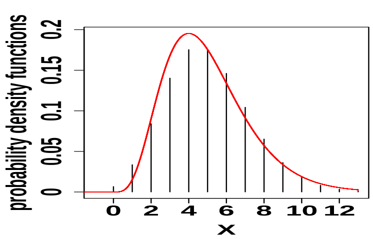
6 The Normal Distribution and Other Continuous Distributions

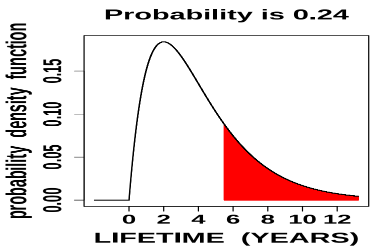
6.1 Continuous Probability Distributions

Rules for a continuous histogram.

1. The area of a histogram is 1.
2. The **probability** of the random variable taking a value in the interval from "a" to "b" is the **area** under the probability distribution curve within this interval.
3. The probability distribution function is nonnegative (cannot have negative probability).



Example: Let X be the lifetime of a computer CPU in years, as shown in the graph below. Determine the probability that a new CPU lasts at least 5.5 years.



□

6.2 The Normal Distribution

Here, we focus on the **normal distribution**, which exists in many applications (at least approximately).

Recall the **empirical rule** from section 3.3.

Empirical Rule

If a large number of observations are sampled from an approximately normal distribution, then (usually)

1. Approximately 68% of the observations fall within **one** standard deviation, σ, of the mean, μ.
2. Approximately 95% of the observations fall within **two** standard deviations, σ, of the mean, μ.
3. Approximately 99.7% of the observations fall within **three** standard deviations, σ, of the mean, μ.

Suppose X has a normal distribution with mean μ and standard deviation σ.

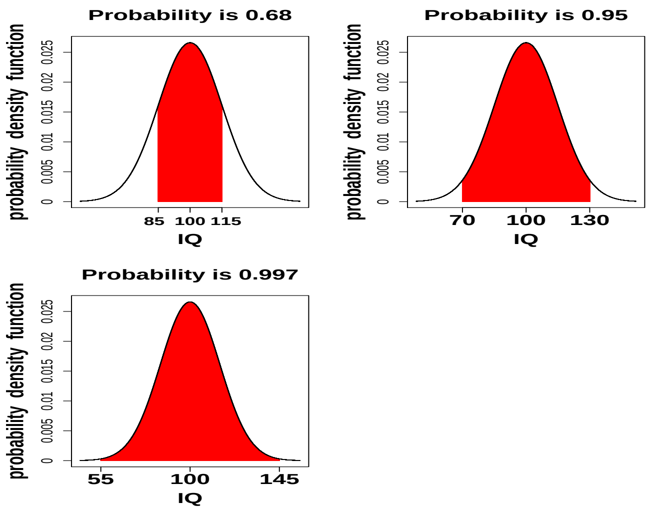
Notation: $X \sim N(\mu, \sigma)$

$$P(\mu - \sigma < X < \mu + \sigma) = 0.68$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.95$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$$

Example: IQ scores of normal adults on the Weschler test have a symmetric bell-shaped distribution with a mean of 100 and standard deviation of 15.



□

The normal distribution is bell-shaped and symmetric.

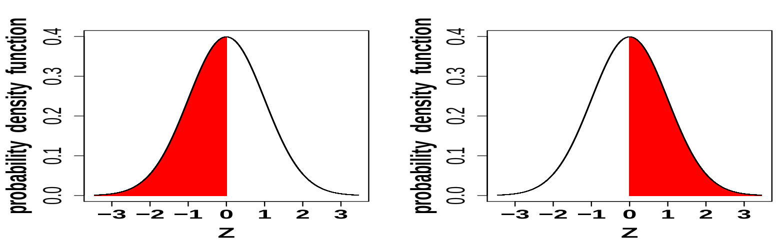
The standard normal distribution

Notation: $Z \sim N(0, 1)$.

