Lecture 9

CONTINUOUS PROBABILITY MODELS

Continuous probability models

We have seen how **discrete** random variables can be modelled by discrete probability distributions such as the **binomial** and **Poisson** distributions.

We now consider how to model **continuous** random variables.

A variable is **discrete** if it takes a **countable** number of values.

For example,

- the number of **blue** cars that I count in a 5 minute period
- the number of **heads** observed when I flip a coin ten times
- Shoe sizes: $1, \ldots, 12, 13, 1, 2, \ldots$
- $r = 0, 0.1, 0.2, \ldots, 0.9, 1.0$

In contrast, the values which a **continuous variable** can take form a **continuous scale**, with no "jumps".

For example,

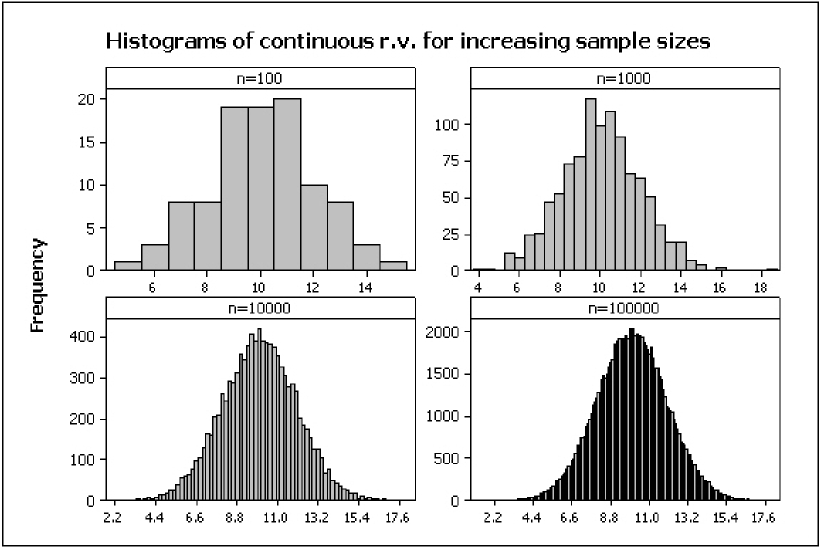
- Height
- Weight
- Temperature

An example

Think about **height**.

- In practice, we might only record height to the **nearest cm**
- If we could measure height *exactly* we'd find that everyone had a **different height**
- This is the essential difference between discrete and continuous variables
- If there are n people on the planet, the probability that someone's height is x would be $\frac{1}{n}$
- As n gets bigger and bigger, this probability tends to zero!!

Consider taking a sample of values from the continuous random variable **X**. This is what we'd observe as our sample got bigger and bigger:



- As the sample size gets bigger, the **interval widths get smaller**
- the jagged profile of the histogram **smooths out** to become a curve
- When the sample size is infinitely large, this curve is known as the **probability density function** (pdf)

Features of the probability density function

The key features of pdfs are:

1. pdfs never take negative values
2. the area under a pdf is one: $P(-\infty < X < \infty) = 1$
3. areas under the curve correspond to probabilities
4. $P(X \leq x) = P(X < x)$ since $P(X = x) = 0$.

Over the next two weeks we will consider some particular probability distributions that are often used to describe continuous random variables.

We start with the most **important**, most **widely–used** statistical distribution of all time...

...wait for it...

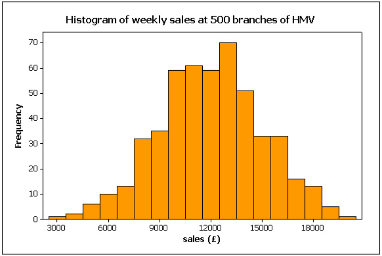
The Normal Distribution

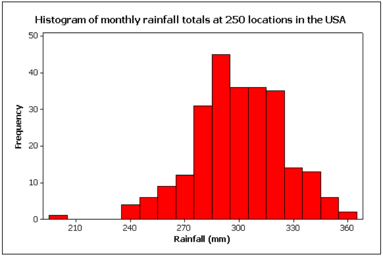
The Normal distribution

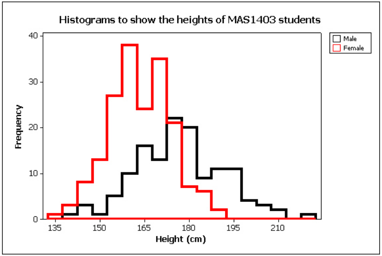
The **Normal distribution** is without doubt the most widely–used statistical distribution in many practical applications:

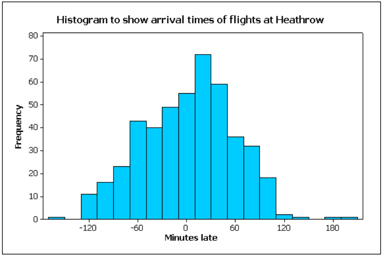
- Normality arises **naturally** in many physical, biological and social measurement situations
- Normality is important in **Statistical inference** *(see Semester 2 material)*
- The normal distribution has many guises:
  - Gaussian distribution
  - Laplacean distribution
  - "bell–shaped curve"

Some real–life examples





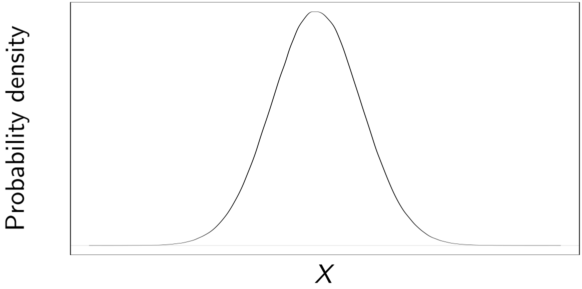




Recall the "parameters" of the binomial and Poisson distributions:

- the **binomial distribution** has two parameters, n and p
- the **Poisson distribution** has one parameter λ
- The **Normal distribution** has two parameters: the mean, μ, and the standard deviation, σ

Its probability density function (pdf) has a "**bell–shaped**" profile (page 105):



The (rather nasty!) formula for this pdf is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

Unlike the binomial and Poisson distributions, there is **no simple formula** for calculating probabilities.

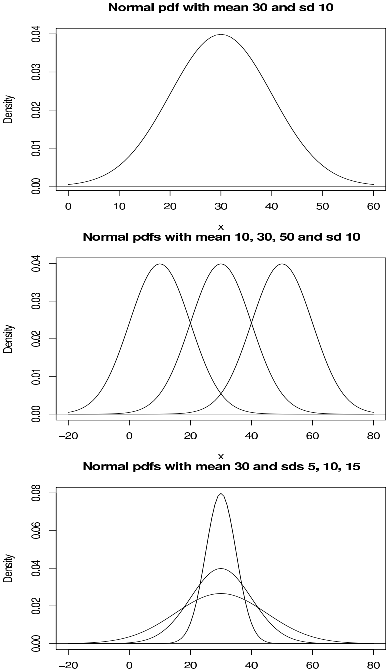
Don't worry though, probabilities from the Normal distribution can be determined using **statistical tables** (see the end of this chapter) or statistical packages such as `Minitab`.

Characteristics of the Normal distribution

There are four important characteristics of the Normal distribution:

1. It is **symmetrical** about its mean, μ.
2. The mean, median and mode all **coincide**.
3. The area under the curve is equal to 1.
4. The curve extends in both directions to infinity (∞).

On the next slide are plots of the pdf for Normal distributions with different values of μ and σ.


Normal pdf with mean 30 and sd 10
Density
0.00
0.01
0.02
0.03
0.04
0
10
20
30
40
50
60
x
Normal pdfs with mean 10, 30, 50 and sd 10
Density
0.00
0.01
0.02
0.03
0.04
−20
0
20
40
60
80
x
Normal pdfs with mean 30 and sds 5, 10, 15
Density
0.00
0.02
0.04
0.06
0.08
−20
0
20
40
60
80

Notation

If a random variable X has a Normal distribution with mean μ and variance σ^2, then we write

$$X \sim N\left(\mu, \sigma^2\right).$$

For example, a random variable X which follows a Normal distribution with mean **10** and variance **25** is written as

$$\begin{aligned} X &\sim N(10, 25) \qquad \text{or} \\ X &\sim N\left(10, 5^2\right). \end{aligned}$$

It is important to note that the second parameter in this notation is the **variance** and not the **standard deviation**.

