Math 1710 Class 10

Normal Distributions and Approximation

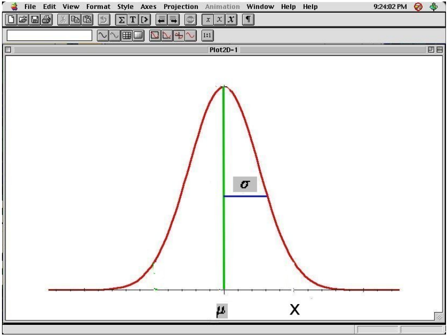
Dr. Back

Sep. 18, 2009

Z-scores

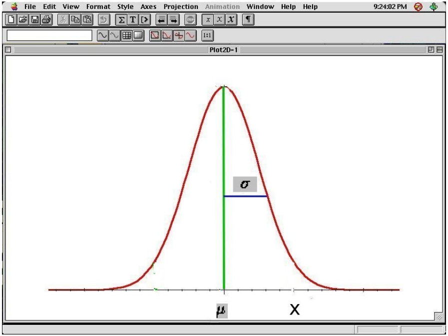
A general normal distribution:

$N(\mu, \sigma)$



Z-scores

$N(\mu, \sigma)$



Can convert from a general $N(\mu, \sigma)$ to $N(0, 1)$ via the Z-score.

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

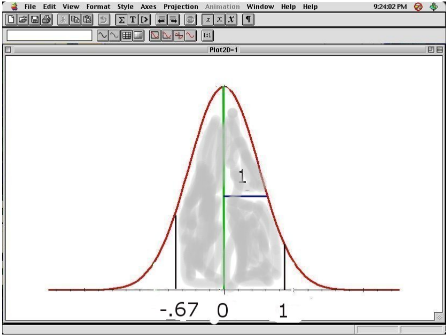
UsingTable Z

For example $P(-.67 < Z < 1) = ?$

Using Table Z

For example $P(-.67 < Z < 1) = ?$

N(0,1)



Using Table Z

For example $P(-.67 < Z < 1) = ?$

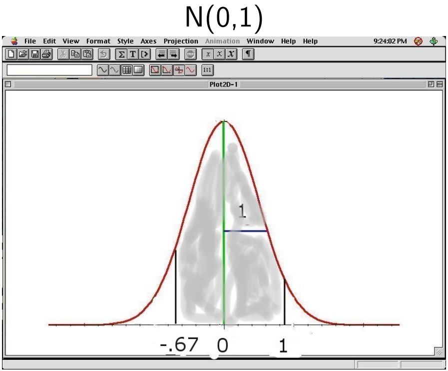


Table Z tells us $P(-.67 < Z < 1) = .8413 - .2514 = .5899.$

Success/Failure Condition

When n is large, binomial distributions can usually be approximated fairly well by normal ones.

Success/Failure Condition

When n is large, binomial distributions can usually be approximated fairly well by normal ones.

Criterion: $np \geq 10$ and $nq \geq 10$.

Success/Failure Condition

When n is large, binomial distributions can usually be approximated fairly well by normal ones.

Criterion: $np \geq 10$ and $nq \geq 10$.

np is the expected number of *successes.*

Success/Failure Condition

When n is large, binomial distributions can usually be approximated fairly well by normal ones.

Criterion: $np \geq 10$ and $nq \geq 10$.

np is the expected number of *successes.*

Explanation is based on some calculus ideas (like $\ln(1+x) \sim x$) applied to estimating ratios of factorials.

Success/Failure Condition

When n is large, binomial distributions can usually be approximated fairly well by normal ones.

Criterion: $np \geq 10$ and $nq \geq 10$.

np is the expected number of *successes.*

Explanation is based on some calculus ideas (like $\ln(1+x) \sim x$) applied to estimating ratios of factorials.

An outline of the case $p = .5$ is at the end of this lecture.

Suppose 70% approve the President . . .

You poll 100 people.

What is the probability that between 60 and 65 report approval?

(including 60 and 65.)

