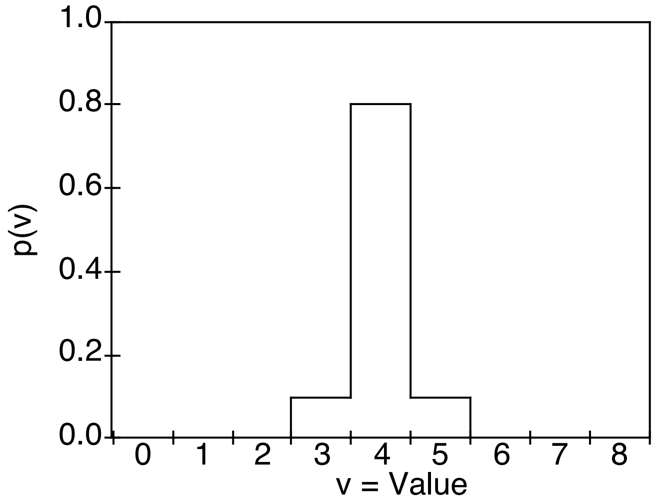
P317: Means and Variances of Probability Distributions: Small Variability

v	p(v)	vp(v)	v^2	$v^2p(v)$
1	0.0	0.0	1.0	0.0
2	0.0	0.0	4.0	0.0
3	0.1	0.3	9.0	0.9
4	0.8	3.2	16.0	12.8
5	0.1	0.5	25.0	2.5
6	0.0	0.0	36.0	0.0
7	0.0	0.0	49.0	0.0
SUMS:	1.0	4.0	140.0	16.2

$E(x) = \Sigma vp(v) = 4.0$

$\text{Variance} = \Sigma v^2p(v) - [E(x)]^2 = 16.200 - 16.000 = 0.200$

$\text{Standard Deviation} = \sqrt{\text{Variance}} = \sqrt{0.200} = 0.447$



P317: Binomial Symmetry

p = 0.1, q = 0.9, N = 12

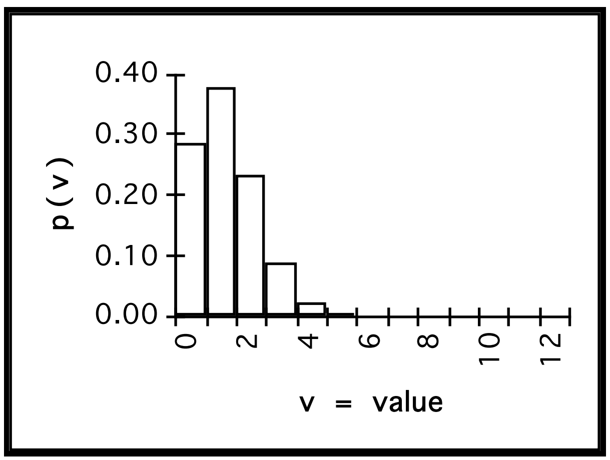
v	p(v)
0	0.28
1	0.38
2	0.23
3	0.09
4	0.02
5	0.00
6	0.00
7	0.00
8	0.00
9	0.00
10	0.00
11	0.00
12	0.00

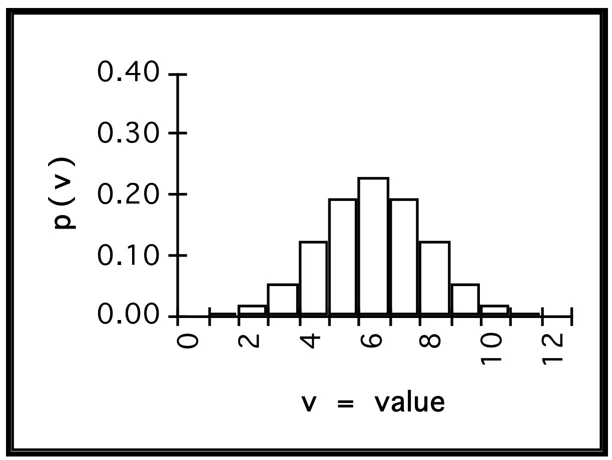
p = 0.5, q = 0.5, N = 12

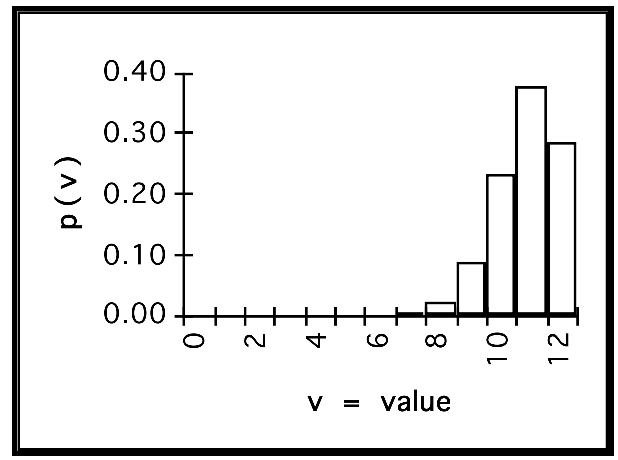
v	p(v)
0	0.00
1	0.00
2	0.02
3	0.05
4	0.12
5	0.19
6	0.23
7	0.19
8	0.12
9	0.05
10	0.02
11	0.00
12	0.00

p = 0.9, q = 0.1, N = 12

v	p(v)
0	0.00
1	0.00
2	0.00
3	0.00
4	0.00
5	0.00
6	0.00
7	0.00
8	0.02
9	0.09
10	0.23
11	0.38
12	0.28