Z represents the number of standard deviations, σ, away from the mean, μ.

Z is the "standardized" variable, known as the Z-score, and has **no units**.

Example: Compute $P(Z < 0)$, $P(Z \leq 0)$, $P(Z > 0)$, and $P(Z \geq 0)$.



□

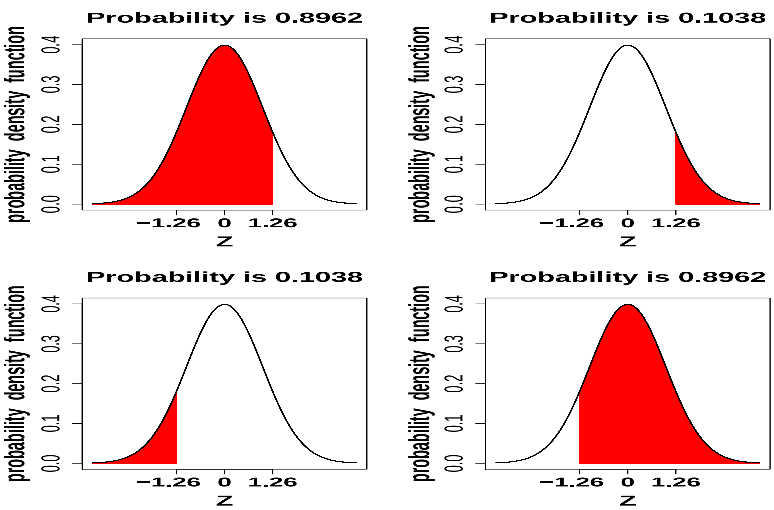
Example: *Using the standard normal table.* Let Z be a standard normal random variable.

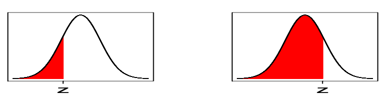
(a) Determine $P(Z < 1.26)$.

(b) Determine $P(Z > 1.26)$.

(c) Determine $P(Z < -1.26)$.

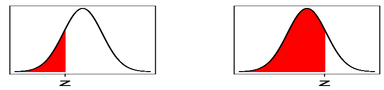
(d) Determine $P(Z > -1.26)$.





The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities										
Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮



The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

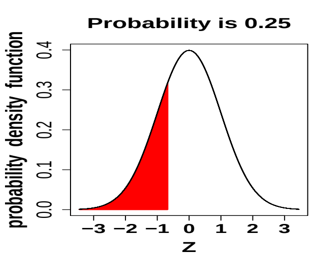
□

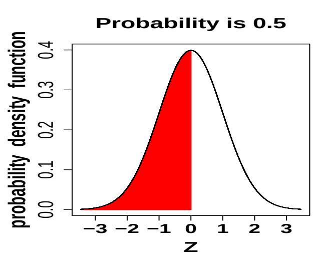
Example: *Using the standard normal table in reverse.* Let Z be a standard normal random variable.

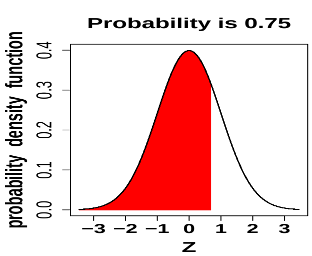
(a) Determine the 25th percentile of Z.

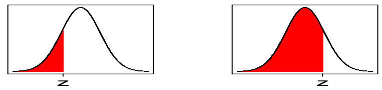
(b) Determine the 50th percentile of Z.

(c) Determine the 75th percentile of Z.





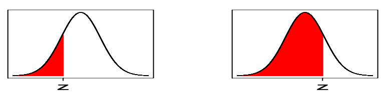




The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities

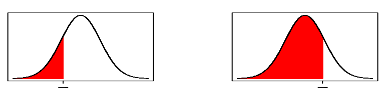
Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮



The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities

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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮



The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities

Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

□

Again consider $X \sim N(\mu, \sigma)$.

