The Normal Curve

Where does the rule of thumb come from?

Making Normal approximation.

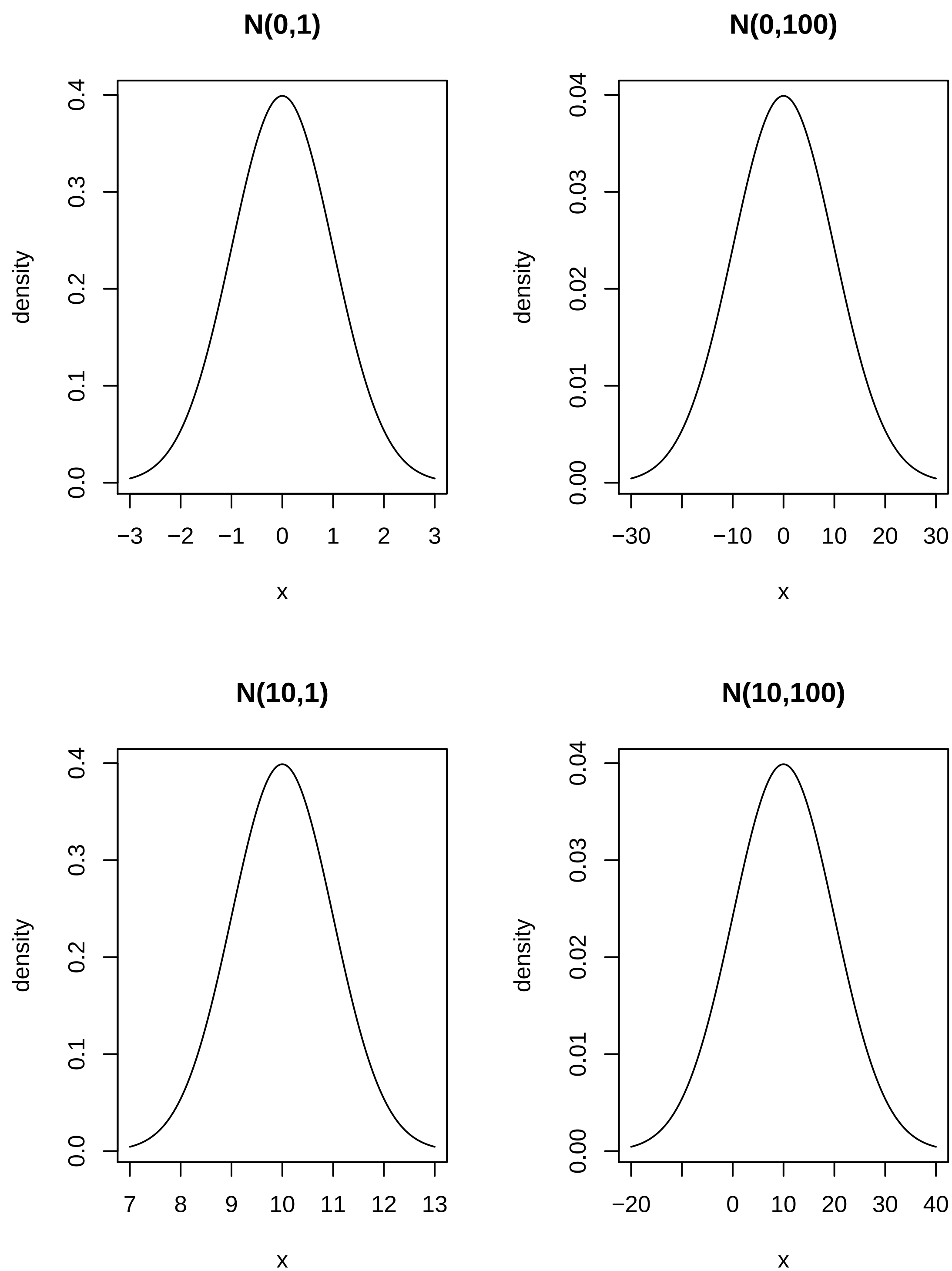
Draw smooth curve on top of histogram.

One curve for each mean and SD.

Formula:

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

Use of formula: NONE in this course. Instead use tables or computers to compute areas under curve (Integrals!)