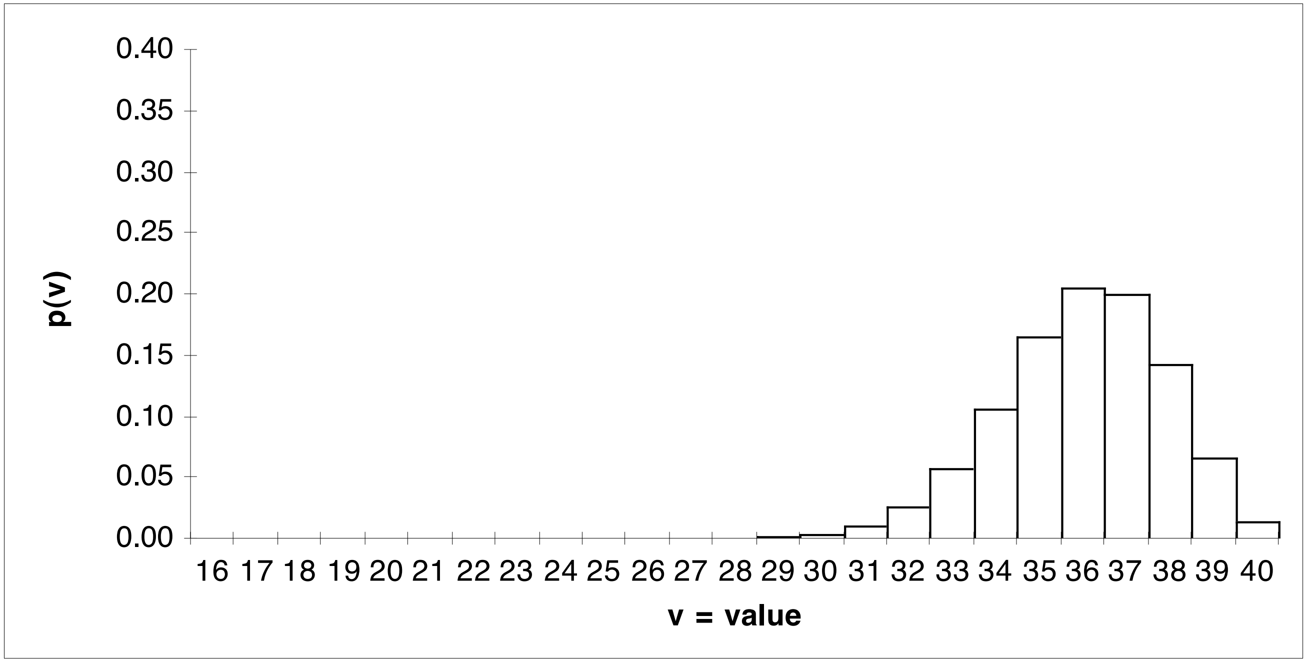




P317: Binomial Symmetry: Large N

v	p(v)
0	0.00
1	0.00
2	0.00
3	0.00
4	0.00
5	0.00
6	0.00
7	0.00
8	0.00
9	0.00
10	0.00
11	0.00
12	0.00
13	0.00
14	0.00
15	0.00
16	0.00
17	0.00
18	0.00
19	0.00
20	0.00
21	0.00
22	0.00
23	0.00
24	0.00
25	0.00
26	0.00
27	0.00
28	0.00
29	0.00
30	0.00
31	0.01
32	0.03
33	0.06
34	0.11
35	0.16
36	0.21
37	0.20
38	0.14
39	0.07
40	0.01

$p = 0.9$
$q = 0.1$
$N = 40$



P317: Binomial Distributions Can be Expressed in terms of either Frequencies or Proportions

Example: $p = 0.5$; $q = 0.5$; $N = 100$

Flip a coin 100 times...How many heads do you expect to get, i.e., what is the <u>frequency</u> of heads that you expect to get (0 - 100)

or, equivalently you can ask...

what is the <u>proportion</u> of heads that you expect to get (0.00 - 1.00)

	Mean: $E(x)$	Variance: σ^2	Standard Deviation: σ
Frequencies	Np (50)	Npq (25)	$\sqrt{Npq}$ (5)
Proportions	p (0.50)	pq/N (0.0025)	$\sqrt{pq/N}$ (0.05)

Thus...to summarize as $E(x) \pm \sigma$

When talking in terms of frequencies, you'd summarize this distribution as:

50 ± 5 is the <u>frequency</u> of heads you expect to get.

When talking in terms of proportions, you'd summarize this distribution as:

0.50 ± 0.05 is the <u>proportion</u> of heads you expect to get.

P317: Sign Test Example: Data

(Entries are rats' weights at adulthood)				
Rat Pair	Saline (Control)	Nicotine (Experimental)	Difference (C - E)	Sign (+/-)
1	500	420	80	+
2	490	510	-20	-
3	480	400	80	+
4	495	490	5	+
5	605	500	105	+
6	500	400	100	+
7	455	454	1	+
8	608	550	58	+

P317: Sign Test Example: Hypothesis Testing

H_0: Independent variable (saline/nicotine) has no effect on dependent variable (weight). This means that any differences between two members of a pair must be just due to random variation.

H_1: Independent variable (saline/nicotine) has an effect on the dependent variable (weight). This means that the nicotine rats weigh systematically less than their control litter-mates.

SUMMARY SCORE: Number of plusses (more plusses = more evidence that H_1 is true)

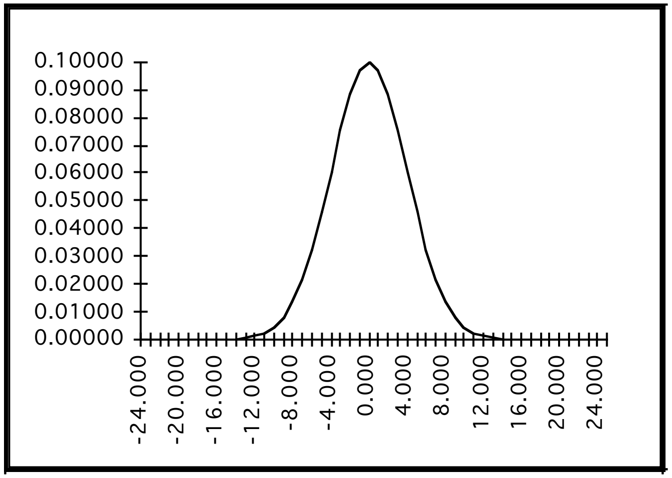
Probability Distribution of Number of Plusses (v)

v	0	1	2	3	4	5	6	7	8
p(v)	0.004	0.031	0.109	0.219	0.273	0.219	0.109	0.031	0.004
p(≥v)	1.000	0.996	0.965	0.855	0.637	0.363	0.145	0.035	0.004

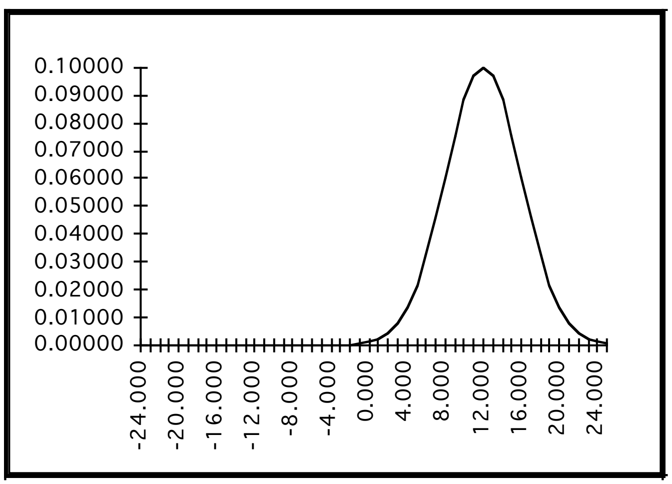
P317: The Mean and Variance of a Normal Distribution

Changing the Mean
Shifts the Distribution...

