Probability and Statistics

Nels Grevstad

Metropolitan State University of Denver

ngrevsta@msudenver.edu

April 3, 2019

Notes

Topics

Notes

Objectives

Objectives:

- Recognize normal random variables.
- Use the normal distribution to find probabilities.
- Compute and interpret standardized values (z-scores).
- State the Empirical Rule.
- Find percentiles of the normal distribution.
- Use the normal distribution to approximate binomial probabilities.

Notes

Normal Random Variables (4.3)

- A random variable is said to follow a ***normal distribution*** with **parameters** μ and σ if its pdf is:

Normal(μ, σ) **Pdf**:

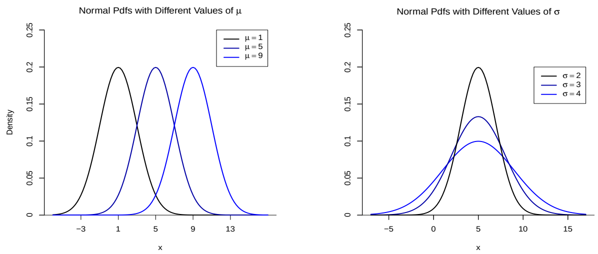
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for } -\infty < x < \infty$$

- We write

$$X \sim \textbf{N}(\mu, \sigma)$$

when X follows a normal distribution.

Notes



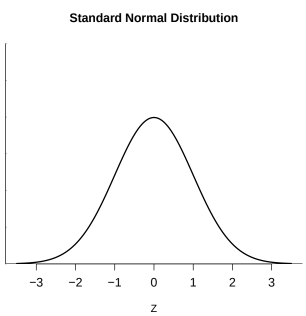
Notes

Normal(μ, σ) Mean and Variance: If $X \sim$ N(μ, σ), then

$$\begin{aligned} E(X) &= \mu \\ V(X) &= \sigma^2 \end{aligned}$$

Notes

- The **N$(0, 1)$** distribution ($\mu = 0$ and $\sigma = 1$) is called the ***standard normal*** distribution.



Notes

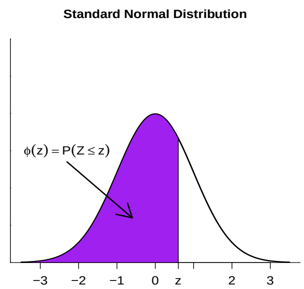
- We use Z to denote a **standard normal** random variable.
- The **pdf** of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad \text{for } -\infty < z < \infty$$

- The **cdf** of Z is denoted by $\phi(z)$. Thus

$$\phi(z) = P(Z \le z),$$

Notes



Notes

- To find probabilities such as $P(a \le Z \le b)$, we **can't integrate** the pdf,

$$\int_a^b f(z)\,dz \;=\; \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz$$

- Instead, the probabilities are obtained from values of $\phi(z) = P(Z \le z)$ given in a ***standard normal table***.

Notes

Example

Suppose

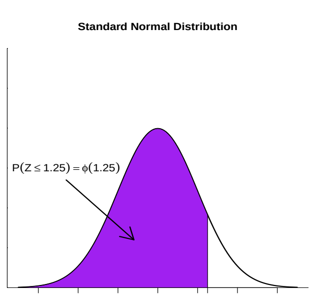
$$Z \sim \mathrm{N}(0,\, 1)$$

From the **standard normal table**,

$$P(Z \le 1.25) \;=\; \phi(1.25) \;=\; \mathbf{0.8944}$$

(1.2 row, 0.05 column of the table).

