7. The Normal Distribution

0

0

Learning Objectives

- Describe features of the Normal distribution
- Describe the effects of changing values of the mean and standard deviation on the normal distribution
- Describe the Empirical Rule and its relationship with the normal distribution
- Describe features of the Standard Normal distribution
- Calculate normal probabilities using z-scores
- Calculate values of a normal random variable given the probability, (using the z-tables in reverse)
- Use R to calculate normal probabilities.

1

1

Continuous Probability Distributions Recap

The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

Note P(X=x) = 0 i.e. there is no area exactly at x !

2

2

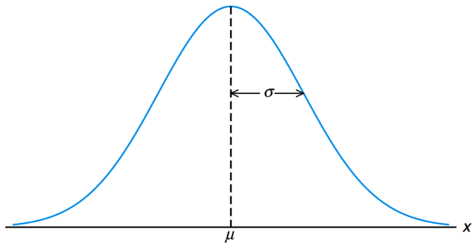
Normal Distribution

- Also called the Gaussian distribution
- pdf is a bell-shaped curve
- The distribution of many types of observations can be approximated by a Normal – eg consider the relative frequency histograms of
  - Heights
  - Weight
  - IQ, …, etc
- Single mode
- Symmetric
- Model for continuous measurements

3

3

The normal distribution



4

4

Normal Distribution

A random variable *X* with probability density function

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

is a normal random variable with parameters μ and σ (where $-\infty < \mu < \infty$ and $\sigma > 0$)

Mean

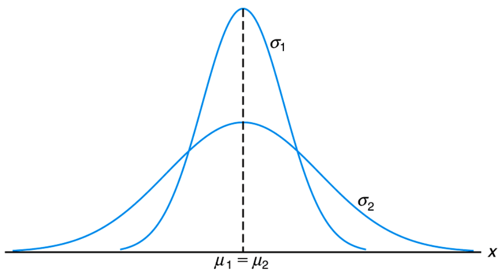
Standard deviation

Write $X \sim N(\mu, \sigma^2)$

5

5

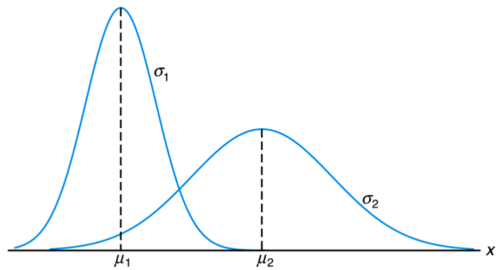
Normal curves with $\mu_1 = \mu_2$ and $\sigma_1 < \sigma_2$



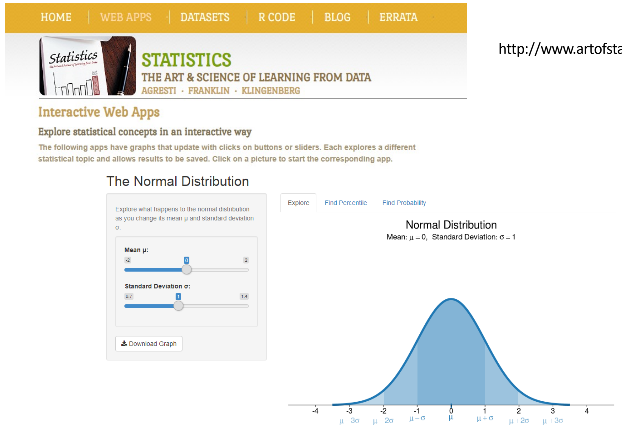
6

6

Normal curves with $\mu_1 < \mu_2$ and $\sigma_1 < \sigma_2$



7



http://www.artofstat.com

8

8

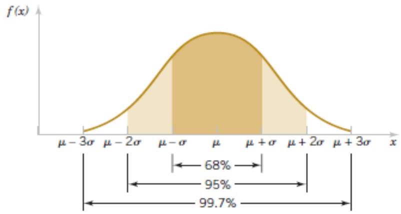
Empirical Rule for a Normal Distribution