The *standard* Normal distribution

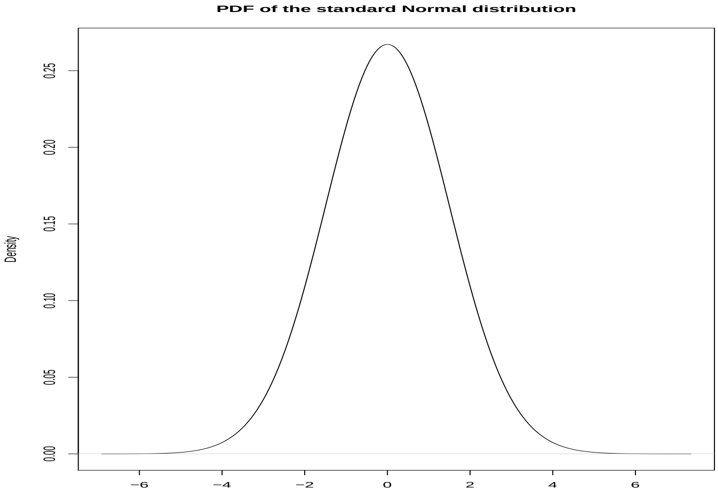
The **Standard** Normal distribution has a mean of $\mathbf{0}$ and a variance of $\mathbf{1}$.

A random variable with this *standard* Normal distribution is usually given the letter Z, and so we say

$$Z \sim N(\mathbf{0}, \mathbf{1}).$$

If our random variable follows a **standard** Normal distribution, then we can obtain *cumulative probabilities* from statistical tables (see the table at the end of this chapter, which give "**less than or equal to**" probabilities).

Probability density function for Z



For example, if $\mathbf{Z \sim N(0,1)}$:

1. The probability that Z is less than -1.46 is $P(Z < -1.46)$. Therefore we look for the probability in tables corresponding to $z = -1.46$: row labelled -1.4, column headed -0.06. This gives $P(Z < -1.46) = \mathbf{0.0721}$.

2. The probability that Z is less than -0.01 is $P(Z < -0.01)$. Therefore we look for the probability in tables corresponding to $z = -0.01$: row labelled 0.0, column headed -0.01. This gives $P(Z < -0.01) = \mathbf{0.4960}$.

3. Similarly, $P(Z < 0.01) = \mathbf{0.5040}$.

z	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	0.00
-2.9	0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019
-2.8	0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026
-2.7	0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035
-2.6	0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047
-2.5	0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062
-2.4	0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082
-2.3	0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107
-2.2	0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139
-2.1	0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179
-2.0	0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228
-1.9	0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287
-1.8	0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359
-1.7	0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446
-1.6	0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548
-1.5	0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668
-1.4	0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808
-1.3	0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968
-1.2	0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151
-1.1	0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357
-1.0	0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587
-0.9	0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841
-0.8	0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119
-0.7	0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420
-0.6	0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743
-0.5	0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085
-0.4	0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446
-0.3	0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821
-0.2	0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207
-0.1	0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602
-0.0	0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000

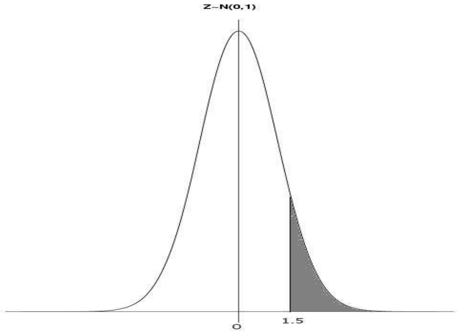
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	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
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

So far so good? Hopefully! But what if we want a "**greater than**" probability? These tables only give "**less than**" probabilities!

Easy! Remember,

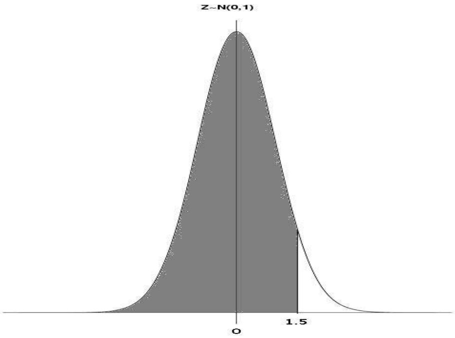
- The area under the entire curve is equal to 1
- So we could find the "**less than**" probability and then subtract from 1 to get what's left over!

4. The probability that Z is greater than 1.5 is $P(Z > 1.5)$. Now our tables give "less than" probabilities, and here we want a "greater than" probability.

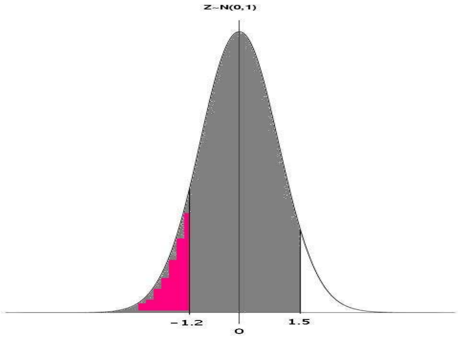


So we find $P(Z < 1.5) = 0.9332$ and subtract this from 1 to give $\mathbf{0.0668}$.