Notes



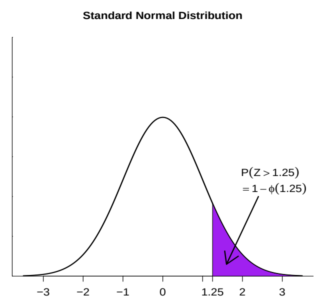
Notes

Example

Also from the **standard normal table**,

$$P(Z > 1.25) = 1 - \phi(1.25) = 1 - 0.8944 = \mathbf{0.1056}.$$

Notes



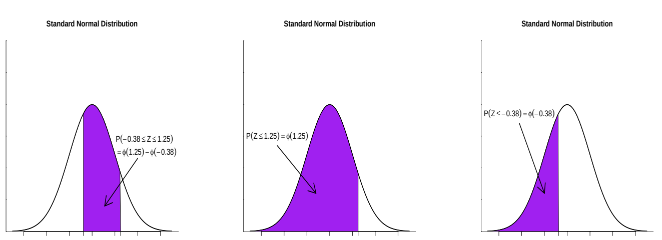
Notes

Example

Also from the **standard normal table**,

$$\begin{aligned} P(-0.38 \le Z \le 1.25) &= \phi(1.25) - \phi(-0.38) \\ &= 0.8944 - 0.3520 \\ &= \mathbf{0.5424}. \end{aligned}$$

Notes



Notes

- We just saw how to find probabilities using a $\mathbf{N}(0,\ 1)$ distribution.
- To find probabilities using normal distributions with **other** values of μ and σ (besides 0 and 1), we use the following proposition.

Notes

Proposition

If $X \sim \mathrm{N}(\mu, \sigma)$ and we define a new random variable Z by

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

then

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

Thus

1. $P(X \leq a) = P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = P\left(Z \leq \frac{a-\mu}{\sigma}\right) = \phi\left(\frac{a-\mu}{\sigma}\right)$

Notes

1. $P(X > b) = P\left(Z > \frac{b-\mu}{\sigma}\right) = 1 - \phi\left(\frac{b-\mu}{\sigma}\right)$
2. $P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$

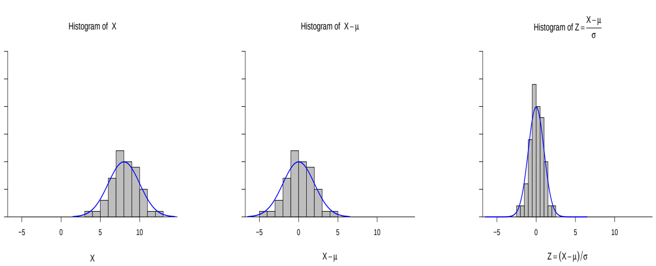
- The variable

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

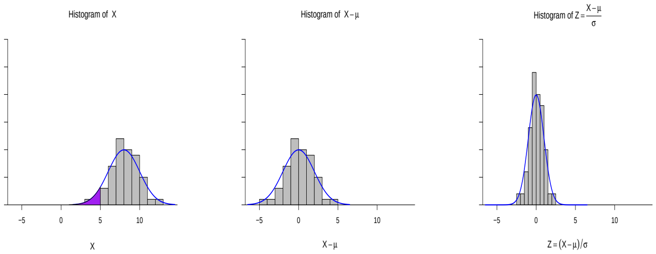
is called a ***standardized*** version of X, or ***z-score***.

It's measured in ***standard units*** (standard deviations away from the mean).

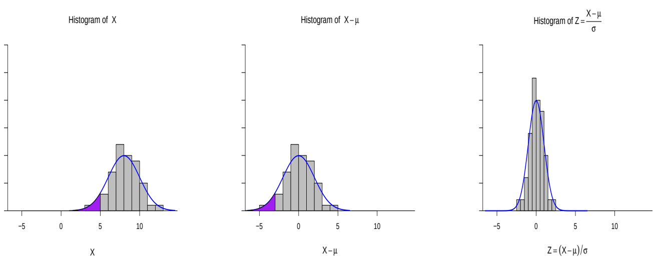
Notes



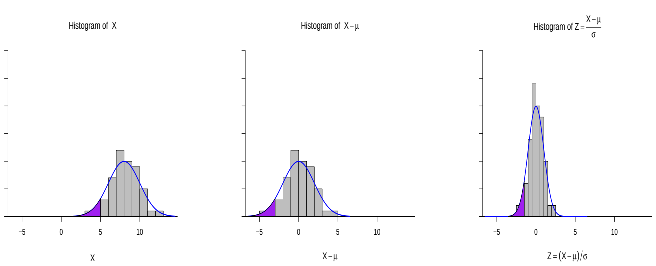
Notes



Notes



Notes

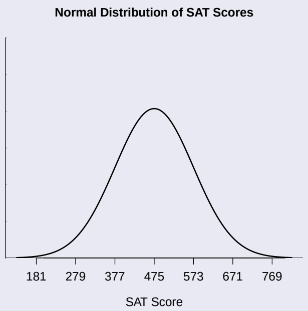


Notes

Example

Scores on the verbal Scholastic Aptitude Test (SAT) follow a **normal distribution** with **mean 475** and **standard deviation 98**.