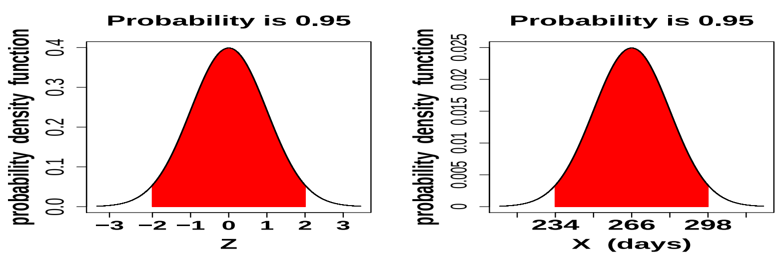
$$Z = \frac{X - \mu}{\sigma}$$

Reverse table look-up uses $X = \mu + \sigma Z$

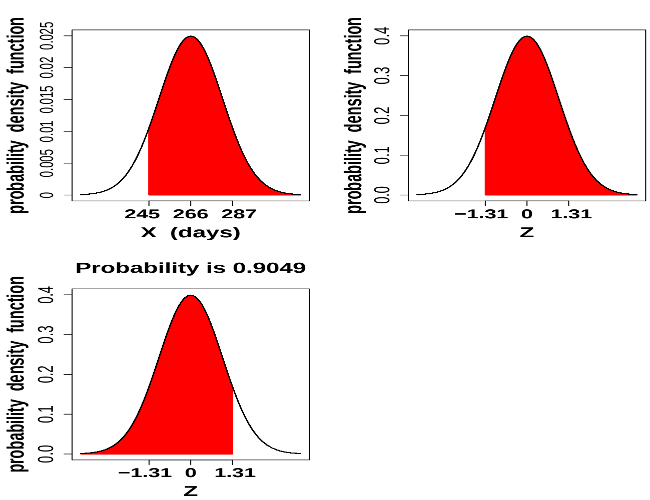
$$X \leftrightarrow Z \leftrightarrow \text{probability}$$

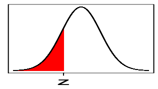
Example: The length of human pregnancies from conception to birth varies according to a distribution which is approximately normal with mean 266 days and standard deviation 16 days.

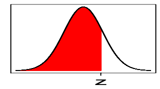
(a) Show the empirical rule regarding 95%.



(b) What proportion of pregnancies last more than 245 days?



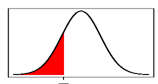


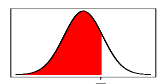


The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮



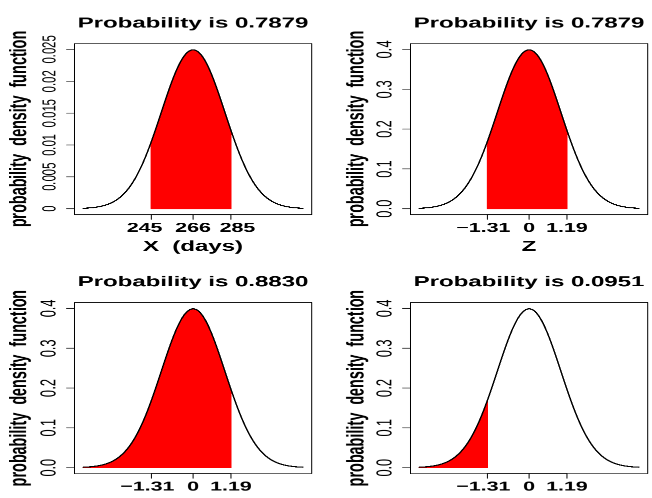


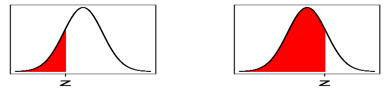
The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(c) What proportion of pregnancies last between 245 and 285 days?