N(0,1)
N(0,100)
N(10,1)
N(10,100)
density
x
0.0 0.1 0.2 0.3 0.4
0.00 0.01 0.02 0.03 0.04
−3 −2 −1 0 1 2 3
−30 −10 0 10 20 30
7 8 9 10 11 12 13
−20 0 10 20 30 40

Notice: all curves look the same when axes drawn to corresponding heights!

Now superimpose normal curve over a data set.

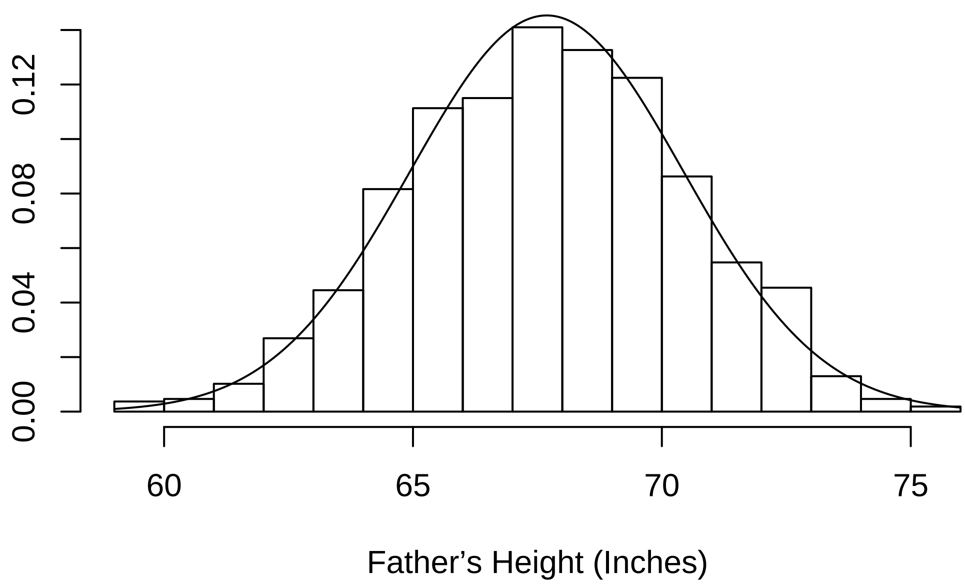
Data come from study of heights of adult men in England in 19th century.

Sample of 1078 Father-Son pairs

Summary statistics – all measurements in inches:

	Fathers	Sons
Minimum	59.01	58.51
1st Quartile	65.79	66.93
Median	67.77	68.62
Mean	67.69	68.68
3rd Quartile	69.60	70.47
Maximum	75.43	78.36
SD	2.74	2.81

Now look at histogram of Fathers' heights with normal curve drawn on top:



Notice general shapes similar.

Use: proportion of fathers with height in given range is AREA under histogram in range.

Approximate this area by area under normal curve.

Total area under histogram is 1 if units on vertical axis chosen as “density” (proportion per x unit).

Total area under normal curve is 1. (Fact from 2nd year calculus.)

Making a normal approximation:

1) Sketch curve.

2) Label x axis and mark desired range.

3) Convert range to standard units: subtract mean from x values and divide by SD.

4) Look up area under standard normal curve using these standardized limits. See Table A in text.

Example: Proportion of father's under 5 feet 10 inches = 70 inches.

SKETCH:

Desired range: area under curve left of 70 inches.

Convert 70 to standard units:

$$\frac{70 - \bar{x}}{s} = \frac{70 - 67.69}{2.74} = 0.84$$

Look up area to left of 0.84 under normal curve.

Get approximately 0.7995 $\approx$ 80% of fathers under 5 foot 10. This is 80% of 1078 or 862 fathers.

Actual number is 856 fathers or 79.4%

Extract of Table A:

z	$\cdots$	.04	.05	$\cdots$
-3.4	$\cdots$	.0003	.0003	$\cdots$
$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$
0.8	$\cdots$	.7995	.8023	$\cdots$
$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$

The numbers in the centre are areas under the normal curve to the left of z.

Value of z is number in column under z followed by last digit in top row.

