Lecture 11: Chapter 7, Secs. 2-3 Binomial and Normal Random Variables

- Definition
- What if Events are Dependent?
- Center, Spread, Shape of Counts, Proportions
- Normal Approximation
- Normal Random Variables

Looking Back: *Review*

- **4 Stages of Statistics**
  - Data Production (discussed in Lectures 1-3)
  - Displaying and Summarizing (Lectures 3-8)
  - Probability
    - Finding Probabilities (discussed in Lectures 9-10)
    - Random Variables (introduced in Lecture 10)
      - Binomial
      - Normal
    - Sampling Distributions
  - Statistical Inference

Definition *(Review)*

- **Discrete Random Variable:** one whose possible values are finite or countably infinite (like the numbers 1, 2, 3, …)

Looking Ahead: *To perform inference about categorical variables, need to understand behavior of sample proportion. A first step is to understand behavior of sample* ***counts****. We will eventually shift from discrete counts to a normal approximation, which is continuous.*

Definition

Binomial Random Variable counts sampled individuals falling into particular category;

- Sample size n is fixed
- Each selection independent of others
- Just 2 possible values for each individual
- Each has same probability p of falling in category of interest

Example: *A Simple Binomial Random Variable*

- **Background**: The random variable *X* is the count of tails in two flips of a coin.
- **Questions:** Why is *X* binomial? What are *n* and *p*?
- **Responses:**
  - Sample size *n* fixed?
  - Each selection independent of others?
  - Just 2 possible values for each?
  - Each has same probability *p*?

 Practice: 7.24 p.307

Example: *A Simple Binomial Random Variable*

- **Background**: The random variable *X* is the count of tails in two flips of a coin.
- **Question:** How do we display *X*?
- **Response:**

Looking Back: *We already discussed and displayed this random variable when learning about probability distributions.*

 Practice: 7.32a p.308

Example: *Determining if R.V. is Binomial*

- **Background**: Consider following R.V.:
  - Pick card from deck of 52, replace, pick another. *X*=no. of cards picked until you get ace.
- **Question:** Is *X* binomial?
- **Response:**

 Practice: 7.25b p.307

Example: *Determining if R.V. is Binomial*

- **Background**: Consider following R.V.:
  - Pick 16 cards without replacement from deck of 52. *X*=no. of red cards picked.
- **Question:** Is *X* binomial?
- **Response:**

 Practice: 7.25d p.307

Example: *Determining if R.V. is Binomial*

- **Background**: Consider following R.V.:
  - Pick 16 cards with replacement from deck of 52. *W*=no. of clubs, *X*=no. of diamonds, *Y*=no. of hearts, *Z*=no. of spades. Goal is to report how frequently each suit is picked.
- **Question:** Are *W*, *X*, *Y*, *Z* binomial?
- **Response:**

Practice: 7.25a p.307

Example: *Determining if R.V. is Binomial*

- **Background**: Consider following R.V.:
  - Pick with replacement from German deck of 32 (doesn't include numbers 2-6), then from deck of 52, back to deck of 32, etc. for 16 selections altogether. *X*=no. of aces picked.
- **Question:** Is *X* binomial?
- **Response:**

Example: *Determining if R.V. is Binomial*

- **Background**: Consider following R.V.:
  - Pick 16 cards with replacement from deck of 52. *X*=no. of hearts picked.
- **Question:** Is *X* binomial?
- **Response:**
  - fixed $n = 16$
  - selections independent (with replacement)
  - just 2 possible values (heart or not)
  - same $p = 0.25$ for all selections

→____

Practice: 7.25c p.307

Requirement of Independence

Snag:

- Binomial theory requires independence
- Actual sampling done *without* replacement so selections are dependent

Resolution: *When sampling without replacement, selections are approximately independent if population is at least 10n.*

Example: *A Binomial Probability Problem*

- **Background**: The proportion of Americans who are left-handed is 0.10. Of 46 presidents, 8 have been left-handed (proportion 0.17).
- **Question:** How can we establish if being left-handed predisposes someone to be president?
- **Response:** Determine if 8 out of 46 (0.17) is ________________ when sampling at random from a population where 0.10 fall in the category of interest.