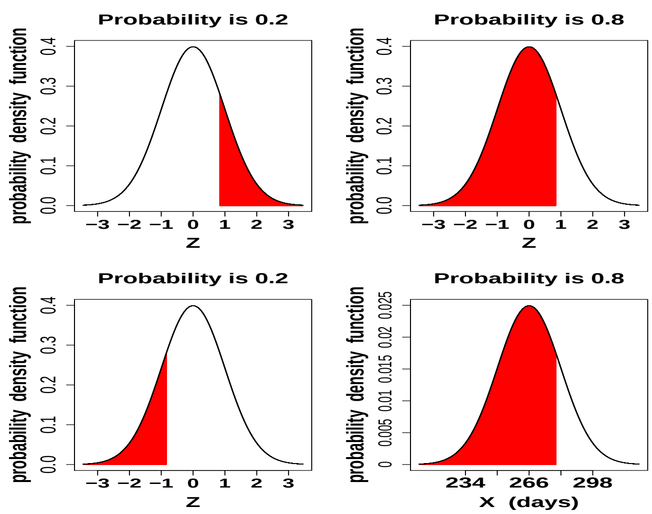


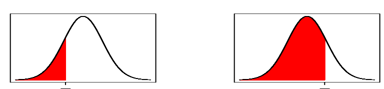


The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities										
Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

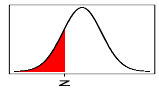
(d) How long do the longest 20% of pregnancies last?

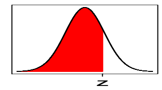




The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities										
Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

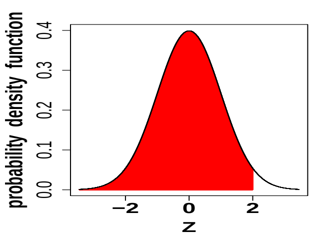


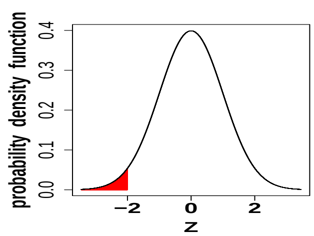


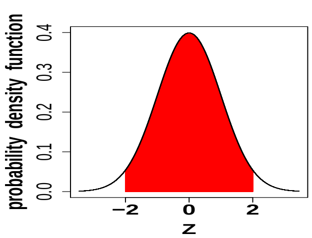
The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

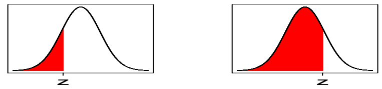
	Cumulative Probabilities									
Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Example: Let $X \sim N(\mu, \sigma)$. Using the standard normal table, verify the empirical rule regarding 95%. In other words, compute $P(\mu - 2\sigma < X < \mu + 2\sigma)$ to four significant digits.