Suppose 70% approve the President . . .

You poll 100 people.

What is the probability that between 60 and 65 report approval?

(including 60 and 65.)

Solution: Let $X = Binomial(100, .7)$

Suppose 70% approve the President . . .

You poll 100 people.
What is the probability that between 60 and 65 report approval?
(including 60 and 65.)
Solution: Let $X = \mathit{Binomial}(100, .7)$

X has mean

$$\mu = (100)(.7) = 70$$

and

$$\sigma = \sqrt{(100)(.7)(.3)} = \sqrt{21} = 4.58.$$

Suppose 70% approve the President . . .

You poll 100 people.

What is the probability that between 60 and 65 report approval?

(including 60 and 65.)

Solution: Let $X = \mathit{Binomial}(100, .7)$

X has mean

$$\mu = (100)(.7) = 70$$

and

$$\sigma = \sqrt{(100)(.7)(.3)} = \sqrt{21} = 4.58.$$

X will then be approximated by a *normal* distribution Y with the same mean and standard deviation.

Suppose 70% approve the President . . .

You poll 100 people.

What is the probability that between 60 and 65 report approval?

(including 60 and 65.)

Solution: Let $X = Binomial(100, .7)$

X has mean

$$\mu = (100)(.7) = 70$$

and

$$\sigma = \sqrt{(100)(.7)(.3)} = \sqrt{21} = 4.58.$$

X will then be approximated by a *normal* distribution Y with the same mean and standard deviation.

$Y = N(70, 4.58)$.

Suppose 70% approve the President ...

You poll 100 people.
What is the probability that between 60 and 65 report approval?
(including 60 and 65.)
Solution: Let $X = \mathit{Binomial}(100, .7)$

X has mean

$$\mu = (100)(.7) = 70$$

and

$$\sigma = \sqrt{(100)(.7)(.3)} = \sqrt{21} = 4.58.$$

X will then be approximated by a *normal* distribution Y with the same mean and standard deviation.

$Y = N(70, 4.58)$.
We can approximate $P(60 \leq X \leq 65)$ by $P(60 \leq Y \leq 65)$.

Suppose 70% approve the President . . .

You poll 100 people.
What is the probability that between 60 and 65 report approval?
(including 60 and 65.)

X will then be approximated by a *normal* distribution Y with the same mean and standard deviation.

$Y = N(70, 4.58)$.
We can approximate $P(60 \leq X \leq 65)$ by $P(60 \leq Y \leq 65)$.
The Z score of 60 is $\frac{-10}{4.58} = -2.18$.
The Z score of 65 is $\frac{-5}{4.58} = -1.09$.
So $P(60 \leq Y \leq 65) = P(Z < -1.09) - P(Z < -2.18) = .1379 - .0146 = .1233$.

Suppose 70% approve the President . . .

You poll 100 people.

What is the probability that between 60 and 65 report approval?

(including 60 and 65.)

We can approximate $P(60 \le X \le 65)$ by $P(60 \le Y \le 65)$.

The Z score of 60 is $\frac{-10}{4.58} = -2.18$.

The Z score of 65 is $\frac{-5}{4.58} = -1.09$.

So $P(60 \le Y \le 65) = P(Z < -1.09) - P(Z < -2.18) = .1379 - .0146 = .1233$.

Actually $P(60 \le X \le 65) = .15036$.

So .1233 is not that good an approximation.

Suppose 70% approve the President . . .

You poll 100 people.

What is the probability that between 60 and 65 report approval?

(including 60 and 65.)

We can approximate $P(60 \leq X \leq 65)$ by $P(60 \leq Y \leq 65)$.

The Z score of 60 is $\frac{-10}{4.58} = -2.18$.

The Z score of 65 is $\frac{-5}{4.58} = -1.09$.

So $P(60 \leq Y \leq 65) = P(Z < -1.09) - P(Z < -2.18) = .1379 - .0146 = .1233$.

