Chapter 4

The Normal Distribution

Yen-Yi Ho

Department of Statistics
University of South Carolina

5.3 The Normal Distribution

- One of the most commonly used continuous r.v.'s has a **bell shaped** probability distribution. It is known as a normal r.v. and its probability distribution is called a **normal distribution**.

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

For a Normal Random Variable X, the probability density function

Normal pdf

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

mean: μ standard deviation: σ

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

For a Normal Random Variable X, the probability density function

Normal pdf

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

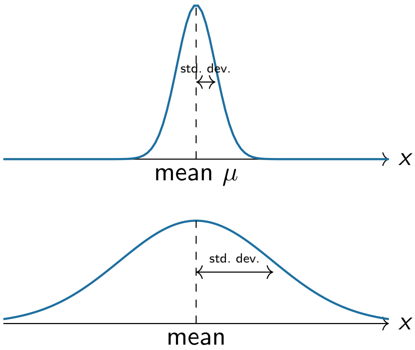
mean: μ standard deviation: σ

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

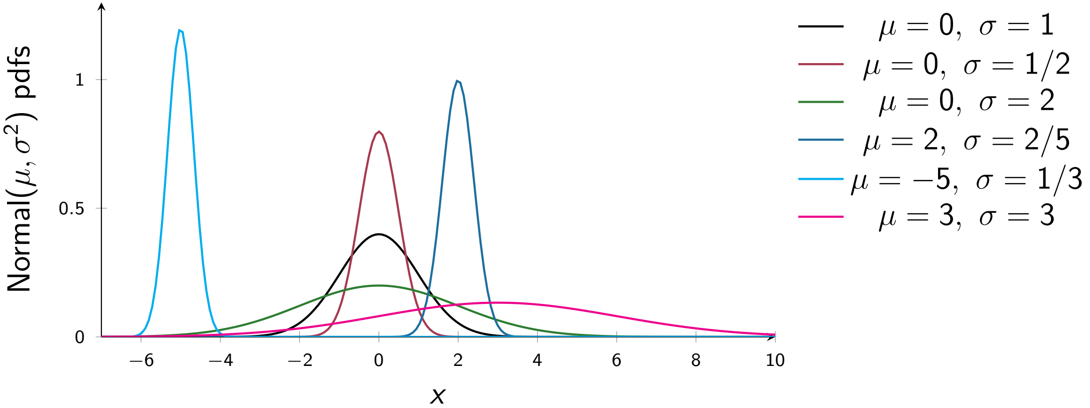
Example:

$X \sim N(1, 4) \implies \sigma = 2$

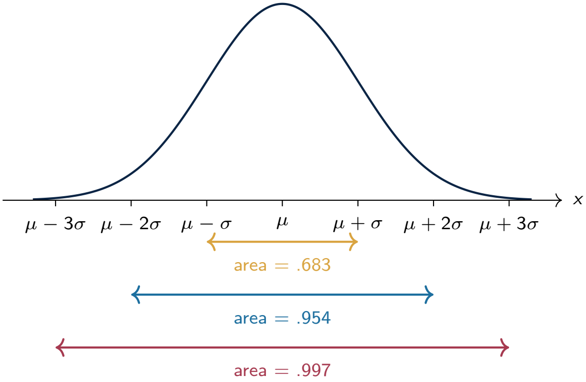
Shape of the Normal Distribution



pdfs of Several Normal Distributions



The pdf of the $\mathcal{N}(\mu, \sigma^2)$ Distribution



Mean and Variance of Normal Distribution

Definition: Mean and Variance

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

- $E(X) = \mu$
- $\mathrm{Var}(X) = \sigma^2$

Example:

Suppose growth in height (ft) of Loblolly pines from age three to five is $\mathcal{N}(\mu = 5, \sigma^2 = 1/4)$. Give the probability that the growth of a randomly selected Loblolly pine is

1. between 4.5 and 5.5 feet.
2. more than 7 feet.
3. less than 5.5 feet.
4. between 3.5 feet and 5.5 feet.