1. $\mu = 0$; $\sigma = 4$; $\sigma^2 = 16$

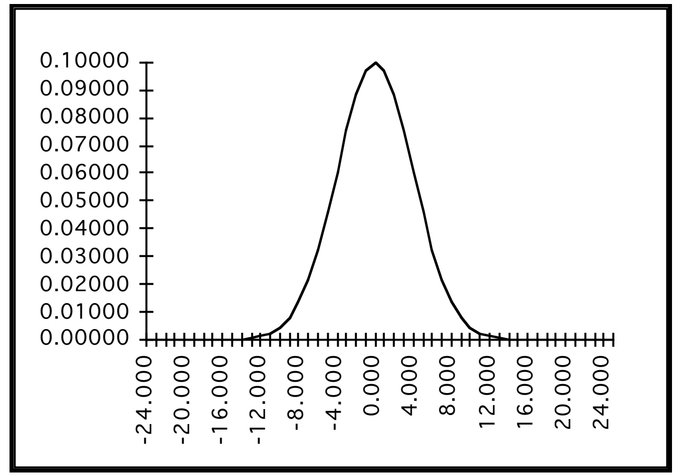


2. $\mu = 12$; $\sigma = 4$; $\sigma^2 = 16$

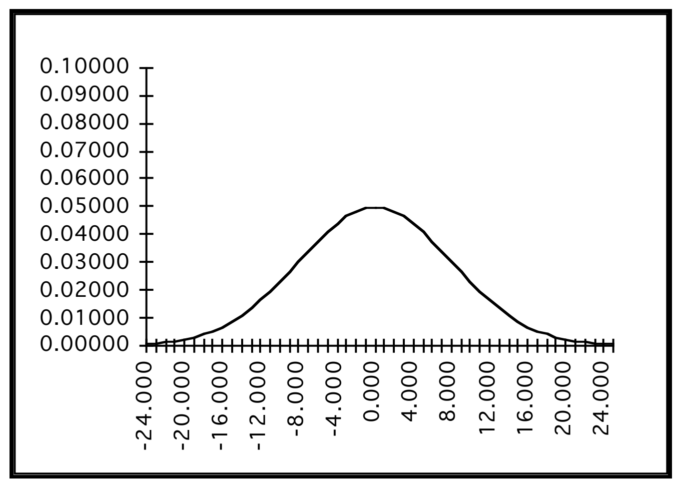


Changing the Variance
Changes the Shape of the Distribution...