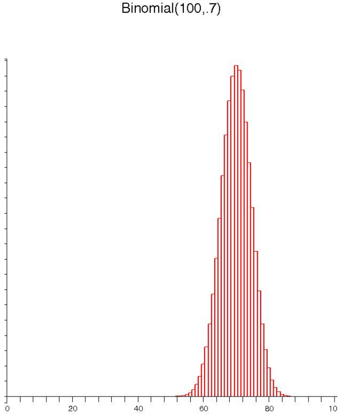
Actually $P(60 \leq X \leq 65) = .15036$.

So .1233 is not that good an approximation.

$P(60 < X < 65) = .09509$. would also be approximated by .1233!

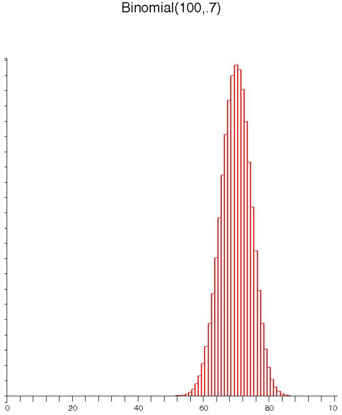
Normal Approximation of X=Binomial(100,.7)

For each $x = 0, 1, 2, \ldots, 100$,
a rectangle with base 1 and height P(X=x) is drawn:



Normal Approximation of X=Binomial(100,.7)

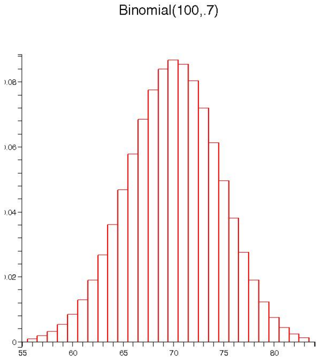
For each $x = 0, 1, 2, \ldots, 100$,
a rectangle with base 1 and height P(X=x) is drawn:



This makes the area of each bar a probability!

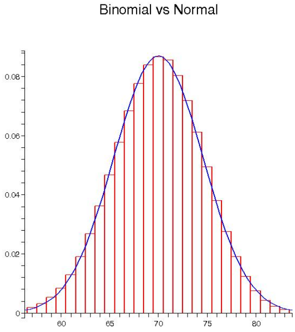
Normal Approximation of X=Binomial(100,.7)

Most of the action between 55 and 85.
Look at that part:



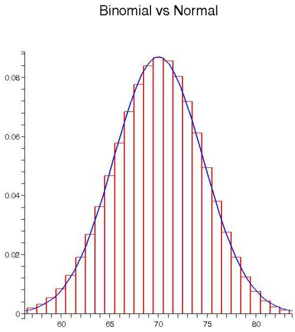
Normal Approximation of X=Binomial(100,.7)

With N(70,4.58) superimposed.



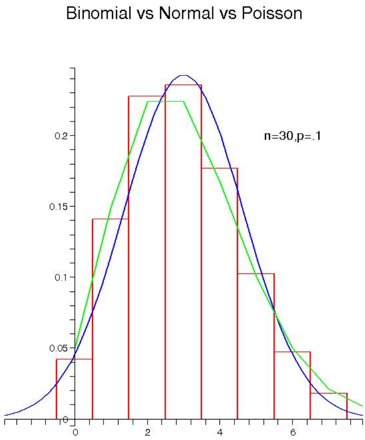
Normal Approximation of X=Binomial(100,.7)

With N(70,4.58) superimposed.

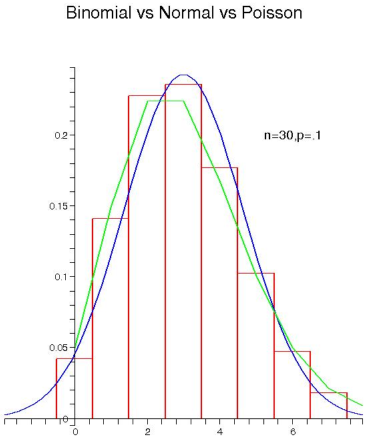


This is the good case when normal approximation works.