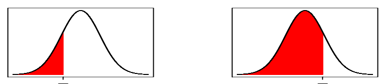




The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities

Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮



The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities

Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

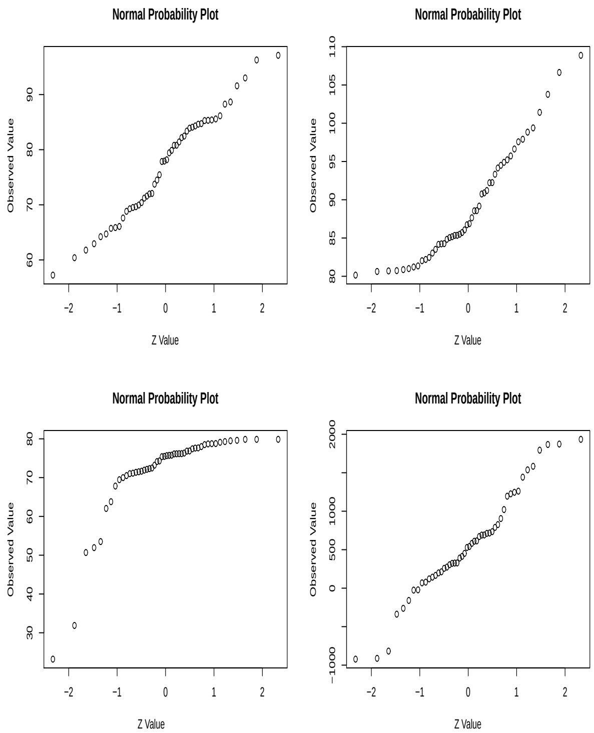
6.3 Evaluating Normality

Normal probability plot or Quantile-Quantile plot

How do we know if a sample is from a population which is approximately normal?

To construct a Q-Q plot, plot the ordered **observations** against *typical* or *quantile* ordered values from a **normal distribution**.

Example: Describe the distributions which likely generated the following Q-Q plots.



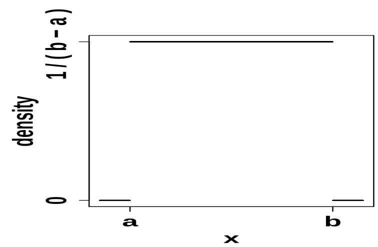
□

6.4 The Uniform Distribution

A **uniform** distribution has pdf

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

for *real* constants a and b such that $a < b$.



What is the **mean** of a uniform distribution?

The **variance** of a uniform distribution can be shown to be $\sigma^2 = (b-a)^2/12$.

Example: Suppose $X \sim \text{Uniform}(a = 30,\ b = 40)$. Determine:

(a) $P(X < 32)$

(b) $P(37 < X < 39)$

(c) $P(31.27 < X < 33.27)$

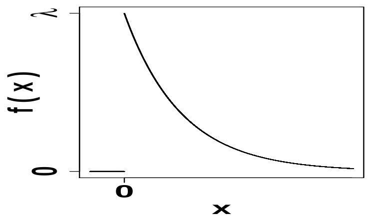
□

6.5 The Exponential Distribution

An **exponential** distribution has pdf

$$f(x) = \begin{cases} \lambda\, e^{-\lambda\, x}, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

with **mean** $= 1/\lambda > 0$.



Example: Let $X \sim$ Exponential(λ).

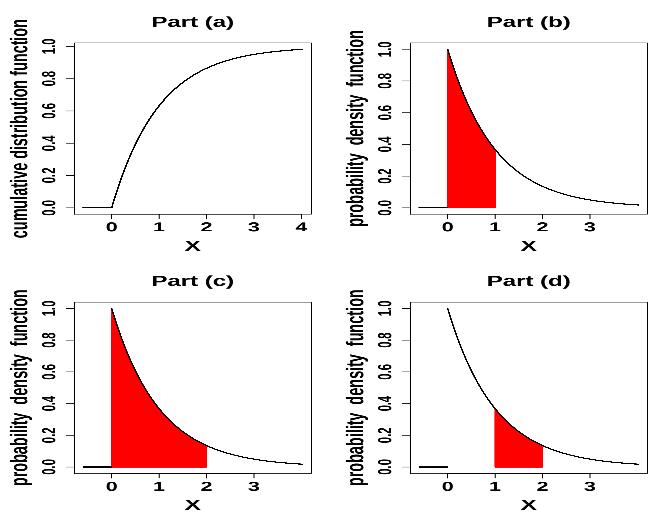
(a) The **cumulative distribution function** of X is the following:

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda\, x}, & \text{if } x \geq 0 \end{cases}$$

(b) Determine $P(X \leq 1)$, for $\lambda = 1$, using $F(\cdot)$ from part (a).

(c) Determine $P(X \leq 2)$, for $\lambda = 1$, using $F(\cdot)$ from part (a).

(d) Determine $P(1 \leq X \leq 2)$, for $\lambda = 1$, using $F(\cdot)$ from part (a).



□

Remark: The **exponential** distribution is **memoryless**.