3. $\mu = 0$; $\sigma = 4$; $\sigma^2 = 16$



4. $\mu = 0$; $\sigma = 8$; $\sigma^2 = 64$

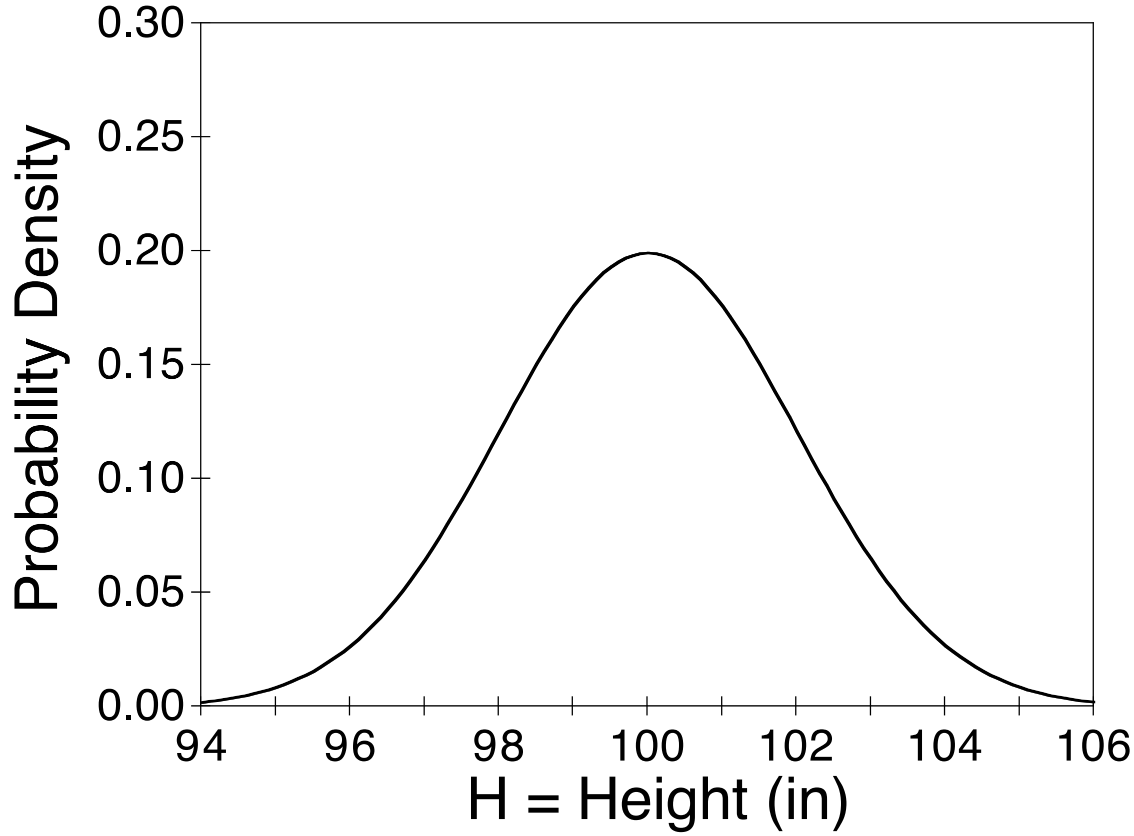


P317: Computing Probabilities Using Normal Distributions

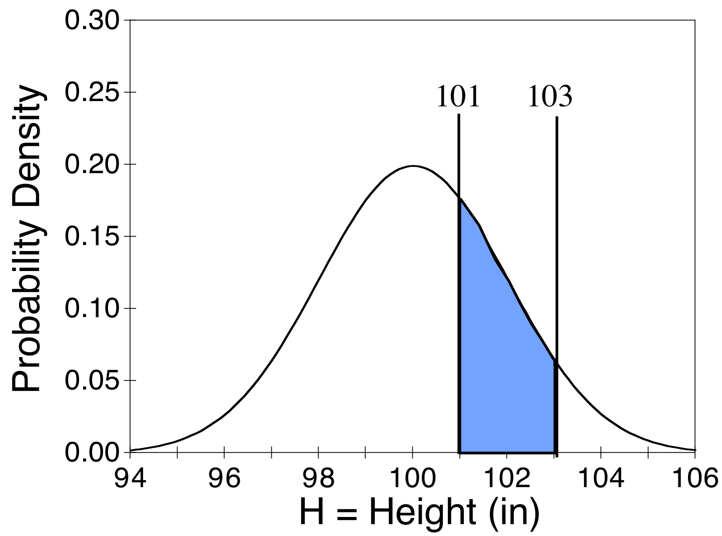
Example: Merkin Plant heights are normally distributed with...

$\mu = 100$ in

$\sigma = \quad 2$ in

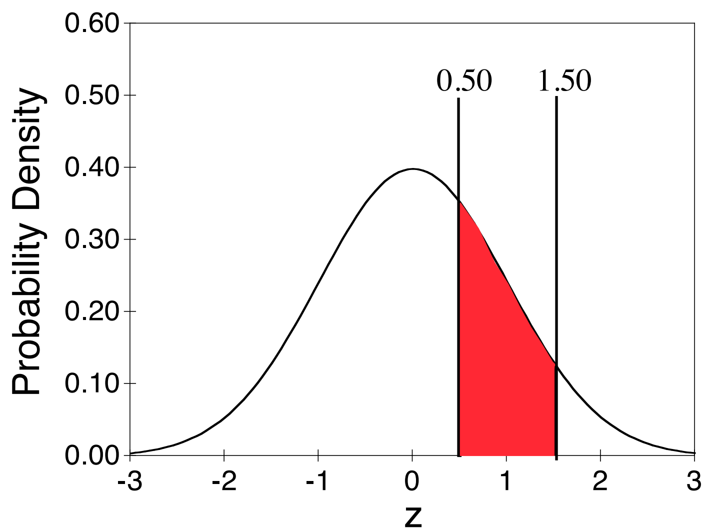


P317: Rescaling to z-scores



Start with original distribution (in H)

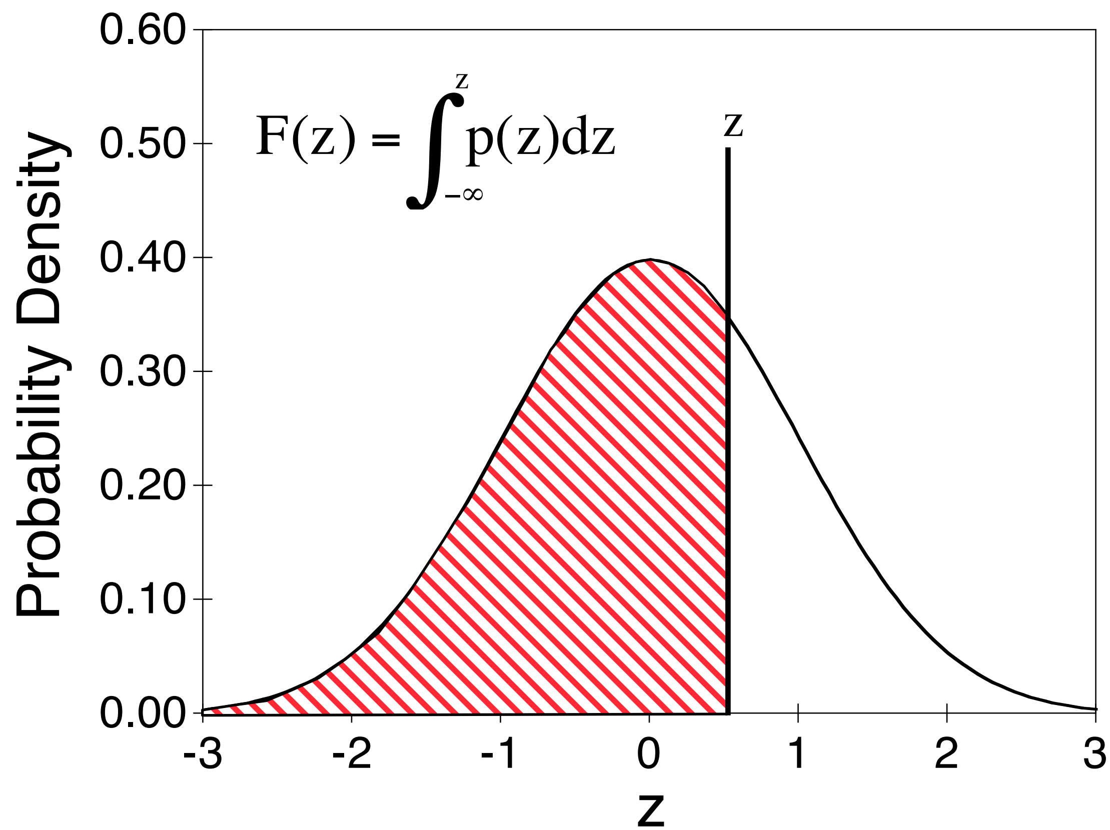
Change to z-scores:



End up with z-distribution

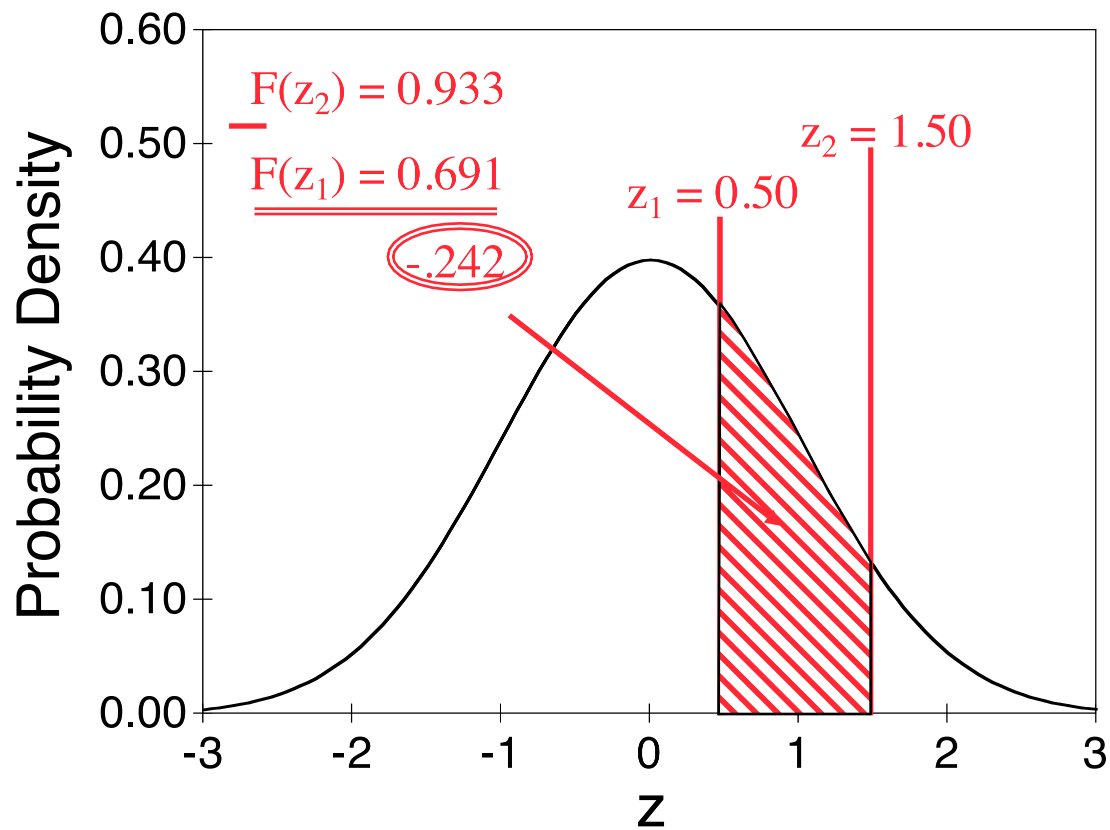
P317: Using Tables to Compute Values of z...

z	F(z)
0.00	.500
0.01	.504
⋮	⋮
0.50	.691
⋮	⋮
1.50	.933
⋮	⋮
5.50	.999



P317: Solution!

z	F(z)
0.00	.500
0.01	.504
⋮	⋮
0.50	.691
⋮	⋮
1.50	.933
⋮	⋮
5.50	.999



The sought-after probablity is the *difference* between $F(z_2)$ and $F(z_1)$

P317: Summary: Computing Areas Under the Normal Distribution

1. Problem: Compute $p(x_1 \leq x \leq x_2)$
2. Compute z-scores
   $z_2 = (x_2-\mu)/\sigma$
   $z_1 = (x_1-\mu)/\sigma$

   This changes problem to: $p(z_1 \leq z \leq z_2)$
3. Compute F(z)'s
   $F(z_2)$ (from table)
   $F(z_1)$ (from table)
4. $p(x_1 \leq x \leq x_2) = p(z_1 \leq z \leq z_2) =$ **$F(z_2) - F(z_1)$**

P317: Suppose *lots* of samples of size n = 16 are drawn from the IQ population...

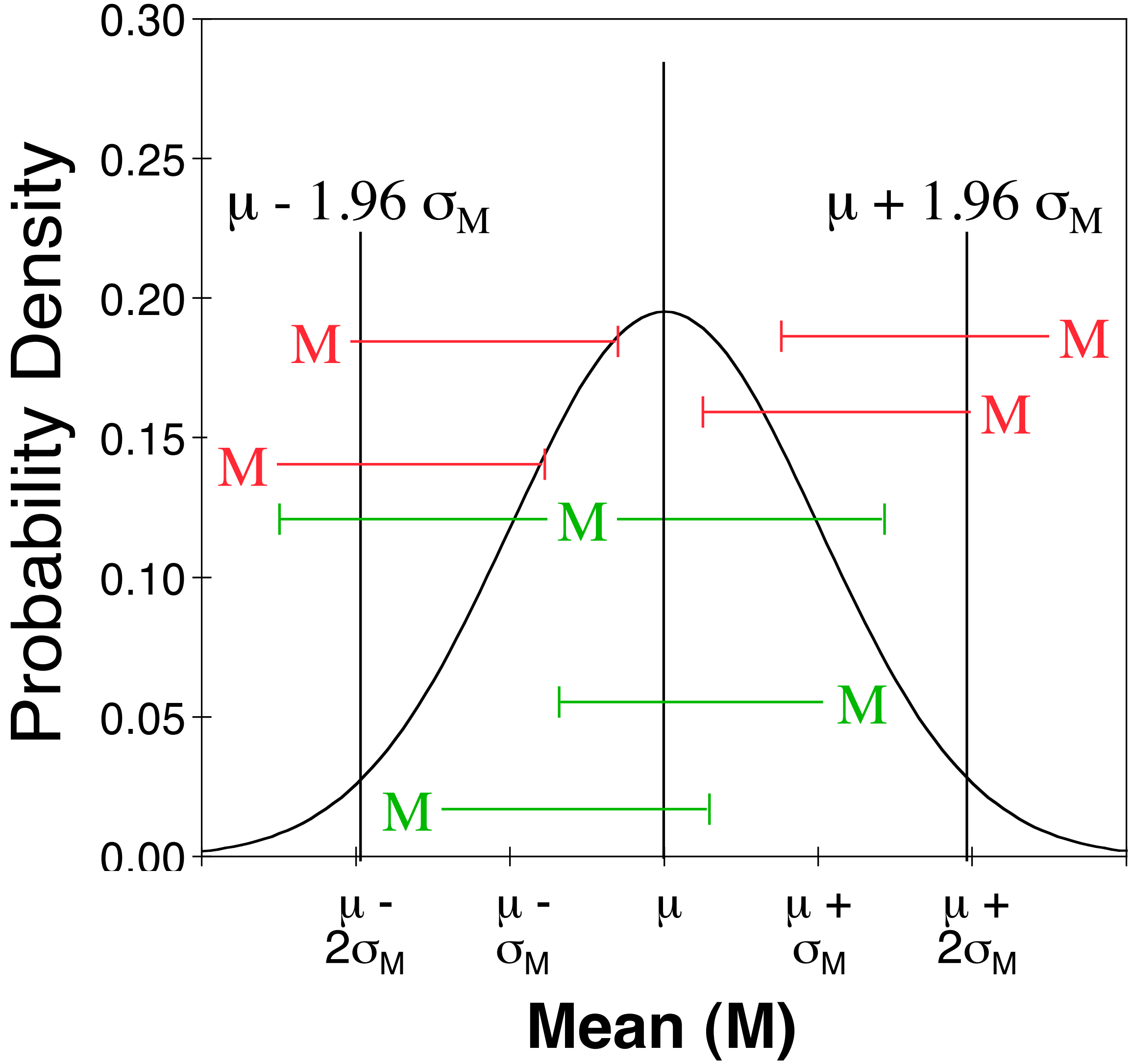
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	…
1	92	107	113	106	88	
2	122	108	115	140	100	
3	111	92	114	117	111	
4	120	89	106	94	72	
5	109	113	101	100	111	
6	88	96	107	122	107	
7	72	94	86	92	127	
8	102	107	73	111	89	
9	100	100	110	110	102	
10	93	92	96	112	86	
11	113	80	100	94	79	
12	85	80	114	127	91	
13	109	124	82	94	101	
14	120	91	100	130	90	
15	110	92	101	77	78	
16	97	102	107	103	89	
M:	102.69	97.93	101.56	108.06	95.06	…
S^2:	187.42	130.64	140.42	254.40	194.04	…
S:	13.69	11.43	11.85	15.95	13.93	…