For any normal random variable,

$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$

$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$

$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$



Probabilities associated with a normal distribution

9

9

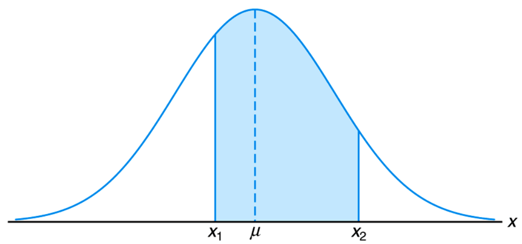
The 68-95-99.7 Rule

- Normal models give us an idea of how extreme a value is by telling us how likely it is to find one that far from the mean.
- It turns out that in a Normal model:
  - about 68% of the values fall within one standard deviation of the mean;
  - about 95% of the values fall within two standard deviations of the mean; and,
  - about 99.7% (almost all!) of the values fall within three standard deviations of the mean.

10

10

$P(x_1 < X < x_2)$ = area of the shaded region



11

11

Areas under the Normal Curve

- Finding an area under a normal distribution in order to calculate probabilities

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

12

12

Standardised Z scores.

To convert a random variable X which follows a N(μ , σ^2) to a random variable Z that follows a standard Normal N(0, 1) calculate Z as

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

Convert X ~ N(100 , 100) to a random variable Z such that Z ~ N(0 , 1)

13

13

Z scores

- A z-score reports the number of standard deviations away from the mean.

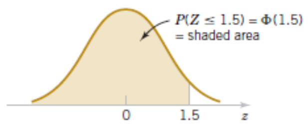
- For example, a Z-score of 2 indicates that the observation is two standard deviations above the mean.

14

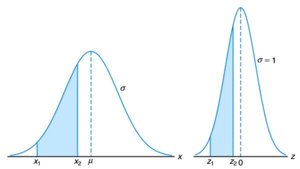
14

$\Phi(z) = P(Z \le z)$

The cumulative distribution function of a standard normal random variable is denoted as $\Phi(z) = P(Z \le z)$



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699



15

15

Calculating Probabilities for N(0,1)

- Left tail – P(Z < 1.8)
  - Directly from table
- Right tail – P(Z > 1.8)
  - By subtraction P(Z>1.8)=1 – P(Z ≤ 1.8)
- Interval Probabilities – P(1 < Z < 1.8)
  - By difference: P(1 < Z < 1.8)=P(Z<1.8)-P(Z<1)

16

16

Normal Probabilities by Hand

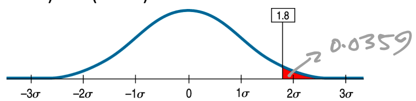
$X \sim N(\mu, \sigma^2)$

- Use a table of the Standard Normal Distribution
- Convert to *z*-scores before using the table. Z

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

- X ~ N(500,100²) ; P(X > 680) = P(Z > 1.8)=1-P(Z<1.8)=1-0.9641=0.0359

$$P(X>680)=P\left(\frac{X-\mu}{\sigma}>\frac{680-\mu}{\sigma}\right)$$
$$=P\left(Z>\frac{680-500}{100}\right)$$
$$=P(Z>1.8)$$

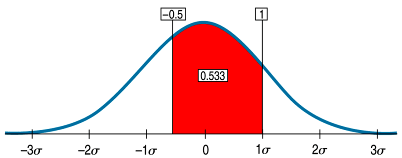


z	.00	.01
⋮	⋮	⋮
1.7	.9554	.9564
1.8	.9641	.9649
1.9	.9713	.9719
⋮	⋮	⋮

17

17

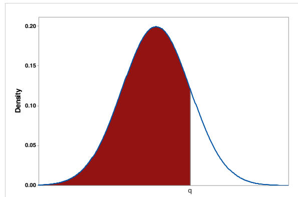
Normal: P(-0.5 < Z < 1)=P(Z<1)-P(Z<-0.5)
=0.8413-0.3085=0.5328



18

18

Using R to calculate probabilities from a Normal Distribution



pnorm(q=?? , mean= ?? , sd= ??)

pnorm returns the integral from -∞ to q for the pdf of the normal distribution with mean μ and standard deviation σ.

19

19

Example: Normal Distribution

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with μ = 10 and σ = 2 mA, what is the probability that the current measurement is less than or equal to 9 mA?

