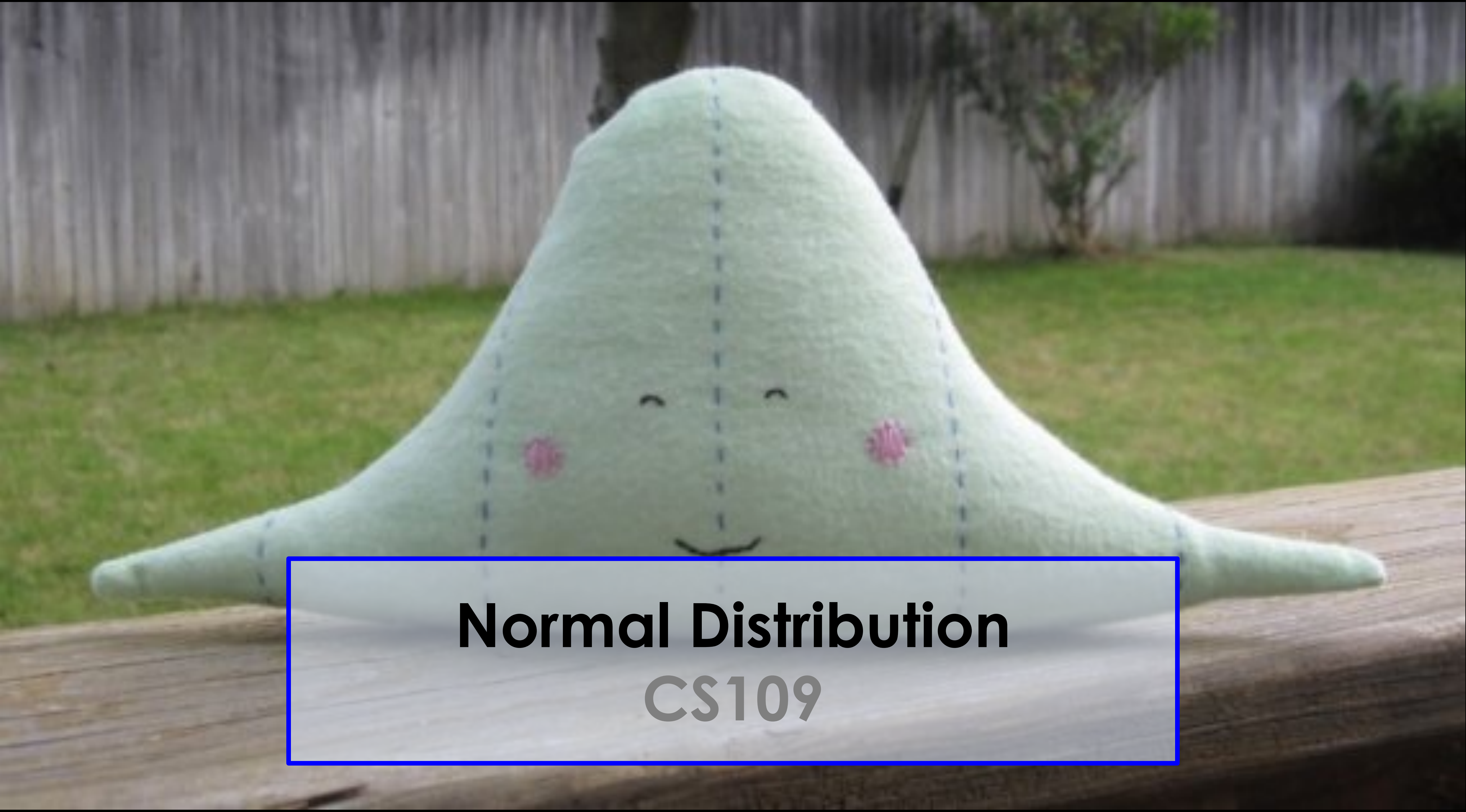


Normal Distribution

CS109

Review

Two Continuous RVs: Uniform and Exponential

Uniform Random Variable

Notation: $X \sim \text{Uni}(\alpha, \beta)$

Description: A continuous random variable that takes on values, with equal likelihood, between α and β

Parameters: $\alpha \in \mathbb{R}$, the minimum value of the variable.
$\beta \in \mathbb{R}, \beta > \alpha$, the maximum value of the variable.

Support: $x \in [\alpha, \beta]$

PDF equation:
$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \text{for } x \in [\alpha, \beta] \\ 0 & \text{else} \end{cases}$$

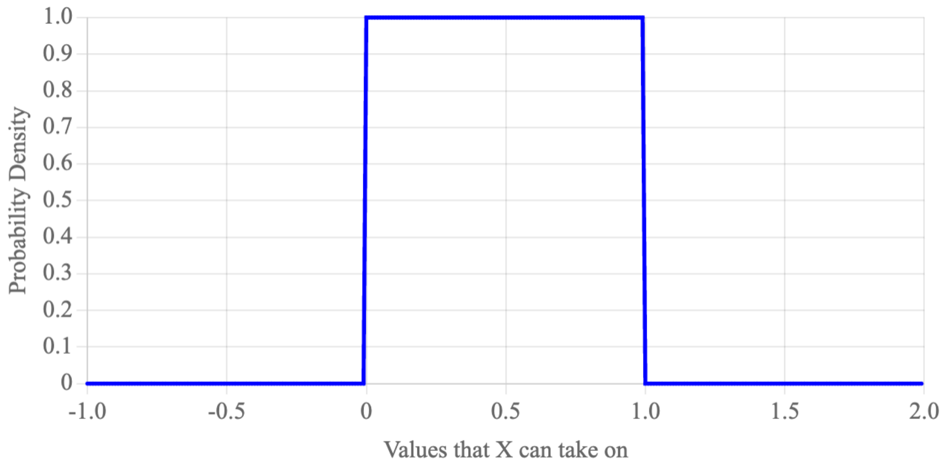
CDF equation:
$$F(x) = \begin{cases} \frac{x-\alpha}{\beta-\alpha} & \text{for } x \in [\alpha, \beta] \\ 0 & \text{for } x < \alpha \\ 1 & \text{for } x > \beta \end{cases}$$

Expectation: $\text{E}[X] = \frac{1}{2}(\alpha + \beta)$

Variance: $\text{Var}(X) = \frac{1}{12}(\beta - \alpha)^2$

PDF graph:

Parameter α: 0 Parameter β: 1



Exponential Random Variable

Notation: $X \sim \text{Exp}(\lambda)$

Description: Time until next events if (a) the events occur with a constant mean rate and (b) they occur independently of time since last event.

Parameters: $\lambda \in \{0, 1, \ldots\}$, the constant average rate.

Support: $x \in \mathbb{R}^+$

PDF equation: $f(x) = \lambda e^{-\lambda x}$

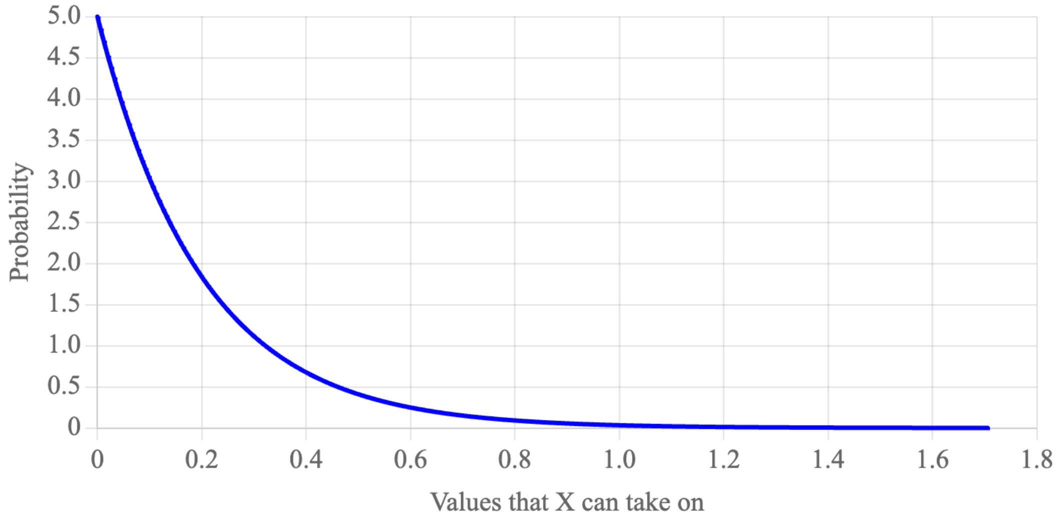
CDF equation: $F(x) = 1 - e^{-\lambda x}$

Expectation: $\text{E}[X] = 1/\lambda$

Variance: $\text{Var}(X) = 1/\lambda^2$

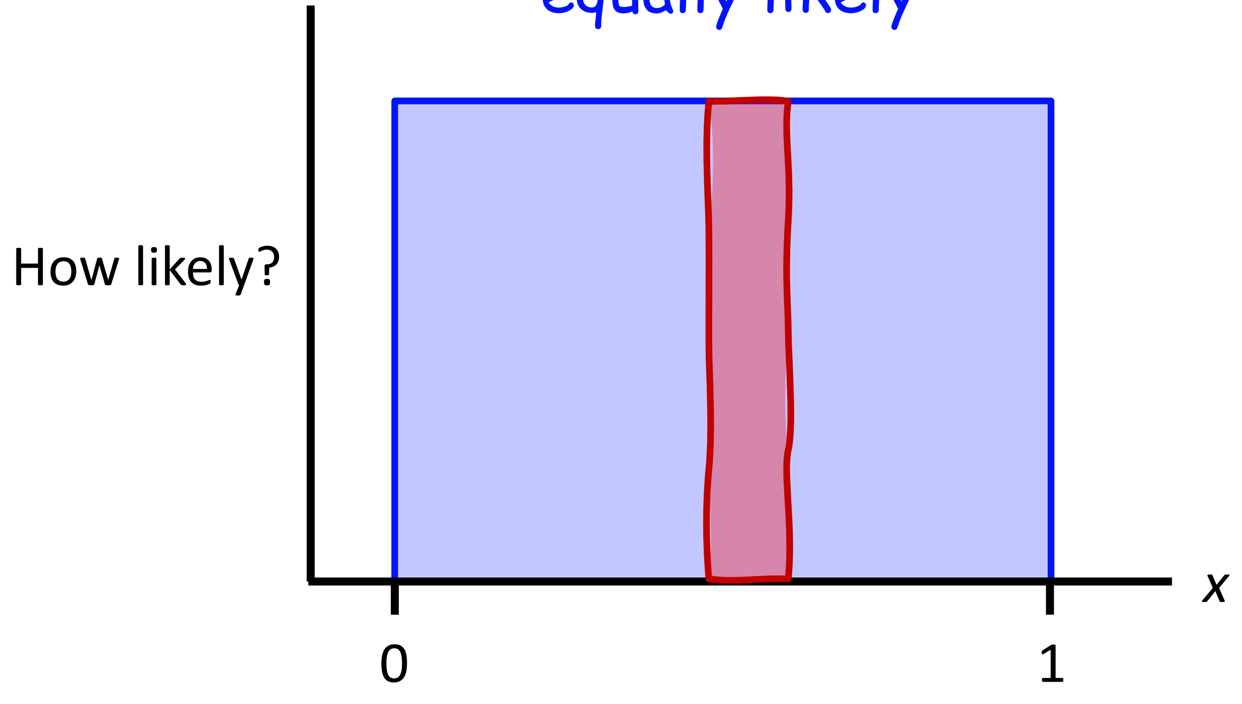
PDF graph:

Parameter λ: 5



$X \sim$ Uniform(0,1): A Continuous Random Variable

All values are equally likely



Possible values are between 0 and 1

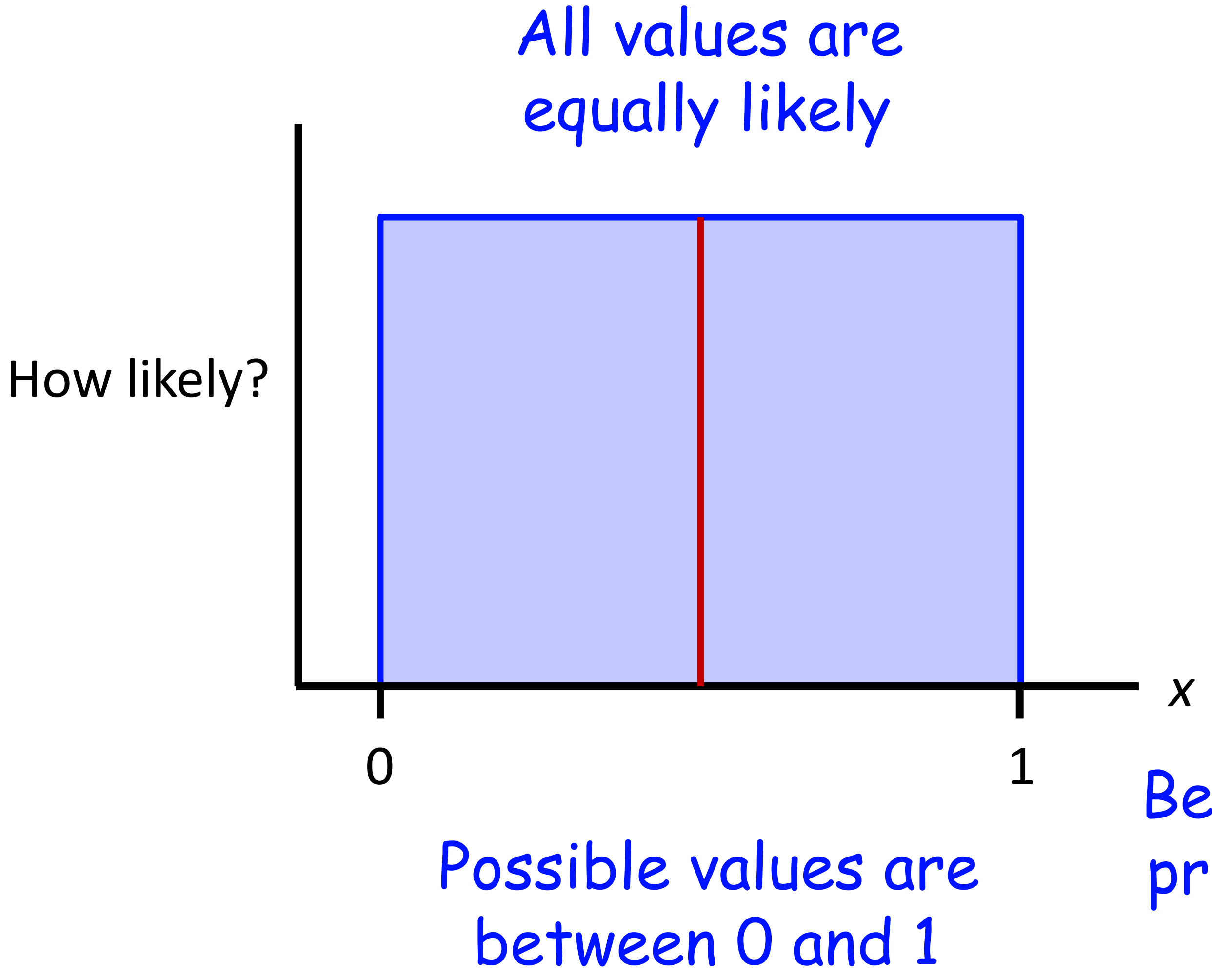
$$P(0 \leq X \leq 1) = 1$$

$$P(0.5 \leq X \leq 1) = 0.5$$

$$P(0.5 \leq X \leq 0.6) = 0.1$$

Finding the probability of a range of values is straightforward!

$X \sim \text{Uniform}(0,1)$: A Continuous Random Variable



$$P(0 \leq X \leq 1) = 1$$

$$P(0.5 \leq X \leq 1) = 0.5$$

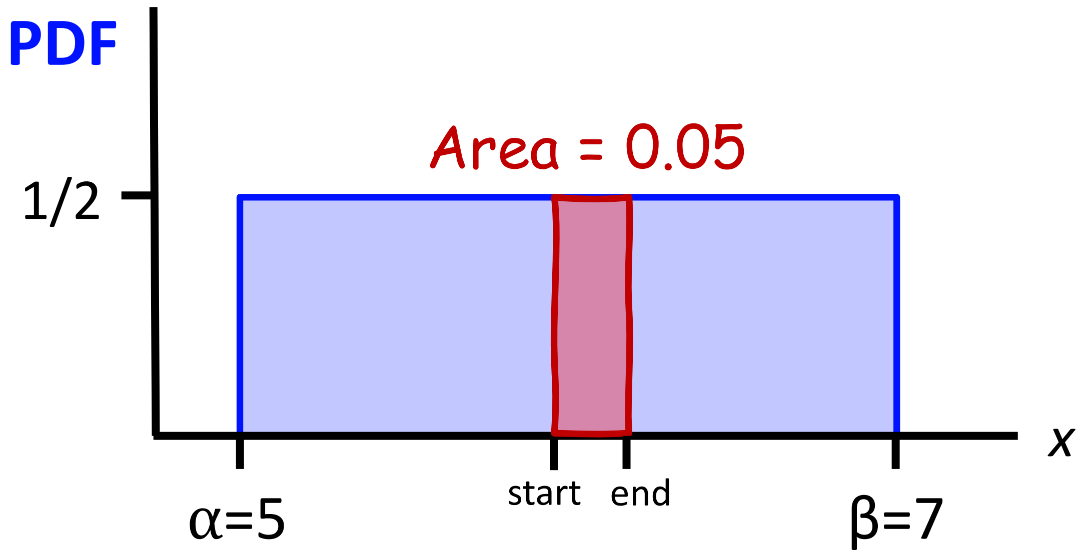
$$P(0.5 \leq X \leq 0.6) = 0.1$$

$$P(X = 0.5) = 0$$

Because of infinitely many outcomes, the probability of any *exact* outcome is zero

No PMFs!

