LECTURE 4: RANDOM VARIABLES

ECON 480 - ECONOMETRICS - FALL 2018

Ryan Safner

September 10, 2018



Random Variables

Expected Value and Variance

Normal Distribution



EXPERIMENTS

 An experiment is any procedure that can (in principle) be repeated infinitely and has a well-defined set of outcomes



EXPERIMENTS

 An experiment is any procedure that can (in principle) be repeated infinitely and has a well-defined set of outcomes

Example Flip a coin 10 times





• A random variable (RV) takes on values that are unknown in advance, but determined by an experiment



- A random variable (RV) takes on values that are unknown in advance, but determined by an experiment
- \cdot A R.V. is a numerical summary of a random outcome



- A random variable (RV) takes on values that are unknown in advance, but determined by an experiment
- · A R.V. is a numerical summary of a random outcome

Example

The number of Heads from 10 coin flips





RANDOM VARIABLES: NOTATION

- Random variable X takes on individual values (X_i) from a set of possible values



RANDOM VARIABLES: NOTATION

- Random variable X takes on individual values (X_i) from a set of possible values
- · Often capital letters to denote RV's



RANDOM VARIABLES: NOTATION

- Random variable X takes on individual values (X_i) from a set of possible values
- $\cdot\,$ Often capital letters to denote RV's

Example

Let X be the number of Heads from 10 coin flips. $x_i \in \{0, 1, 2, ..., 10\}$



· A discrete random variable takes on a finite/countable set of possible values

Example

Let X be the number of times your computer crashes this semester, $x_i \in \{0, 1, 2, 3, 4\}$





DISCRETE RANDOM VARIABLES: PROBABILITY DISTRIBUTION

• The probability distribution of a R.V. fully lists all the possible values of X and their associated probabilities

Example

Xi	$P(X=x_i)$
0	0.80
1	0.10
2	0.06
3	0.03
4	0.01



• The probability distribution function (pdf) summarizes the possible outcomes of *X* and their probabilities

$$f_X = p_i, \quad i = 1, 2, ..., k$$



- The probability distribution function (pdf) summarizes the possible outcomes of *X* and their probabilities
- Notation: f_X is the pdf of X:

$$f_X = p_i, \quad i = 1, 2, ..., k$$



- The probability distribution function (pdf) summarizes the possible outcomes of *X* and their probabilities
- Notation: f_X is the pdf of X:

$$f_X = p_i, \quad i = 1, 2, ..., k$$

· For any real number x_i , $f(x_i)$ is the probability that $X = x_i$



DISCRETE RANDOM VARIABLES: PDF EXAMPLE

Example

Xi	$P(X=x_i)$
0	0.80
1	0.10
2	0.06
3	0.03
4	0.01



DISCRETE RANDOM VARIABLES: PDF EXAMPLE

Example

Xi	$P(X=x_i)$
0	0.80
1	0.10
2	0.06
3	0.03
4	0.01

• What is f(0)?



DISCRETE RANDOM VARIABLES: PDF EXAMPLE

Example

Xi	$P(X=x_i)$
0	0.80
1	0.10
2	0.06
3	0.03
4	0.01

- What is f(0)?What is f(3)?



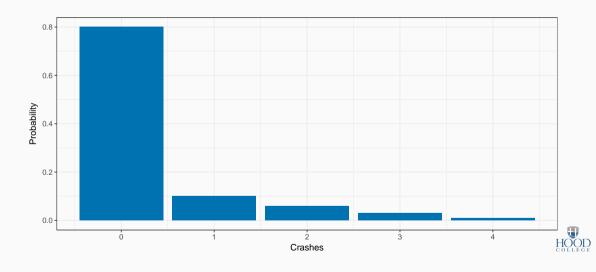
DISCRETE RANDOM VARIABLES: PDF GRAPH

```
crashes<-c(0,1,2,3,4)
prob<-c(0.8,0.1,0.06,0.03,0.01)
crashes<-data.frame(crashes, prob)

crashes.pdf<-ggplot(crashes, aes(x=crashes, y=prob))+
   geom_bar(stat="identity", fill="#0072B2")+xlab("Crashes")+ylab("Probability")</pre>
```