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

$X \sim N(10, 2^2)$

Plot:

$P(X<9) = P\left(\frac{X-\mu}{\sigma} < \frac{9-\mu}{\sigma}\right)$

$= P\left(Z < \frac{9-10}{2}\right)$

$= P(Z < -0.5) = 0.3085$

pnorm(9, mean=10, sd=2) =

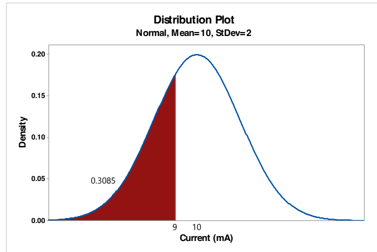
20

20

Example: Normal Distribution

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with μ = 10 and σ = 2 mA, what is the probability that the current measurement is less than or equal to 9 mA?

Plot:



21

21

Example: Normal Distribution

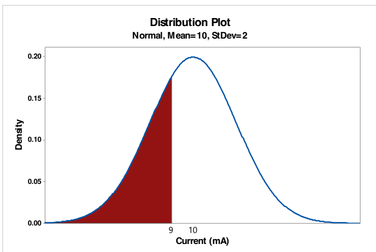
Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with μ = 10 and σ = 2 mA, what is the probability that the current measurement is less than or equal to 9 mA?

Area:

$P(-\infty < X \le 9)$

$= \int_{-\infty}^{9} \frac{1}{\sqrt{2\pi 2^2}} exp^{\frac{(x-10)^2}{2(2^2)}} dx$

$= P\left(\frac{X-10}{2} < \frac{9-10}{2}\right) = P(Z < -0.5)$



22

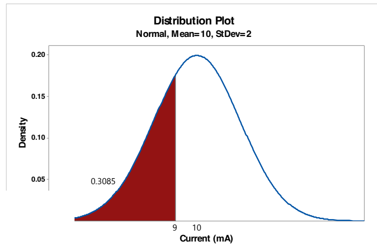
22

Example: Normal Distribution

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with μ = 10 and σ = 2 mA, what is the probability that the current measurement is less than or equal to 9 mA?

Area:

```
> pnorm(q=9, mean=10, sd=2, lower.tail = TRUE)
[1] 0.3085375
> pnorm(-0.5, mean=0, sd=1, lower.tail = T)
[1] 0.3085375
> pnorm(-0.5)
[1] 0.3085375
```



23

23

Using R to calculate probabilities from a Normal Distribution

pnorm

pnorm(q=?? , mean= ?? , sd= ??, lower.tail = ??)

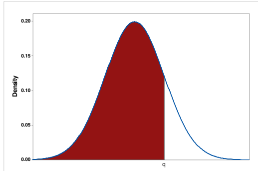
pnorm returns the integral from $-\infty$ to q for the pdf of the normal distribution with mean μ and standard deviation σ.

Note: the default is a standardised normal. It means

pnorm(q=??)=pnorm(q=?? , mean= 0, sd= 1, lower.tail = ??)

24

24



TRUE is the default

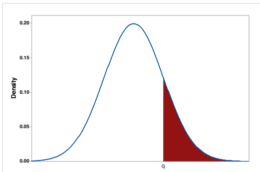
pnorm(q=?? , mean= 0, sd= 1, lower.tail = TRUE)

Defaults

Which equals to: pnorm(q=??)

25

25



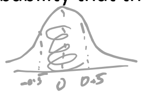
pnorm(q=?? , mean= 0, sd= 1, lower.tail = FALSE)

26

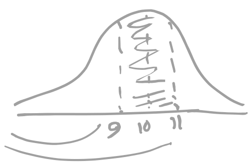
26

Example: Normal Distribution

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with $\mu = 10$ and $\sigma = 2$ mA, what is the probability that the current measurement is between 9 and 11 mA?