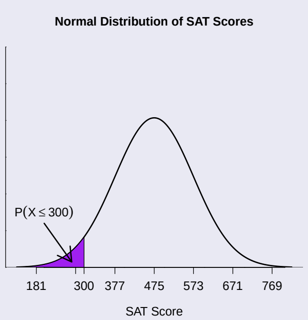
Notes



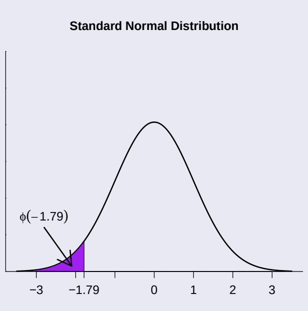
Notes

First we'll find $P(X \leq 300)$, the probability that an SAT score will be **less than 300**.

Notes



Notes



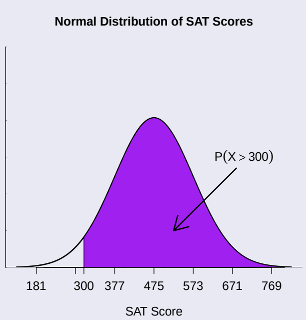
Notes

$$\begin{aligned} P(X \leq 300) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{300-\mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{300-475}{98}\right) \\ &= P(Z \leq -1.79) \\ &= \phi(-1.79) \\ &= \mathbf{0.0367}. \end{aligned}$$

Notes

Now we'll find $\boldsymbol{P(X > 300)}$, the probability that an SAT score will be **greater than 300**.

Notes



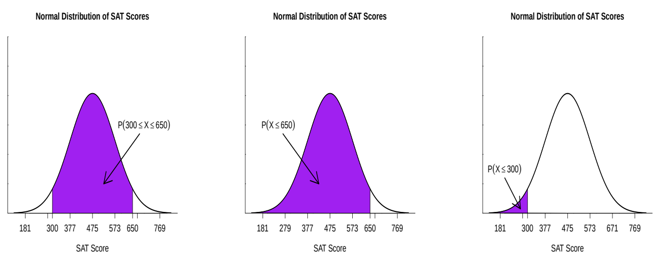
Notes

$$\begin{aligned} P(X > 300) &= 1 - P(X \leq 300) \\ &\vdots \\ &= 1 - \phi(-1.79) \\ &= 1 - 0.0367 \\ &= \mathbf{0.9633}. \end{aligned}$$

Notes

Lastly, we'll find $P(300 \leq X \leq 650)$, the probability that an SAT score will be **between 300 and 650**.

Notes



Notes

$$
\begin{aligned}
P(300 \leq X \leq 650) &= P\left(\frac{300-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{650-\mu}{\sigma}\right) \\
&= P\left(\frac{300-475}{98} \leq Z \leq \frac{650-475}{98}\right) \\
&= P(-1.79 \leq Z \leq 1.79) \\
&= \phi(1.79) - \phi(-1.79) \\
&= 0.9633 - 0.0367 \\
&= \mathbf{0.9266}.
\end{aligned}
$$

Notes

- A **standardized value** (or **z-score**) can be used to indicate an individual's standing **relative to others** in the population.

Example

Suppose you score **70** on your **Math** test, for which the **mean** is **65** and **standard deviation** is **5**.

Suppose also you score **80** on your **English** test, for which the **mean** is **75** and **standard deviation** is **7**.

On which test did you perform better relative to the rest of the class?

Notes

The **standardized** Math score is

$$Z = \frac{70 - 65}{5} = \mathbf{1}.0,$$

so it's **1.0** standard deviation above the mean.