Use the picture on the previous slide.

Conversion to the Standard Normal Distribution

Get probabilities for $X \sim \mathcal{N}(\mu, \sigma^2)$ like

$$P(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx$$

😭

Definition: Conversion to the Standard Normal Distribution

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

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

The $\mathcal{N}(0,1)$ dist. is called the **Standard Normal distribution** and its pdf is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

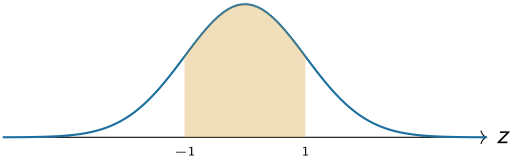
Can look up integrals over this pdf in a table.

The Standard Normal Distribution

- The **standard normal distribution** is a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.
- A random variable with a standard normal distribution, denoted by the symbol Z, is called a **standard normal r.v.**

$$Z \sim N(0, 1)$$

The 68-95-99.7 Rule



- 68% of the data falls within 1 std deviation of the mean: $P(-1 < Z < 1) \approx .68$
- 95% of the data falls within 2 std deviations of the mean: $P(-2 < Z < 2) \approx .95$
- 99.7% of the data falls within 3 std deviations of the mean: $P(-3 < Z < 3) \approx .997$

Exercises: Standard Normal

Example:

Suppose that Z is a standard normal random variable.

Exercises: Standard Normal

Example:

Suppose that Z is a standard normal random variable.

a) Find the probability that Z falls between -1.33 and 1.33.

$$P(-1.33 < Z < 1.33) = 2 \cdot P(0 < Z < 1.33) = 2(.4082) = .8164$$

Exercises: Standard Normal

Example:

Suppose that Z is a standard normal random variable.

a) Find the probability that Z falls between -1.33 and 1.33.

$$P(-1.33 < Z < 1.33) = 2 \cdot P(0 < Z < 1.33) = 2(.4082) = .8164$$

b) Find the probability that Z exceeds 1.64.

$$P(Z > 1.64) = 0.5 - P(0 < Z < 1.64) = .5 - .4495 = .0505$$

Exercises: Standard Normal

Example:

Suppose that Z is a standard normal random variable.

a) Find the probability that Z falls between -1.33 and 1.33.

$$P(-1.33 < Z < 1.33) = 2 \cdot P(0 < Z < 1.33) = 2(.4082) = .8164$$

b) Find the probability that Z exceeds 1.64.

$$P(Z > 1.64) = 0.5 - P(0 < Z < 1.64) = .5 - .4495 = .0505$$

c) Find the probability that Z lies to the left of $.67$.

$$P(Z < .67) = .5 + P(0 < Z < .67) = .5 + .2487 = .7487$$

Exercises: Standard Normal (cont'd)

d) Find the probability that Z exceeds 1.96 in absolute value.

d) Find the probability that Z exceeds 1.96 in absolute value.

$$\begin{aligned}P(|Z| > 1.96) &= P(Z > 1.96) + P(Z < -1.96)\\ &= 2\,P(Z > 1.96)\\ &= 2(.5 - P(0 < Z < 1.96))\\ &= 2(.5 - .475) = .05\end{aligned}$$

Finding $P(a < X < b)$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, we can find $P(a < X < b)$ in two steps:

Two Steps

1. Transform a and b to the Z-world (# of standard deviations world):

$$a \mapsto \frac{a-\mu}{\sigma} \quad \text{and} \quad b \mapsto \frac{b-\mu}{\sigma},$$

2. Find

$$P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

by using a "Z-table"—a table of Standard Normal probabilities.

Exercise: Loblolly Pines

Example:

Suppose growth in height (ft) of Loblolly pines from age three to five is $\mathcal{N}(5, 1/4)$. Give the probability that the growth of a randomly selected Loblolly pine is

1. between 5.25 and 6.25 feet.
2. more than 7.8 feet.
3. less than 5.31 feet.
4. between 4.1 feet and 5.2 feet.