DISCRETE RANDOM VARIABLES: PDF GRAPH II



• The *cumulative* density function (cdf) lists probability X will be at most (less than or equal to) a given value x_i



- The *cumulative* density function (cdf) lists probability X will be at most (less than or equal to) a given value x_i
- Notation: $F_X = P(X \le x_i)$



- The *cumulative* density function (cdf) lists probability X will be at most (less than or equal to) a given value x_i
- Notation: $F_X = P(X \le x_i)$

Example

Xi	f(X)	F(X)
0	0.80	0.80
1	0.10	0.90
2	0.06	0.96
3	0.03	0.99
4	0.01	1.00



- The *cumulative* density function (cdf) lists probability X will be at most (less than or equal to) a given value x_i
- Notation: $F_X = P(X \le x_i)$

Example

Xi	f(X)	F(X)
0	0.80	0.80
1	0.10	0.90
2	0.06	0.96
3	0.03	0.99
4	0.01	1.00



- The *cumulative* density function (cdf) lists probability X will be at most (less than or equal to) a given value x_i
- · Notation: $F_X = P(X \le x_i)$

Example

Xį	f(X)	F(X)
0	0.80	0.80
1	0.10	0.90
2	0.06	0.96
3	0.03	0.99
4	0.01	1.00

• What is the probability your computer will crash at most once, F(1)? three times, F(3)?





3 0.03 0.99 4 0.01 1.00

4

5

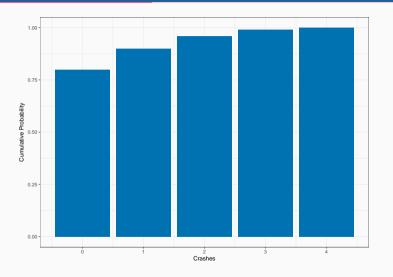
```
crashes$cprob<-(cumsum(crashes$prob))

crashes #show table

## crashes prob cprob
## 1     0 0.80     0.80
## 2     1 0.10     0.90
## 3     2 0.06     0.96</pre>
```

```
crashes.cdf<-ggplot(crashes, aes(x=crashes, y=cprob))+
  geom_bar(stat="identity", fill="#0072B2")+xlab("Crashes")+
  ylab("Cumulative Probability")</pre>
```

DISCRETE RANDOM VARIABLES: CDF GRAPH II





EXPECTED VALUE AND VARIANCE





$$E(X) = p_1 x_1 + p_2 x_2 + \dots + p_k x_k$$



$$E(X) = p_1 x_1 + p_2 x_2 + \dots + p_k x_k = \sum_{i=1}^k p_i x_i$$

- · A probability-weighted average of X, with each x_i weighted by its associated probability p_i
- Also called the mean or expectation of X, also denoted μ_{X}



EXPECTED VALUE: EXAMPLE

Example

Suppose you lend your friend \$100 at 10% interest. If the loan is repaid, you receive \$110 . Your friend is 99% likely to repay, but there is a default risk of 1% where you get nothing. What is the expected value of repayment?



EXPECTED VALUE: EXAMPLE II

Example

Let X be a random variable that is described by the following pdf:

Xi	$P(X=x_i)$
1	0.50
2	0.25
3	0.15
4	0.10

• What is E(X)?



EXPECTED VALUE: EXAMPLE III

```
x<-c(1,2,3,4) #x values
p.x<-c(0.50,0.25,0.15,0.10) #probabilities
rv<-data.frame(x,p.x) #combine into data frame
sum(rv$x*rv$p.x) #sum of each value times its probability</pre>
```

[1] 1.85



VARIANCE OF A RANDOM VARIABLE

• The variance of a random variable X, denoted var(X) or σ_X^2 is:

$$\sigma_X^2 = \sum_{i=1}^n (x_i - \mu_X)^2 p_i$$



VARIANCE OF A RANDOM VARIABLE

• The variance of a random variable X, denoted var(X) or σ_X^2 is:

$$\sigma_X^2 = \sum_{i=1}^n (x_i - \mu_X)^2 p_i$$



VARIANCE OF A RANDOM VARIABLE

• The variance of a random variable X, denoted var(X) or σ_X^2 is:

$$\sigma_X^2 = \sum_{i=1}^n (x_i - \mu_X)^2 p_i = E[(X - \mu_X)^2]$$

- This is the expected value of the squared deviations from the mean
 - $\boldsymbol{\cdot}\,$ i.e. the probability-weighted average of the squared deviations