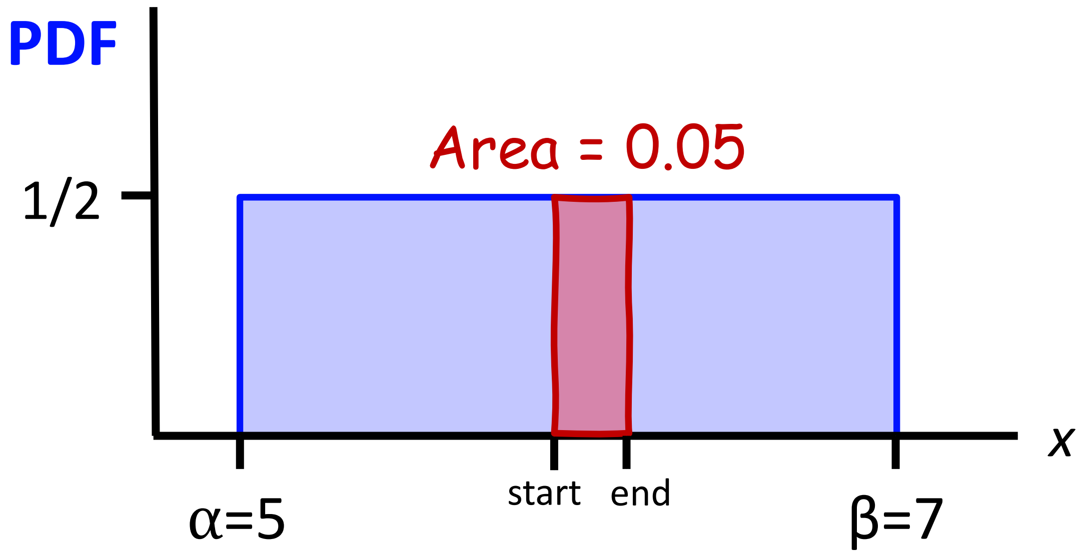
Probability Density Functions



The **probability density function** (PDF) of a continuous random variable represents the relative likelihood of various values.

Units: probability *divided by units of X, or the derivative of the probability of x.*
Integrate it to get probabilities!

Probability Density Functions



The **probability density function** (PDF) of a continuous random variable represents the relative likelihood of various values.

Units: probability *divided by units of X, or the derivative of the probability of x.*
Integrate it to get probabilities!

$$P(a < X < b) = \int_{x=a}^{b} f(X = x)dx$$

PDFs - $f(X = x)$ vs. PMFs - $P(X = x)$

$$P(X = x)$$

"The probability that a **discrete** random variable X takes on the value *x*."

$$f(X = x)$$

"The ***derivative*** of the probability that a **continuous** random variable X takes at the value *x*."

PDFs - $f(X = x)$ vs. PMFs - $P(X = x)$

$$P(X = x)$$

"The probability that a **discrete** random variable X takes on the value *x*."

$$f(X = x)$$

"The ***derivative*** of the probability that a **continuous** random variable X takes at the value *x*."

What do you get if you integrate over a probability *density* function?

A probability!

They are *both* measures of how **likely** *X* is to take on the value *x*.

Cumulative Density Functions

A *cumulative* density function (CDF) is a "closed-form" equation for the probability that a continuous random variable is less than a given value.

$$F(x) = P(X < x)$$

$$P(X < x) = \int_{y=-\infty}^{x} f(y)\, dy$$

Cumulative Density Functions

A *cumulative* density function (CDF) is a "closed-form" equation for the probability that a continuous random variable is less than a given value.

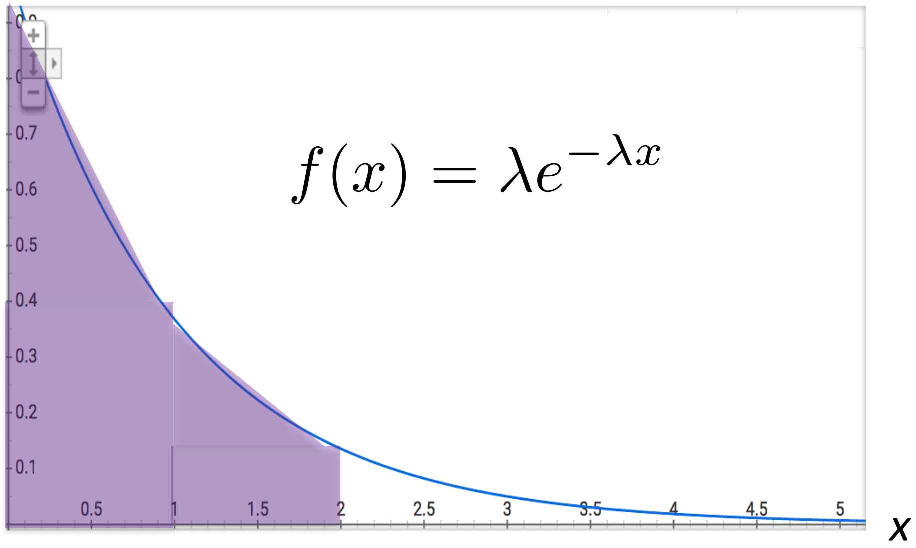
$$F(x) = P(X < x)$$

$$P(X < x) = \int_{y=-\infty}^{x} f(y) \, dy$$

For random variables that have cumulative density functions, we can avoid integrals!

Example: $X \sim \text{Exp}(\lambda = 1)$

Probability Density Function

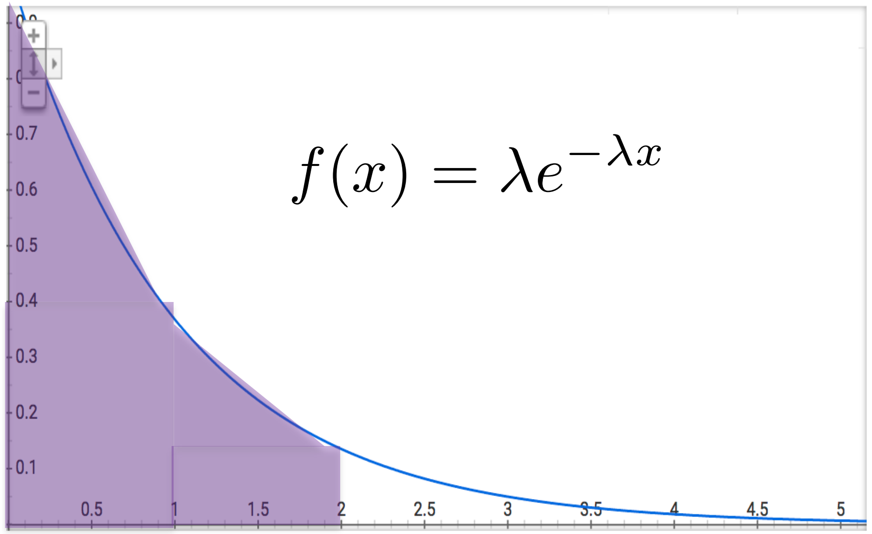


$$P(X < 2)$$

$$= \int_{x=-\infty}^{2} f(x)\, dx$$

Example: $X \sim \text{Exp}(\lambda = 1)$

Probability
Density
Function



$P(X < 2)$

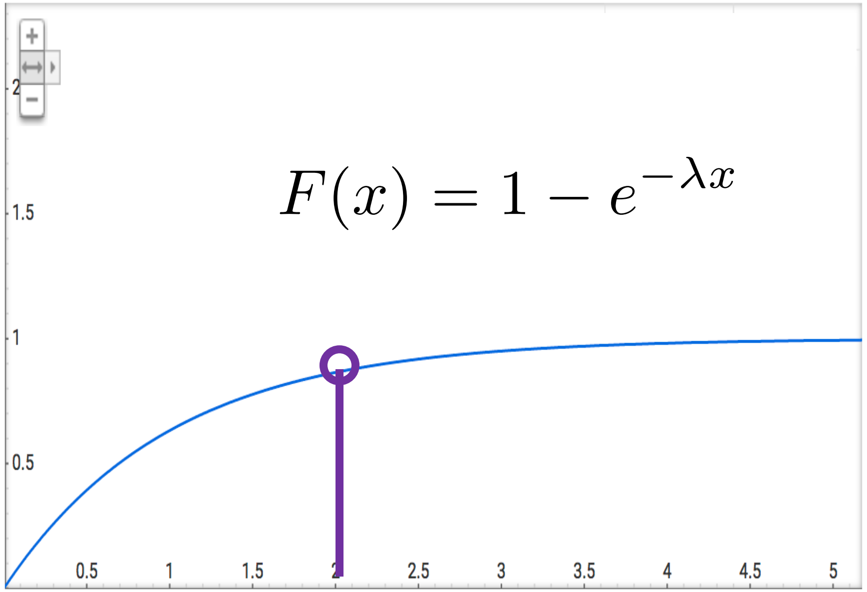
$$= \int_{x=-\infty}^{2} f(x)\, dx$$

or

Cumulative
Density
Function

$$F_X(x) = P(X < x)$$

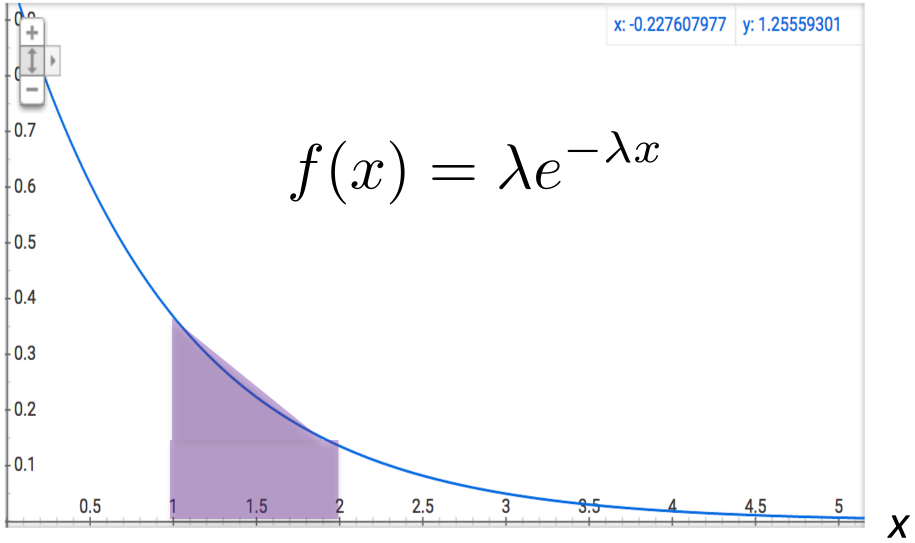
$$= \int_{y=-\infty}^{x} f(y)\, dy$$



$$F(2) = 1 - e^{-2}$$
$$\approx 0.84$$

Example: $X \sim \text{Exp}(\lambda = 1)$

Probability Density Function

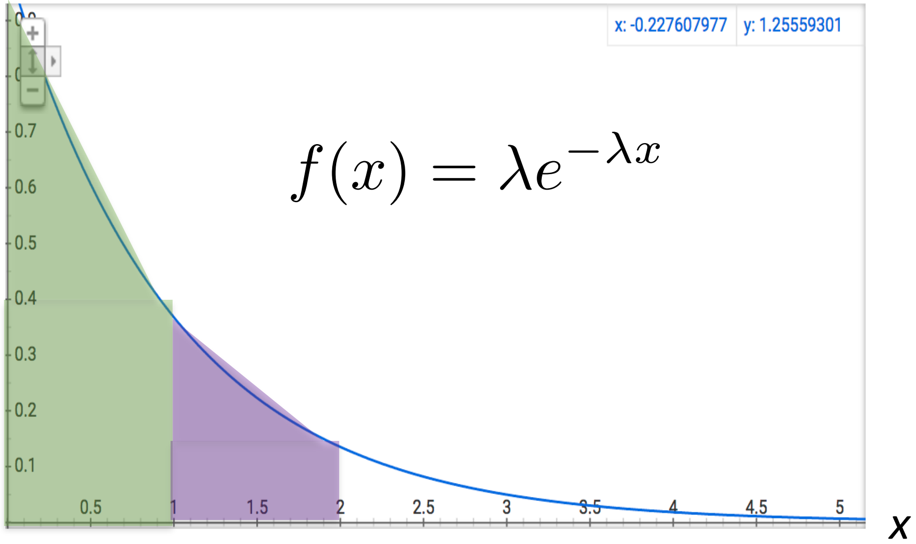


$$P(1 < X < 2)$$

$$= \int_{x=1}^{2} f(x)\, dx$$

Example: $X \sim \text{Exp}(\lambda = 1)$

Probability
Density
Function



$P(1 < X < 2)$

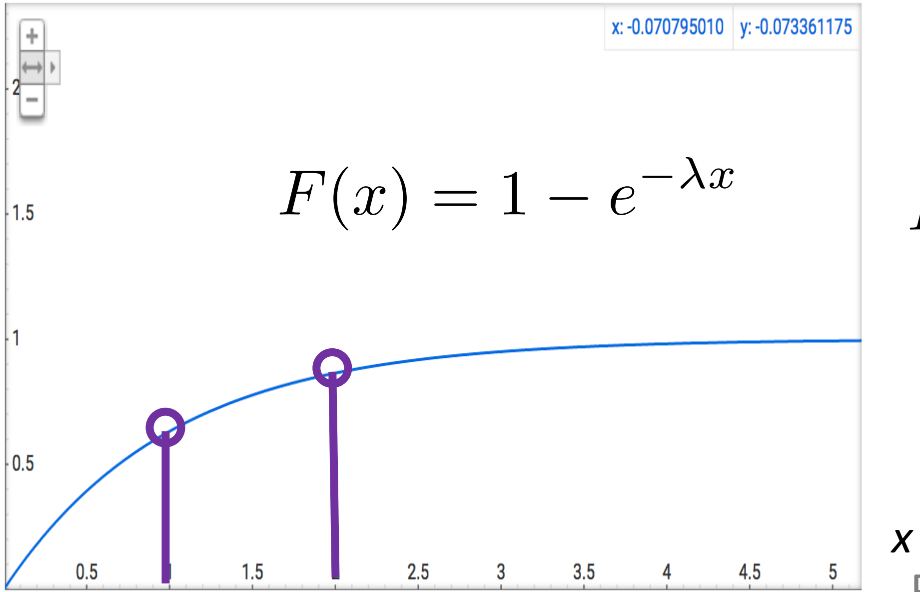
$$= \int_{x=1}^{2} f(x)\, dx$$

or

Cumulative
Density
Function

$$F_X(x) = P(X < x)$$

$$= \int_{y=-\infty}^{x} f(y)\, dy$$



$$F(2) - F(1) = (1 - e^{-2}) - (1 - e^{-1})$$

$$\approx 0.23$$

How Long Until the Next Big Earthquake?

Based on historical data, major earthquakes (with magnitude 8.0+) happen at a **rate of 0.002** per year*.
What is the probability of **a major earthquake in the next 30 years?**

Let *Y* be years until the next earthquake of magnitude 8.0+.

Exponential PDF:

$$f_Y(y) = \lambda e^{-\lambda y}$$

$$Y \sim \text{Exp}(\lambda = \frac{1}{500})$$

Exponential CDF:

$$F_Y(y) = 1 - e^{-\lambda y}$$

$$P(Y < 30) = \int_0^{30} \frac{1}{500} e^{-\frac{y}{500}} \, dy$$

$$= \left[-e^{-\frac{y}{500}} \right]_0^{30} = -e^{\frac{30}{500}} + e^0 \approx 0.058$$

$$P(Y < 30) = 1 - e^{-\frac{30}{500}} \approx 0.058$$

*In California, according to the USGS, 2015

End Review

The most famous continuous random variable

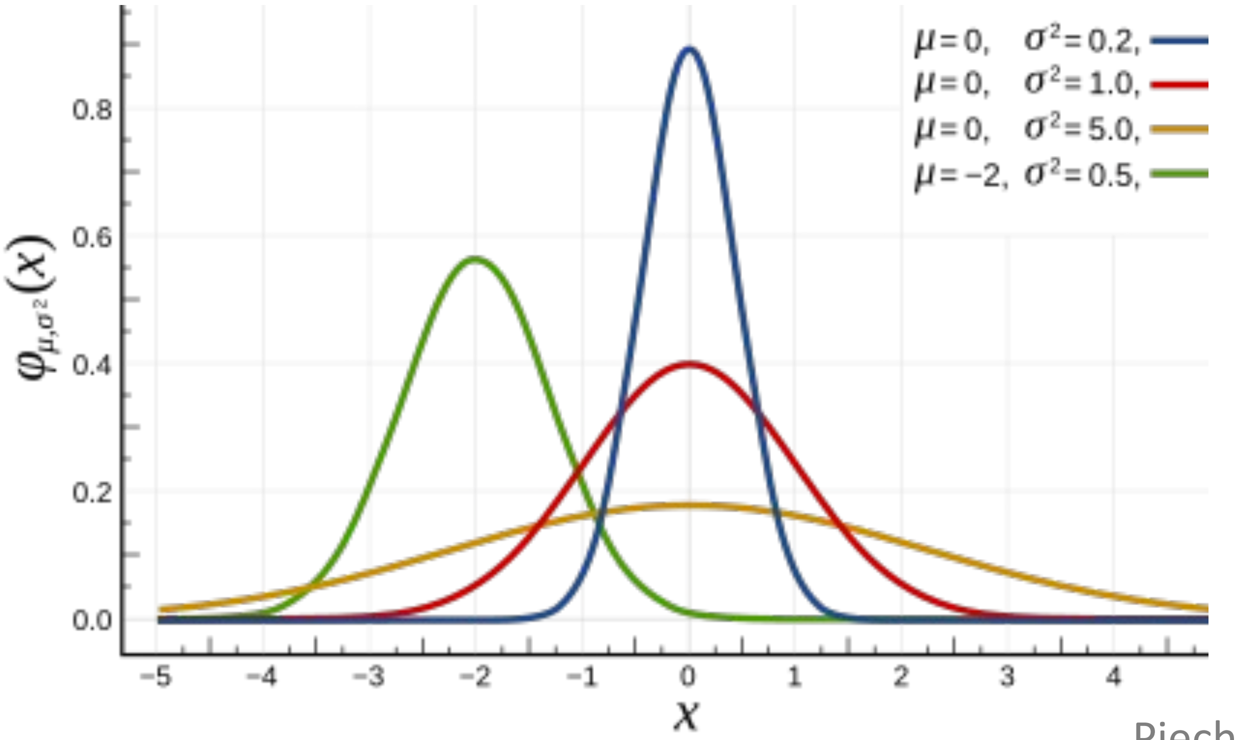
Normal (Gaussian) Random Variable

Support: $(-\infty, \infty)$

mean

variance

$$X \sim \mathcal{N}(\mu, \sigma^2)$$



Normal (Gaussian) Random Variable

Support:
$(-\infty, \infty)$

mean

variance

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

PDF:

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

Normal (Gaussian) Random Variable

Support:
$(-\infty, \infty)$

mean, variance

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

PDF:

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

Anatomy of a The Normal PDF

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

Anatomy of a The Normal PDF

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

a constant:
makes the integral
over all possible
outcomes sum to 1

Anatomy of a The Normal PDF

distance to the mean
(makes the PDF symmetric
around the mean)

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

a constant:
makes the integral
over all possible
outcomes sum to 1

Anatomy of a The Normal PDF

distance to the mean
(makes the PDF symmetric
around the mean)

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

a constant:
makes the integral
over all possible
outcomes sum to 1

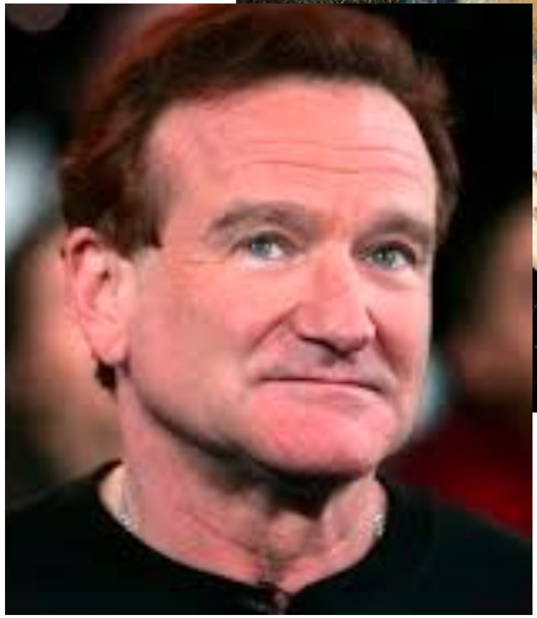
...normalized by
the variance

Carl Friedrich Gauss (1777-1855)

- German mathematician
- Sort-of invented the normal distribution
- Also astronomer, geologist, physicist
- Super influential in a lot of fields

Carl Friedrich Gauss (1777-1855)

- German mathematician
- Sort-of invented the normal distribution
- Also astronomer, geologist, physicist
- Super influential in a lot of fields

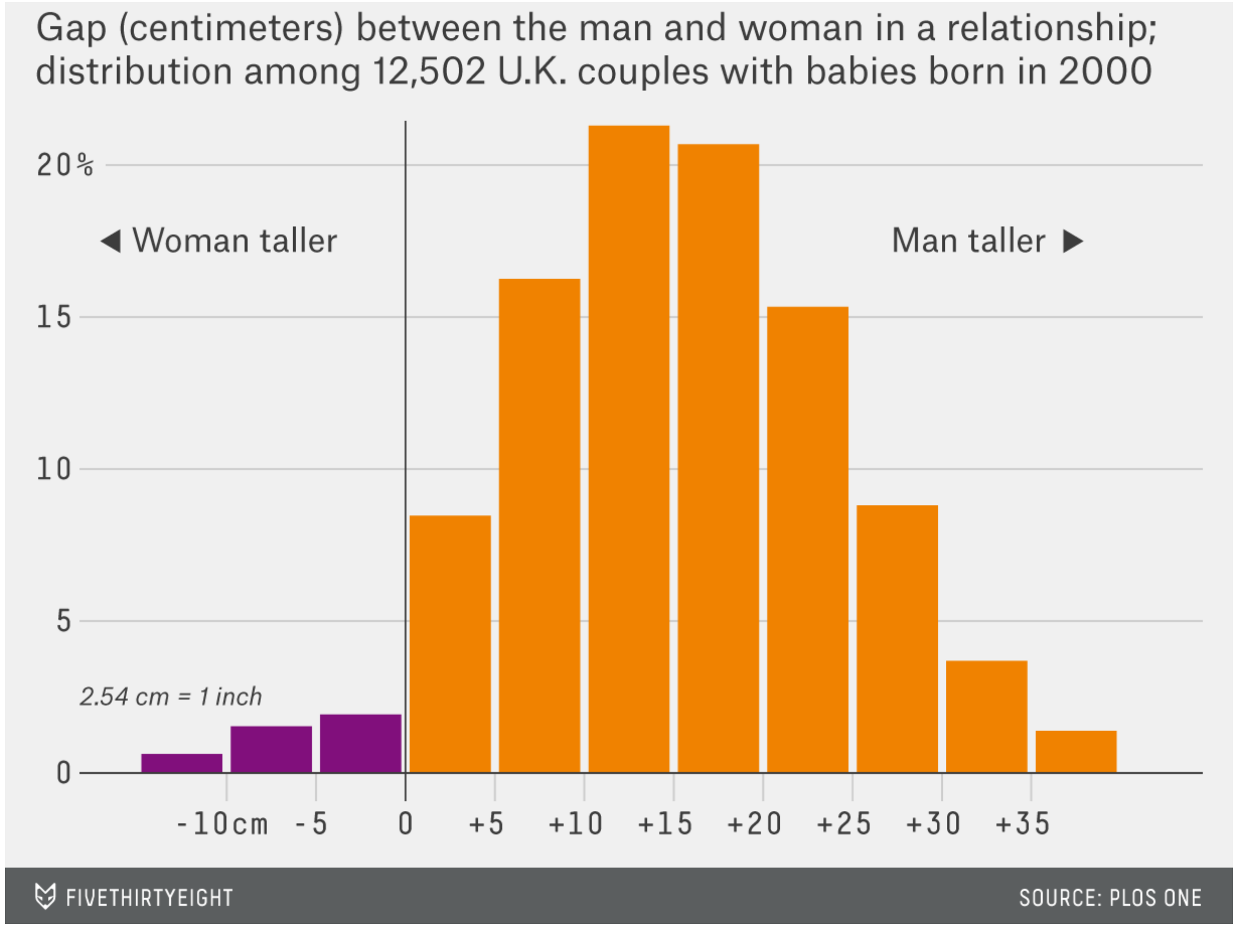


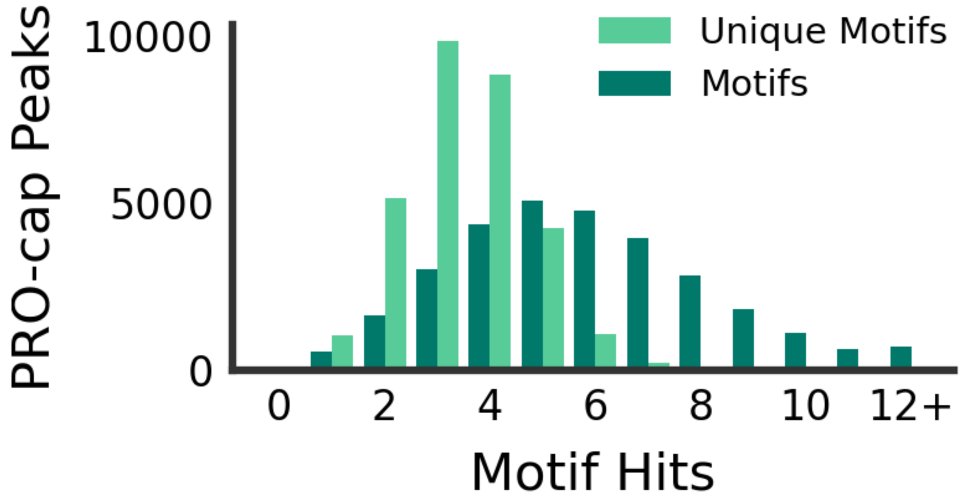
Looks like
Robin Williams

Why the Normal?

Why the Normal?

- Common for natural phenomena: human height, weight, shoe sizes, etc.

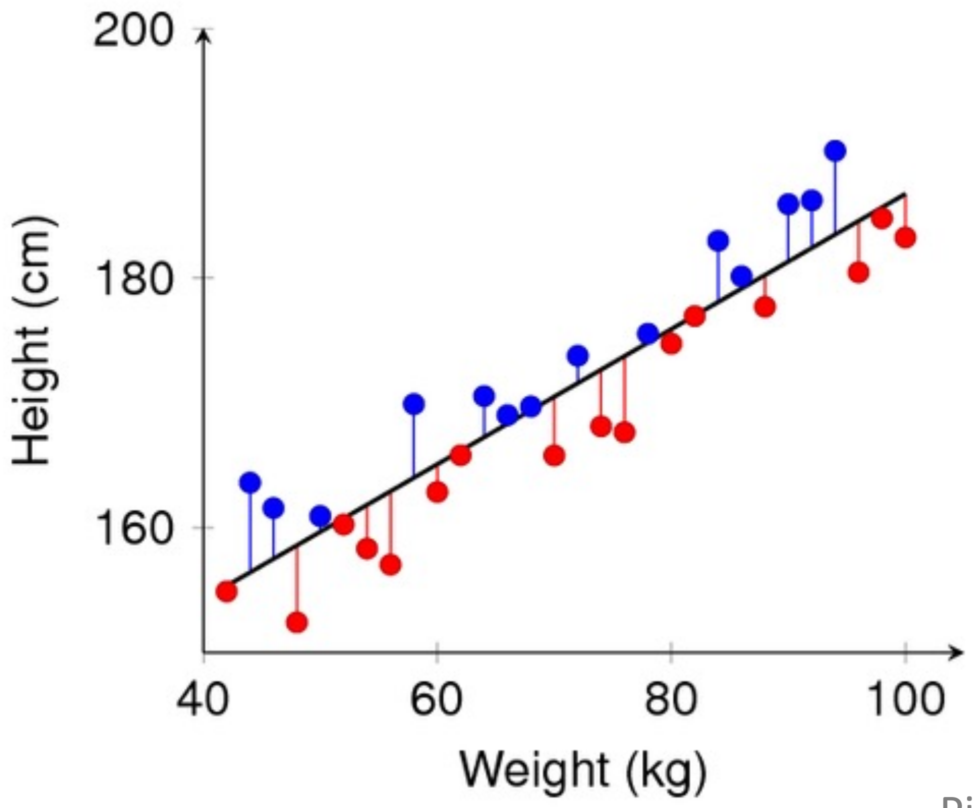




(random example from Kelly's research)

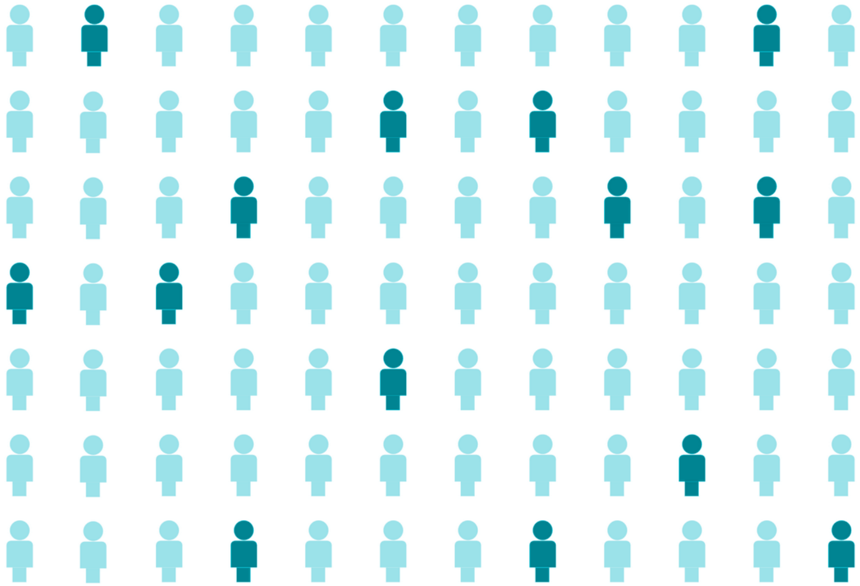
Why the Normal?

- Common for natural phenomena: human height, weight, shoe sizes, etc.
- A lot of noise in the world is Normal
  - E.g. random errors in measurements, residuals in linear regression



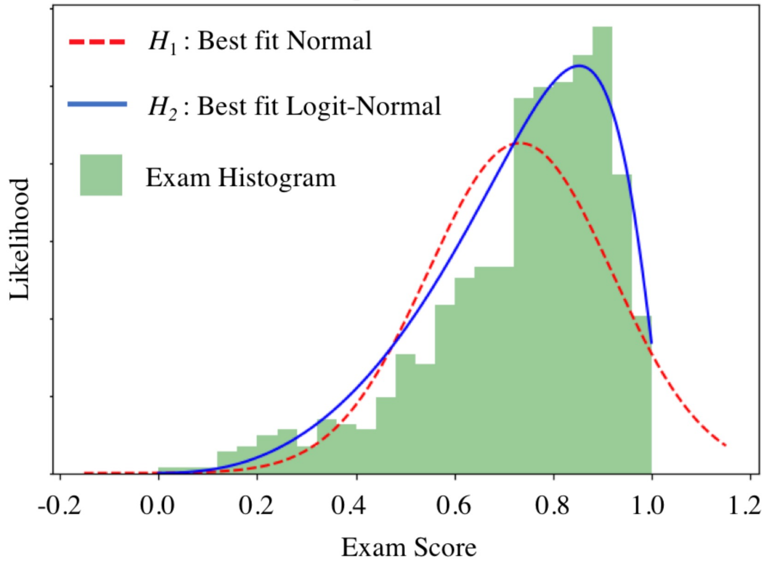
Why the Normal?

- Common for natural phenomena: human height, weight, shoe sizes, etc.
- A lot of noise in the world is Normal
  - E.g. random errors in measurements, residuals in linear regression
- The sum of many random variables often looks Normal (spoilers)
- Sample means are distributed normally – important for statistics



Why the Normal?

- Common for natural phenomena: human height, weight, shoe sizes, etc.
- A lot of noise in the world is Normal
  - E.g. random errors in measurements, residuals in linear regression
- The sum of many random variables often looks Normal (spoilers)
- Sample means are distributed normally – important for statistics
- Even things that aren't Normal might fit a normal-related distribution



Why the Normal?

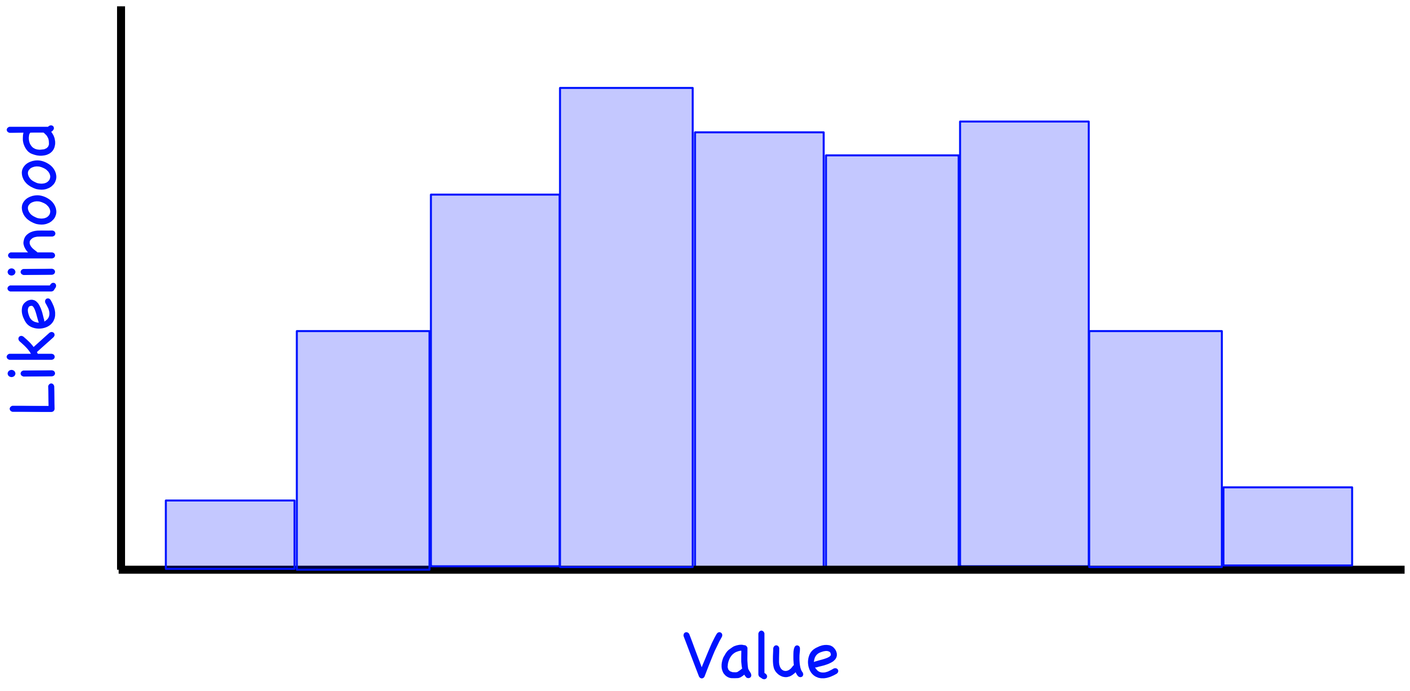
- Common for natural phenomena: human height, weight, shoe sizes, etc.
- A lot of noise in the world is Normal
  - E.g. random errors in measurements, residuals in linear regression
- The sum of many random variables often looks Normal (spoilers)
- Sample means are distributed normally – important for statistics
- Even things that aren't Normal might fit a normal-related distribution

People also just assume things are normally distributed a lot.