If $X \sim$ Exponential, then it can be shown that
$P(X \geq t + t_0 | X \geq t_0) = P(X \geq t)$, for any $t, t_0 \geq 0$.

Example: Radioactive decay.

Example: Failure time of computer chip.

Remark: The **exponential** distribution is related to the **Poisson** distribution.

The amounts of time separating responses in a **Poisson** process are *independent and identically distributed* **exponential** random variables.

Example: Radioactive decay.

Let Y be the number of **responses** in one day.

Let X be the number of **days** in between responses.

6.6 The Normal Approximation to the Binomial Distribution

Suppose $X \sim \text{Binomial}(n,\ \pi)$.

Then, $\mu_x = n\pi$ and $\sigma_x = \sqrt{n\pi(1-\pi)}$.

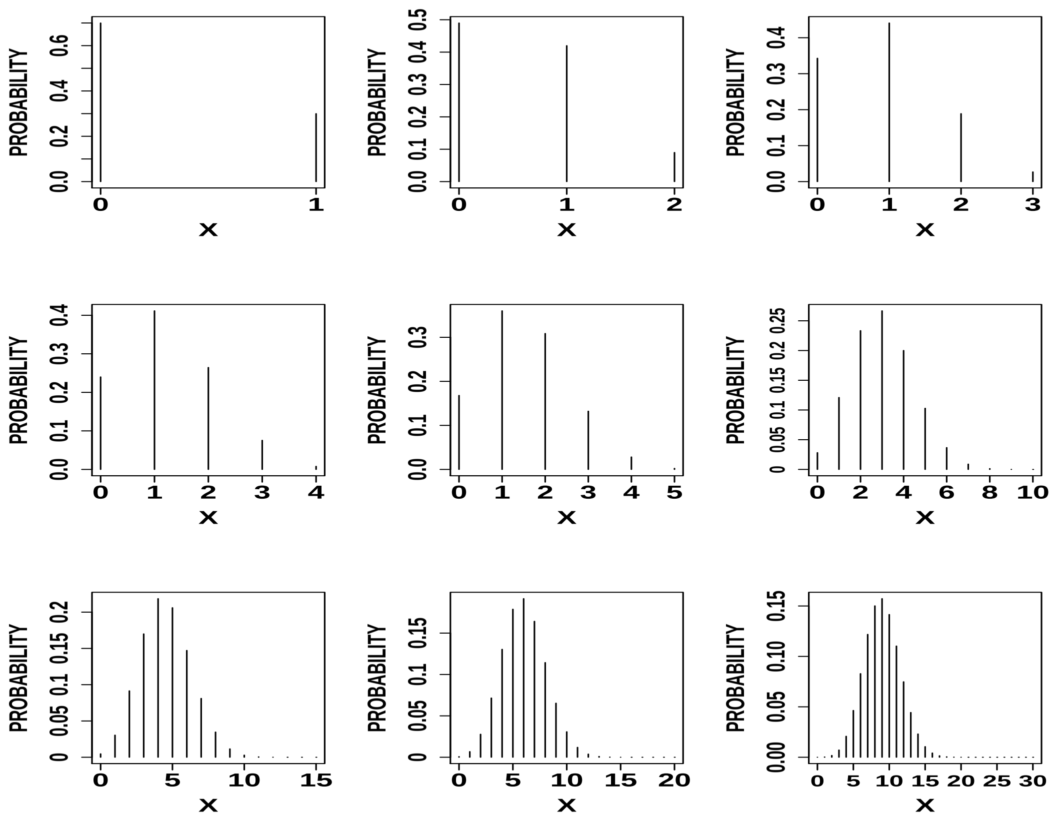
Rule of thumb: If $\min\{n\pi,\ n(1-\pi)\}$ is sufficiently large, say, at least **5**, then X is approximately $N(\mu_x,\ \sigma_x)$.

This result follows from the Central Limit Theorem (as defined in section 5.3), since a Binomial random variable is a sample sum of Bernoulli random variables.