STANDARD DEVIATION OF A RANDOM VARIABLE

• The standard deviation of a random variable X, denoted sd(X) or σ_X is:

$$\sigma_{X} = \sqrt{\sigma_{X}^{2}}$$



STANDARD DEVIATION: EXAMPLE

Example What is the standard deviation of computer crashes?

	Number of Computer Crashes				
Xi	0	1	2	3	4
$P(X=x_i)$	0.80	0.10	0.06	0.03	0.01



STANDARD DEVIATION: EXAMPLE

Example

What is the standard deviation of computer crashes?

	Number of Computer Crashes				
Xi	0	1	2	3	4
$P(X=x_i)$	0.80	0.10	0.06	0.03	0.01

• First, calculate the mean:

$$E(X) = 0(0.80) + 1(0.10) + 2(0.06) + 3(0.03) + 4(0.01)$$
$$= 0 + 0.1 + 0.12 + 0.09 + 0.04$$
$$= 0.35$$



STANDARD DEVIATION: EXAMPLE II

Example What is the standard deviation of computer crashes?



STANDARD DEVIATION: EXAMPLE II

Example

What is the standard deviation of computer crashes?

• Next, find the deviations from the mean and square them:

	Number of Crashes ($\mu=$ 0.35)				
Xi	0	1	2	3	4
$P(X=x_i)$	0.80	0.10	0.06	0.03	0.01
$(x_i - \mu)$	-0.35	0.65	1.65	2.65	3.65
$(x_i\mu)^2$	0.1225	0.4225	2.7225	7.0225	13.3225



Example

What is the standard deviation of computer crashes?

· Next, find the deviations from the mean and square them:

	Number of Crashes ($\mu=0.35$)				
Xi	0	1	2	3	4
$P(X=x_i)$	0.80	0.10	0.06	0.03	0.01
$(x_i - \mu)$	-0.35	0.65	1.65	2.65	3.65
$(x_i\mu)^2$	0.1225	0.4225	2.7225	7.0225	13.3225

· Take the expectation of the squared deviations to get variance

$$\begin{split} \sigma_{\chi}^2 &= 0.1225(0.80) + 0.4225(0.10) + 2.7225(0.06) + 7.0225(0.03) + 13.3225(0.01) \\ &= 0.098 + 0.04225 + 0.16335 + 0.210675 + 0.133225 \\ &= 0.6475 \end{split}$$



Example

What is the standard deviation of computer crashes?

· To get standard deviation, σ , take the square root:

$$\sigma = \sqrt{\sigma^2}$$

$$= \sqrt{0.6475}$$

$$= 0.8047$$



STANDARD DEVIATION: EXAMPLE IV

Example For the random variable we saw before, what is it's standard deviation?

1 0.50 2 0.25	(_i	$P(X=x_i)$
2 0.25	1	0.50
	2	0.25
3 0.15	3	0.15
4 0.10	4	0.10



STANDARD DEVIAIONT: EXAMPLE V

```
rv$deviations<-(rv$x-1.85) #take deviations
rv$sqdeviations<-(rv$deviations^2) #square deviations</pre>
```

rv #show table

sum(rv\$sqdeviations*rv\$p.x) #add squared deviations times probabilities



STANDARD DEVIATION: EXAMPLE VI

sqrt(sum(rv\$sqdeviations*rv\$p.x)) #take sqaure root of variance for sd

[1] 1.013657



· Continuous random variables can take on an uncountable (infinite) number of values





- · Continuous random variables can take on an uncountable (infinite) number of values
- \cdot So many values that the probability of any specific value is infinitely small o 0.