Solving Binomial Probability Problems

- Use binomial formula or tables

  Only practical for small sample sizes
- Use software

  Won't take this approach until later
- Use normal approximation for count X

  Not quite: more interested in proportions
- Use normal approximation for proportion

  Need mean and standard deviation…

Example: *Mean of Binomial Count, Proportion*

- **Background**: Based on long-run observed outcomes, probability of being left-handed is approx. 0.1. Randomly sample 100 people.
- **Questions:** On average, what should be the
  - **count** of lefties?
  - **proportion** of lefties?
- **Responses:** On average, we should get
  - **count** of lefties _____
  - **proportion** of lefties _____

Mean and S.D. of Counts, Proportions

Count X binomial with parameters n, p has:

- **Mean** np
- **Standard deviation** $\sqrt{np(1-p)}$

Sample proportion $\hat{p} = \frac{X}{n}$ has:

- **Mean** p
- **Standard deviation** $\sqrt{\frac{p(1-p)}{n}}$

Looking Back: *Formulas for s.d. require independence: population at least 10n.*

Example: *Standard Deviation of Sample Count*

- **Background**: Probability of being left-handed is approx. 0.1. Randomly sample 100 people. Sample **count** has mean 100(0.1)= 10, standard deviation $\sqrt{100(0.1)(1-0.1)} = 3$
- **Question:** How do we interpret these?
- **Response:** On average, expect

  sample count =____ lefties.

  Counts vary; typical distance from 10 is ____.

 Practice: 7.29c-d p.308

Example: *S.D. of Sample Proportion*

- **Background**: Probability of being left-handed is approx. 01. Randomly sample 100 people. Sample **proportion** has mean 0.l, standard deviation $\sqrt{\frac{0.1(1-0.1)}{100}} = 0.03$
- **Question:** How do we interpret these?
- **Response:** On average, expect

  sample proportion = ____ lefties.

  Proportions vary; typical distance from 0.1 is ____

 Practice: 7.29g-h p.308

Example: *Role of Sample Size in Spread*

- **Background**: Consider proportion of tails in various sample sizes *n* of coinflips.
- **Questions:** What is the standard deviation for
  - ***n*=1? *n*=4? *n*=16?**
- **Responses:**
  - ***n*=1:** s.d.=
  - ***n*=4:** s.d.=
  - ***n*=16:** s.d.=

***A Closer Look:** Due to n in the denominator of formula for standard deviation, spread of sample proportion____________ as n increases.*

 Practice: 7.34b p.309

Shape of Distribution of Count, Proportion

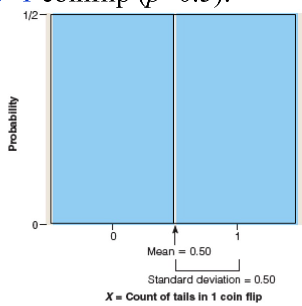
Binomial count X or proportion $\hat{p} = \frac{X}{n}$ for repeated random samples has shape approximately normal if samples are large enough to offset underlying skewness.

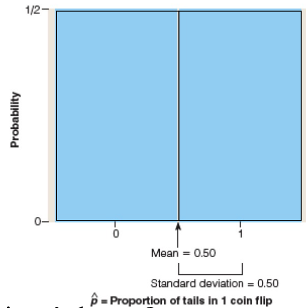
(Central Limit Theorem)

For a given sample size n, shapes are identical for count and proportion.

Example: *Underlying Coinflip Distribution*

- **Background**: Distribution of count or proportion of tails in n=1 coinflip (p=0.5):



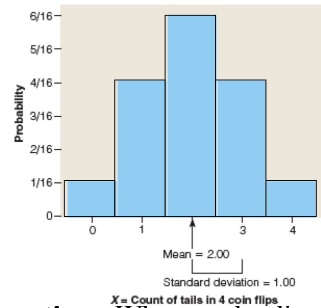


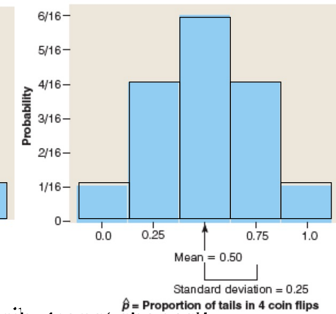
- **Question:** What are the distributions' shapes?
- **Response:**