Normal Approximation of X=Binomial(30,.1)

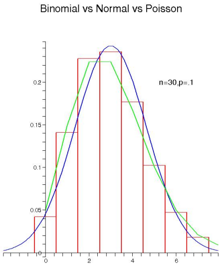


Normal Approximation of X=Binomial(30,.1)



$\mu = 3$ and $\sigma = \sqrt{30(.1)(.9} = 1.64.$

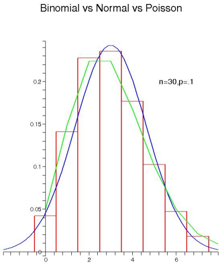
Normal Approximation of X=Binomial(30,.1)



$\mu = 3$ and $\sigma = \sqrt{30(.1)(.9} = 1.64.$

Two standard deviations to the left of the mean and our normal model is reporting negative values for the number of successes!

Normal Approximation of X=Binomial(30,.1)



$\mu = 3$ and $\sigma = \sqrt{30(.1)(.9} = 1.64$.

This is typical of what happens when success/failure is not satisfied.

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$

$P(60 \leq Y \leq 65) = .1233$
is not that good an approximation to either.

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$

$P(60 \leq Y \leq 65) = .1233$
is not that good an approximation to either.

The problem is that for a continuous model like Y,
$P(Y = 65) = 0$.
But $P(X = 65) = .04678$.

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

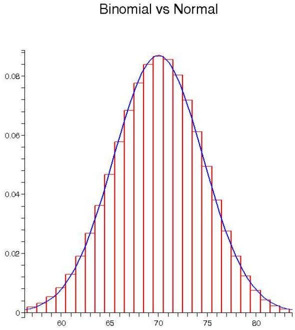
$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$

$P(60 \leq Y \leq 65) = .1233$
So Y doesn't care about $<$ vs. $\leq$. But X does!

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

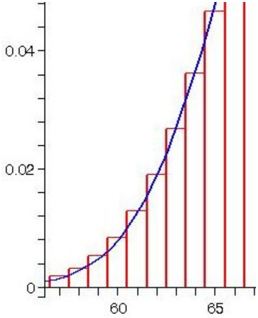
$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$
$P(60 \leq Y \leq 65) = .1233$

Looking closely at the picture provides the key:



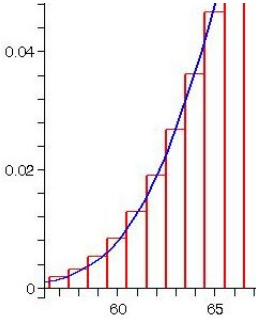
$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$
$P(60 \leq Y \leq 65) = .1233$



$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$
$P(60 \leq Y \leq 65) = .1233$



It is really the area under the normal curve between 64.5 and 65.5 which approximates P(X=65).

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$
$P(60 \leq Y \leq 65) = .1233$

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$
$P(60 \leq Y \leq 65) = .1233$
So $P(60 \leq X \leq 65) = .15036$ should be approximated by $P(59.5 \leq Y \leq 65.5)$.

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$
$P(60 \leq Y \leq 65) = .1233$
So $P(60 \leq X \leq 65) = .15036$ should be approximated by $P(59.5 \leq Y \leq 65.5)$.
The z-scores of 59.5 and 65.5 are
$\frac{-10.5}{4.58} = -2.29$ and $\frac{-4.5}{4.58} = -.98$ resp.

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$
$P(60 \leq Y \leq 65) = .1233$
So $P(60 \leq X \leq 65) = .15036$ should be approximated by $P(59.5 \leq Y \leq 65.5)$.
The z-scores of 59.5 and 65.5 are
$\frac{-10.5}{4.58} = -2.29$ and $\frac{-4.5}{4.58} = -.98$ resp.
So $P(59.5 \leq Y \leq 65.5) = P(-2.29 \leq Z \leq -.98) = .1635 - .0110 = .1525.$
Much closer to $P(60 \leq X \leq 65) = .15036$!