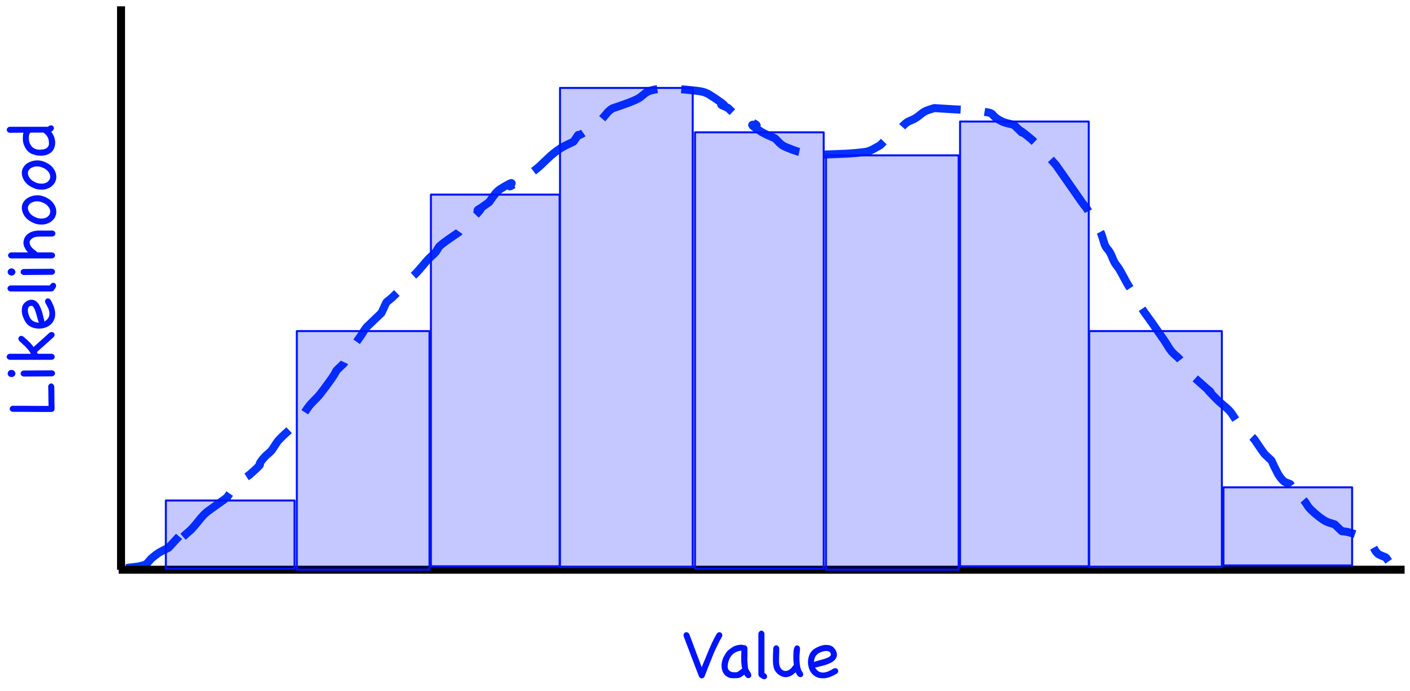
- They can do this in part because the Normal is so common
- But there's a deeper reason to it…

"The simplest explanation is usually the best one"

When We Fit Models To Data, We Try To Keep It Simple

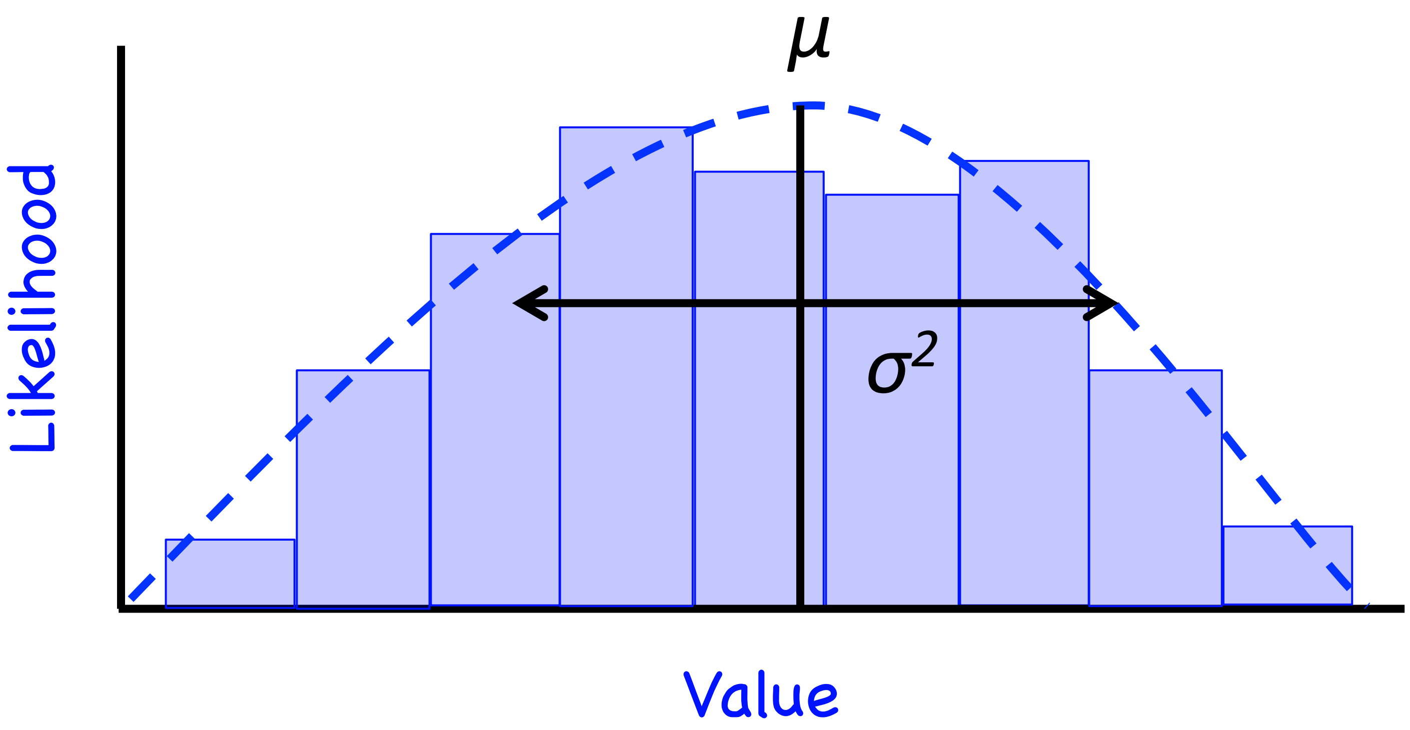


When We Fit Models To Data, We Try To Keep It Simple



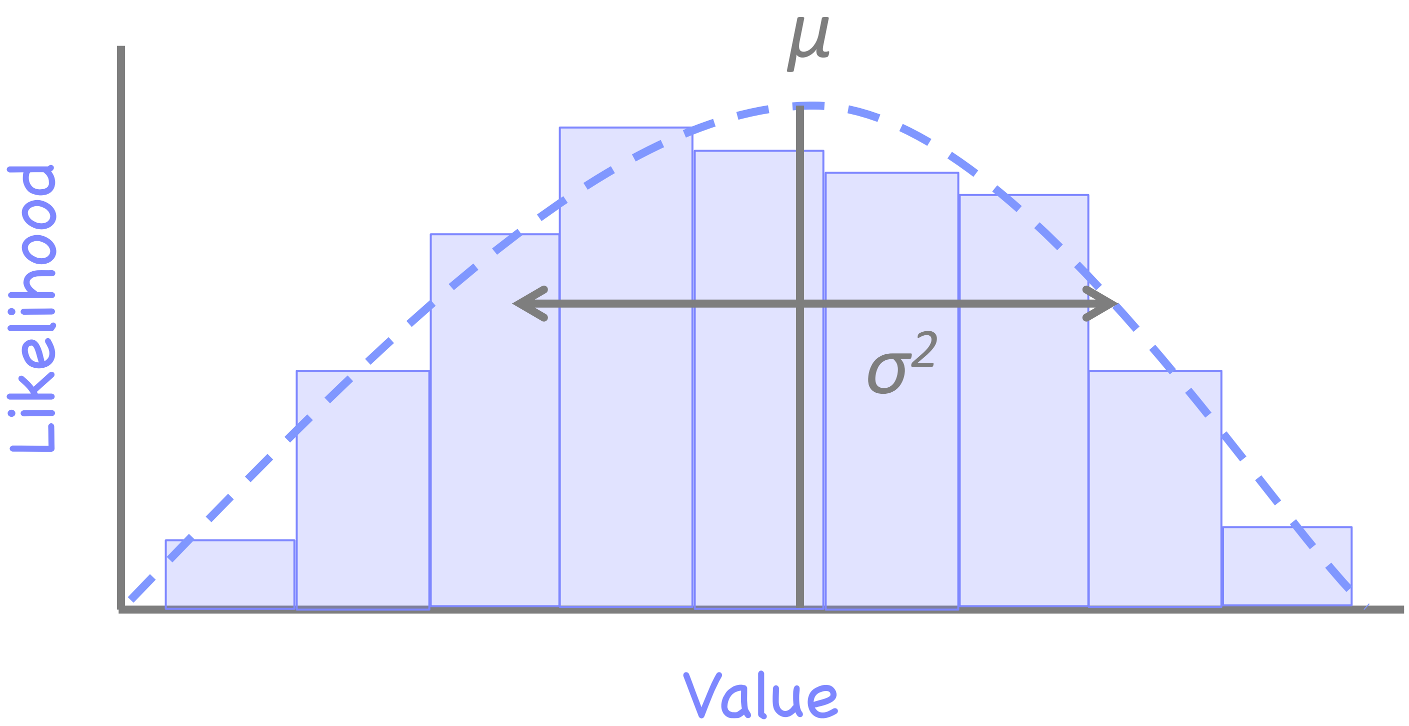
This curve fits the data well, but does it really represent the distribution?
Or is it “overfit”, so that the curve captures too much of the noise?

When We Fit Models To Data, We Try To Keep It Simple



This curve fits the data about as well, but appears to overfit less.
We could say that this simpler distribution makes fewer assumptions.
The formal concept for this idea is **entropy**

When We Fit Models To Data, We Try To Keep It Simple

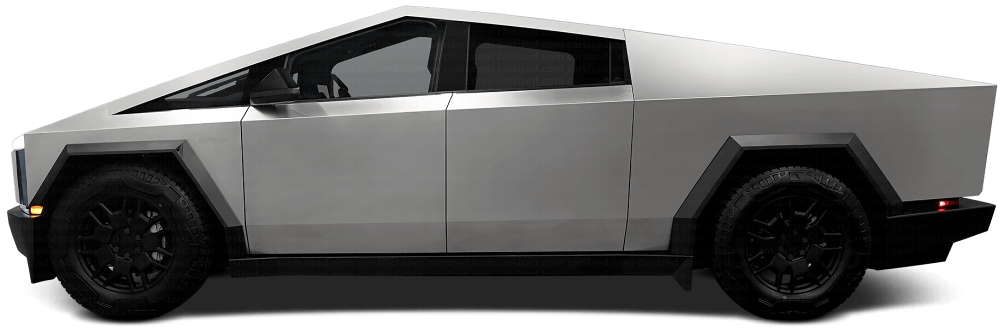


This cu… …ess.
We could… …tions.

The Normal distribution is the simplest distribution, that makes the fewest assumptions (has maximum entropy), for a given mean and variance.

Let’s Try It Out: Cybertruck Manufacturing

Your team is tasked with producing the side panels for cybertrucks. Elon Musk requires all panels to be built ”accurate within 10 microns”. You check how precise your manufacturing is, and find these stats:

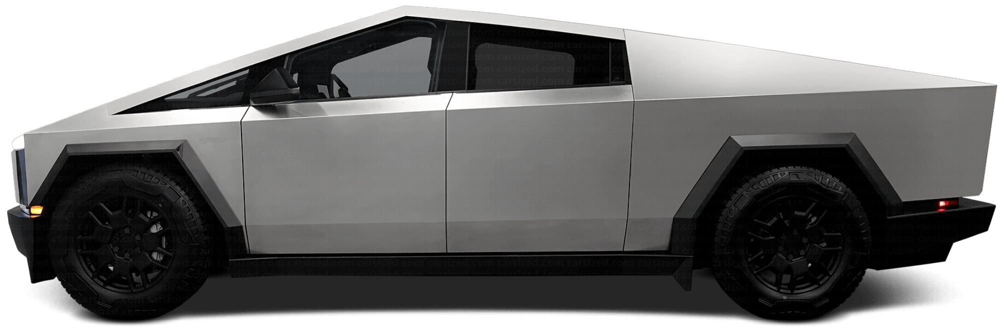


- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

What fraction of the panels you manufacture will meet Elon’s standards?

Let’s Try It Out: Cybertruck Manufacturing

Your team is tasked with producing the side panels for cybertrucks. Elon Musk requires all panels to be built ”accurate within 10 microns”. You check how precise your manufacturing is, and find these stats:



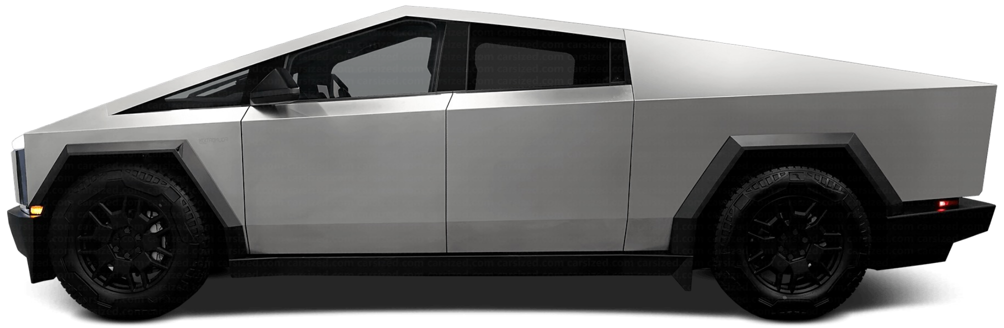
- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

What fraction of the panels you manufacture will meet Elon’s standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

Let's Try It Out: Cybertruck Manufacturing

Your team is tasked with producing the side panels for cybertrucks. Elon Musk requires all panels to be built "accurate within 10 microns". You check how precise your manufacturing is, and find these stats:



- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

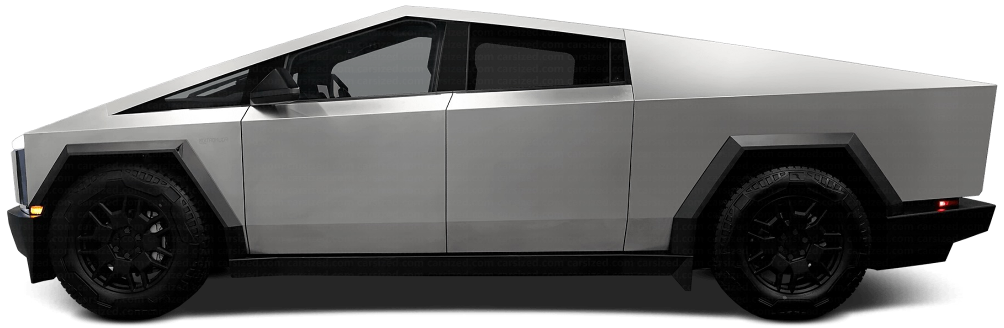
What fraction of the panels you manufacture will meet Elon's standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$$P(490 \leq X \leq 510) = \int_{490}^{510} f(X = x)dx = \int_{490}^{510} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Let's Try It Out: Cybertruck Manufacturing

Your team is tasked with producing the side panels for cybertrucks. Elon Musk requires all panels to be built ”accurate within 10 microns”. You check how precise your manufacturing is, and find these stats:



- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

What fraction of the panels you manufacture will meet Elon's standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