 Practice: 7.32a p.308

Example: *Distribution for 4 Coinflips*

- **Background**: Distribution of count or proportion of tails in n=4 coinflips (p=0.5):





- **Question:** What are the distributions' shapes?
- **Response:**

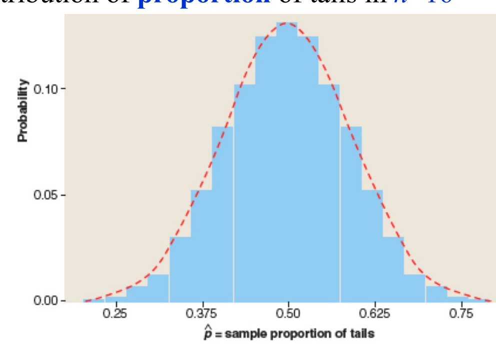
 Practice: 7.32b p.308

Shift from Counts to Proportions

- Binomial Theory begins with counts
- Inference will be about proportions

Example: *Distribution of $\hat{p}$ for 16 Coinflips*

- **Background**: Distribution of **proportion** of tails in n=16 coinflips (p=0.5):

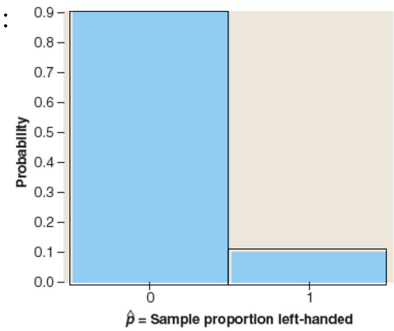


- **Question:** What is the shape?
- **Response:**

 Practice: 7.34c p.309

Example: *Underlying Distribution of Lefties*

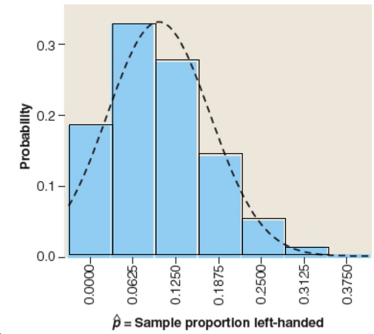
- **Background**: Distribution of **proportion** of lefties (p=0.1) for samples of n=1:



- **Question:** What is the shape?
- **Response:**

Example: *Dist of* $\hat{p}$ *of Lefties for n= 16*

- **Background**: Distribution of proportion of lefties (p=0.1) for n=16:

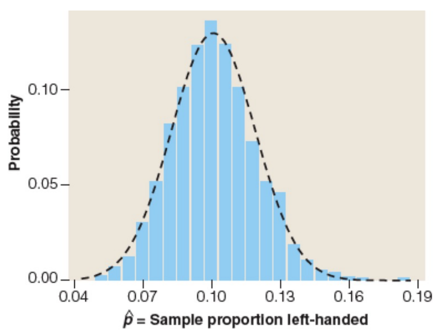


- **Question:** What is the shape?
- **Response:**

 Practice: 7.32c p.309

Example: *Dist of* $\hat{p}$ *of Lefties for n=100*

- **Background**: Distribution of proportion of lefties (p=0.1) for n=100:



- **Question:** What is the shape?
- **Response:**

 Practice: 7.35a,b p.309

Rule of Thumb: Sample Proportion Approximately Normal

Distribution of $\hat{p}$ is approximately normal if sample size n is large enough relative to shape, determined by population proportion p.

Require $np \geq 10$ and $n(1-p) \geq 10$

Together, these require us to have larger n for p close to 0 or 1 (underlying distribution skewed right or left).

Example: *Applying Rule of Thumb*

- **Background**: Consider distribution of sample proportion for various n and p:

n=4, p=0.5; n=20, p=0.5; n=20, p=0.1; n=20, p=0.9; n=100, p=0.