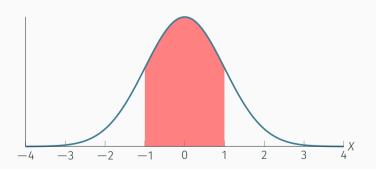


- · Continuous random variables can take on an uncountable (infinite) number of values
- \cdot So many values that the probability of any specific value is infinitely small o 0.
- · Instead, we focus on a range of values it might take on





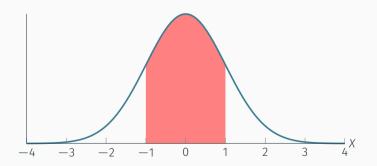
• The probability *density* function (pdf) of a continuous variable represents the probability between two values as the area under a curve



HOOD

 $P(-1 \le X \le 1)$: area under the curve between -1 and 1.

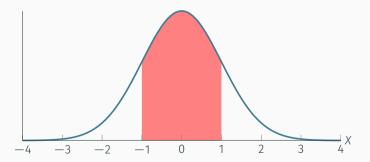
- The probability *density* function (pdf) of a continuous variable represents the probability between two values as the area under a curve
- · The total area under the curve is 1



HOOD

 $P(-1 \le X \le 1)$: area under the curve between -1 and 1.

- The probability *density* function (pdf) of a continuous variable represents the probability between two values as the area under a curve
- · The total area under the curve is 1
- Since P(a) = 0 and P(b) = 0, $P(a < X < b) = P(a \le X \le b)$

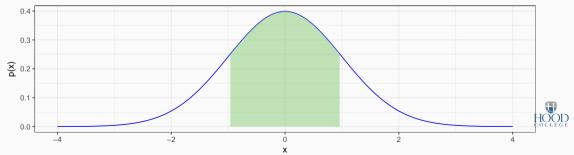




 $P(-1 \le X \le 1)$: area under the curve between -1 and 1.

FYI using calculus:

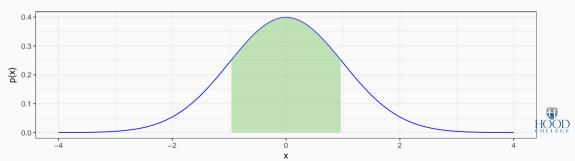
$$P(a \le X \le b) = \int_a^b f(x) dx$$



FYI using calculus:

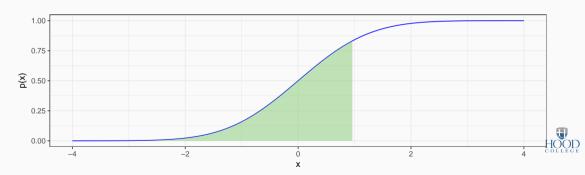
$$P(a \le X \le b) = \int_a^b f(x) dx$$

· Complicated: software or (old fashioned!) probability tables to calculate



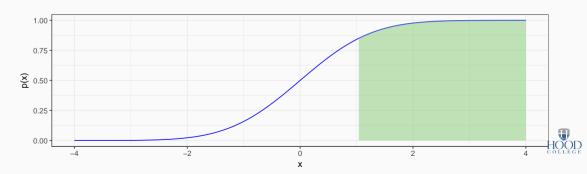
• The *cumulative* density function (cdf) describes the area under the pdf for all values less than or equal to (i.e. to the left of) a given value, *k*

$$P(X \le k)$$



• Note: to find the probability of values *greater* than or equal to (to the right of) a given value k:

$$P(X \ge k) = 1 - P(X \le k)$$

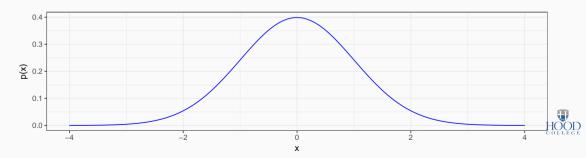




THE NORMAL DISTRIBUTION

• The Gaussian or normal distribution is the most useful type of probability distribution

$$X \sim N(\mu, \sigma)$$

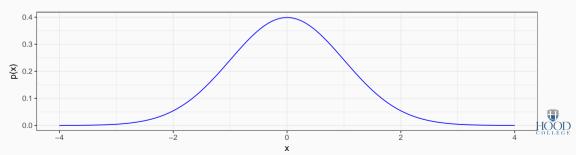


THE NORMAL DISTRIBUTION

• The Gaussian or normal distribution is the most useful type of probability distribution

$$X \sim N(\mu, \sigma)$$

- Continuous, symmetric, unimodal, with mean μ and standard deviation σ



THE NORMAL PDF

. The pdf of ${\it X} \sim {\it N}(\mu,\sigma)$ is

$$P(X = k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \left(\frac{(k-\mu)}{\sigma}\right)^2}$$