Plot:



$P(9 < X < 11) = P(X < 11) - P(X < 9)$

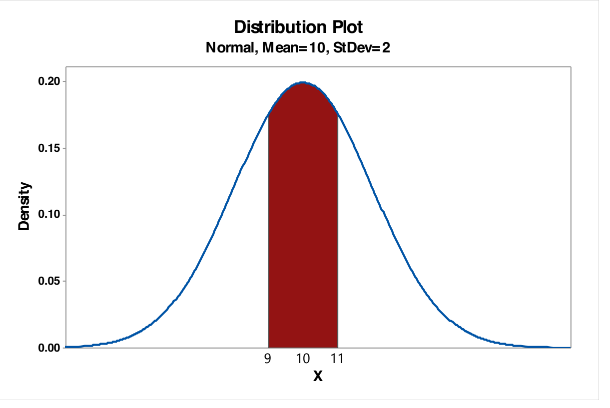
$= P\left(\frac{X-\mu}{\sigma} < \frac{11-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} < \frac{9-\mu}{\sigma}\right)$

$= P\left(Z < \frac{11-10}{2}\right) - P\left(Z < \frac{9-10}{2}\right) = P(Z < 0.5) - P(Z < -0.5)$

$= 0.6915 - 0.3085$

27

27

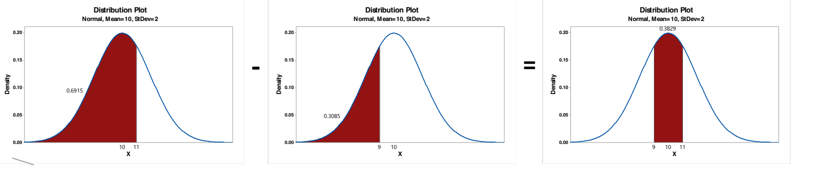


28

28

Example: Normal Distribution

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with $\mu = 10$ and $\sigma = 2$ mA, what is the probability that the current measurement is between 9 and 11 mA?



29

29

Example: Normal Distribution

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with $\mu = 10$ and $\sigma = 2$ mA, what is the probability that the current measurement is between 9 and 11 mA?

pnorm(q=11, mean=10, sd=2) - pnorm(q=9, mean=10, sd=2)=
pnorm(q=0.5, mean=0, sd=1) - pnorm(q=-0.5, mean=0, sd=1)=
pnorm(q=0.5) - pnorm(q=-0.5)

Probability: 0.3829

30

30

Example: Normal Distribution determine percentiles …

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with $\mu = 10$ and $\sigma = 2$ mA.

Determine the value for which the probability that a current measurement is below 0.98.

Plot:

31

31

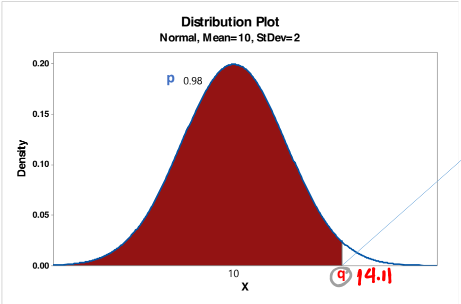
Example: Normal Distribution determine percentiles …

Determine the value for which the probability that a current measurement is below 0.98.

Plot:

P(X<q)=0.98

q=?



Supply p to find q

32

Example: Normal Distribution determine percentiles …

$$P(X < k) = 0.98$$

$$P\left(\frac{X-10}{2} < \frac{k-10}{2}\right) = 0.98$$

$$P\left(Z < \frac{k-10}{2}\right) = 0.98$$

2.05 2.06

0.9798 0.9803

0.98

$\frac{2.05+2.06}{2} = 2.055$

$P(Z < 2.055) = 0.98$

We also know from the normal table that:

$\cancel{P(Z < 2.05)} = 0.98$

$P(Z < 2.055) = 0.98$

Therefore:

$P\left(Z < \frac{k-10}{2}\right) = P(Z < 2.05)$ which means $\frac{k-10}{2} = 2.055$

Then: $k = 2 * 2.055 + 10 = 14.110$

33

Using R to calculate percentiles from a Normal Distribution

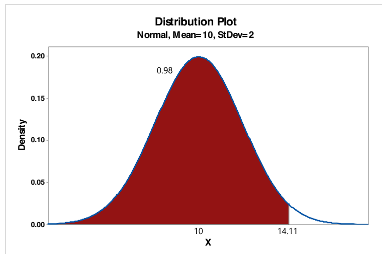
qnorm

qnorm(p=0.??, mean= ??, sd=?? , lower.tail = ??)

qnorm is the inverse of the cdf, which you can also think of as the inverse of pnorm. Use qnorm to determine the x corresponding to the p[th] quantile of the normal distribution?