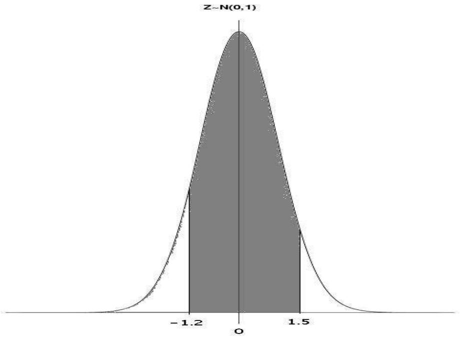
5. What about the probability that Z lies between -1.2 and 1.5? It often helps to think about this graphically.



5. What about the probability that Z lies between -1.2 and 1.5? It often helps to think about this graphically.



5. What about the probability that Z lies between -1.2 and 1.5? It often helps to think about this graphically.



Doing so, gives

$$\begin{aligned} P(-1.2 < Z < 1.5) &= P(Z < 1.5) - P(Z \le -1.2) \\ &= 0.9332 - 0.1151 \\ &= 0.8181. \end{aligned}$$

So how do we calculate probabilities for **any** Normal distribution, not just the **standard** Normal distribution – for which we have tables?

Idea: “make” the Normal distribution that we have “look like” the standard Normal distribution, and then we can just use the tables as before!

But how?! Use the **slide–squash** technique!!

Two examples (pages 108–111)

1. IQs of 18–19 year olds
2. Vitamin C content of tomato juice

The "slide–squash" technique

The formula which changes **any** Normal random variable X into the **standard** Normal random variable Z is given by