THE NORMAL PDF

• The pdf of $X \sim N(\mu, \sigma)$ is

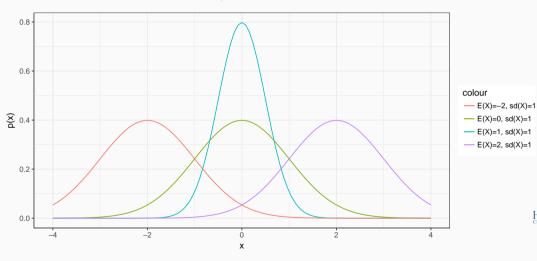
$$P(X=k) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{(k-\mu)}{\sigma}\right)^2}$$

 \cdot Do not try and learn this, we have software and (previously tables) to calculate pdfs and cdfs



THE EFFECTS OF PARAMETER CHANGES I

 \cdot The pdf moves left/right based on μ

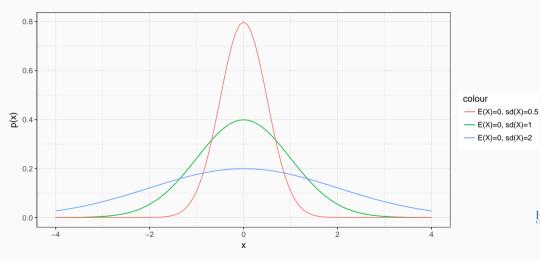




-E(X)=0, sd(X)=1-E(X)=1, sd(X)=1

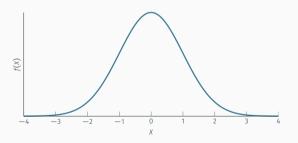
THE EFFECTS OF PARAMETER CHANGES II

 \cdot The pdf gets fatter/skinnier based on σ





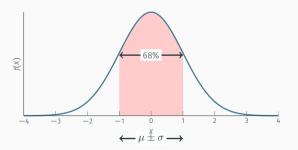
• The 68-95-99.7% (empirical) rule: for a normal distribution:





• The 68-95-99.7% (empirical) rule: for a normal distribution:

•
$$P(\mu - 1\sigma \le X \le \mu + 1\sigma) \approx 68\%$$

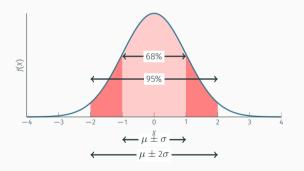




• The 68-95-99.7% (empirical) rule: for a normal distribution:

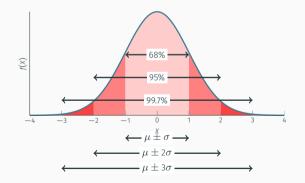
•
$$P(\mu - 1\sigma \le X \le \mu + 1\sigma) \approx 68\%$$

•
$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) \approx 95\%$$





- The 68-95-99.7% (empirical) rule: for a normal distribution:
 - $P(\mu 1\sigma \le X \le \mu + 1\sigma) \approx 68\%$
 - $P(\mu 2\sigma \le X \le \mu + 2\sigma) \approx 95\%$
 - $P(\mu 3\sigma \le X \le \mu + 3\sigma) \approx 99.7\%$





THE STANDARD NORMAL DISTRIBUTION

• We standardize a variable by calculating its Z-score and converting to the standard normal distribution

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



THE STANDARD NORMAL DISTRIBUTION

 We standardize a variable by calculating its Z-score and converting to the standard normal distribution

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

· Z is the number of standard deviations a value is above/below of its mean



THE STANDARD NORMAL DISTRIBUTION

• We standardize a variable by calculating its Z-score and converting to the standard normal distribution

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

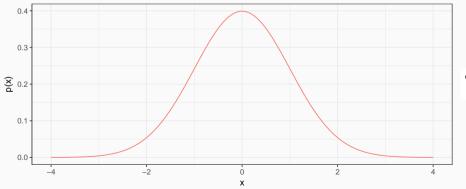
- · Z is the number of standard deviations a value is above/below of its mean
- · Can compare distributions of variables with very different units!