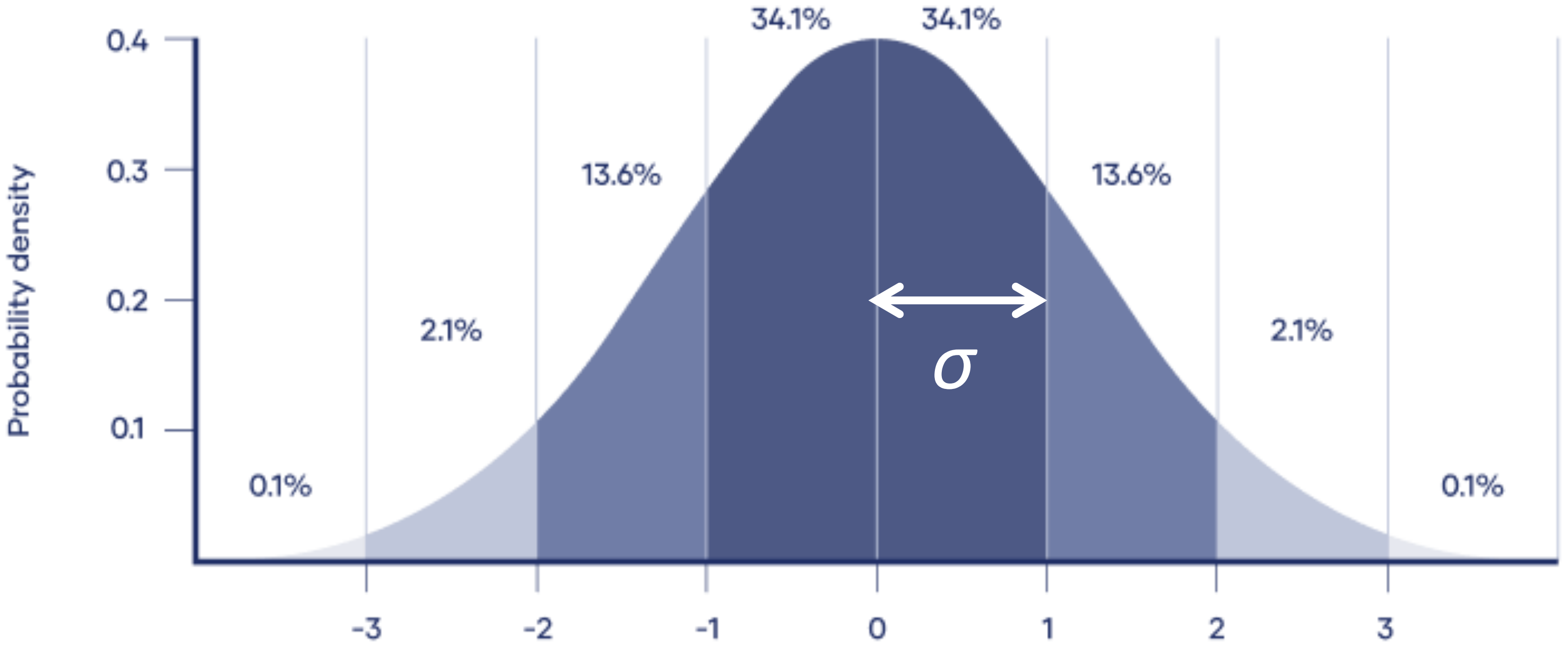
$$P(490 \leq X \leq 510) = \int_{490}^{510} f(X = x)dx = \int_{490}^{510} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

There is no closed form for the integral of this PDF

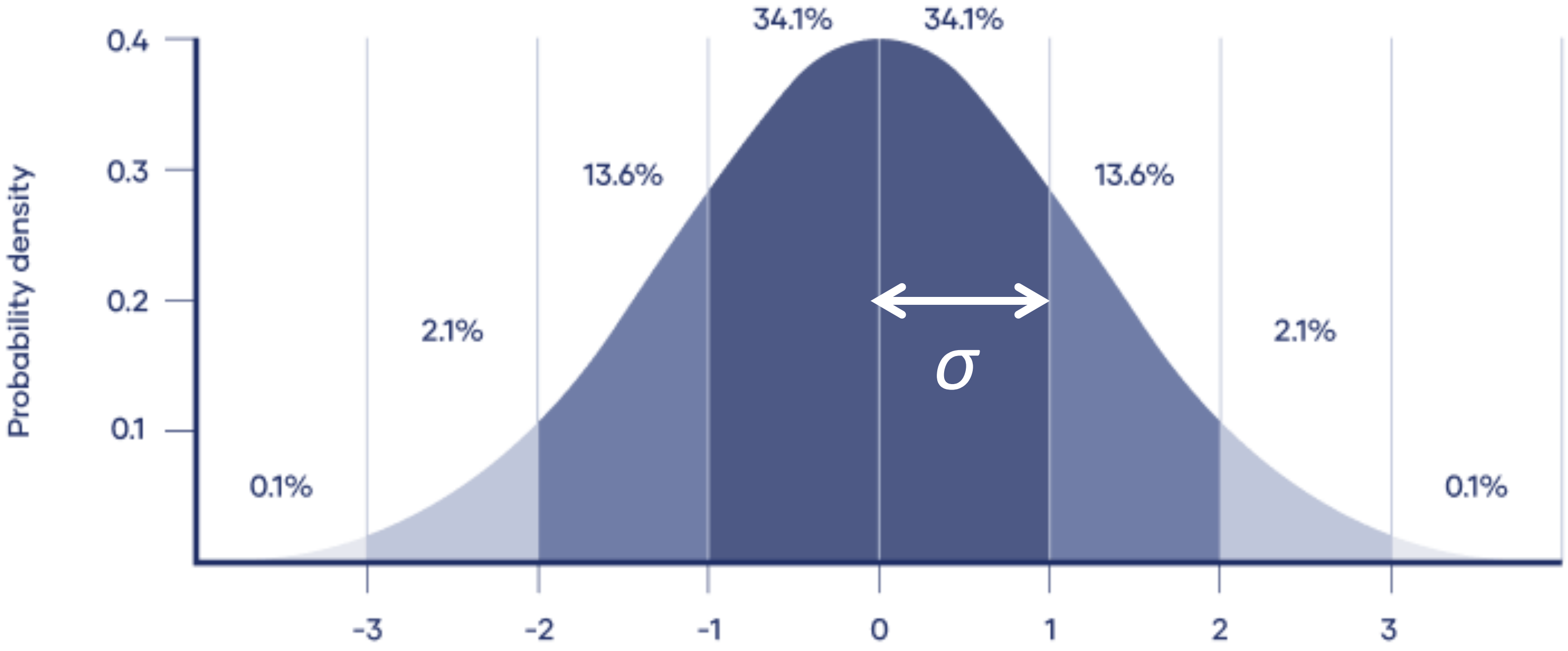
There is no closed form for the integral of this PDF

So no CDF???

The Standard Normal: $Z \sim N(\mu = 0, \sigma^2 = 1)$

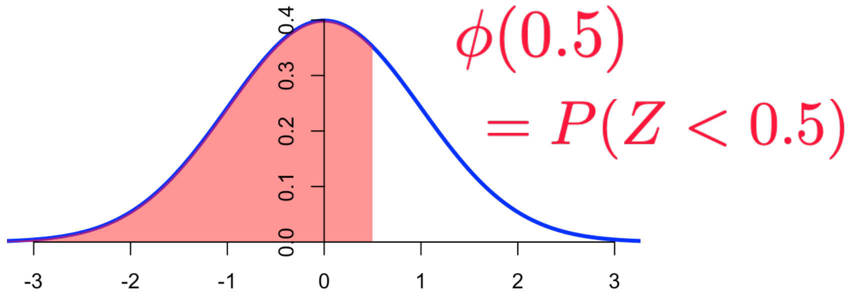


The Standard Normal: $Z \sim N(\mu = 0, \sigma^2 = 1)$



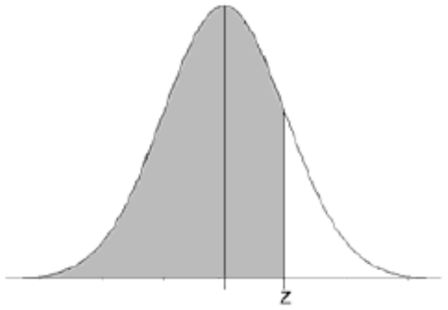
For the Standard Normal, we have a CDF!

$$F(x) = \phi(x)$$



What Does The Phi Function Look Like? Oh

Standard Normal Cumulative Probability Table



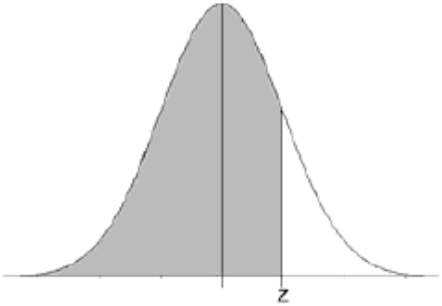
Cumulative probabilities for **POSITIVE** z-values are shown in the following table:

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	0.5753
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	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	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	0.7995	0.8023	0.8051	0.8078	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
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

What Does The Phi Function Look Like? Oh

Standard Normal Cumulative Probability Table

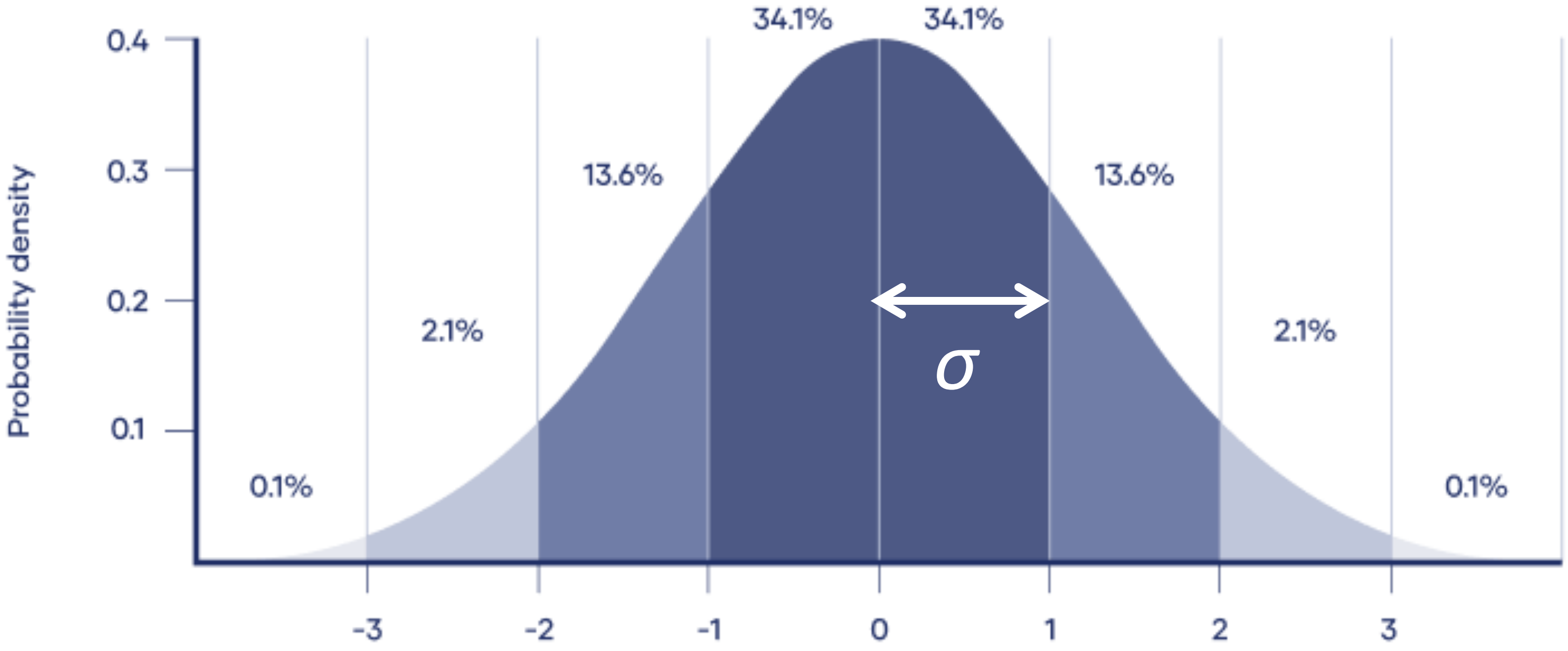
$\Phi(0.54) = 0.7054$



Cumulative probabilities for **POSITIVE** z-values are shown in the following table:

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	0.5753
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	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	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	0.7995	0.8023	0.8051	0.8078	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
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

The Standard Normal: $Z \sim N(\mu = 0, \sigma^2 = 1)$



For the Standard Normal, we have a CDF!

$$F(x) = \phi(x)$$

A function that has been solved for us numerically

Our probability ancestors did the work of solving for the CDF of the standard normal.

How do we use this for *any* normal distribution?

Fun Fact: The Linear Transform of a Normal Is…Normal

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ ⟹ $Y = aX + b$

is also Normal.

Fun Fact: The Linear Transform of a Normal Is…Normal

$$\text{Let } X \sim \mathcal{N}(\mu, \sigma^2) \quad \longrightarrow \quad Y = aX + b$$

is also Normal.

What would the mean and variance of *Y* be?

Fun Fact: The Linear Transform of a Normal Is…Normal

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ → $Y = aX + b$

is also Normal.

What would the mean and variance of *Y* be?

$$\begin{aligned} E[Y] &= E[aX + b] \\ &= aE[X] + b \\ &= a\mu + b \end{aligned}$$

Linearity property of expectation!

Fun Fact: The Linear Transform of a Normal Is…Normal

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ → $Y = aX + b$

is also Normal.

What would the mean and variance of *Y* be?

$$\begin{aligned} E[Y] &= E[aX + b] \\ &= aE[X] + b \\ &= a\mu + b \end{aligned}$$

$$\begin{aligned} Var(Y) &= Var(aX + b) \\ &= a^2 Var(X) \\ &= a^2\sigma^2 \end{aligned}$$

Fun Fact: The Linear Transform of a Normal Is…Normal

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ → $Y = aX + b$

is also Normal.

What would the mean and variance of *Y* be?

$$\begin{aligned} E[Y] &= E[aX + b] \\ &= aE[X] + b \\ &= a\mu + b \end{aligned}$$

$$\begin{aligned} Var(Y) &= Var(aX + b) \\ &= a^2 Var(X) \\ &= a^2\sigma^2 \end{aligned}$$

$$Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

Let’s Linear-Transform X into Z, The Standard Normal

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ ➡ $Y = aX + b$ ➡ $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

is also Normal

What linear transform of *X* would get us to *Z*?

Let's Linear-Transform X into Z, The Standard Normal

$$\text{Let } X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow Y = aX + b \Rightarrow Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

is also Normal

What linear transform of *X* would get us to *Z*?

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$$

Let's Linear-Transform X into Z, The Standard Normal

$$\text{Let } X \sim \mathcal{N}(\mu, \sigma^2) \quad \Rightarrow \quad Y = aX + b \quad \Rightarrow \quad Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

is also Normal

What linear transform of *X* would get us to *Z*?

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma} \qquad a = \frac{1}{\sigma} \quad b = -\frac{\mu}{\sigma}$$

Let's Linear-Transform X into Z, The Standard Normal

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ ➡ $Y = aX + b$ ➡ $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

is also Normal

What linear transform of *X* would get us to *Z*?

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma} \qquad a = \frac{1}{\sigma} \quad b = -\frac{\mu}{\sigma}$$

If we plug in these values for *a* and *b*, we get the standard normal:

$$Z \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

$$\sim \mathcal{N}(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2})$$

$$\sim \mathcal{N}(0, 1)$$

Let's Linear-Transform X into Z, The Standard Normal

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ ➡ $Y = aX + b$ ➡ $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

is also Normal

What linear transform of *X* would get us to *Z*?

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$$

$$a = \frac{1}{\sigma} \qquad b = -\frac{\mu}{\sigma}$$

If we plug in these values for *a* and *b*, we get the standard normal:

$$Z \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

$$\sim \mathcal{N}(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2})$$

$$\sim \mathcal{N}(0, 1)$$

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

How Do We Use This?

$\text{Let } X \sim \mathcal{N}(\mu, \sigma^2)$. Use the fact that $Z = \dfrac{X - \mu}{\sigma}$ to compute the CDF for X.

$$F_X(x) = P(X \leq x)$$

How Do We Use This?

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Use the fact that $Z = \dfrac{X - \mu}{\sigma}$ to compute the CDF for X.

$$
\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(X - \mu \leq x - \mu) \\
&= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right)
\end{aligned}
$$

Apply linear transform
to both sides

How Do We Use This?

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Use the fact that $Z = \frac{X - \mu}{\sigma}$ to compute the CDF for X.

$$F_X(x) = P(X \le x)$$

$$= P(X - \mu \le x - \mu)$$

Apply linear transform to both sides

$$= P\left(\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}\right)$$

Recognize that left-hand side is Z

$$= P\left(Z \le \frac{x - \mu}{\sigma}\right)$$

How Do We Use This?

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Use the fact that $Z = \frac{X - \mu}{\sigma}$ to compute the CDF for X.

$$F_X(x) = P(X \leq x)$$

$$= P(X - \mu \leq x - \mu)$$

Apply linear transform to both sides

$$= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right)$$

Recognize that left-hand side is Z

$$= P\left(Z \leq \frac{x - \mu}{\sigma}\right)$$

Recognize that the whole expression is the CDF

$$= \Phi\left(\frac{x - \mu}{\sigma}\right)$$

General CDF For Any Normal Random Variable

The cumulative density function of *any* normal, $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

To calculate P(*X* < *x*), for any normally distributed *X*,
we transform *X* to the standard normal, *Z*, and then use phi.

General CDF For Any Normal Random Variable

The cumulative density function of *any* normal, $X \sim \mathcal{N}(\mu, \sigma^2)$:

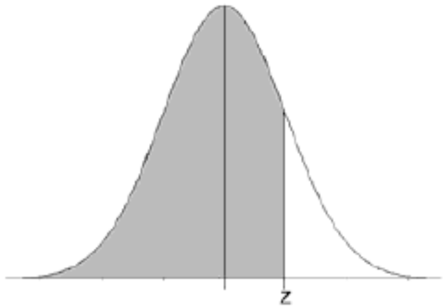
$$F(x) = \Phi\left(\frac{x - \mu}{\sigma^{\square}}\right)$$

not variance!

To calculate P(*X* < *x*), for any normally distributed *X*,
we transform *X* to the standard normal, *Z*, and then use phi.

Do We Have To Use The Table??