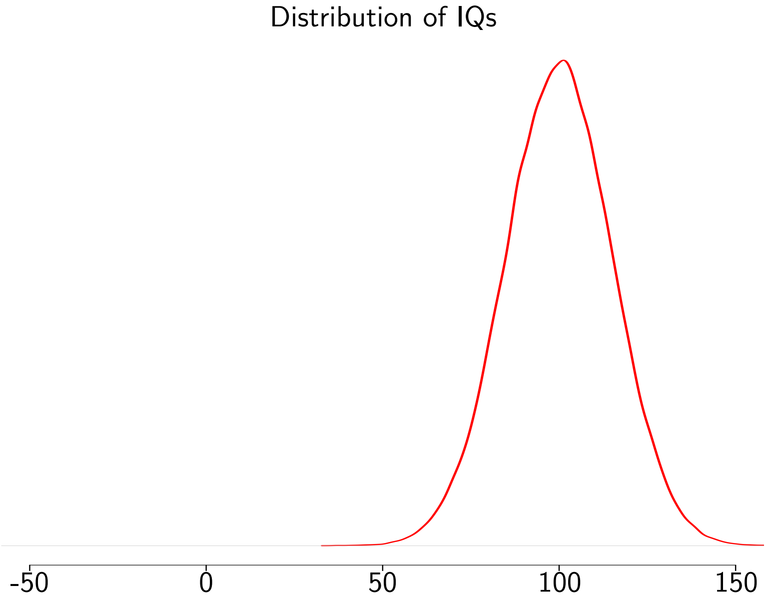
$$Z = \frac{X-\mu}{\sigma},$$

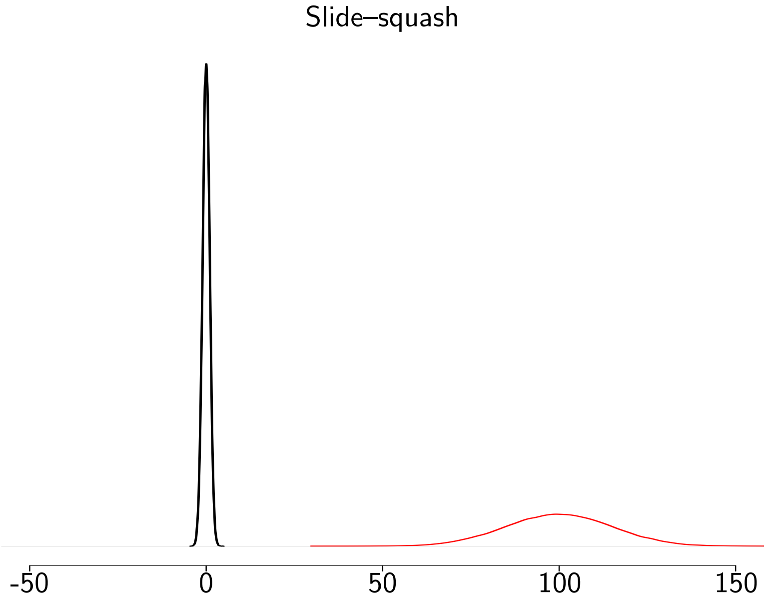
where

- μ is the mean
- σ is the standard deviation

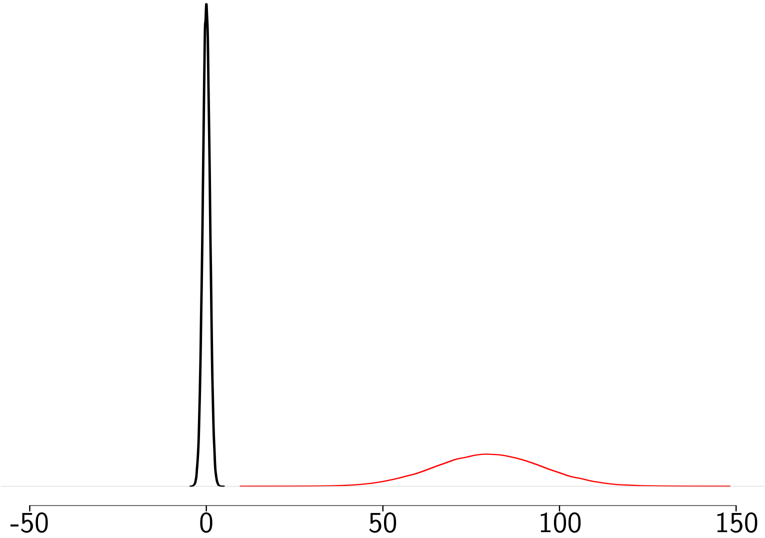
This can be translated into probability statements:

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

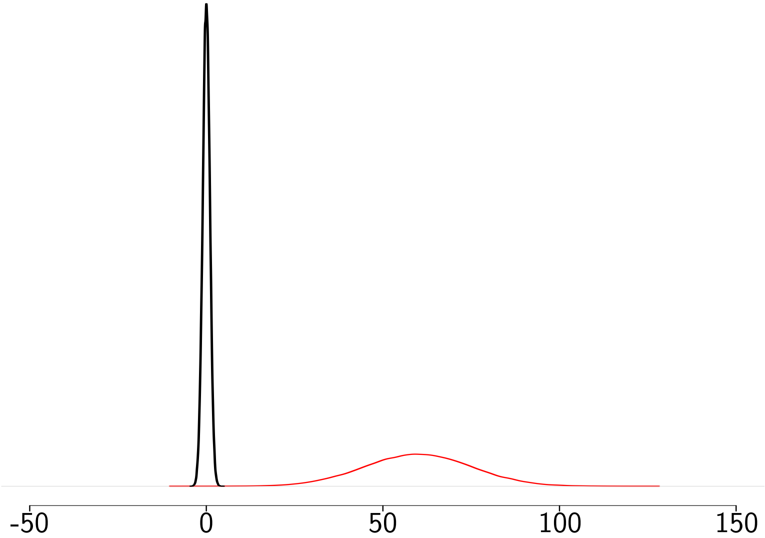

Distribution of IQs
-50
0
50
100
150


Slide–squash
-50
0
50
100
150

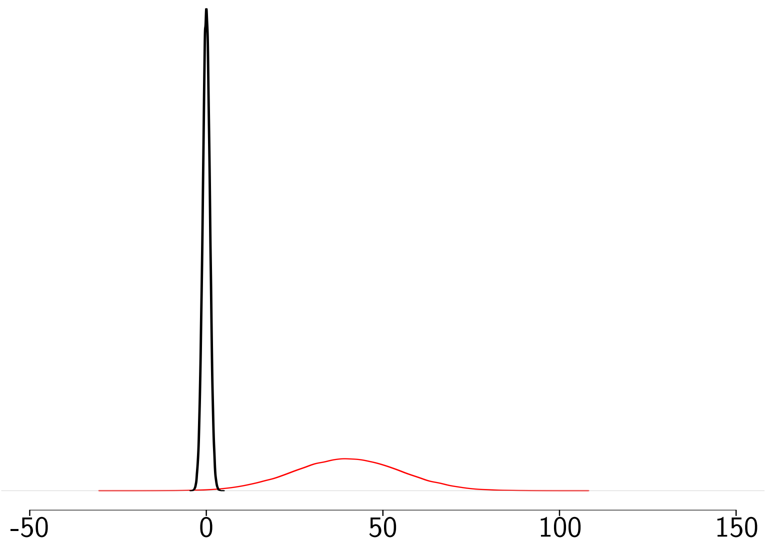
Slide–squash



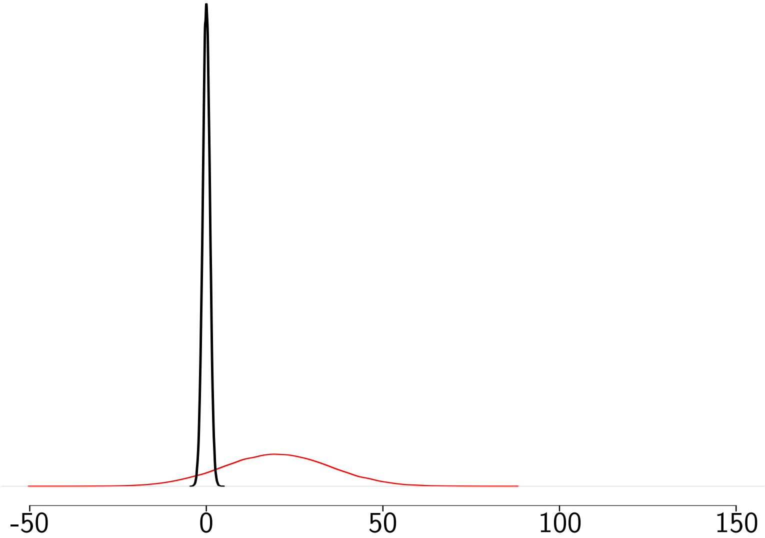
Slide–squash



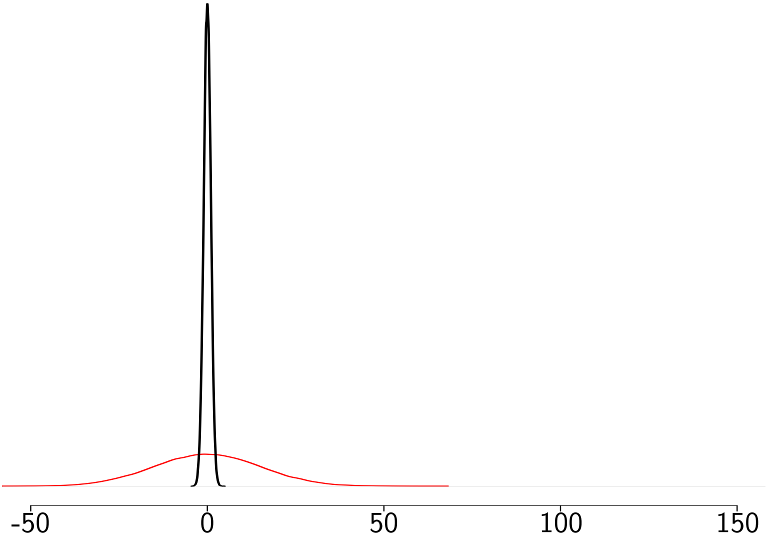
Slide–squash

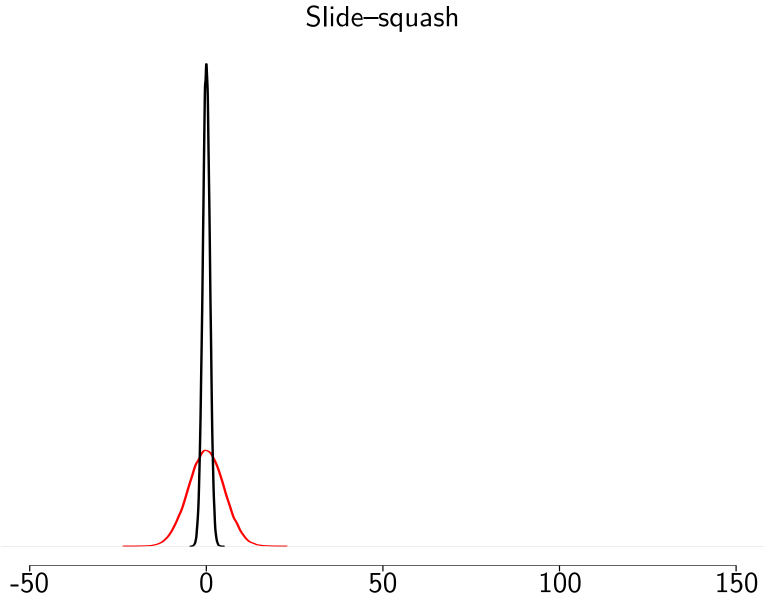


Slide–squash

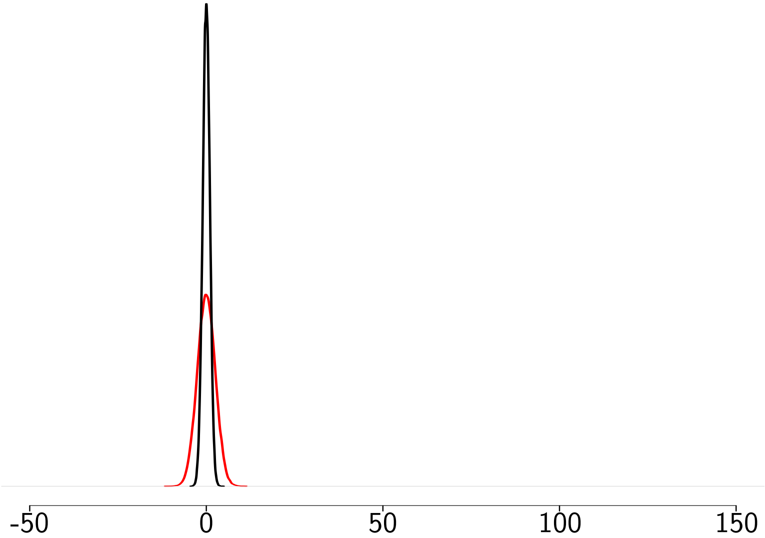


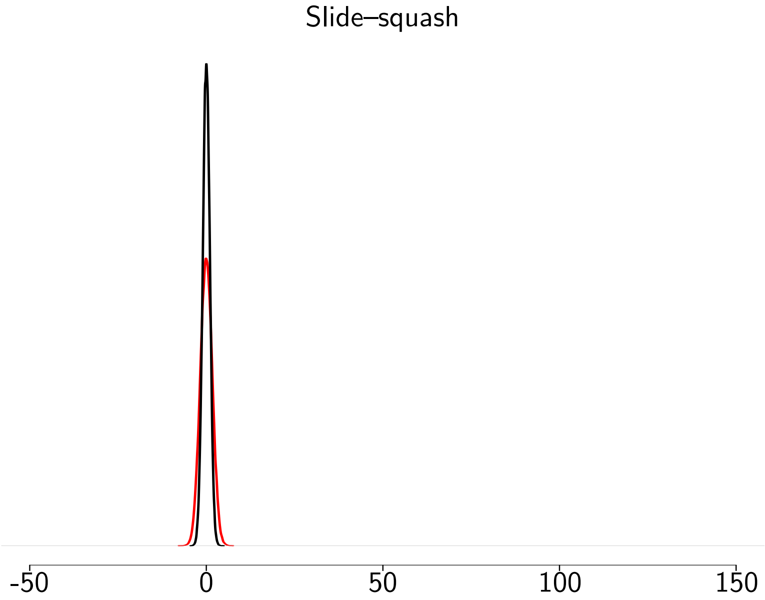
Slide–squash



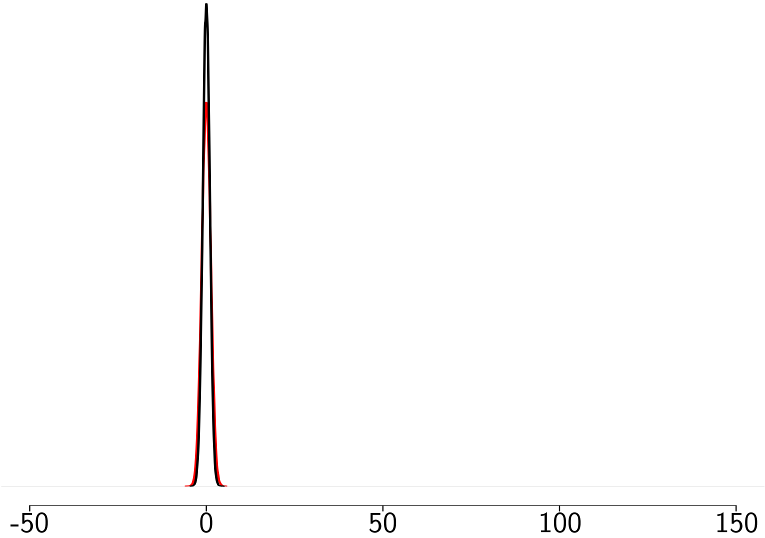

Slide–squash
-50
0
50
100
150

Slide–squash

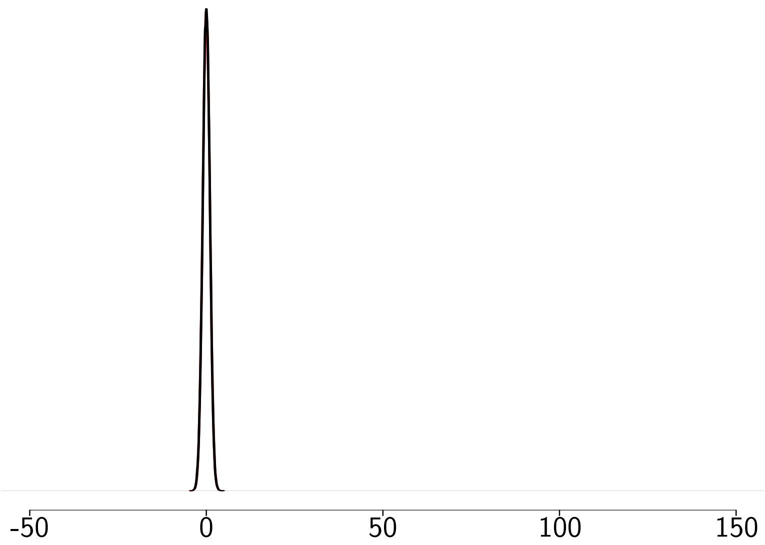



Slide–squash
-50
0
50
100
150

Slide–squash



Slide–squash



IQ example

What is the probability that an 18–19 year old has an IQ less than 85?

$$
\begin{aligned}
P(X < 85) &= P\left(Z < \frac{X-\mu}{\sigma}\right) \\
&= P\left(Z < \frac{85-100}{15}\right) \\
&= P(Z < -1) \\
&= 0.1587.
\end{aligned}
$$

IQ example (i)

