Continuous Random Variables

Ryan Miller

Introduction

- Previously, we introduced the concept of a **random variable**, or a variable representing the *numeric outcome of a random process*
  - Our discussion focused on *discrete random variables*, or those with countable/integer outcomes

Introduction

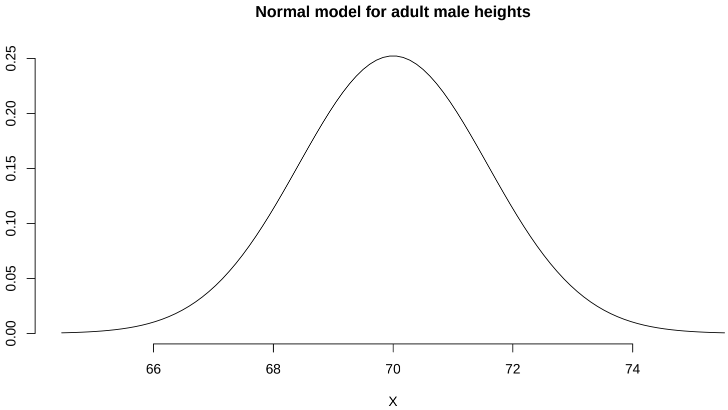
- Previously, we introduced the concept of a **random variable**, or a variable representing the *numeric outcome of a random process*
    - Our discussion focused on *discrete random variables*, or those with countable/integer outcomes
- This presentation will focus on *continuous random variables*, or those with infinitely many real-valued outcomes

Introduction

- Previously, we introduced the concept of a **random variable**, or a variable representing the *numeric outcome of a random process*
  - Our discussion focused on *discrete random variables*, or those with countable/integer outcomes
- This presentation will focus on *continuous random variables*, or those with infinitely many real-valued outcomes
- For example, consider sampling a random adult male and let X be their height (inches)
  - Depending on our measurement tool, X could take on infinitely many values (70.0 inches, 70.01 inches, 70.001 inches, etc.)
  - This means we cannot use a table to express a probability model for X

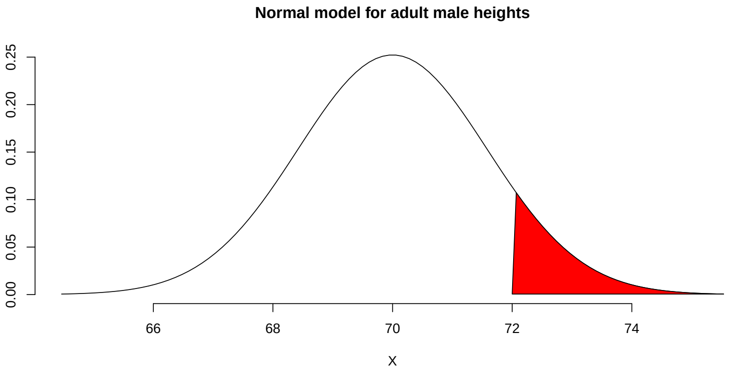
The Normal Model

- Continuous random variables necessitate continuous probability models
  - One of the most widely used is the **Normal model**



The Normal Model (cont)

- Under a continuous probability model, the probability of any single value of X is zero (as there are infinitely many possible values)
  - Thus, probabilities only make sense for intervals, for example we can represent $P(X > 72)$ using the *shaded area* shown below:



The Normal Model (cont)

- The Normal probability model is defined by the curve:

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

- The parameter μ is a constant that defines the *center* of the bell-curve
- The parameter σ is a constant that defines the *spread* of the bell-curve (how tall or flat it appears)

The Normal Model (cont)

- The Normal probability model is defined by the curve:

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

- The parameter μ is a constant that defines the *center* of the bell-curve
- The parameter σ is a constant that defines the *spread* of the bell-curve (how tall or flat it appears)
- Note that there are infinitely many Normal curves, one for each combination of μ and σ
  - We will denote these curves as $N(\mu, \sigma)$, for example $N(70, 2.5)$ might apply to our height example

Standardization

- Historically, it was really inconvenient for statisticians to consider infinitely many different probability models
  - This inspired the use of a transformation known as **standardization** to move their data (regardless of its original units) on a unitless, standardized scale

Standardization

- Historically, it was really inconvenient for statisticians to consider infinitely many different probability models
  - This inspired the use of a transformation known as **standardization** to move their data (regardless of its original units) on a unitless, standardized scale
- A common use of standardization is the calculation of Z-scores
  - Consider a random variable X and a Normal model defined by μ and σ, under this model Z-scores are calculated:

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

Z-scores

- A Z-score can be interpreted as how many standard deviations an observation is above or below its expected value

Z-scores

- A Z-score can be interpreted as how many standard deviations an observation is above or below its expected value
- For example, suppose X is a random variable from a $N(\mu = 70, \sigma = 2.5)$ distribution and we observe $x = 72$
  - The corresponding Z-score is $z = (72 - 70)/2.5 = 0.8$
  - So a height of 72 inches is 0.8 standard deviations above what would expected (under this probability model)

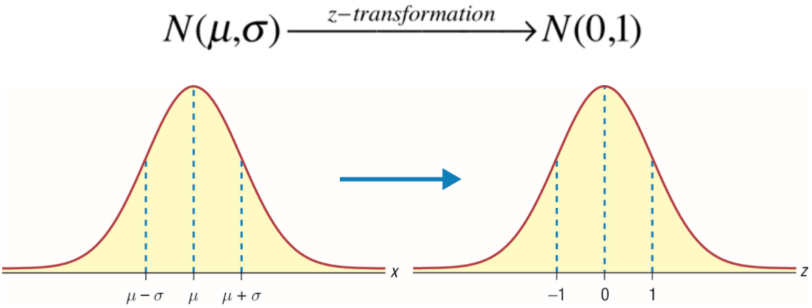
The Standard Normal Distribution

- Standardization allows us to use the **Standard Normal distribution** as a probability model for a wide variety of settings

The Standard Normal Distribution

- Standardization allows us to use the **Standard Normal distribution** as a probability model for a wide variety of settings
- For example, suppose adult male heights follow a Normal distribution centered at 70 inches with a standard deviation of 2.5 inches
  - This means, $X \sim N(70, 2.5)$
  - After standardization, $Z = \frac{X-70}{2.5} \sim N(0, 1)$

The Standard Normal Distribution



Example

Let X denote the height of a randomly chosen adult male, and assume the probability model $X \sim N(70, 2.5)$

1) Determine the probability that this male's height is between 5'10 and 6'0 directly from this model
2) Determine the same probability using *standardization*, and the Standard Normal model

For each of these tasks, we'll utilize a new program: StatKey

Example (solution)

Using Statkey:

1) On the N(70,2.5) curve, the area to the left of 70 inches (5'10) is 0.5, while the area to the left of 72 inches (6'0) is 0.788; thus, there is a 28% probability of a random adult male being between 5'10 and 6'0 under this model
2) To use the Standard Normal model, we'd the same thing, but with the preliminary step of calculating Z-scores. The Z-score for 70 inches is 0, while the Z-score for 72 increases is 0.8. On the Standard Normal curve, the area to the left of 0 is 0.5, while the area to the left of 0.8 is 0.788; again we find a 28% probability that a random adult male is between 5'10 and 6'0 under this model

Conclusion

- In this presentation, we learned about probability models for continuous random variables
  - More specifically, we discussed the **Normal probability model**

Conclusion

- In this presentation, we learned about probability models for continuous random variables
  - More specifically, we discussed the **Normal probability model**
- In general, any function, $f(X)$, that maps a continuous random variable, X, to a curve which integrates to 1 is a valid probability model
  - Thankfully, in this class we will focus on the Normal probability model and avoid dealing with complicated functions

Conclusion

- In this presentation, we learned about probability models for continuous random variables
  - More specifically, we discussed the **Normal probability model**
- In general, any function, $f(X)$, that maps a continuous random variable, X, to a curve which integrates to 1 is a valid probability model
  - Thankfully, in this class we will focus on the Normal probability model and avoid dealing with complicated functions
- The Normal curve is defined by a center, μ, and a standard deviation, σ
  - By **standardizing**, we can transform our variable onto the unitless Z-scale, and thereby use the Standard Normal model regardless of the specific situation

Conclusion

- In this presentation, we learned about probability models for continuous random variables
  - More specifically, we discussed the **Normal probability model**
- In general, any function, $f(X)$, that maps a continuous random variable, X, to a curve which integrates to 1 is a valid probability model
  - Thankfully, in this class we will focus on the Normal probability model and avoid dealing with complicated functions
- The Normal curve is defined by a center, μ, and a standard deviation, σ
  - By **standardizing**, we can transform our variable onto the unitless Z-scale, and thereby use the Standard Normal model regardless of the specific situation
- The next presentation will further explore when the Normal model can be used