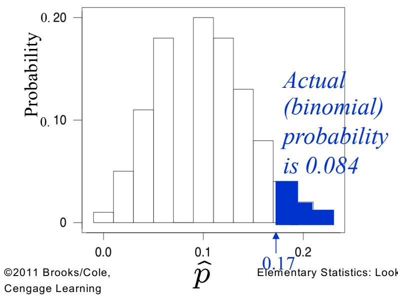
- **Question:** Is shape approximately normal?
- **Response:** Normal?
  - n=4, p=0.5 ________ [np=4(0.5)=2<10]
  - n=20, p=0.5 ________ [np=20(0.5)=10=n(1-p)]
  - n=20, p=0.1 No [______________________]
  - n=20, p=0.9 No [______________________]
  - n=100, p=0.1 ________

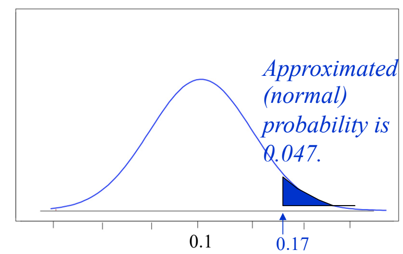
  [np=100(0.1)=10, n(1-p)=100(0.9)=90 both ≥ 10]

 Practice: 7.38 p.310

Example: *Solving the Left-handed Problem*

- **Background**: The proportion of Americans who are lefties is 0.1. Consider P($\hat{p}$ ≥8/46=0.17) for a sample of 46 presidents.
- **Question:** Can we use a normal approximation to find the probability that at least 8 of 46 (0.17) are left-handed?
- **Response:**





 Practice: 7.35b p.309

Example: *From Count to Proportion and Vice Versa*

- **Background**: Consider these reports:
  - In a sample of 87 assaults on police, 23 used weapons.
  - 0.44 in sample of 25 bankruptcies were due to med. bills
- **Question:** In each case, what are n, X, and $\hat{p}$?
- **Response:**
  - First has n = ____, X = ____, $\hat{p}$ = _____
  - Second has n = ____, $\hat{p}$ = _____, X = __________

 Practice: 7.29a-b p.308

Looking Back: *Review*

- **4 Stages of Statistics**
  - Data Production (discussed in Lectures 1-3)
  - Displaying and Summarizing (Lectures 3-8)
  - Probability
    - Finding Probabilities (discussed in Lectures 9-10)
    - Random Variables (introduced in Lecture 10)
      - Binomial (just discussed in Lecture 11)
      - Normal
    - Sampling Distributions
  - Statistical Inference

Role of Normal Distribution in Inference

- **Goal:** Perform inference about unknown population proportion, based on sample proportion
- **Strategy:** Determine behavior of sample proportion in random samples with known population proportion
- **Key Result:** Sample proportion follows normal curve for large enough samples.

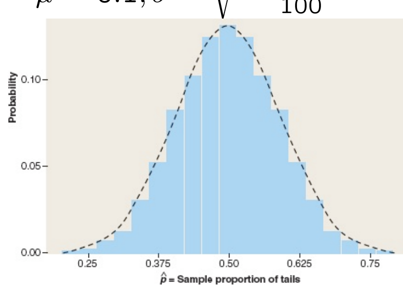
***Looking Ahead:** Similar approach will be taken with means.*

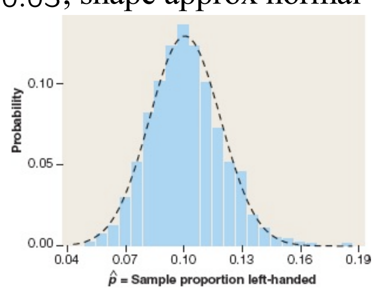
Discrete vs. Continuous Distributions

- **Binomial Count X**
  - **discrete** (distinct possible values like numbers 1, 2, 3, …)
- **Sample Proportion** $\hat{p} = \frac{X}{n}$
  - **also discrete** (distinct values like count)
- **Normal Approx. to Sample Proportion**
  - **continuous** (follows normal curve)
  - Mean p, standard deviation $\sqrt{\frac{p(1-p)}{n}}$

Sample Proportions Approx. Normal *(Review)*

- Proportion of tails in n=16 coinflips (p=0.5) has $\mu = 0.5, \sigma = \sqrt{\frac{0.5(1-0.5)}{16}} = 0.125$, shape approx normal
- Proportion of lefties (p=0.1) in n=100 people has $\mu = 0.1, \sigma = \sqrt{\frac{0.1(1-0.1)}{100}} = 0.03$, shape approx normal