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$
$P(60 \leq Y \leq 65) = .1233$
So $P(60 \leq X \leq 65) = .15036$ should be approximated by $P(59.5 \leq Y \leq 65.5)$.
Similarly $P(60.5 \leq Y \leq 64.5)$
$= P(Z < -1.20) - P(Z < -2.07) = .1151 - .0192 = .0959$ is a great approximation to $P(60 < X < 65) = .09509$.

$X = Binomial(100, .7)$ and $Y = N(70, 4.58)$.

$P(60 \leq X \leq 65) = .15036$
$P(60 < X < 65) = .09509$
$P(60 \leq Y \leq 65) = .1233$
So $P(60 \leq X \leq 65) = .15036$ should be approximated by $P(59.5 \leq Y \leq 65.5)$.
This is called the continuity approximation.
You needn't use it on exams or homework!
(We suggest you not.)

Nearly Normal Condition

Your textbook uses this as a synonym for a

unimodal and symmetric distribution.

Nearly Normal Condition

Your textbook uses this as a synonym for a

unimodal and symmetric distribution.

This doesn't mean unimodal and symmetric distributions are necessarily well approximated by normal ones.

Nearly Normal Condition

Your textbook uses this as a synonym for a

unimodal and symmetric distribution.

This doesn't mean unimodal and symmetric distributions are necessarily well approximated by normal ones.

Here's a symmetric and unimodal example to illustrate this:

Nearly Normal Condition

n=318: -10,-9,-8, . . .-2, 100 -1's, 100 0's, . . .

tl+mid

Frequency breakdown of tl+mid

No Selector

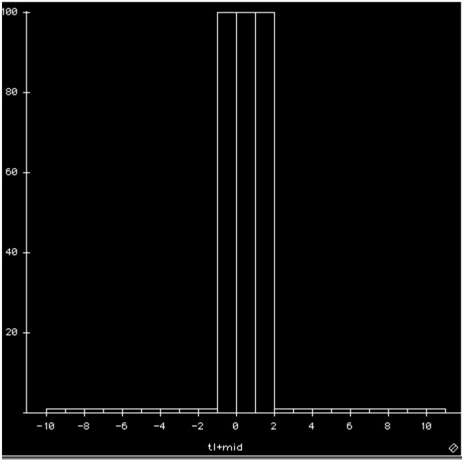
Total Cases 318

Number of Categories 21

Group	Count
-10	1
-9	1
-8	1
-7	1
-6	1
-5	1
-4	1
-3	1
-2	1
-1	100
0	100
1	100
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1

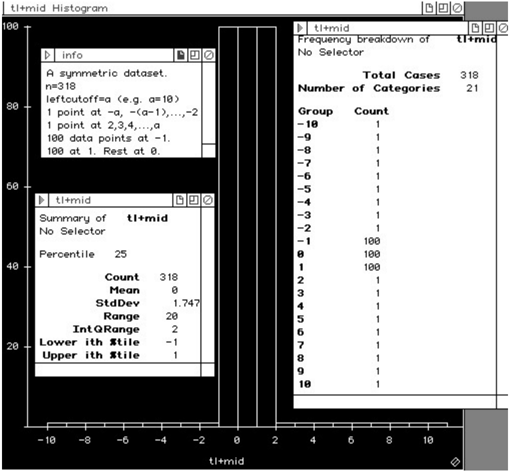
Nearly Normal Condition

A Symmetric and Unimodal Histogram



Nearly Normal Condition

Summaries



Nearly Normal Condition

Summary Statistics: $n = 318$, $\bar{x} = 0$, $s = 1.747$

-10,-9,-8, . . . -2, 100 -1's, 100 0's, . . .

tl+mid

Summary of **tl+mid**
No Selector

Percentile 25

Count	318
Mean	0
StdDev	1.747
Range	20
IntQRange	2
Lower ith %tile	-1
Upper ith %tile	1

Nearly Normal Condition

Summary Statistics: $n = 318$, $\bar{x} = 0$, $s = 1.747$

-10,-9,-8, . . . -2, 100 -1's, 100 0's, . . .