Find the probabilities in the "Z-world" using a Z-table.

Quantiles of a Continuous Random Variable

Definition: Quantile

For a continuous rv X with a strictly increasing cdf, the p **quantile** of X is the value q which satisfies

$$P(X \leq q) = p.$$

A quantile is like a percentile, but not expressed as a percentage.

Draw pictures.

Finding Quantiles

If $X \sim \mathcal{N}(\mu, \sigma^2)$, we can find q such that $P(X \leq q) = p$ in two steps:

Two Steps

1. Find q_Z such that $P(Z < q_Z) = \theta$ using a "Z table".
2. Get the corresponding quantile in the X world as $q = \mu + \sigma q_Z$

Example:

Suppose growth in height (ft) of Loblolly pines from age three to five is $\mathcal{N}(5, 1/4)$. Let X denote the height of a randomly selected Loblolly pine and find

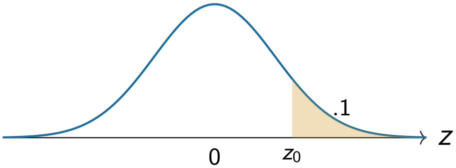
1. the 75%-tile of growth.
2. the median of the growths, i.e. the 50%-tile of X.
3. an interval, centered at the mean, within which X lies with probability 0.50.

Exercises: Finding z_0

e) Find z_0 such that only 10% of the time Z will exceed z_0.

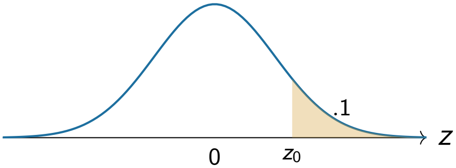
Exercises: Finding z_0

e) Find z_0 such that only 10% of the time Z will exceed z_0.
$z_0 = .9$th quantile $= 90$th percentile



Exercises: Finding z_0

e) Find z_0 such that only 10% of the time Z will exceed z_0.
$z_0 = .9$th quantile $= 90$th percentile



By the Z-table, the closest area to .9 is .8997, therefore the corresponding $z_0 = 1.28$.

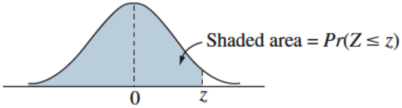


TABLE 1

Standard normal curve areas

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143

TABLE 1
(continued)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

Property of Normal Distributions

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

- If X is a normal r.v. with mean μ and standard deviation σ, then the random variable Z, defined by the formula

Standardization

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

has a standard normal distribution.

Exercise: Cell Phone Charge Time

Example:

Assume that the length of time, X, between charges of a cellular phone is normally distributed with a mean of 10 hours and a standard deviation of 1.5 hours.

$$X \sim N(10, 1.5^2)$$

Exercise: Cell Phone Charge Time

Example:

Assume that the length of time, X, between charges of a cellular phone is normally distributed with a mean of 10 hours and a standard deviation of 1.5 hours.

$$X \sim N(10, 1.5^2)$$

a) Find the probability that the cell phone will last between 8 and 12 hours between charges.

$$\begin{aligned} P(8 < X < 12) &= P\left(\frac{8-10}{1.5} < \frac{X-10}{1.5} < \frac{12-10}{1.5}\right) \\ &= P(-1.33 < Z < 1.33) = .8164 \end{aligned}$$

Exercise: Cell Phone Charge Time (cont'd)

b) Find the probability that the cell phone will last at least 11 hours.

Exercise: Cell Phone Charge Time (cont'd)

b) Find the probability that the cell phone will last at least 11 hours.

$$
\begin{aligned}
P(X \geq 11) = 1 - P(X \leq 11) = P\left(Z \geq \frac{11 - 10}{1.5}\right) \\
&= P(Z \geq .67) \\
&= .5 - P(0 < Z < .67) \\
&= .5 - .2486 = .2514
\end{aligned}
$$

Property of Normal Distributions

Definition: Standardizing a Normal Random Variable

If X is a normal r.v. with mean μ and standard deviation σ, then the random variable Z, defined by the formula

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

has a **standard normal distribution**.