THE STANDARD NORMAL DISTRIBUTION II

• The standard normal distribution has mean 0 and standard deviation 1

$$Z \sim N(0,1)$$



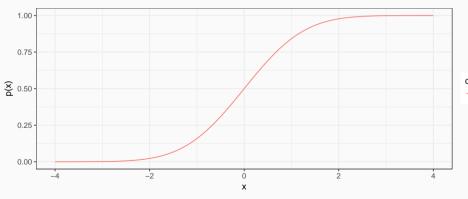
colour
— E(X)=0, sd(X)=1



THE STANDARD NORMAL CDF

· Standard normal cdf:

$$\Phi(\mathbf{k}) = \mathbf{P}(\mathbf{Z} \le \mathbf{k})$$



colour
— E(X)=0, sd(X)=1



Example

On August 8, 2011, the Dow dropped 634.8 points, sending shock waves through the financial community. Assume that during mid-2011 to mid-2012 the daily change for the Dow is normally distributed, with the mean daily change of 1.87 points and a standard deviation of 155.28 points. What is the *Z*-score?



Example

On August 8, 2011, the Dow dropped 634.8 points, sending shock waves through the financial community. Assume that during mid-2011 to mid-2012 the daily change for the Dow is normally distributed, with the mean daily change of 1.87 points and a standard deviation of 155.28 points. What is the *Z*-score?

• Convert to *Z*-score:

$$Z = \frac{X - \mu}{\sigma}$$



Example

On August 8, 2011, the Dow dropped 634.8 points, sending shock waves through the financial community. Assume that during mid-2011 to mid-2012 the daily change for the Dow is normally distributed, with the mean daily change of 1.87 points and a standard deviation of 155.28 points. What is the *Z*-score?

• Convert to *Z*-score:

$$Z = \frac{X - \mu}{\sigma} = \frac{634.8 - 1.87}{155.28}$$



Example

On August 8, 2011, the Dow dropped 634.8 points, sending shock waves through the financial community. Assume that during mid-2011 to mid-2012 the daily change for the Dow is normally distributed, with the mean daily change of 1.87 points and a standard deviation of 155.28 points. What is the *Z*-score?

• Convert to *Z*-score:

$$Z = \frac{X - \mu}{\sigma} = \frac{634.8 - 1.87}{155.28} = -4.1$$

• This is 4.1 standard deviations (σ) beneath the mean.



Example



Example

In the last quarter of 2015, a group of 64 mutual funds had a mean return of 2.4% with a standard deviation of 5.6%. These returns can be approximated by a normal distribution.

• What percent of the funds would you expect to be earning between -3.2% and 8.0% returns?



Standardizing Variables: Example: $X \rightarrow Z$



Standardizing Variables: Example: $X \rightarrow Z$

$$P(-3.2 < X < 8)$$



Standardizing Variables: Example: $X \rightarrow Z$

$$P(-3.2 < X < 8)$$

$$P(\frac{-3.2 - 2.4}{5.6} < \frac{X - 2.4}{5.6} < \frac{8 - 2.4}{5.6})$$



$$P(-3.2 < X < 8)$$

$$P(\frac{-3.2 - 2.4}{5.6} < \frac{X - 2.4}{5.6} < \frac{8 - 2.4}{5.6})$$

$$P(-1 < Z < 1)$$



Standardizing Variables: Example: X o Z

$$P(-3.2 < X < 8)$$

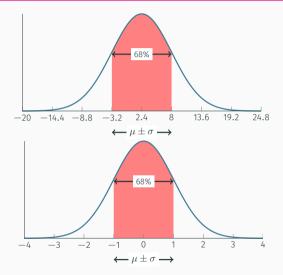
$$P(\frac{-3.2 - 2.4}{5.6} < \frac{X - 2.4}{5.6} < \frac{8 - 2.4}{5.6})$$