P317: Conceptualization of Sampling Distribution Matrix

<u>Sample Statistic</u>

Sample Size (n)	M = Mean	S^2 = Variance	S = Standard Deviation	Md = Median	…
n = 1					
n = 2					
⋮	⋮	⋮	⋮	⋮	⋮
n = 16					
⋮	⋮	⋮	⋮	⋮	⋮
n = ∞					

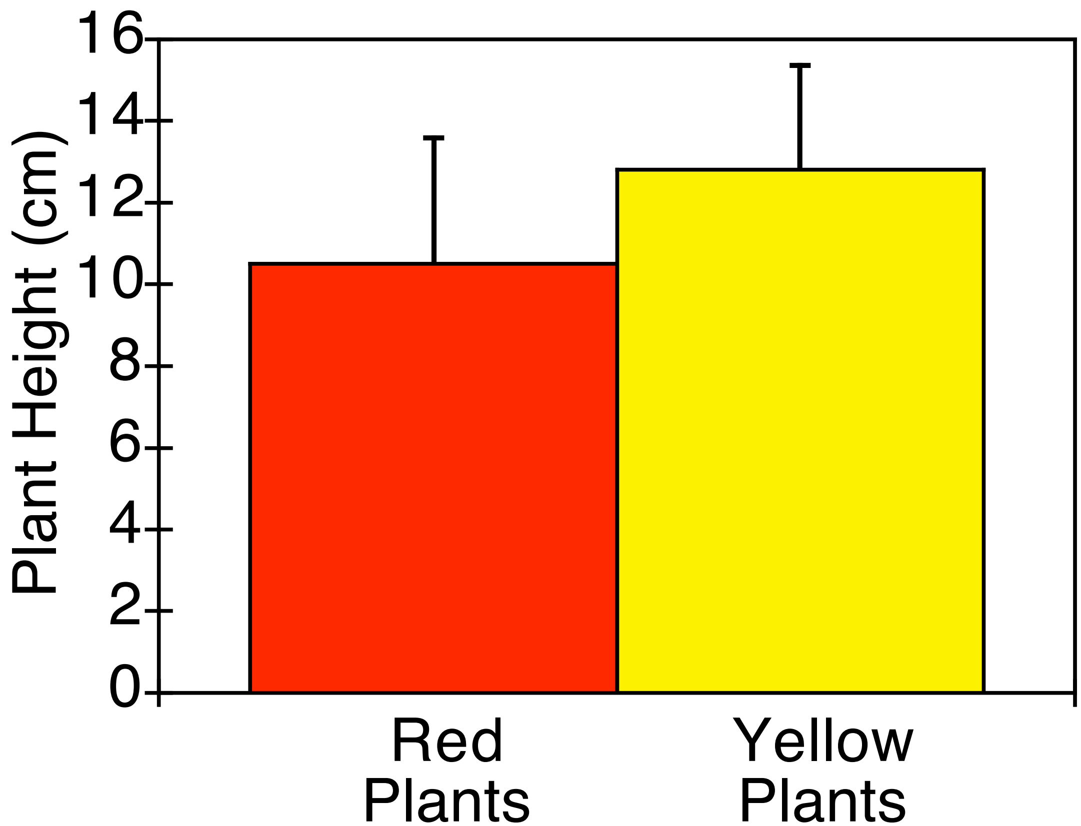
Only those 5% of Sample Means falling outside these limits *fail* to capture the population mean:



P317: Confidence Intervals around Plant Heights

	Group 1 (Red Plants)	Group 2 (Yellow Plants)
n	$n_1 = 4$	$n_2 = 6$
Sample Mean	$M_1 = 10.5$ cm	$M_2 = 12.8$ cm
Population Variance	$\sigma^2 = 10$	$\sigma^2 = 10$
SEM	$\sigma_{M_1} = \sqrt{10/6} = 1.581$	$\sigma_{M_2} = \sqrt{10/4} = 1.291$
95% CI	$10.5 \pm (1.581)1.96$ $= 10.5 \pm 3.100$	$12.8 \pm (1.291)1.96$ $= 12.8 \pm 2.530$

P317: Plotted Data:



P317: Confidence Intervals and Sample Size

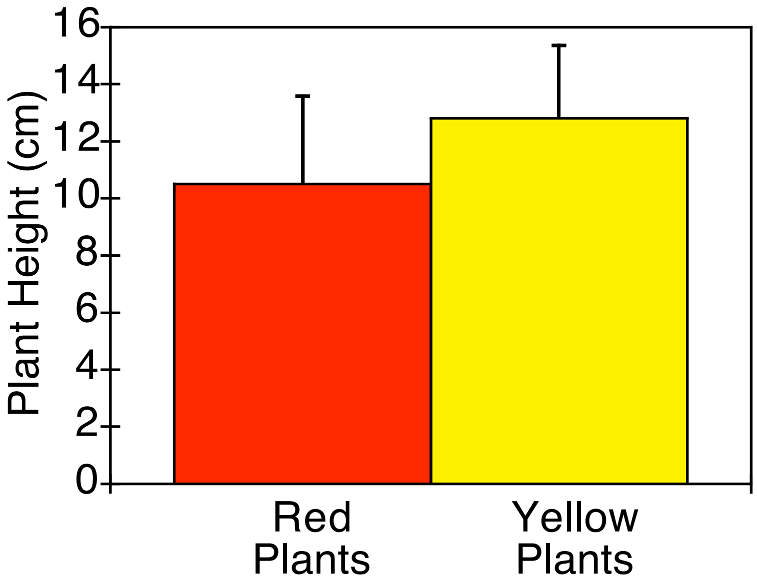
	Group 1 (Red Plants)	Group 2 (Yellow Plants)
n	$n_1 = 4$	$n_2 = 6$
Sample Mean	$M_1 = 10.5$ cm	$M_2 = 12.8$ cm
Population Variance	$\sigma^2 = 10$	$\sigma^2 = 10$
SEM	$\sigma_M = 1.581$	$\sigma_M = 1.291$
95% CI	$10.5 \pm (1.581)1.96$ $= 10.5 \pm 3.100$	$12.8 \pm (1.291)1.96$ $= 12.8 \pm 2.530$

	Group 1 (Red Plants)	Group 2 (Yellow Plants)
n	$n_1 = 50$	$n_2 = 200$
Sample Mean	$M_1 = 10.5$ cm	$M_2 = 12.8$ cm
Population Variance	$\sigma^2 = 10$	$\sigma^2 = 10$
SEM	$\sigma_M = 0.447$	$\sigma_M = 0.224$
95% CI	$10.5 \pm (0.447)1.96$ $= 10.5 \pm 0.876$	$12.8 \pm (0.224)1.96$ $= 12.8 \pm 0.439$

P317: Graphical Comparison...

$n_1 = 4$

$n_2 = 6$

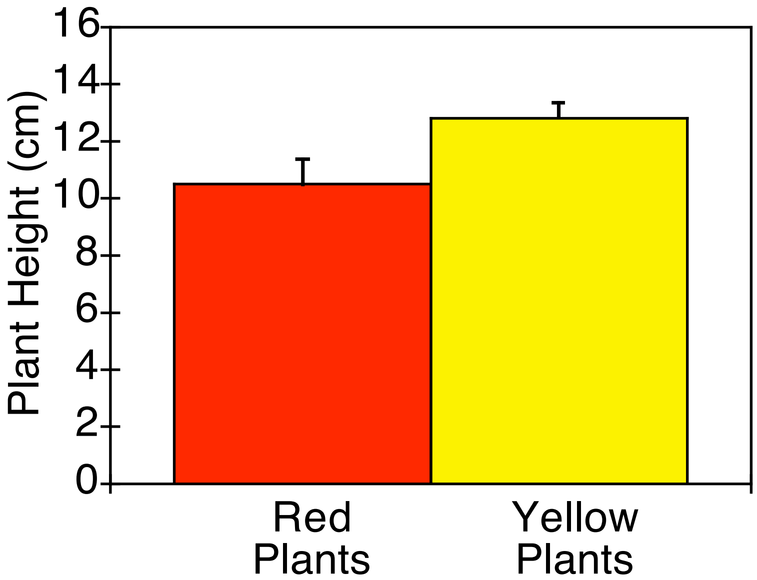


$(M_2 - M_1)$:

95% CI = 2.3 ± 4.000

$n_1 = 50$

$n_2 = 200$



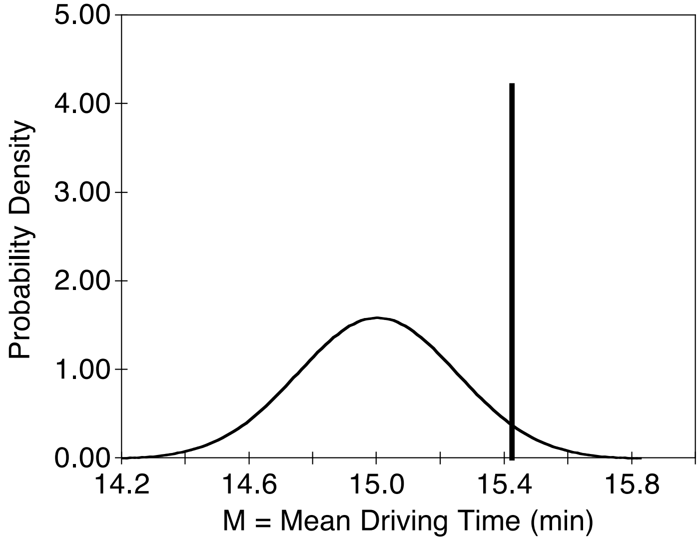
$(M_2 - M_1)$:

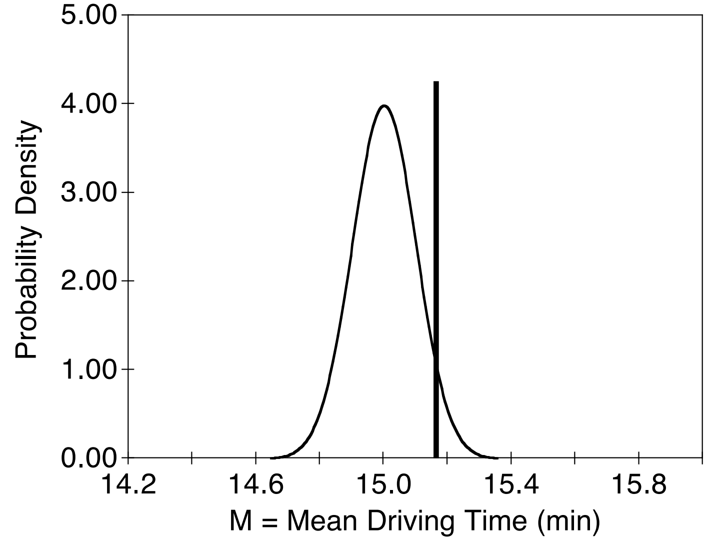
95% CI = 2.3 ± 0.98

P317: As n Increases, Power Increases

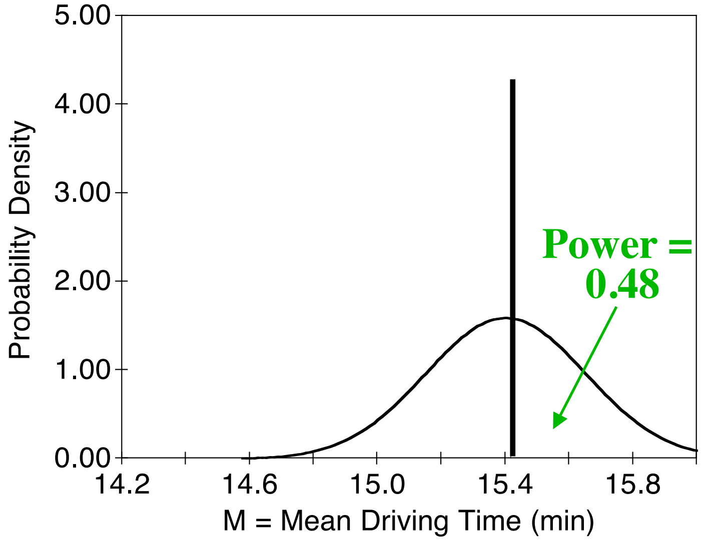
n = 16

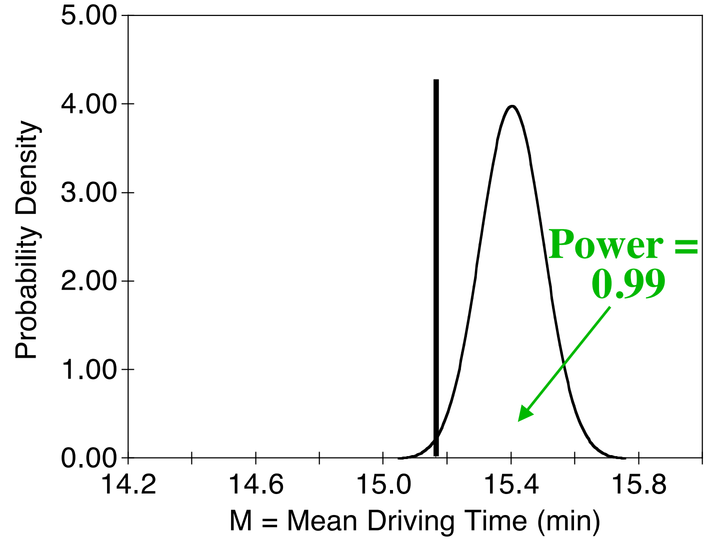
n = 100





H_0 True



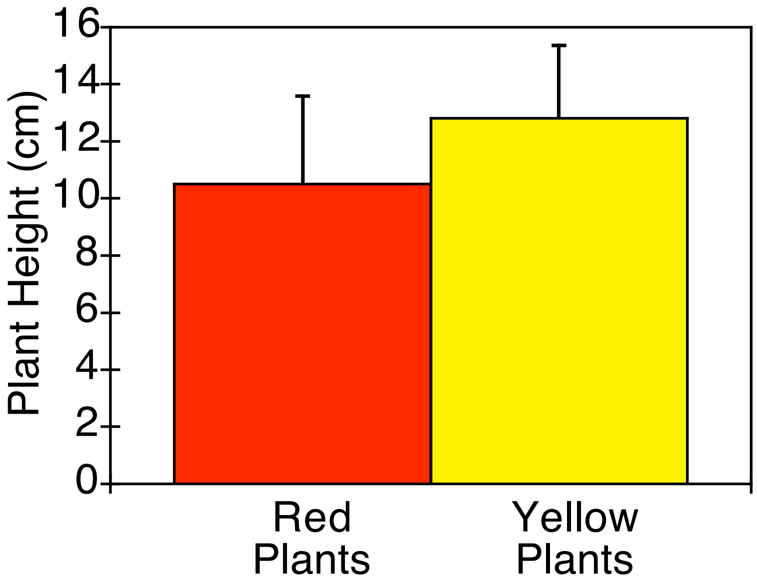


H_1 True

P317: As n Increases, Sizes of the Associated Confidence Intervals Decrease

$n_1 = 4$

$n_2 = 6$

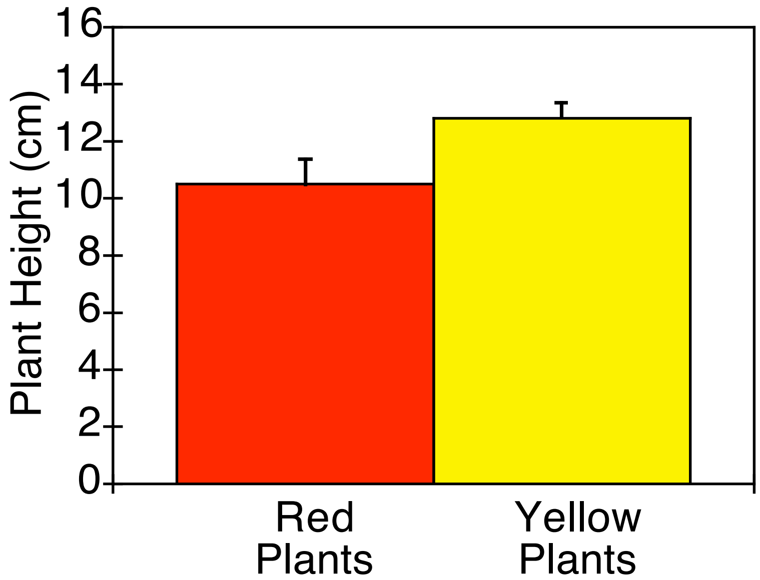


$(M_2 - M_1)$:

95% CI = 2.3 ± 4.000

$n_1 = 50$

$n_2 = 200$



$(M_2 - M_1)$:

95% CI = 2.3 ± 0.98