- $Z \sim N(0, 1)$
- Standardizing lets us use a single table (or a single set of R functions) for *any* normal distribution.

R Functions for the Normal Distribution

The Four Normal Functions in R

`dnorm(x,` μ`,` σ`)`	calculate density prob. at x
`pnorm(x,` μ`,` σ`)`	calculate cumulative prob. at x, $P(X \leq x)$
`qnorm(p,` μ`,` σ`)`	calculate the quantile x of the $N(\mu, \sigma)$ s.t. $P(X \leq x) = p$
`rnorm(n,` μ`,` σ`)`	generate n random values from the $N(\mu, \sigma)$

Note: `pnorm(.,0,1)` is the standard normal CDF, i.e. $P(Z \leq z)$.

Example:

Redo some exercises with `pnorm` and `qnorm` functions in R.

Exercise 1: Cellular Phone Battery Life

Example: Cell Phone Charges

Assume that the length of time, X, between charges of a cellular phone is normally distributed with a mean of 10 hours and a standard deviation of 1.5 hours. $X \sim N(\mu = 10, \sigma = 1.5)$.

a.) Find the probability that the cell phone will last between 8 and 12 hours between charges.

b.) Find the probability that the cell phone will last at least 11 hours.

Exercise 1(a): Solution

$$P(8 < X < 12) = P(X \leq 12) - P(X \leq 8)$$

Exercise 1(a): Solution

$$P(8 < X < 12) = P(X \le 12) - P(X \le 8)$$

$$= \texttt{pnorm(12,10,1.5)} - \texttt{pnorm(8,10,1.5)}$$

Exercise 1(a): Solution

$$P(8 < X < 12) = P(X \leq 12) - P(X \leq 8)$$

$$= \texttt{pnorm(12,10,1.5)} - \texttt{pnorm(8,10,1.5)}$$

Standardizing:

$$= P\left(\frac{8-10}{1.5} < Z < \frac{12-10}{1.5}\right) = P(-1.33 < Z < 1.33)$$

Exercise 1(b): Solution

$$P(X \geq 11) = 1 - P(X < 11) = 1 - \texttt{pnorm(11,10,1.5)}$$

Exercise 1(b): Solution

$$P(X \geq 11) = 1 - P(X < 11) = 1 - \texttt{pnorm(11,10,1.5)}$$

Equivalently, standardizing first:

$$P\left(Z \geq \frac{11 - 10}{1.5}\right) = P(Z \geq 0.67)$$

Exercise 2: College Entrance Exam Scores

Example: College Entrance Exam

Suppose the score, X, on a college entrance examination is normally distributed with a mean of 550 and a standard deviation of 100. $X \sim N(550, 100)$. A certain university will consider for admission only those applicants whose scores exceed the 90th percentile of the distribution. Find the minimum score an applicant must achieve in order to receive consideration for admission to the university.

Exercise 2: College Entrance Exam Scores

Example: College Entrance Exam

Suppose the score, X, on a college entrance examination is normally distributed with a mean of 550 and a standard deviation of 100. $X \sim N(550, 100)$. A certain university will consider for admission only those applicants whose scores exceed the 90th percentile of the distribution. Find the minimum score an applicant must achieve in order to receive consideration for admission to the university.

Recall:

$$p\text{th quantile} = (100 \times p)\text{th percentile}, \qquad .5\text{th quantile} = 50\text{th percentile}$$

Exercise 2: Solution

We need the .9th quantile (90th percentile) of $N(550, 100)$:

$$x_{.9} = \texttt{qnorm(.9, 550, 100)}$$

Exercise 2: Solution

We need the .9th quantile (90th percentile) of $N(550, 100)$:

$$x_{.9} = \texttt{qnorm(.9, 550, 100)}$$

Or by hand, using $Z = \dfrac{X - \mu}{\sigma} \iff X = \mu + \sigma Z$:

Exercise 2: Solution

We need the .9th quantile (90th percentile) of $N(550, 100)$:

$$x_{.9} = \texttt{qnorm(.9, 550, 100)}$$

Or by hand, using $Z = \dfrac{X - \mu}{\sigma} \iff X = \mu + \sigma Z$:

$$z_{.9} = 1.28 \quad \text{(from the standard normal table)}$$

Exercise 2: Solution

We need the .9th quantile (90th percentile) of $N(550, 100)$:

$$x_{.9} = \texttt{qnorm(.9, 550, 100)}$$

Or by hand, using $Z = \dfrac{X - \mu}{\sigma} \iff X = \mu + \sigma Z$:

$$z_{.9} = 1.28 \quad \text{(from the standard normal table)}$$

$$x_{.9} = 550 + 100(1.28) = 678$$

Exercise 2: Solution

We need the .9th quantile (90th percentile) of $N(550, 100)$:

$$x_{.9} = \texttt{qnorm(.9, 550, 100)}$$

Or by hand, using $Z = \dfrac{X - \mu}{\sigma} \iff X = \mu + \sigma Z$:

$$z_{.9} = 1.28 \quad \text{(from the standard normal table)}$$

$$x_{.9} = 550 + 100(1.28) = 678$$

An applicant must score at least **678** to be considered for admission.

Do My Data Come from a Normal Distribution?

Example: Commute Times

These are the commute times (sec) to class of a sample of students.

1832	1440	1620	1362	577	934	928	998	1062	900
1380	913	654	878	172	773	1171	1574	900	900

Check with a histogram and a Q-Q plot whether the data quantiles match those of a Normal distribution.

Assessing Normality: Steps

1. Construct a histogram and note the shape of the graph.

Assessing Normality: Steps

1. Construct a histogram and note the shape of the graph.
2. Compute the intervals $\bar{x} \pm s$, $\bar{x} \pm 2s$, $\bar{x} \pm 3s$, and determine the percentage of measurements falling into each.

Assessing Normality: Steps

1. Construct a histogram and note the shape of the graph.
2. Compute the intervals $\bar{x} \pm s$, $\bar{x} \pm 2s$, $\bar{x} \pm 3s$, and determine the percentage of measurements falling into each.

Interval	Empirical Rule
$\bar{x} \pm s$	$\approx 68\%$
$\bar{x} \pm 2s$	$\approx 95\%$
$\bar{x} \pm 3s$	$\approx 99.7\%$

Assessing Normality: Steps

1. Construct a histogram and note the shape of the graph.
2. Compute the intervals $\bar{x} \pm s$, $\bar{x} \pm 2s$, $\bar{x} \pm 3s$, and determine the percentage of measurements falling into each.

Interval	Empirical Rule
$\bar{x} \pm s$	$\approx 68\%$
$\bar{x} \pm 2s$	$\approx 95\%$
$\bar{x} \pm 3s$	$\approx 99.7\%$

3. Find the IQR and s, and then compare the ratio IQR$/s$ to 1.3.

Assessing Normality: Steps

1. Construct a histogram and note the shape of the graph.
2. Compute the intervals $\bar{x} \pm s$, $\bar{x} \pm 2s$, $\bar{x} \pm 3s$, and determine the percentage of measurements falling into each.

Interval	Empirical Rule
$\bar{x} \pm s$	$\approx 68\%$
$\bar{x} \pm 2s$	$\approx 95\%$
$\bar{x} \pm 3s$	$\approx 99.7\%$

3. Find the IQR and s, and then compare the ratio IQR/s to 1.3.

Interquartile range: $\mathrm{IQR} = Q_3 - Q_1 =$ 75th percentile $-$ 25th percentile.

R Code: Checking Normality

```r
data = rnorm(100, 100, 15)
# data = rexp(100, rate = 1/20)

hist(data)
mean(data)
sd(data)

quantile(data, c(0, .25, .5, .75, 1), type = 6)
IQR(data, type = 6)
IQR(data, type = 6) / sd(data)

qqnorm(data)
qqline(data)
```

- `rnorm(n, μ, σ)` helps to generate n random values from the $N(\mu, \sigma)$.