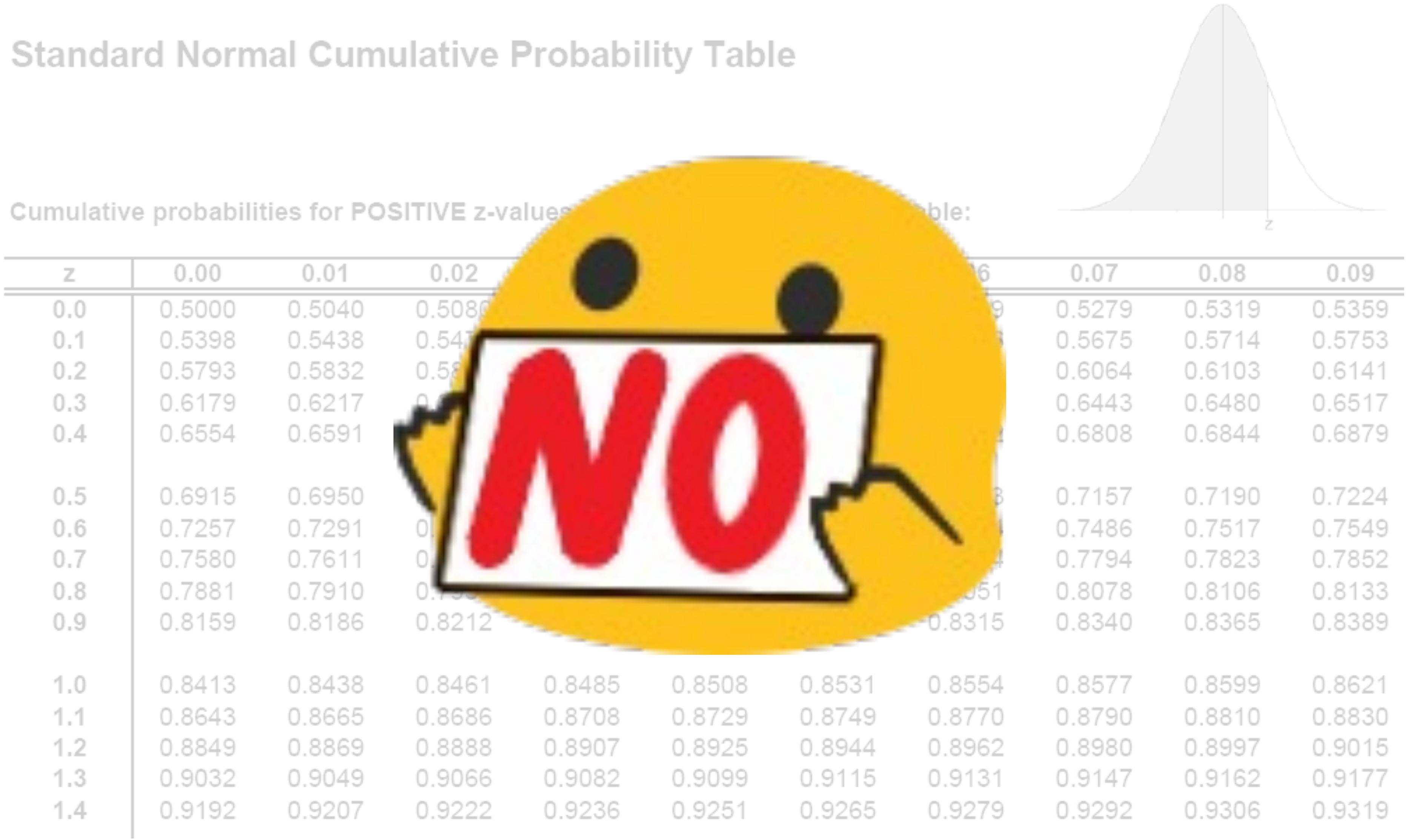
Standard Normal Cumulative Probability Table



Cumulative probabilities for **POSITIVE** z-values are shown in the following table:

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	0.5753
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	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	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	0.7995	0.8023	0.8051	0.8078	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
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Do We Have To Use The Table??



Standard Normal Cumulative Probability Table

Cumulative probabilities for POSITIVE z-values [illegible]

z	0.00	0.01	0.02	[illegible]	[illegible]	[illegible]	[illegible]	0.07	0.08	0.09
0.0	0.5000	0.5040	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	[illegible]	[illegible]	[illegible]	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

We Are Computer Scientists!

Every modern programming language has phi stored in a library:

```
from scipy import stats

stats.norm.cdf(x, mean, std)
```

$$= P(X < x) \text{ where } X \sim \mathcal{N}(\mu, \sigma^2)$$

We Are Computer Scientists!

Every modern programming language has phi stored in a library:

```
from scipy import stats

stats.norm.cdf(x, mean, std)
```

$$= P(X < x) \text{ where } X \sim \mathcal{N}(\mu, \sigma^2)$$

not variance!!!

We Are Computer Scientists!

Every modern programming language has phi stored in a library:

```
from scipy import stats

stats.norm.cdf(x, mean, std)
```

$$= P(X < x) \text{ where } X \sim \mathcal{N}(\mu, \sigma^2)$$

not variance!!!

The course reader also has a calculator:

Norm CDF Calculator

x: 0.0

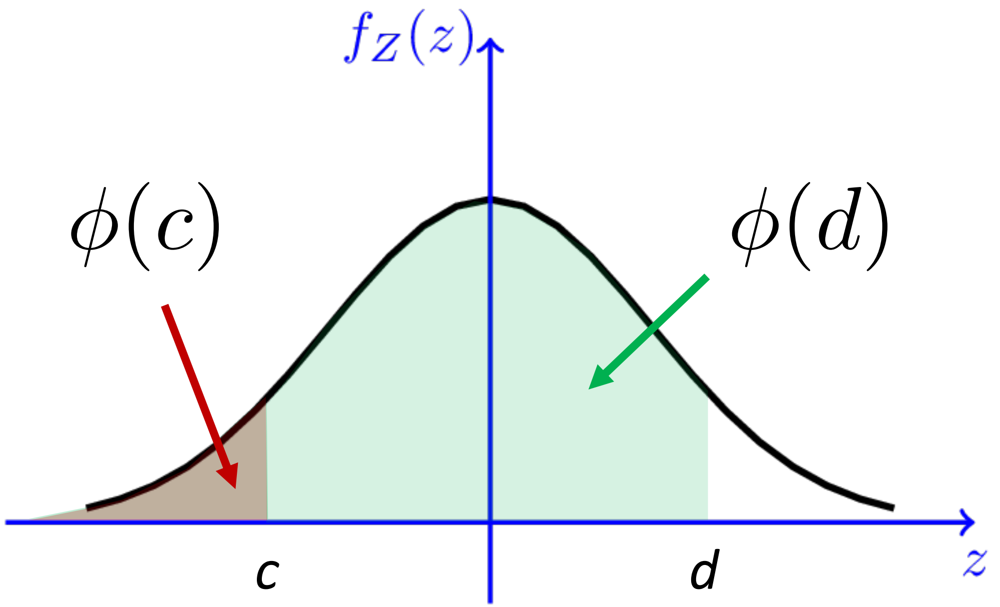
mu: 0

std: 1

`norm.cdf(x, mu, std)`

Fun Ways To Use Phi

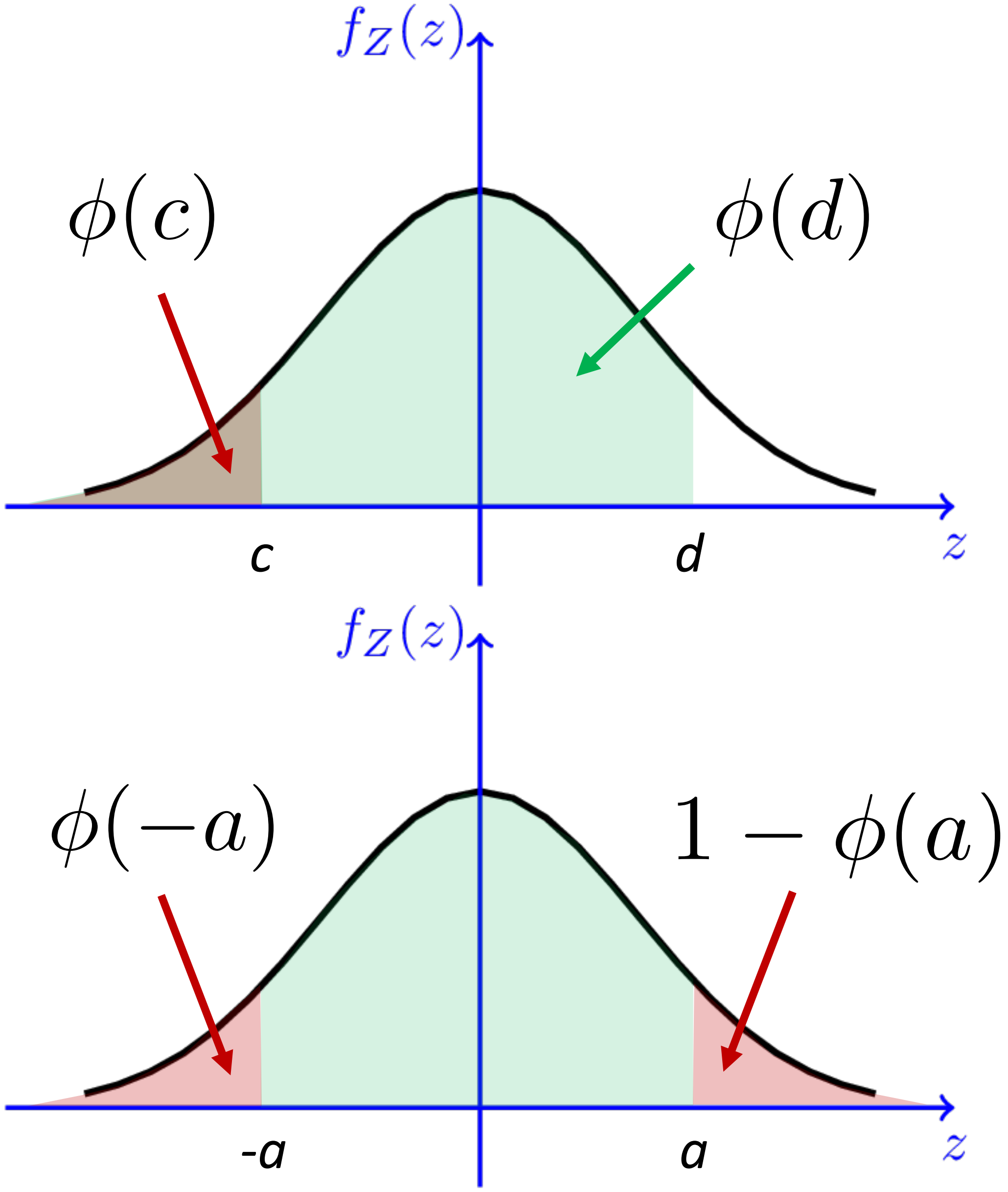
$$P(c < Z < d) = \phi(d) - \phi(c)$$



Fun Ways To Use Phi

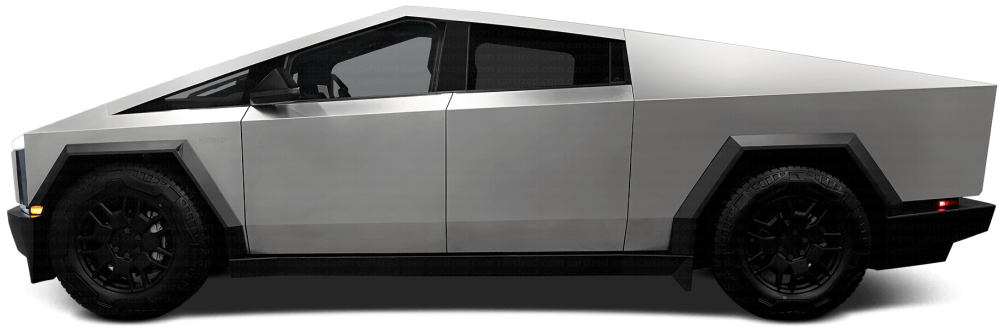
$$P(c < Z < d) = \phi(d) - \phi(c)$$

$$\phi(-a) = 1 - \phi(a)$$



Practice: Cybertruck Manufacturing

Your team is tasked with producing the side panels for cybertrucks. Elon Musk requires all panels to be built ”accurate within 10 microns”. You check how precise your manufacturing is, and find these stats:



- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

What fraction of the panels you manufacture will meet Elon’s standards?

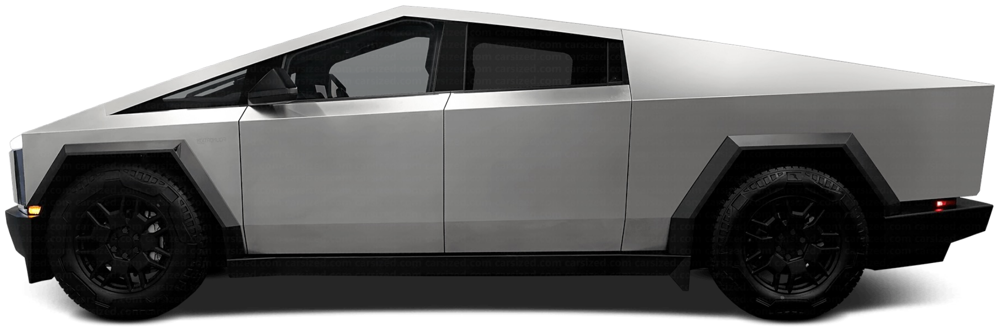
$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$$P(490 \leq X \leq 510) = \int_{490}^{510} f(X = x)dx$$

Practice: Cybertruck Manufacturing

If $X \sim \mathcal{N}(\mu, \sigma^2)$, $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Your team is tasked with producing the side panels for cybertrucks. Elon Musk requires all panels to be built ”accurate within 10 microns”. You check how precise your manufacturing is, and find these stats:



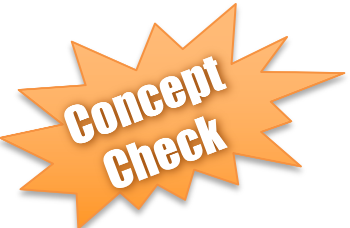
- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

Now using the CDF!

What fraction of the panels you manufacture will meet Elon's standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

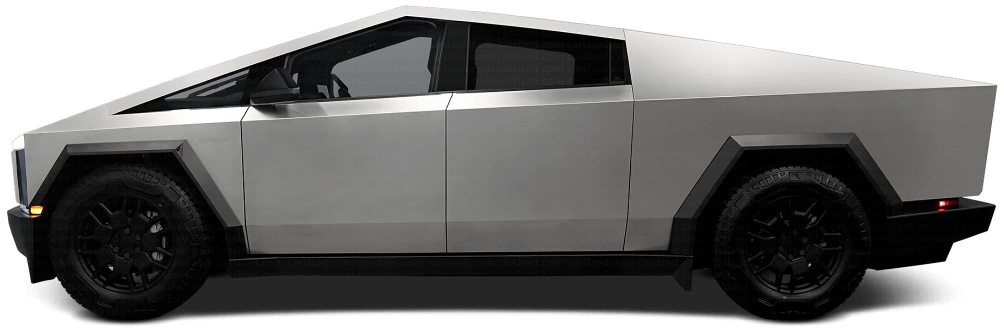
$P(490 \leq X \leq 510) = ?$



Practice: Cybertruck Manufacturing

If $X \sim \mathcal{N}(\mu, \sigma^2)$, $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Your team is tasked with producing the side panels for cybertrucks. Elon Musk requires all panels to be built ”accurate within 10 microns”. You check how precise your manufacturing is, and find these stats:



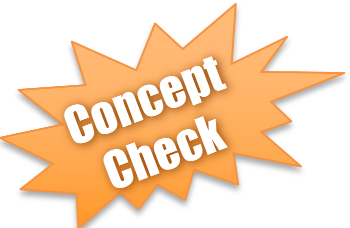
- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

Now using the CDF!

What fraction of the panels you manufacture will meet Elon’s standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$$P(490 \le X \le 510) = P(X < 510) - P(X < 490) = \Phi\left(\frac{510 - 500}{6}\right) - \Phi\left(\frac{490 - 500}{6}\right)$$

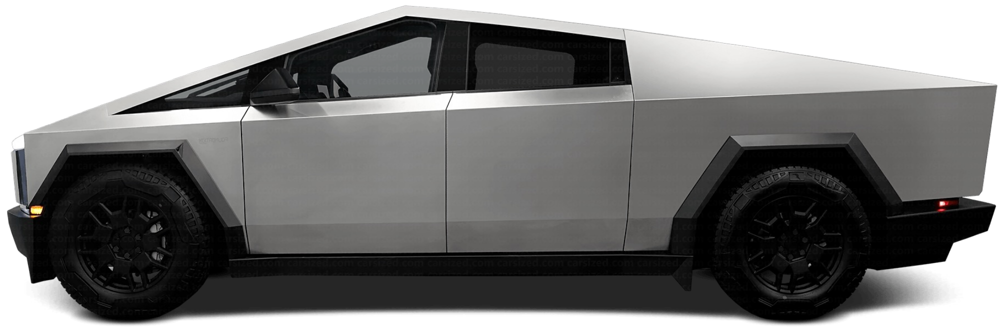


subtract mean, divide by std. dev.

Practice: Cybertruck Manufacturing

If $X \sim \mathcal{N}(\mu, \sigma^2)$, $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Your team is tasked with producing the side panels for cybertrucks. Elon Musk requires all panels to be built "accurate within 10 microns". You check how precise your manufacturing is, and find these stats:



- Average panel thickness: $\mu = 500$ microns
- Variance of thickness: $\sigma^2 = 36$ microns^2

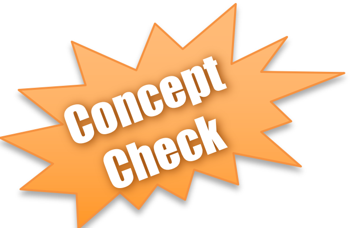
Now using the CDF!

What fraction of the panels you manufacture will meet Elon's standards?

$$X \sim \mathcal{N}(\mu = 500, \sigma^2 = 36)$$

$$P(490 \leq X \leq 510) = P(X < 510) - P(X < 490) = \Phi\left(\frac{510 - 500}{6}\right) - \Phi\left(\frac{490 - 500}{6}\right)$$

$$= \Phi\left(\frac{5}{3}\right) - \left(1 - \Phi\left(\frac{5}{3}\right)\right) = 2\ \Phi\left(\frac{5}{3}\right) - 1 \approx 0.904$$



Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$. Std deviation $\sigma = 4$.