Example: *Variable Types*

- **Background**: Variables in survey excerpt:

age	breakfast?	comp	credits	…
19.67	no	120	15	
20.08	no	120	16	
19.08	yes	40	14	
…	…	…	…	

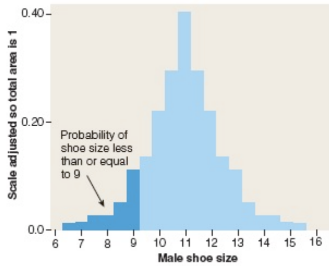
- **Question:** Identify type (cat, discrete quan, continuous quan)
  - Age? Breakfast? Comp (daily min. on computer)? Credits?
- **Response:**
  - Age:
  - Breakfast:
  - Comp (daily time in min. on computer):
  - Credits:

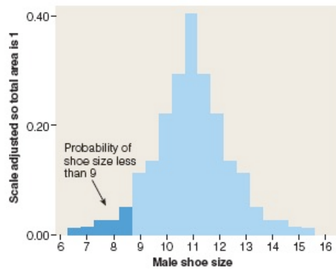
 Practice: 7.43 p.330

Probability Histogram for Discrete R.V.

Histogram for male shoe size X represents probability by area of bars

- $P(X \leq 9)$ (on left)
- $P(X < 9)$ (on right)



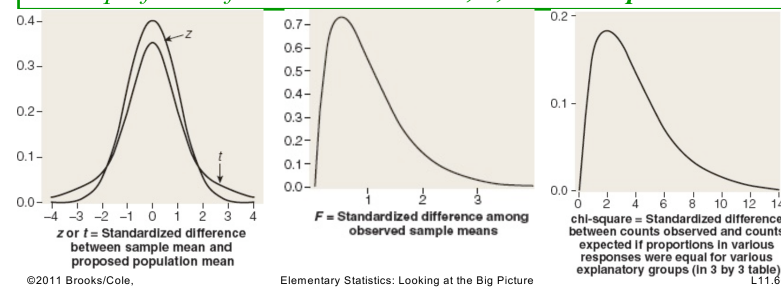


For discrete R.V., strict inequality or not matters.

Definition

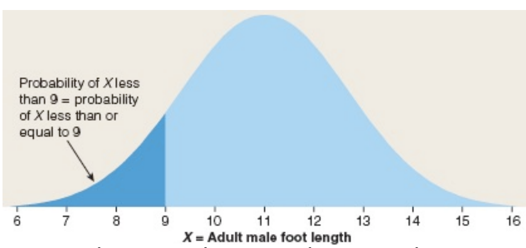
Density curve: smooth curve showing prob. dist. of continuous R.V. Area under curve shows prob. that R.V. takes value in given interval.

Looking Ahead: *Most commonly used density curve is* ***normal z*** *but to perform inference we also use* ***t, F, and chi-square*** *curves.*



Density Curve for Continuous R.V.

Density curve for male foot length X represents probability by area under curve.



$P(X \leq 9) = P(X < 9)$

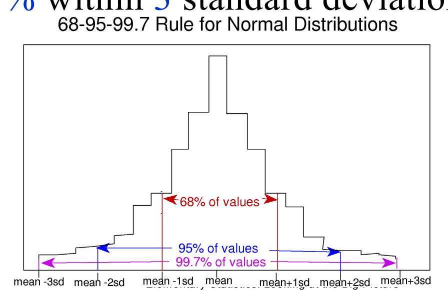
Continuous RV: strict inequality or not doesn't matter.

A Closer Look: *Shoe sizes are discrete; foot lengths are continuous.*

68-95-99.7 Rule for Normal Data *(Review)*

Values of a normal **data set** have

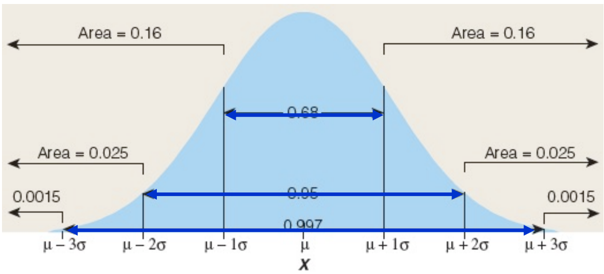
- 68% within 1 standard deviation of mean
- 95% within 2 standard deviations of mean
- 99.7% within 3 standard deviations of mean