34

Determine the value for which the probability that a current measurement is below 0.98.




```
> qnorm(p=0.98, mean=10, sd=2, lower.tail = TRUE)
[1] 14.1075
```

≈ 14.11

35

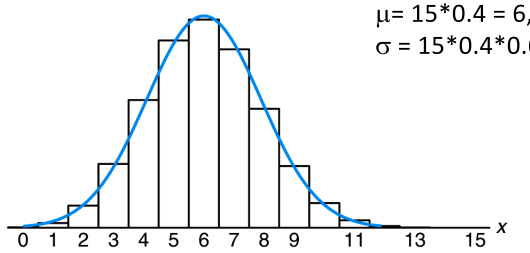
Normal Approximations

- The binomial and Poisson distributions become more bell-shaped and symmetric as their mean value increase.
  - If X ~Binomial(n, p) then X ~ $N(np, np(1-p))$
  - If X ~Poisson(λ) then X ~ $N(\lambda, \lambda)$
- The normal distribution is a good approximation for:
  - Binomial if $np > 5$ and $n(1-p) > 5$. $X \sim B(n,p) \Rightarrow E(X) = np$, $Var(X) = np(1-p)$
  - Poisson if $\lambda > 5$. $X \sim P(\lambda) \Rightarrow E(X) = \lambda$, $Var(X) = \lambda$
- For manual calculations, the normal approximation is practical – use R for exact probabilities of the binomial and Poisson.

36

36

Normal approximation of $b(x;\ n=15,\ p=0.4)$



$\mu = 15*0.4 = 6$,
$\sigma = 15*0.4*0.6 = 3.6$

37

Normal Approximation to the Poisson

If X is a Poisson random variable with $E(X) = \lambda$ and $V(X) = \lambda$,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

The approximation is good for $\lambda \geq 5$

38

38

Continuity Correction

Using the normal distribution to approximate a discrete distribution (e.g. binomial) we need to take into account the fact that the normal distribution is continuous.

Discrete		**Continuous**
$P(X > k)$	$\longrightarrow$	$P\left(X > k + \frac{1}{2}\right)$
$P(X \geq k)$	$\longrightarrow$	$P\left(X > k - \frac{1}{2}\right)$
$P(X < k)$	$\longrightarrow$	$P\left(X < k - \frac{1}{2}\right)$
$P(X \leq k)$	$\longrightarrow$	$P\left(X < k + \frac{1}{2}\right)$
$P(k_1 < X < k_2)$	$\longrightarrow$	$P\left(k_1 + \frac{1}{2} < X < k_2 - \frac{1}{2}\right)$
$P(k_1 \leq X \leq k_2)$	$\longrightarrow$	$P\left(k_1 - \frac{1}{2} < X < k_2 + \frac{1}{2}\right)$

39

39

The number of phone calls at a call centre is Poisson distributed with mean 64 per hour. $X \sim P(\lambda = 64) \Rightarrow X \approx N(64, 64)$

1. What is the probability of 70 or more calls in a given hour?

2. What is the probability of less than 240 calls in a 4 hour period?

40

40

The number of phone calls at a call centre is Poisson distributed with mean 64 per hour.

1. What is the probability of 70 or more calls in a given hour?
   By using normal approximation to the poisson:

   $$X \approx N(64, 64)$$

   Pnorm(q = 69.5, mean = 64, sd = 64, Lower.tail = F)

   $P(X \geq 70) = P(X > 70 - \frac{1}{2}) = P(X > 69.5) = P(\frac{X-64}{\sqrt{64}} > \frac{69.5-64}{\sqrt{64}}) = P(Z > 0.69) = 1 - P(Z < 0.69) = 1 - 0.7549 = 0.2451$

2. What is the probability of less than 240 calls in a 4 hour period? In four hours period $X_{4hrs} \sim \text{Poisson}(4 \times 64) = P(256)$

   $$X_{4hrs} \approx N(4 \times 64, 4 \times 64) \equiv N(256, 256)$$

   $P(X_{4hrs} < 240) = P(X_{4hrs} < 240 - \frac{1}{2}) = P(X_{4hrs} < 239.5) = P(\frac{X_{4hrs}-256}{\sqrt{256}} > \frac{239.5-256}{\sqrt{256}}) = P(Z < -1.03) = 0.1515$

41

41