1. $P(X > 0)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-z) = 1 - \Phi(z)$

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$.
Note standard deviation $\sigma = 4$.

How would you write each of the below probabilities as a function of the standard normal CDF, Φ?

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-z) = 1 - \Phi(z)$

1. $P(X > 0)$ (we just did this)
2. $P(2 < X < 5)$
3. $P(|X - 3| > 6)$

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$. Std deviation $\sigma = 4$.

1. $P(X > 0)$
2. $P(2 < X < 5)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-z) = 1 - \Phi(z)$

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$. Std deviation $\sigma = 4$.

1. $P(X > 0)$
2. $P(2 < X < 5)$
3. $P(|X - 3| > 6)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-x) = 1 - \Phi(x)$

Compute $z = \frac{(x-\mu)}{\sigma}$

Look up $\Phi(z)$ in table

$$P(X < -3) + P(X > 9)$$

$$= F(-3) + \left(1 - F(9)\right)$$

$$= \Phi\left(\frac{-3-3}{4}\right) + \left(1 - \Phi\left(\frac{9-3}{4}\right)\right)$$

Get your Gaussian On

Let $X \sim \mathcal{N}(\mu = 3, \sigma^2 = 16)$. Std deviation $\sigma = 4$.

1. $P(X > 0)$
2. $P(2 < X < 5)$
3. $P(|X - 3| > 6)$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- Symmetry of the PDF of Normal RV implies $\Phi(-x) = 1 - \Phi(x)$

Compute $z = \frac{(x-\mu)}{\sigma}$

$$P(X < -3) + P(X > 9)$$

$$= F(-3) + (1 - F(9))$$

$$= \Phi\left(\frac{-3-3}{4}\right) + \left(1 - \Phi\left(\frac{9-3}{4}\right)\right)$$

Look up $\Phi(z)$ in table

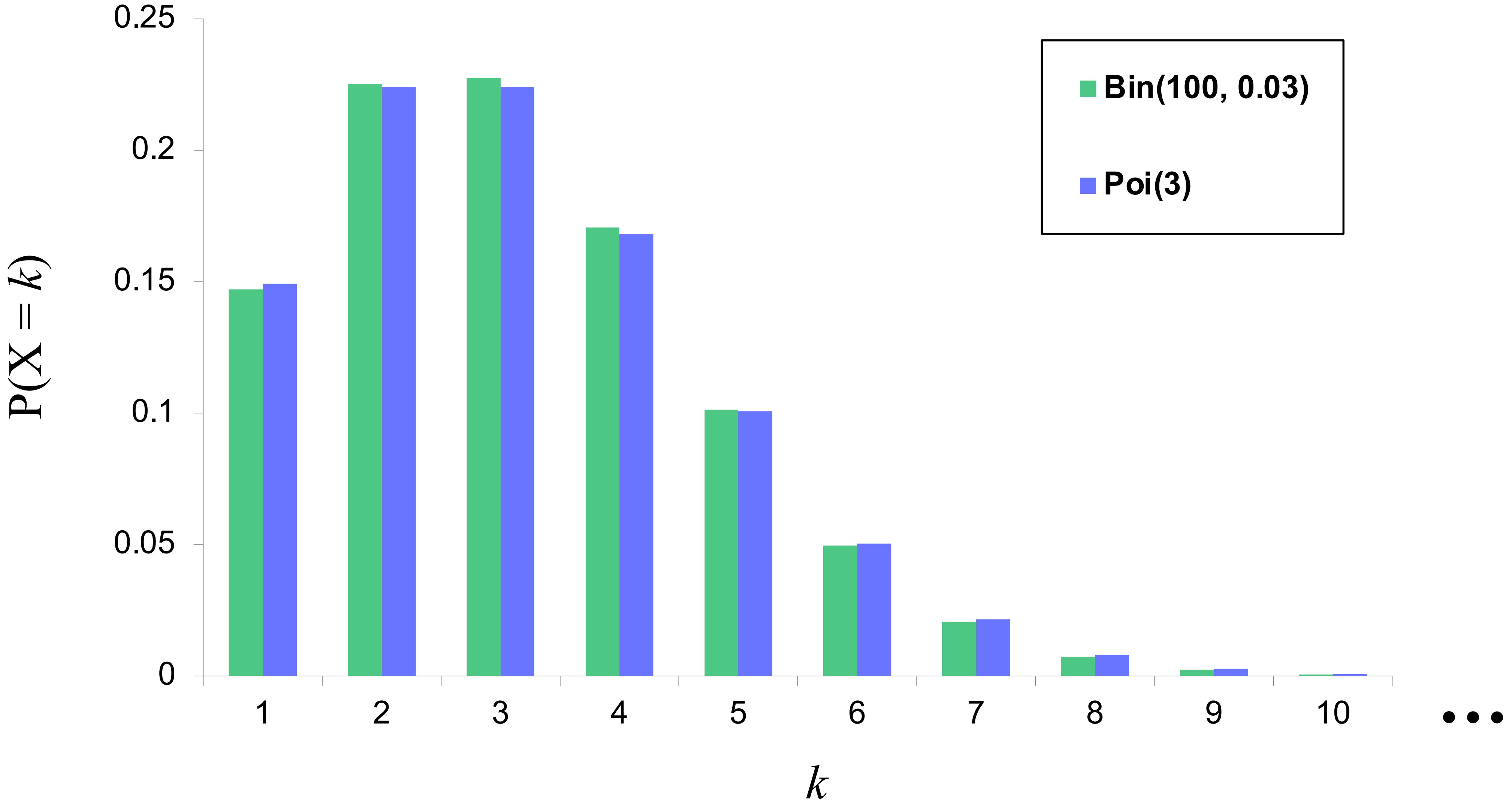
$$= \Phi\left(-\frac{3}{2}\right) + \left(1 - \Phi\left(\frac{3}{2}\right)\right)$$

$$= 2\left(1 - \Phi\left(\frac{3}{2}\right)\right)$$

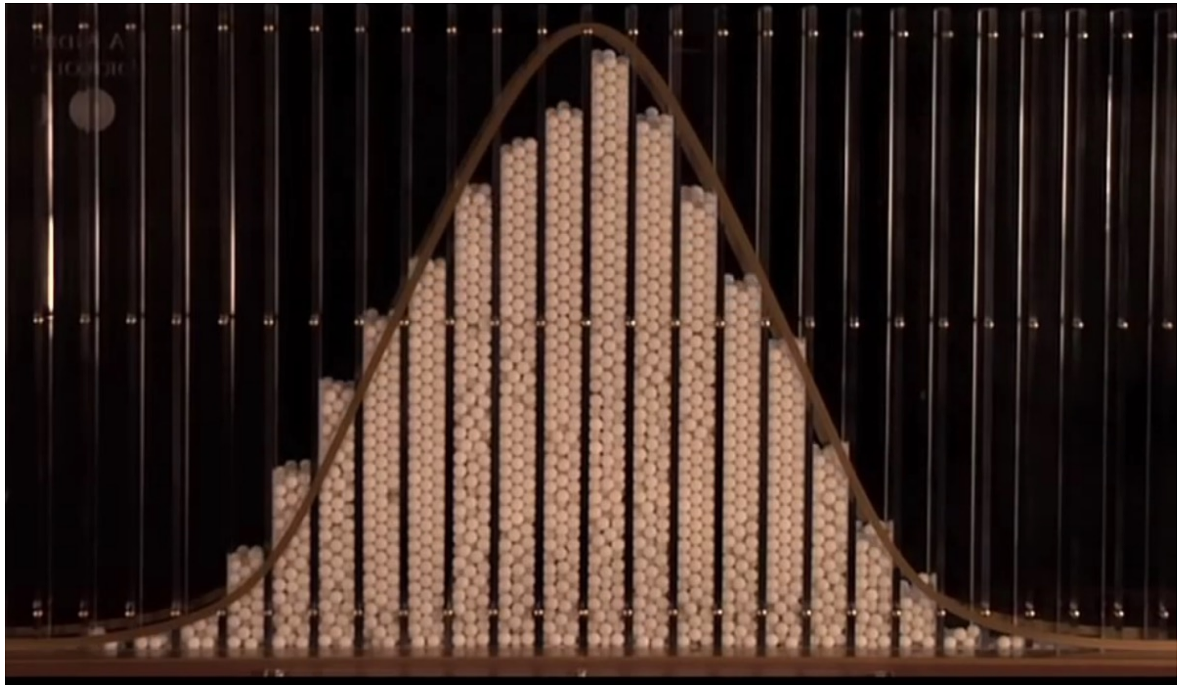
$$\approx 0.1337$$

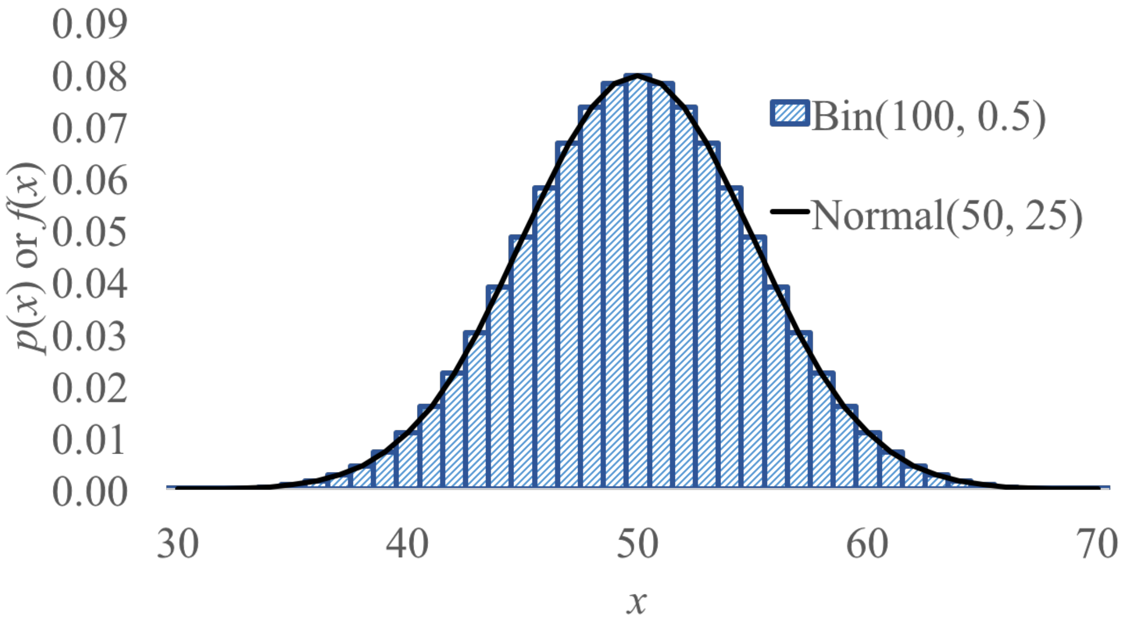
The Normal can *also* approximate the Binomial

Poisson Approximates Binomial, With Extreme *n* and *p*



Normal Approximates Binomial, With Moderate p





The shapes are the same!

Just set the normal's μ, σ^2 to be the mean and variance of the binomial.

Two Ways To Approximate The Binomial

$$X \sim \text{Bin}(n, p)$$
$$E[X] = np$$
$$\text{Var}(X) = np(1-p)$$

$$Y \sim \text{Poi}(\lambda)$$
$$\lambda = np$$

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$
$$\mu = np$$
$$\sigma^2 = np(1-p)$$

Poisson approximation for big n, small p.

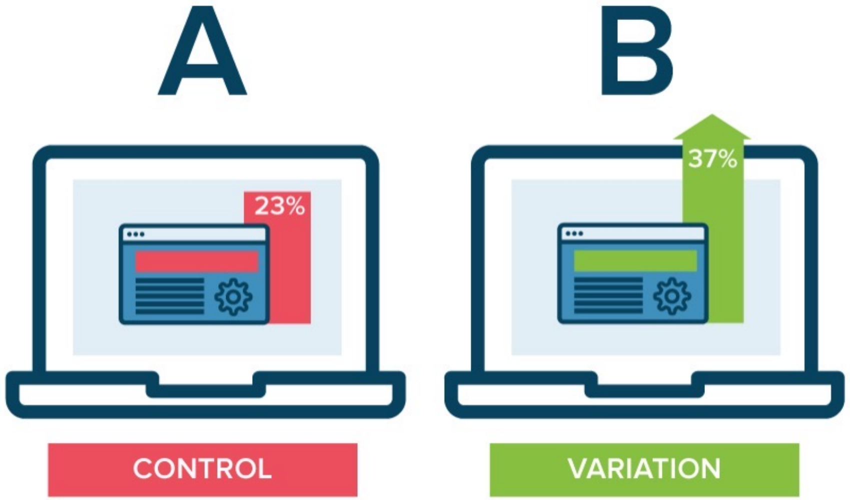
Normal approximation for big n, medium p.

Website Testing

A new website design is tested out on 100 users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if $X \geq 65$.

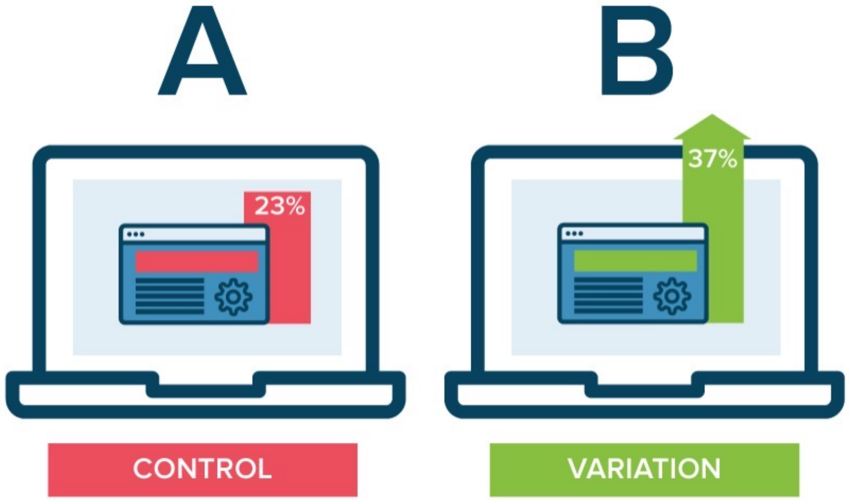
What is P(CEO endorses change| it has no effect)?



Website Testing

A new website design is tested out on 100 users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 65.



What is P(CEO endorses change| it has no effect)?

Without approximation: $X \sim \text{Bin}(n = 100, p = 0.5)$

$$P(X \geq 65) = \sum_{i=65}^{100} \binom{100}{i} (0.5)^i (1 - 0.5)^{100-i} \approx 0.0018$$

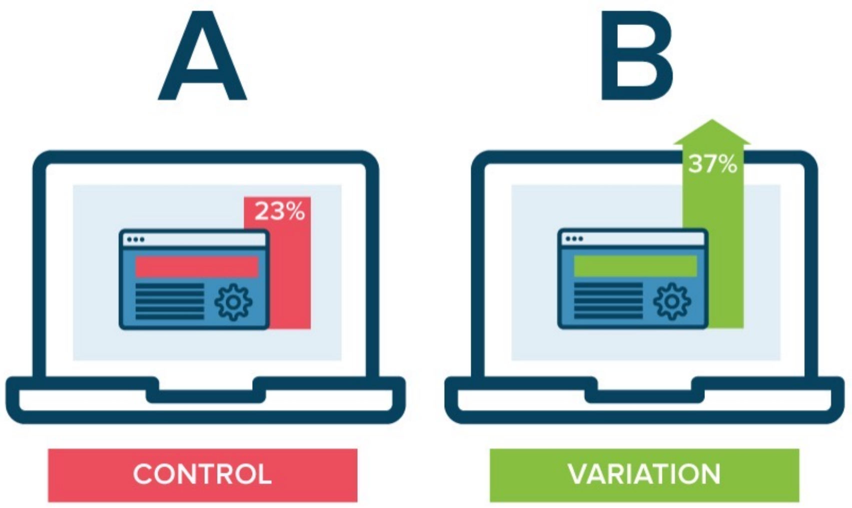
Website Testing

A new website design is tested out on 100 users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if $X \geq 65$.

What is P(CEO endorses change| it has no effect)?

Without approximation: $X \sim \text{Bin}(n = 100, p = 0.5)$

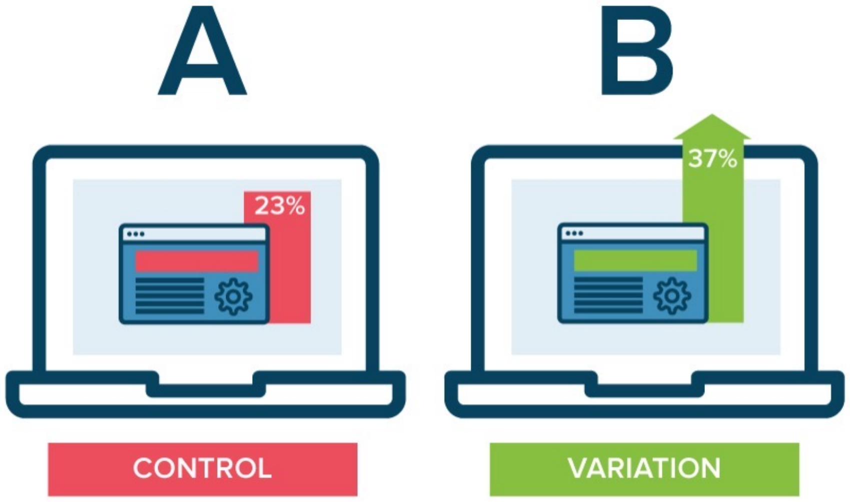


Website Testing

A new website design is tested out on 100 users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 65.