4. Normal Quantile–Quantile (Q–Q) Plot

Definition: Normal Q–Q Plot

Plot of the data in order, against the corresponding quantiles (percentiles) from a normal distribution.

4. Normal Quantile–Quantile (Q–Q) Plot

Definition: Normal Q–Q Plot

Plot of the data in order, against the corresponding quantiles (percentiles) from a normal distribution.

- Horizontal axis: **theoretical quantiles** from the $N(0, 1)$ distribution.
- Vertical axis: **sample quantiles** of the observed data.

4. Normal Quantile–Quantile (Q–Q) Plot

Definition: Normal Q–Q Plot

Plot of the data in order, against the corresponding quantiles (percentiles) from a normal distribution.

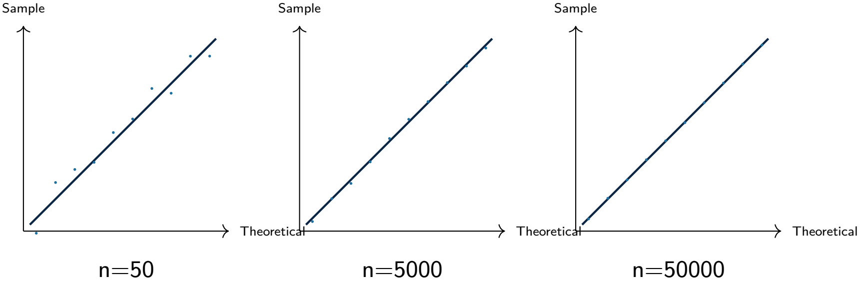
- Horizontal axis: **theoretical quantiles** from the $N(0, 1)$ distribution.
- Vertical axis: **sample quantiles** of the observed data.

If the data are approximately normal, the points will fall (approximately) on a straight line.

Q–Q Plots: Effect of Sample Size

Example: Simulated Normal Data

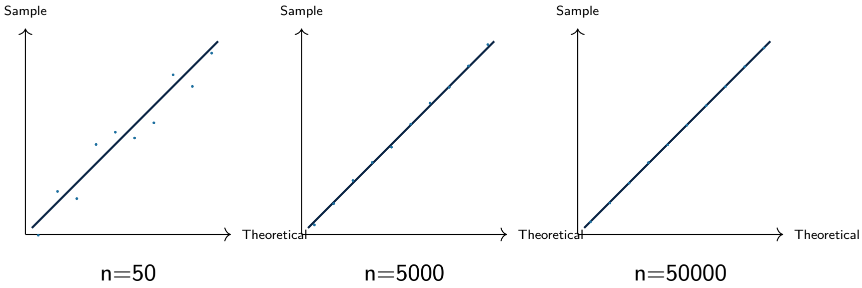
When data truly come from a normal distribution, the Q–Q plot hugs the 45° reference line more and more closely as the sample size n grows.



Q–Q Plots: Effect of Sample Size

Example: Simulated Normal Data

When data truly come from a normal distribution, the Q–Q plot hugs the 45° reference line more and more closely as the sample size n grows.



Small samples show more scatter around the line even when the population is exactly normal.

Q–Q Plots: Repeated Samples, $n = 50$

Example:

Six different random samples of size $n = 50$, all drawn from the *same* normal population, will each produce a slightly different Q–Q plot.

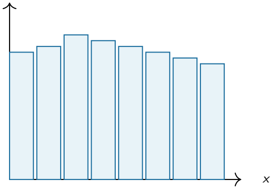
Q–Q Plots: Repeated Samples, $n = 50$

Example:

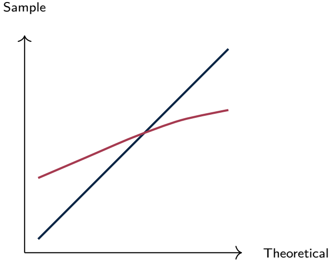
Six different random samples of size $n = 50$, all drawn from the *same* normal population, will each produce a slightly different Q–Q plot.