tl+mid

Summary of **tl+mid**
No Selector

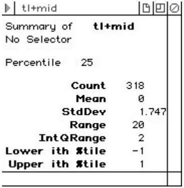
Percentile 25

Count	318
Mean	0
StdDev	1.747
Range	20
IntQRange	2
Lower ith %tile	-1
Upper ith %tile	1

Nearly Normal Condition

Summary Statistics: $n = 318$, $\bar{x} = 0$, $s = 1.747$

-10,-9,-8, . . . -2, 100 -1's, 100 0's, . . .



tl+mid

Summary of tl+mid
No Selector

Percentile 25

Count	318
Mean	0
StdDev	1.747
Range	20
IntQRange	2
Lower ith %tile	-1
Upper ith %tile	1

Only $18 = 2 \cdot 9$ of 318 values
are more than 1 std. dev. (1.747) from $\bar{x} = 0$.

Nearly Normal Condition

Summary Statistics: $n = 318$, $\bar{x} = 0$, $s = 1.747$

-10,-9,-8, . . . -2, 100 -1's, 100 0's, . . .

tl+mid

Summary of tl+mid
No Selector

Percentile 25

Count	318
Mean	0
StdDev	1.747
Range	20
IntQRange	2
Lower ith %tile	-1
Upper ith %tile	1

Only $18 = 2 \cdot 9$ of 318 values
are more than 1 std. dev. (1.747) from $\bar{x} = 0$.
94% within 1 std dev vs. 68% for normal!

Nearly Normal Condition

Summary Statistics: $n = 318$, $\bar{x} = 0$, $s = 1.747$
-10,-9,-8, . . . -2, 100 -1's, 100 0's, . . .

tl+mid

Summary of **tl+mid**
No Selector

Percentile 25

Count	318
Mean	0
StdDev	1.747
Range	20
IntQRange	2
Lower ith %tile	-1
Upper ith %tile	1

94% within 1 std dev vs. 68% for normal!
4^{+}% more than 3 std dev vs. 0.3% for normal!

Nearly Normal Condition

Your textbook uses this as a synonym for a

unimodal and symmetric distribution.

This doesn't mean unimodal and symmetric distributions are necessarily well approximated by normal ones.

It is true that unimodal and symmetric does mean normality might be at least roughly indicative.

Some statistical methods derived under assumptions of normality can be shown to usually work more generally. I think this is what the authors really have in mind.

Chapter 17 #31 (Lefties)

- 200 seats
- 30 for left-handers
- class size 188
- on avg., 13% left handers

Chapter 17 #31 (Lefties)

- 200 seats
- 30 for left-handers
- class size 188
- on avg., 13% left handers

Prob. run out of right hand seats?

Chapter 17 #31 (Lefties)

- 200 seats
- 30 for left-handers
- class size 188
- on avg., 13% left handers

Prob. run out of right hand seats?

Model num *left* hnd students by X=Binomial(188,.13).

Chapter 17 #31 (Lefties)

- 200 seats
- 30 for left-handers
- class size 188
- on avg., 13% left handers

Prob. run out of right hand seats?

Model num *left* hnd students by X=Binomial(188,.13).

$\mu = 188 \cdot .13 = 24.44$, $\sigma = \sqrt{188 \cdot .13 \cdot .87} = 4.61$.

Chapter 17 #31 (Lefties)

- 200 seats
- 30 for left-handers
- class size 188
- on avg., 13% left handers

Prob. run out of right hand seats?

Model num *left* hnd students by X=Binomial(188,.13).

$\mu = 188 \cdot .13 = 24.44$, $\sigma = \sqrt{188 \cdot .13 \cdot .87} = 4.61$.

Approximate by $Y = N(24.44, 4.61)$.

Chapter 17 #31 (Lefties)

- 200 seats
- 30 for left-handers
- class size 188
- on avg., 13% left handers

Prob. run out of right hand seats?

Model num *left* hnd students by X=Binomial(188,.13).