What is P(CEO endorses change| it has no effect)?

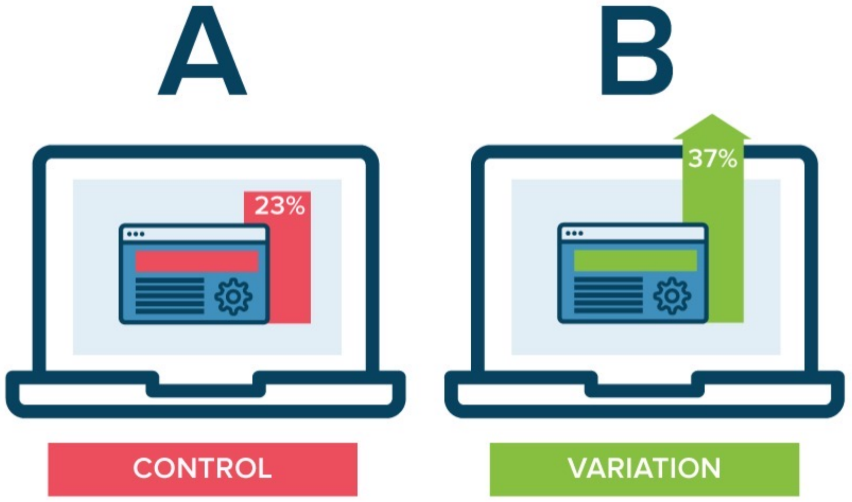


With approximation: $Y \sim \mathcal{N}(\mu, \sigma^2)$

Website Testing

A new website design is tested out on 100 users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if $X \geq 65$.

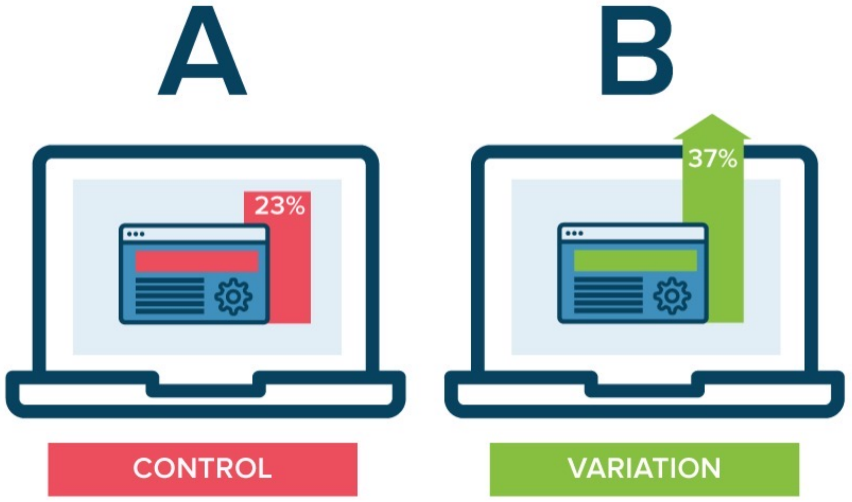


What is P(CEO endorses change| it has no effect)?

$$\mu = np = 50$$
$$\sigma^2 = np(1-p) = 25$$
$$\sigma = \sqrt{25} = 5$$

With approximation: $Y \sim \mathcal{N}(\mu, \sigma^2)$

Website Testing



A new website design is tested out on 100 users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 65.

What is P(CEO endorses change| it has no effect)?

$\mu = np = 50$
$\sigma^2 = np(1-p) = 25$
$\sigma = \sqrt{25} = 5$

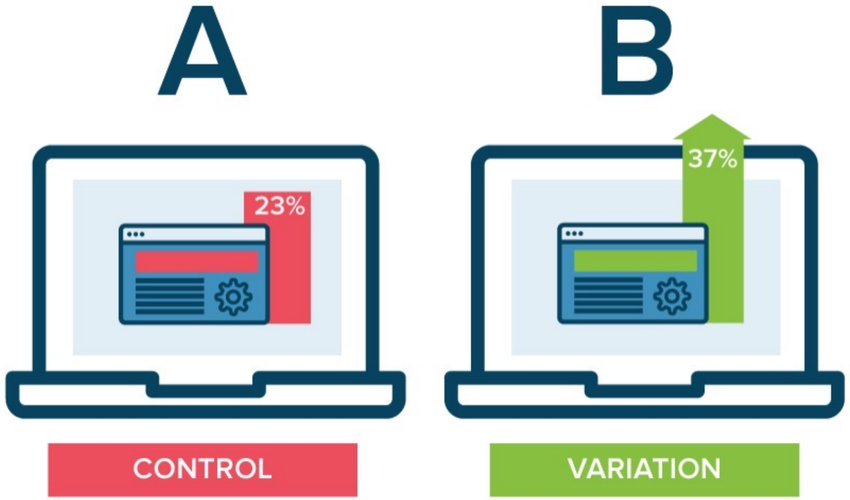
With approximation: $Y \sim \mathcal{N}(\mu, \sigma^2)$

$$P(X \geq 65) \approx P(Y \geq 65) = 1 - F_Y(65)$$

Website Testing

A new website design is tested out on 100 users.

- Let X be the number of users whose time on the site increases with the new design.
- The CEO will endorse the new design if X ≥ 65.



What is P(CEO endorses change| it has no effect)?

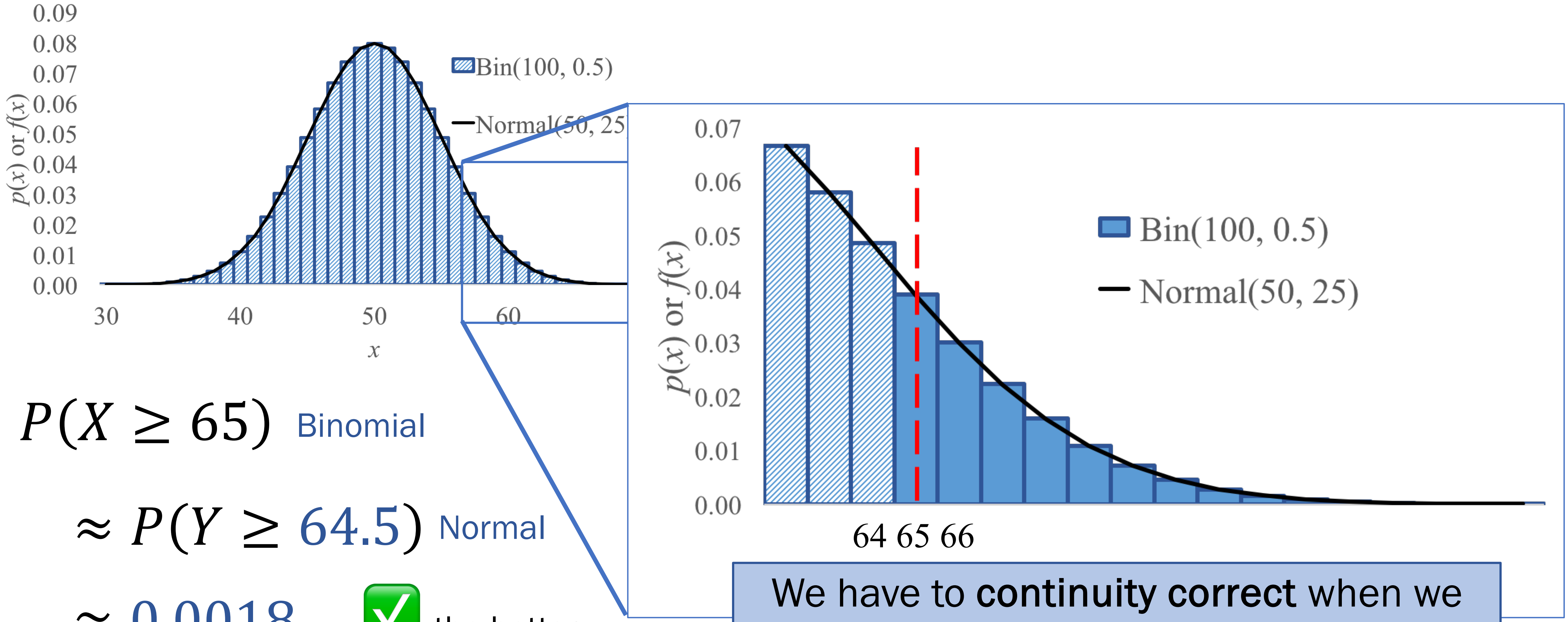
$$\mu = np = 50$$
$$\sigma^2 = np(1-p) = 25$$
$$\sigma = \sqrt{25} = 5$$

With approximation: $Y \sim \mathcal{N}(\mu, \sigma^2)$

$$P(X \geq 65) \approx P(Y \geq 65) = 1 - F_Y(65)$$
$$= 1 - \Phi\left(\frac{65-50}{5}\right) = 1 - \Phi(3) \approx 0.0013?$$

Website Testing, With Continuity Correction

$Y \sim \mathcal{N}(50, 25)$ approximates $X \sim \text{Bin}(100, 0.5)$, but $P(X \geq 65) \neq P(Y \geq 65)$?



$P(X \geq 65)$ Binomial

$\approx P(Y \geq 64.5)$ Normal

≈ 0.0018 ✅ the better Approach 2

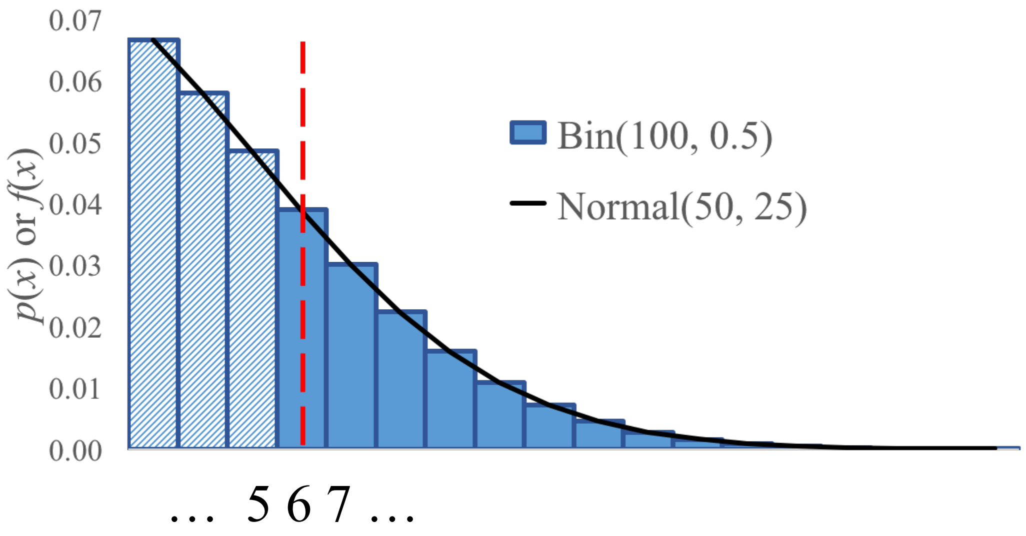
We have to **continuity correct** when we approximate a Binomial using a Normal.

Continuity Correction Practice

$Y \sim \mathcal{N}(np, np(1-p))$ approximates $X \sim \text{Bin}(n, p)$.

How do we approximate the following probabilities?

Discrete (e.g., Binomial) probability question	Continuous (Normal) probability question
$P(X = 6)$	
$P(X \geq 6)$	
$P(X > 6)$	
$P(X < 6)$	
$P(X \leq 6)$	

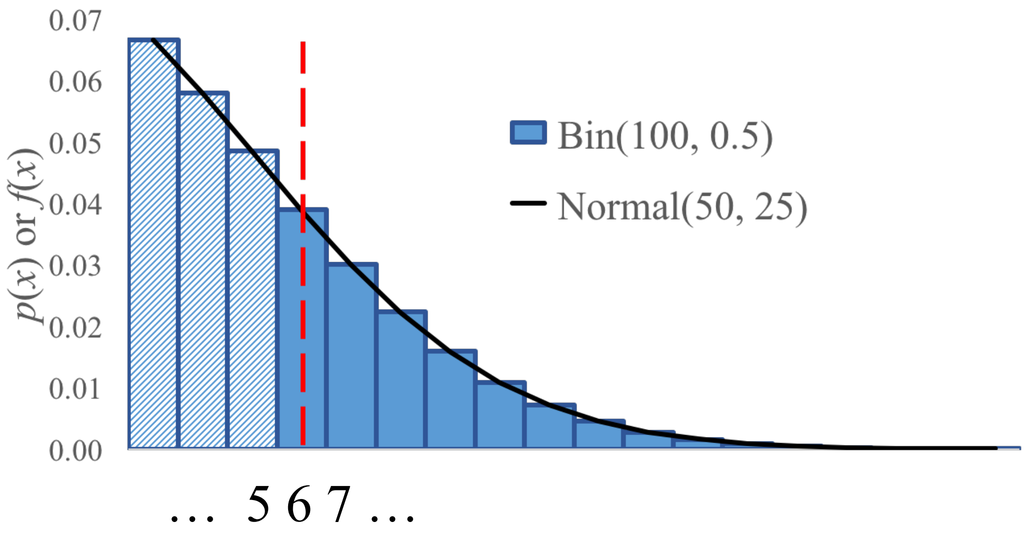


Continuity Correction Practice

$Y \sim \mathcal{N}(np, np(1-p))$ approximates $X \sim \text{Bin}(n, p)$.

How do we approximate the following probabilities?

Discrete (e.g., Binomial) probability question	Continuous (Normal) probability question
$P(X = 6)$	$P(5.5 \le Y \le 6.5)$
$P(X \ge 6)$	$P(Y \ge 5.5)$
$P(X > 6)$	$P(Y \ge 6.5)$
$P(X < 6)$	$P(Y \le 5.5)$
$P(X \le 6)$	$P(Y \le 6.5)$



Practice: Stanford Admissions

Stanford accepts 2480 students.

- Each admitted student independently matriculates with probability 0.68.
- Let X be the number of students who will attend.

What is $P(X > 1745)$? *Give a numerical approximation*.

Strategy:

A. Just Binomial
B. Poisson
C. Normal
D. None/other

Practice: Stanford Admissions

Stanford accepts 2480 students.

- Each admitted student independently matriculates with probability 0.68.
- Let X be the number of students who will attend.

What is $P(X > 1745)$? *Give a numerical approximation*.

Strategy:

A. Just Binomial
B. Poisson
C. Normal
D. None/other

Practice: Stanford Admissions

Stanford accepts 2480 students.

- Each admitted student independently matriculates with probability 0.68.
- Let X be the number of students who will attend.

What is $P(X > 1745)$? *Give a numerical approximation*.

$$X \sim \mathcal{N}(n = 2480, p = 0.68)$$

Step 1: define binomial,
like you normally would

Practice: Stanford Admissions

Stanford accepts 2480 students.

- Each admitted student independently matriculates with probability 0.68.
- Let X be the number of students who will attend.

What is $P(X > 1745)$? *Give a numerical approximation*.

$X \sim \mathcal{N}(n = 2480, p = 0.68)$ ⟹ Let $Y \sim \mathcal{N}\left(E[X], \text{Var}(X)\right)$

Step 1: define binomial, like you normally would

Step 2: define the normal that will approximate X

Practice: Stanford Admissions

Stanford accepts 2480 students.

- Each admitted student independently matriculates with probability 0.68.
- Let X be the number of students who will attend.

What is $P(X > 1745)$? *Give a numerical approximation*.

$X \sim \mathcal{N}(n = 2480, p = 0.68)$ ⇒ Let $Y \sim \mathcal{N}\big(E[X], \text{Var}(X)\big)$

Step 3: find parameters for the normal

$$E[X] = np = 1686$$

$$\text{Var}(X) = np(1-p) \approx 540 \rightarrow \sigma = 23.3$$

Practice: Stanford Admissions

Stanford accepts 2480 students.

- Each admitted student independently matriculates with probability 0.68.
- Let X be the number of students who will attend.

What is $P(X > 1745)$? *Give a numerical approximation*.

$$X \sim \mathcal{N}(n = 2480, p = 0.68) \quad \Rightarrow \quad \text{Let } Y \sim \mathcal{N}\left(E[X], \text{Var}(X)\right)$$

$$E[X] = np = 1686$$

$$\text{Var}(X) = np(1-p) \approx 540 \rightarrow \sigma = 23.3$$

Step 4: figure out what probability you want, then continuity correct

$$P(X > 1745) \approx P(Y \geq 1745.5)$$

Practice: Stanford Admissions

Stanford accepts 2480 students.

- Each admitted student independently matriculates with probability 0.68.
- Let X be the number of students who will attend.

What is $P(X > 1745)$? *Give a numerical approximation*.

$$X \sim \mathcal{N}(n = 2480, p = 0.68) \quad \Rightarrow \quad \text{Let } Y \sim \mathcal{N}\big(E[X], \text{Var}(X)\big)$$

$$E[X] = np = 1686$$

$$\text{Var}(X) = np(1-p) \approx 540 \rightarrow \sigma = 23.3$$

$$P(X > 1745) \approx P(Y \geq 1745.5)$$

Step 5: solve!

$$P(Y \geq 1745.5) = 1 - F(1745.5) = 1 - \Phi\left(\frac{1745.5 - 1686}{23.3}\right) = 1 - \Phi(2.54) \approx 0.0055$$

Challenge Problem


福
11.11
光棍节
双11
SINGLE'S DAY
SALE

How Many Servers Is Enough?

At the busiest minute of the shopping rush, your website receives R pings:

$$R \sim N(\mu = 10^6, \sigma = 10^4)$$

To anticipate the rush, you plan to buy *N* servers. Each server can handle 10,000 pings per minute, but if it receives any more, it will drop customers.

What is the smallest value of *N* such that P(drop) < 0.0001?

Ponder Before Wednesday!