- This sample-to-sample variability is expected — don't over-interpret small deviations from the line in a single Q–Q plot.
- Look for *systematic* patterns (curvature, S-shapes) rather than random wiggle.

Pattern: Light Tails

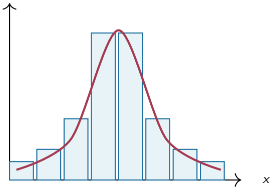


Histogram: roughly flat / uniform

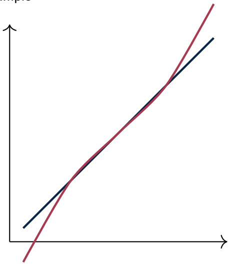


Q–Q plot flattens at the ends (light tails)

Pattern: Heavy Tails

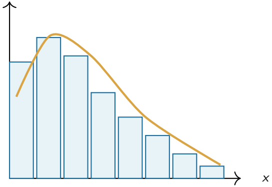


Histogram: sharply peaked with heavy tails

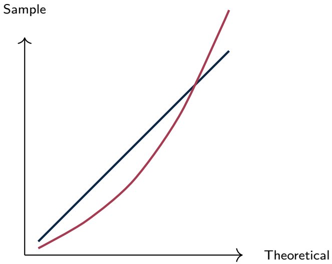


Q–Q plot has an S-shape (heavy tails)

Pattern: Skewed Right

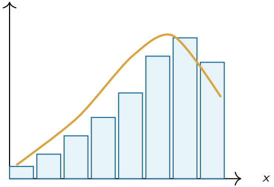


Long tail to the right

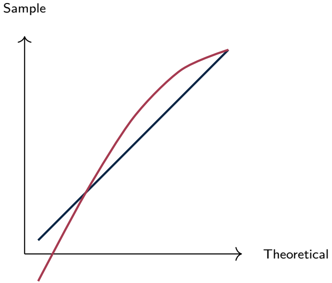


Q–Q plot bends up at the right end

Pattern: Skewed Left

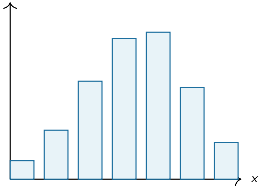


Long tail to the left

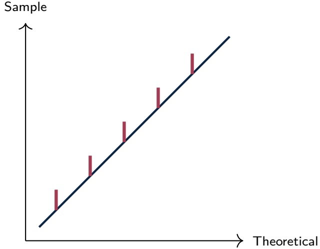


Q–Q plot bends down at the left end

Pattern: Discrete Data



Histogram: few distinct values



Q–Q plot: “staircase” of horizontal clusters

Exercise: Baby Food Jars

Example:

You sell jars of baby food labelled as weighing 4oz $\approx$ 113g. Suppose your process results in jar weights with the $\mathcal{N}(120, 4^2)$ distribution and that you will be fined if more than 2% of your jars weigh less than 113g.

1. What proportion of your jars weigh less than 113g?
2. To what must you increase μ to avoid being fined?
3. Keeping $\mu = 120$g, to what must you reduce σ to avoid being fined?

Sum of Independent Normal Random Variables

Sum of Independent Normal Random Variables

If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \ldots, X_n \sim \mathcal{N}(\mu_1, \sigma_n^2)$ are independent random variables, then

$$\sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i,\ \sum_{i=1}^{n} \sigma_i^2\right).$$

In the above, **independent** means that the values of the rvs don't affect one other.

Example:

Consider boxes containing 25 jars of baby food (from previous).

1. What is the expected weight of the boxes?
2. What is the standard deviation of the box weights?
3. Give the probability that the box weighs less than 2,975 grams.

Definition: Setup

$X \sim \text{binomial}(n, p)$, $x = 0, 1, 2, \ldots, n$.

5.5 Approximating a Binomial Distribution with a Normal Distribution

Definition: Setup

$X \sim \text{binomial}(n, p)$, $x = 0, 1, 2, \ldots, n$.