The **standardized** English score is

$$Z = \frac{80 - 75}{7} = \mathbf{0}.7,$$

so it's **0.7** of a standard deviation above the mean.

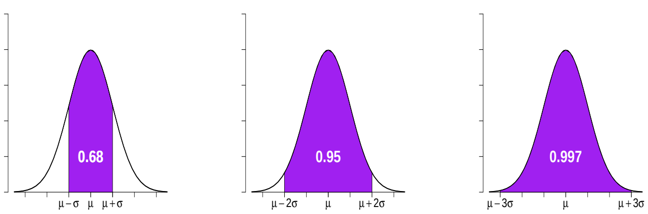
You did better on the **Math** test.

Notes

Empirical Rule (or 68-95-99.7 Rule): For **any normal distribution**,

1. Approximately **68%** of the distribution lies within **one σ of μ**.
2. Approximately **95%** of the distribution lies within **two σ's of μ**.
3. Approximately **99.7%** of the distribution lies within **three σ's of μ**.

Notes



Notes

- The $(100p)$***th percentile*** of a normal distribution is a value below which $(100p)$**%** of the distribution lies.
- For example, the **90th percentile** is the value below which 90% of the distribution lies.

Notes

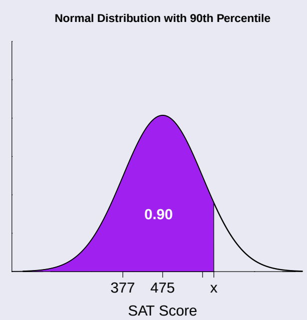
Example (Cont'd)

Recall that scores on the verbal SAT follow a **normal distribution** with **mean 475** and **standard deviation 98**.

The **90th percentile** is the score below which **90%** of all scores lie.

It's marked x on the horizontal axis below.

Notes

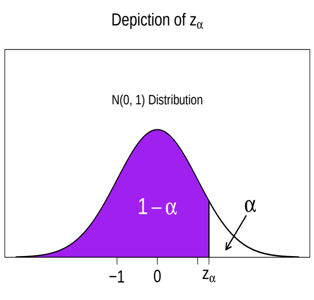


Notes

- We'll see how to find a **percentile** of a normal distribution with mean μ and standard deviation σ.
- First, though, we need to look at how to find a **percentile** of the **standard normal** distribution.

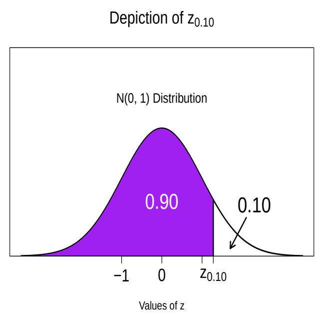
Notes

- We use z_α to denote the value that has area α to its **right** under the **N**$(0,\ 1)$ curve.



Notes

- For example, $z_{0.10}$ has area **0.10** to its **right** under the N$(0,\ 1)$ curve:



Notes

- z_α is the $\mathbf{100(1-\alpha)}$**th percentile** of the N$(0, 1)$ distribution.

  For example, $z_{0.10}$ is the **90th percentile** of the N$(0, 1)$ distribution.

- z_α is called a z ***critical value***.

Notes

Example

To find a z_α value, we search the body of the standard normal table for $1-\alpha$, then get the corresponding z value (from the table margin). We find that

$$\begin{aligned} z_{0.10} &= 1.28 \\ z_{0.05} &= 1.64 \\ z_{0.025} &= 1.96 \\ z_{0.005} &= 2.58 \end{aligned}$$

Notes

The $\mathbf{100p}$**th percentile** of a N$(\mu,\ \sigma)$ distribution, x, is obtained by "unstandardizing" the $100p$th percentile z of the N$(0,\ 1)$ distribution:

$$x = \mu + z\,\sigma$$

The above expression was obtained by solving

$$z = \frac{x-\mu}{\sigma}$$

for x.

Notes

Example (Cont'd)

Recall that verbal SAT scores follow a **normal distribution** with **mean 475** and **standard deviation 98**.