Example: *Viewing the normal approximation to the binomial distribution.*

Consider the graphs below for **binomial** random variables, using $\pi = 0.3$ and

$n = 1, 2, 3, 4, 5, 10, 15, 20$, and 30.



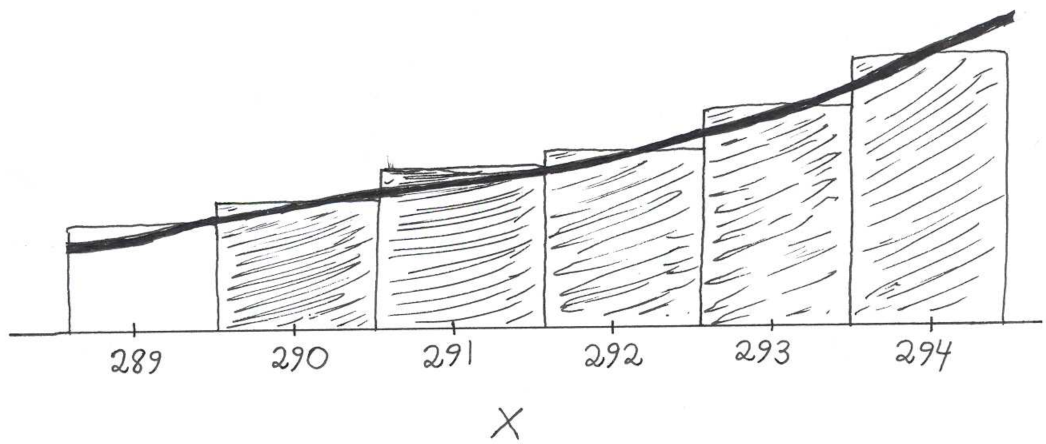
Example: Toss a coin 1000 times where $P(\text{heads}) = 0.3$, and let X be the number of heads.

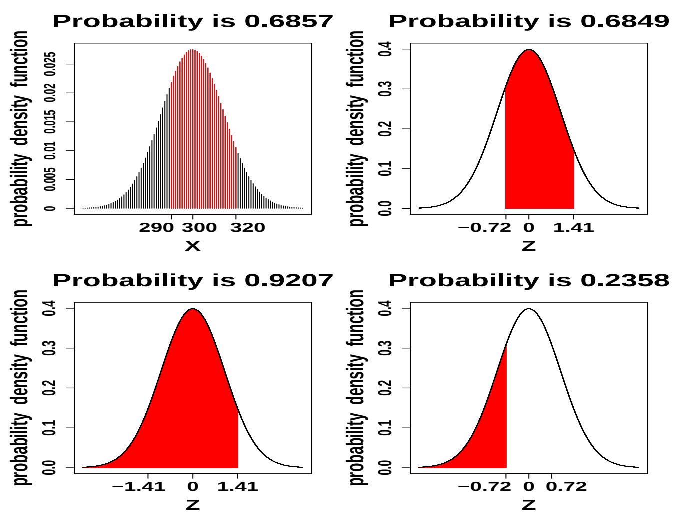
(a) State the **exact** distribution of X.

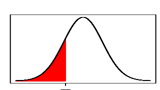
(b) Compute the **mean** and **standard deviation** of X.

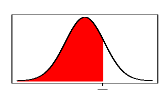
(c) Check the **rule of thumb**.

(d) Calculate $P(290 \leq X \leq 320)$ using the normal approximation **with continuity correction**.





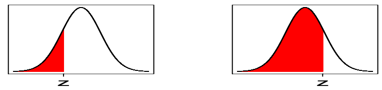




The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
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
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮



The Cumulative Standard Normal Distribution, pp. 914–915, Table E.2

Cumulative Probabilities										
Z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

□

Read p. 252, Appendix E6, Using Microsoft Excel to Compute Probabilities from the Normal Distribution and Other Continuous Distributions.