$$\begin{aligned} P(X < 110) &= P\left(Z < \frac{X-\mu}{\sigma}\right) \\ &= P\left(Z < \frac{110-100}{15}\right) \\ &= P(Z < 0.67) \\ &= 0.7486. \end{aligned}$$

IQ example (ii)

$$\begin{aligned} P(X > 110) &= 1 - P(X < 110) \\ &= 1 - 0.7486 \\ &= 0.2514. \end{aligned}$$

IQ example (iii)

$$
\begin{aligned}
P(X > 125) &= 1 - P(X < 125) \\
&= 1 - P\left(Z < \frac{125 - 100}{15}\right) \\
&= 1 - P(Z < 1.67) \\
&= 1 - 0.9525 \\
&= 0.0475.
\end{aligned}
$$

IQ example (iv)

$$
\begin{aligned}
P(95 < X < 115) &= P(X < 115) - P(X < 95) \\
&= P\left(Z < \frac{115-100}{15}\right) - P\left(Z < \frac{95-100}{15}\right) \\
&= P(Z < 1) - P(Z < -0.33) \\
&= 0.8413 - 0.3707 \\
&= 0.4706.
\end{aligned}
$$

Using tables in reverse

We can also use tables for the standard Normal distribution in **reverse**.

For example, suppose $Z \sim N(0, 1)$. Below what value are 95% of the population?

This time, we **know** the probability, and we want the value from tables which gives this probability!

More precisely, we want the value **?** such that

$$P(Z < ?) \quad = \quad 0.95$$

From tables, we can see that

$$\begin{aligned} P(Z < 1.64) &= 0.9495 \qquad \text{and} \\ P(Z < 1.65) &= 0.9505. \end{aligned}$$

Therefore, the value we want for $?$ lies between 1.64 and 1.65.

In fact, the value we want lies about half–way between 1.64 and 1.65! And so we can say that $? = \mathbf{1.645}$.

Another example: back to IQs

What IQ score identifies the bottom 10% of 18–19 year olds?

So we need the value z that satisfies $P(Z < z) = 0.1$.

A quick examination of tables gives the two key probabilities as

$$P(Z < -1.28) = 0.1003 \quad \textbf{and} \quad P(Z < -1.29) = 0.0985$$

0.1003 is closer to 0.1 than 0.0985, and so we take z to be **-1.28**.

But! This was on the **standard** Normal distribution scale! Remember that

$$\begin{aligned}
Z &= \frac{X-\mu}{\sigma}, \qquad \text{and so} \\
-1.28 &= \frac{X-100}{15} \\
-1.28 \times 15 &= X - 100 \\
-19.2 &= X - 100 \\
-19.2 + 100 &= X \\
80.8 &= X
\end{aligned}$$