68-95-99.7 Rule: Normal Random Variable

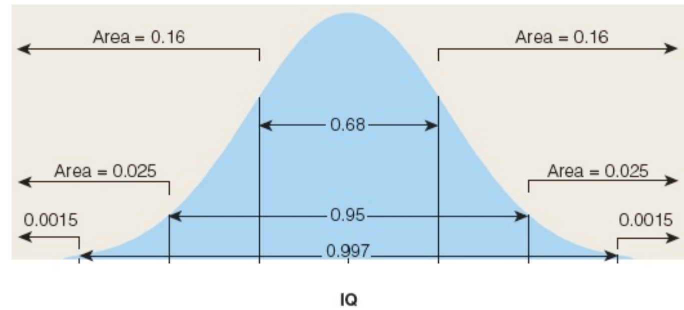
Sample at **random** from normal **population**; for sampled value X (a R.V.), probability is

- 68% that X is within 1 standard deviation of mean
- 95% that X is within 2 standard deviations of mean
- 99.7% that X is within 3 standard deviations of mean



Example: *68-95-99.7 Rule for Normal R.V.*

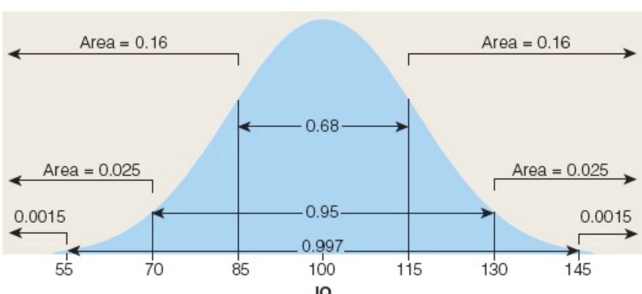
- **Background**: IQ for randomly chosen adult is normal R.V. X with $\mu = 100,\ \sigma = 15$
- **Question:** What does Rule tell us about distribution of X?
- **Response:** We can sketch distribution of X:



 Practice: 7.47 p.331

Example: *Finding Probabilities with Rule*

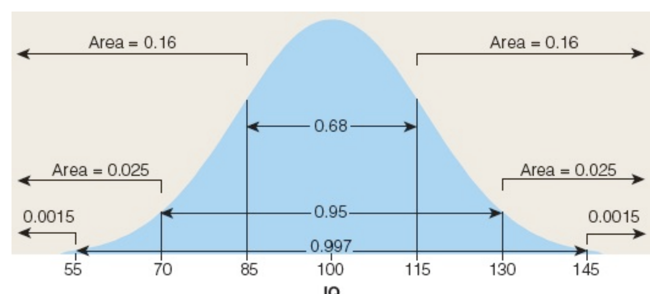
- **Background**: IQ for randomly chosen adult is normal R.V. X with $\mu = 100,\ \sigma = 15$
- **Question:** Prob. of IQ between 70 and 130 = ?
- **Response:**



 Practice: 7.48a p.331

Example: *Finding Probabilities with Rule*

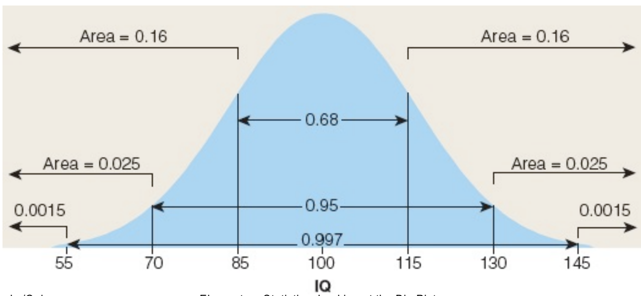
- **Background**: IQ for randomly chosen adult is normal R.V. X with $\mu = 100,\ \sigma = 15$
- **Question:** Prob. of IQ less than 70 = ?
- **Response:**