For example: area to left of -3.45 is under .05 and across from -3.4 (so is .0003).

Some areas under the normal curve from the tables:

Left of 0	50%
Right of 0	50%
Between -1 and 1	68.3% $\approx 2/3$
Between -2 and 2	95.4% $\approx$ 95%
Between -3 and 3	99.7%
Between -4 and 4	99.994%
Between -6 and 6	$1\text{-}1.97\times 10^{-9}$

Notice source of rule of thumb.

Finding areas:

Tables show areas to left of standard value:

Get other areas by subtracting:

Area to left of 2 is 97.72%

Area to left of 0 is 50.00%

So: area from 0 to 2 is difference: 47.72%

One more example: fathers between 5 foot 2 and 5 foot 10

Convert 62 inches and 70 inches to standard units.

$$\frac{62-\bar{x}}{s} = -2.07 \quad \frac{70-\bar{x}}{s} = 0.84$$

Area to left of 0.84 is 0.7995.

Area to left of -2.07 is 0.0192.

Subtract to get 0.7803≈78%

Exact answer is 77.6%.

Normal approximation applied to income: poor.

Proportion of adults earning under \$30,000?

Mean income is \$29,250 approximately.

SD of income is \$23,600 approximately.

Convert \$30,000 to standard units:

$$\frac{30000 - 29250}{23600} = 0.03$$

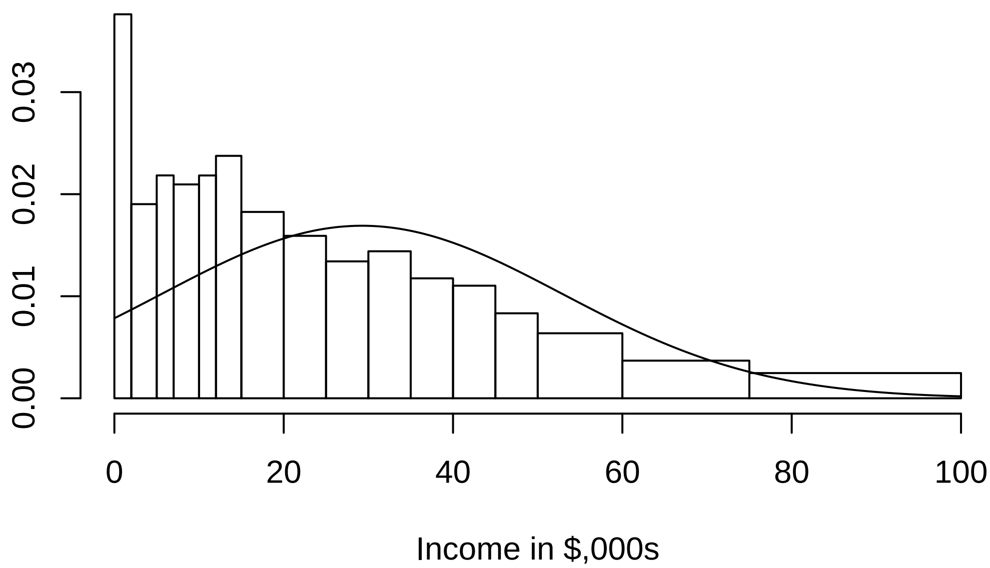
Area to left of 0.03 is 51%

Correct percentage is 59%.

Income distribution is "skewed to the right".

It has a 'long right hand tail'.

Incomes with normal curve on top:



Notice that normal curve extends below 0.

Normal approximation predicts many negative incomes!

Reversing the process. What is the IQR of fathers heights?

First quartile of standard normal curve: -0.67

Third quartile is 0.67.

Convert back to original units: multiply standard units by s and add $\bar{x}$.

So: -0.67 Standard units is

$$-0.67 * 2.74 + 67.69 = 65.85$$

and 0.67 Standard units is

$$0.67 * 2.74 + 67.69 = 69.53$$

So IQR is approximately 69.53-65.85=3.68. Actual value 3.81.

Summary:

Standard units:

$$z = \frac{x - \bar{x}}{s}$$

Look up areas by converting limits to standard units and using standard normal curve.

From standard units z back to original x values:

$$x = s * z + \bar{x}$$