- When n is large, a normal probability distribution may be used to provide a good approximation to the probability distribution of a binomial random variable.

Definition: Setup

$X \sim \text{binomial}(n, p)$, $x = 0, 1, 2, \ldots, n$.

- When n is large, a normal probability distribution may be used to provide a good approximation to the probability distribution of a binomial random variable.

Normal Approximation to the Binomial

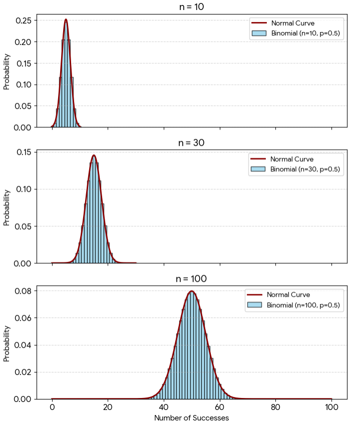
$$E(X) = np, \qquad \sigma_X = \sqrt{np(1-p)}$$

$$X \mathrel{\dot\sim} N\Big(\mu = np,\ \sigma = \sqrt{np(1-p)}\Big)$$

Illustration: Effect of n on the Approximation

Example: Binomial(n, $p = 0.5$)

As n grows (e.g. $n = 10, 30, 100$), the binomial histogram becomes increasingly bell-shaped and is better approximated by a normal curve.



Example: Defective Items

Example: Defective Items in a Large Lot

Ten percent of items in a large lot are defective. If a random sample of 400 items is selected, what is the probability that fewer than 25 defectives were be selected?

Example: Defective Items

Example: Defective Items in a Large Lot

Ten percent of items in a large lot are defective. If a random sample of 400 items is selected, what is the probability that fewer than 25 defectives were be selected?

X = the # of defectives among 400 items

$$X \sim \text{binomial}(n = 400,\ p = .1), \qquad x = 0, 1, 2, \ldots, 400$$

Example: Defective Items

Example: Defective Items in a Large Lot

Ten percent of items in a large lot are defective. If a random sample of 400 items is selected, what is the probability that fewer than 25 defectives were be selected?

X = the # of defectives among 400 items

$$X \sim \text{binomial}(n = 400,\ p = .1), \qquad x = 0, 1, 2, \ldots, 400$$

$$P(X \leq 25) = P(X = 0) + P(X = 1) + \cdots + P(X = 25)$$

Example: Defective Items — Solution

Normal approximation:

$$\mu = np = 400(.1) = 40, \qquad \sigma = \sqrt{np(1-p)} = \sqrt{400(.1)(.9)} = 6$$

Example: Defective Items — Solution

Normal approximation:

$$\mu = np = 400(.1) = 40, \qquad \sigma = \sqrt{np(1-p)} = \sqrt{400(.1)(.9)} = 6$$

$$X \dot{\sim} N(\mu = 40,\ \sigma = 6)$$

Example: Defective Items — Solution

Normal approximation:

$$\mu = np = 400(.1) = 40, \qquad \sigma = \sqrt{np(1-p)} = \sqrt{400(.1)(.9)} = 6$$

$$X \mathrel{\dot\sim} N(\mu = 40,\ \sigma = 6)$$

Using the continuity correction ($25 \to 25.5$):

$$P(X \le 25) = P\left(\frac{X-40}{6} \le \frac{25-40}{6}\right) = P(Z \le -2.5)$$

Example: Defective Items — Solution

Normal approximation:

$$\mu = np = 400(.1) = 40, \qquad \sigma = \sqrt{np(1-p)} = \sqrt{400(.1)(.9)} = 6$$

$$X \dot{\sim} N(\mu = 40,\ \sigma = 6)$$

Using the continuity correction ($25 \rightarrow 25.5$):

$$P(X \leq 25) = P\left(\frac{X - 40}{6} \leq \frac{25 - 40}{6}\right) = P(Z \leq -2.5)$$

$$\approx \texttt{pnorm(-2.5)}$$