$$P(-1 < Z < 1)$$

$$\cdot P(X \pm 1\sigma) = 0.68$$



Standardizing Variables: Example: $X \rightarrow Z$ II





Example

In the last quarter of 2015, a group of 64 mutual funds had a mean return of 2.4% with a standard deviation of 5.6%. These returns can be approximated by a normal distribution.

• What percent of the funds would you expect to be earning between -3.2% and 8.0% returns?



Example

In the last quarter of 2015, a group of 64 mutual funds had a mean return of 2.4% with a standard deviation of 5.6%. These returns can be approximated by a normal distribution.

• What percent of the funds would you expect to be earning between -3.2% and 8.0% returns?



Example

- · What percent of the funds would you expect to be earning between -3.2% and 8.0% returns?
- · What percent of the funds would you expect to be earning 2.4% or less?



Example

- · What percent of the funds would you expect to be earning between -3.2% and 8.0% returns?
- What percent of the funds would you expect to be earning 2.4% or less?
- What percent of the funds would you expect to be earning between -8.8% and 13.6%?



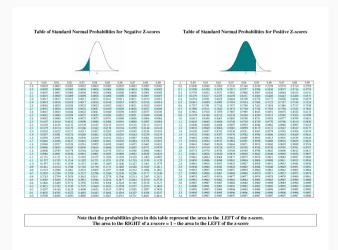
Example

- · What percent of the funds would you expect to be earning between -3.2% and 8.0% returns?
- · What percent of the funds would you expect to be earning 2.4% or less?
- What percent of the funds would you expect to be earning between -8.8% and 13.6%?
- What percent of the funds would you expect to be earning returns greater than 13.6%?



FINDING Z-Score Probabilities

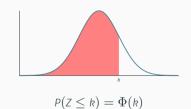
• How do we actually find the probabilities for Z—scores?



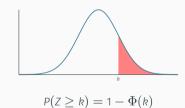


FINDING Z-Score Probabilities II

Probability to Left of k



Probability to the Right of *k*



Probability Between j and k

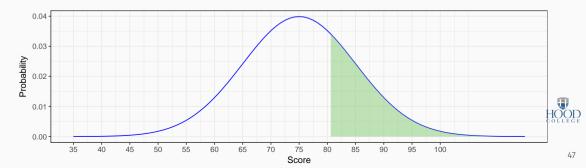


$$P(j \ge Z \ge k) = \Phi(k) - \Phi(j)$$



FINDING Z-Score Probabilities with R

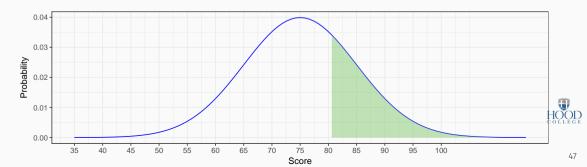
• Let the distribution of grades be normal, with mean 75 and standard deviation 10.



FINDING Z-Score Probabilities with R

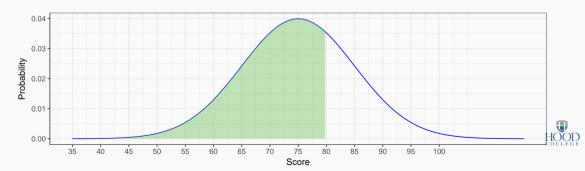
- Let the distribution of grades be normal, with mean 75 and standard deviation 10.
- · Probability a student gets at least an 80:

pnorm(80, mean=75, sd=10, lower.tail=FALSE)



FINDING Z-Score Probabilities with R II

· Probability a student gets at most an 80:



FINDING Z-Score Probabilities with R III

• Probability a student gets between a 65 and 85:

```
pnorm(85, mean=75, sd=10, lower.tail=TRUE)-pnorm(
65, mean=75, sd=10, lower.tail=TRUE)
```