$\mu = 188 \cdot .13 = 24.44$, $\sigma = \sqrt{188 \cdot .13 \cdot .87} = 4.61$.

Approximate by $Y = N(24.44, 4.61)$.

Calculate $P(Y < 18)$.

Chapter 17 #33 (Telemarketers)

- Story: 12% success rate on calls.
- Experience: 10 successes out of 200 calls.
- (first week)

Chapter 17 #33 (Telemarketers)

- Story: 12% success rate on calls.
- Experience: 10 successes out of 200 calls.
- (first week)

Likely Misled?

Chapter 17 #33 (Telemarketers)

- Story: 12% success rate on calls.
- Experience: 10 successes out of 200 calls.
- (first week)

Likely Misled?

Chapter 17 #33 (Telemarketers)

- Story: 12% success rate on calls.
- Experience: 10 successes out of 200 calls.
- (first week)

Likely Misled?

Model by X=Binomial(200,.12)

Reasonable?

Chapter 17 #33 (Telemarketers)

- Story: 12% success rate on calls.
- Experience: 10 successes out of 200 calls.
- (first week)

Likely Misled?

Model by X=Binomial(200,.12)

$\mu = 24$, $\sigma = \sqrt{200 \cdot .12 \cdot .88} = 4.60$.

Chapter 17 #33 (Telemarketers)

- Story: 12% success rate on calls.
- Experience: 10 successes out of 200 calls.
- (first week)

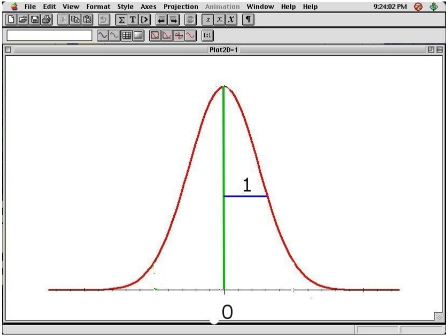
Likely Misled?
Model by X=Binomial(200,.12)
$\mu = 24$, $\sigma = \sqrt{200 \cdot .12 \cdot .88} = 4.60$.
10 is more than 3 standard deviations below the mean, so not very likely.
So model (story) likely inappropriate.

Chapter 17 #33 (Telemarketers)

- Story: 12% success rate on calls.
- Experience: 10 successes out of 200 calls.
- (first week)

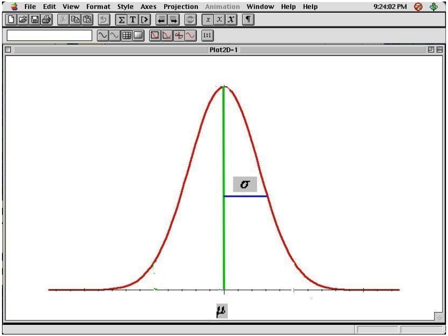
Likely Misled?
Model by X=Binomial(200,.12)
$\mu = 24$, $\sigma = \sqrt{200 \cdot .12 \cdot .88} = 4.60$.
So model (story) likely inappropriate.
But for 1000 salesmen starting the same week, it wouldn't be surprising for one to have this experience even if the model is right.

Normal Distribution Formula

N(0,1) 

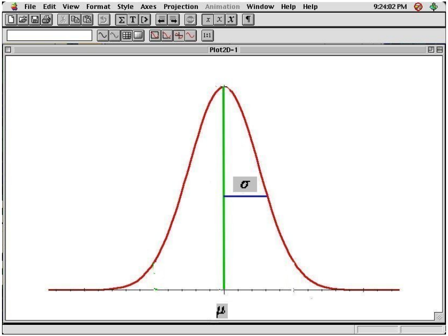
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Normal Distribution Formula

$N(\mu, \sigma)$ 

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

Normal Distribution Formula

$N(\mu, \sigma)$ 

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

These formulas and the following argument are far above the basic level of our course.

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$
$\mu = m$ and $\sigma = \sqrt{.5m}$.