 Practice: 7.49a p.331

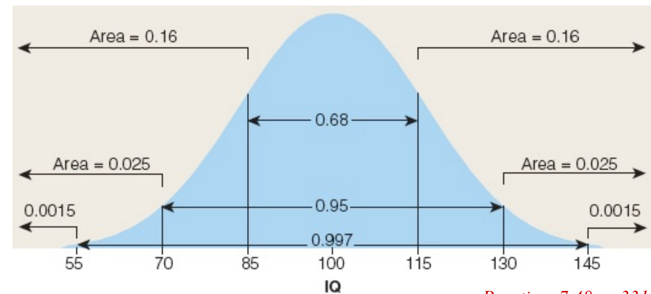
Example: *Finding Probabilities with Rule*

- **Background**: IQ for randomly chosen adult is normal R.V. X with $\mu = 100, \ \sigma = 15$
- **Question:** Prob. of IQ less than 100 = ?
- **Response:**



Example: *Finding Values of X with Rule*

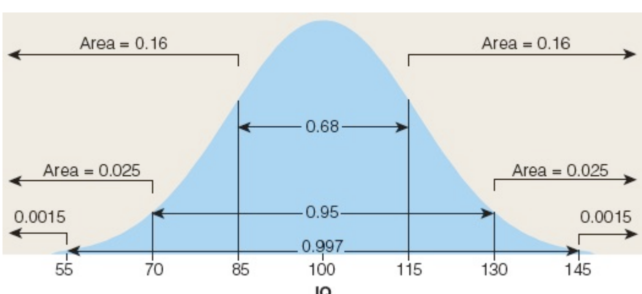
- **Background**: IQ for randomly chosen adult is normal R.V. X with $\mu = 100, \ \sigma = 15$
- **Question:** Prob. is 0.997 that IQ is between…?
- **Response:**



 Practice: 7.48c p.331

Example: *Finding Values of X with Rule*

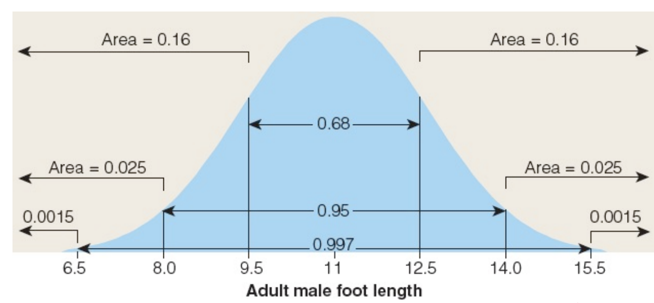
- **Background**: IQ for randomly chosen adult is normal R.V. X with $\mu = 100, \ \sigma = 15$
- **Question:** Prob. is 0.025 that IQ is above…?
- **Response:**



 Practice: 7.48d p.331

Example: *Using Rule to Evaluate Probabilities*

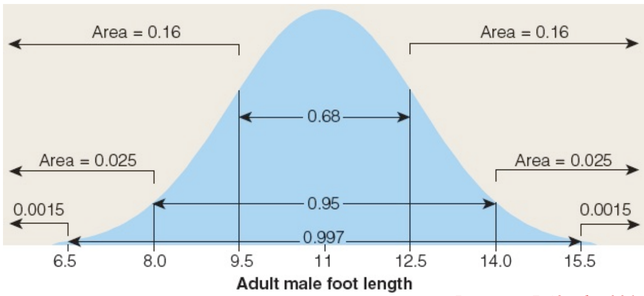
- **Background**: Foot length of randomly chosen adult male is normal R.V. X with $\mu = 11, \ \sigma = 1.5$(in.)
- **Question:** How unusual is foot less than 6.5 inches?
- **Response:**



 Practice: 7.49b p.331

Example: *Using Rule to Estimate Probabilities*

- **Background**: Foot length of randomly chosen adult male is normal R.V. X with $\mu = 11,\ \sigma = 1.5$ (in.)
- **Question:** How unusual is foot more than 13 inches?
- **Response:**



 Practice: 7.49c-d p.331

Lecture Summary
(Binomial Random Variables)

- Definition; 4 requirements for binomial
- R.V.s that do or don't conform to requirements
- Relaxing requirement of independence
- Binomial counts, proportions
  - Mean
  - Standard deviation
  - Shape
- Normal approximation to binomial

Lecture Summary
(Normal Random Variables)

- Relevance of normal distribution
- Continuous random variables; density curves
- 68-95-99.7 Rule for normal R.V.s