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Set

$$a_k = P(X = m + k) = \binom{2m}{m+k} (.5)^{2m}$$

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Set

$$a_k = P(X = m + k) = \binom{2m}{m+k} (.5)^{2m}$$

The z-score of $m + k$ is $\frac{k\sqrt{2}}{\sqrt{m}}$.

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Set

$$a_k = P(X = m + k) = \binom{2m}{m+k} (.5)^{2m}$$

The z-score of $m + k$ is $\frac{k\sqrt{2}}{\sqrt{m}}$.

So we want to show $a_k \sim ce^{-\frac{k^2}{m}}$.

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Set

$$a_k = P(X = m + k) = \binom{2m}{m+k} (.5)^{2m}$$

So we want to show $a_k \sim ce^{-\frac{k^2}{m}}$.

i.e. $\ln a_k \sim \ln c - \frac{k^2}{m}$.

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Set

$$a_k = P(X = m + k) = \binom{2m}{m+k} (.5)^{2m}$$

i.e. $\ln a_k \sim \ln c - \frac{k^2}{m}$.

Strategy: Compare a_k to a_0 using the approximation $\ln(1 + x) \sim x$ for x small.

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Set

$$a_k = P(X = m + k) = \binom{2m}{m+k}(.5)^{2m}$$

i.e. $\ln a_k \sim \ln c - \frac{k^2}{m}$.

$$a_k = \frac{(2m)!(.5)^{2m}}{(m+k)!(m-k)!} = a_0 \frac{(m)(m-1)\ldots(m-k+1)}{(m+k)(m+k-1)\ldots(m+1)}$$

$$= a_0 \frac{(1)(1-\frac{1}{m})\ldots(1-\frac{k-1}{m})}{(1+\frac{k}{m})(1+\frac{k-1}{m})\ldots(1+\frac{1}{m})}$$

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Set

$$a_k = P(X = m+k) = \binom{2m}{m+k}(.5)^{2m}$$

i.e. $\ln a_k \sim \ln c - \frac{k^2}{m}$.

$$a_k = \frac{(2m)!(.5)^{2m}}{(m+k)!(m-k)!} = a_0 \frac{(m)(m-1)\ldots(m-k+1)}{(m+k)(m+k-1)\ldots(m+1)}$$

$$= a_0 \frac{(1)(1-\frac{1}{m})\ldots(1-\frac{k-1}{m})}{(1+\frac{k}{m})(1+\frac{k-1}{m})\ldots(1+\frac{1}{m})}$$

So using $\ln(1+x) \sim x$,

$$\ln a_k \sim \ln a_0 - 2\left(\frac{1}{m} + \frac{2}{m} + \ldots + \frac{k-1}{m}\right) - \frac{k}{m}$$

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Set

$$a_k = P(X = m + k) = \binom{2m}{m+k} (.5)^{2m}$$

i.e. $\ln a_k \sim \ln c - \frac{k^2}{m}$.

So using $\ln(1 + x) \sim x$,

$$\ln a_k \sim \ln a_0 - 2\left(\frac{1}{m} + \frac{2}{m} + \ldots + \frac{k-1}{m}\right) - \frac{k}{m}$$

But $1 + 2 + \ldots k - 1 = \frac{k(k-1)}{2}$, so

Why Normal Out of Binomial?

Suppose $X \sim Binom(2m, .5)$

We want to understand why $X \sim N(m, \sqrt{.5m})$ approximately.

Set

$$a_k = P(X = m + k) = \binom{2m}{m+k} (.5)^{2m}$$

i.e. $\ln a_k \sim \ln c - \frac{k^2}{m}$.

So using $\ln(1 + x) \sim x$,

$$\ln a_k \sim \ln a_0 - 2\left(\frac{1}{m} + \frac{2}{m} + \ldots + \frac{k-1}{m}\right) - \frac{k}{m}$$

But $1 + 2 + \ldots k - 1 = \frac{k(k-1)}{2}$, so

$$\ln a_k \sim \ln a_0 - \frac{k^2}{m}$$

as desired.

C

C