We'll find the **90th percentile** of the distribution.

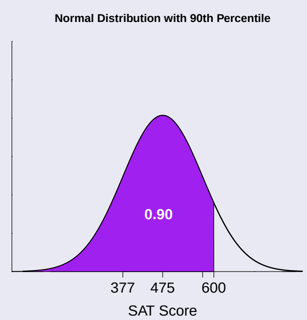
The 90th percentile of the **N**$(0,\ 1)$ distribution is

$$\boldsymbol{z = 1.28},$$

so the **90th percentile** of the distribution of **SAT scores** is

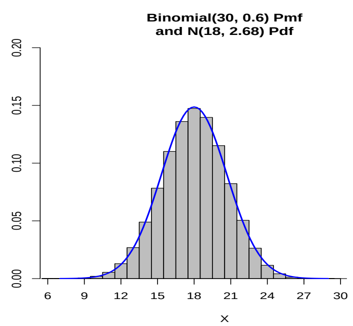
$$\begin{aligned} x &= \mu + z\sigma \\ &= 475 + 1.28(98) \\ &= \mathbf{600}. \end{aligned}$$

Notes



Notes

- The **normal** distribution can be used to **approximate** a **binomial** distribution when the number of trials n **is large**.



Notes

Proposition

Suppose

$$\boldsymbol{X \sim \textbf{binomial}(n, p)},$$

so that $\mu_x = np$ and $\sigma_x = \sqrt{np(1-p)}$. Then **if n is large**,

$$\boldsymbol{X \sim \textbf{N}\left(np,\ \sqrt{np(1-p)}\right)} \qquad \textbf{approximately}.$$

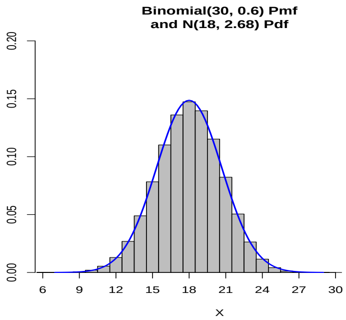
In particular, for each possible value x of X,

$$\begin{aligned} P(X \le x) &\approx \left(\begin{array}{l}\text{Area under the normal curve} \\ \text{to the left of } \boldsymbol{x + 0.5}\end{array}\right) \\ &= \phi\left(\frac{\boldsymbol{x + 0.5} - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

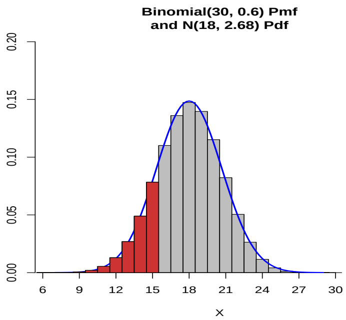
Notes

- Adding **0.5** to x is referred to as a ***continuity correction***.

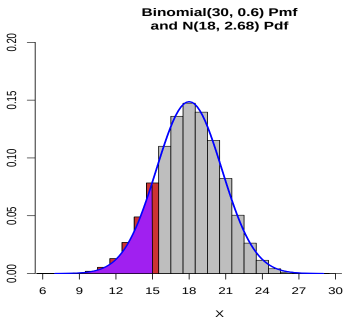
Notes



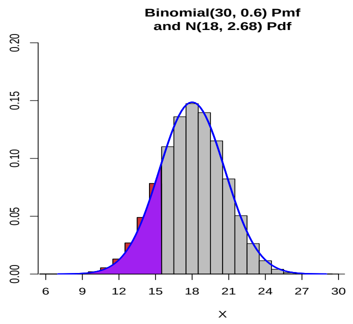
Notes



Notes



Notes



Notes

Example

A student is taking a true/false test with **100** questions. Suppose she has a probability $p = 3/4$ of getting each question right.

Let

$$X \;=\; \text{The number of questions she gets right.}$$

Then

$$X \sim \text{binomial}(100,\, 3/4)$$

We'll use the **normal distribution** to **approximate** the probability $P(X \leq 70)$ that she'll get **70 or fewer** right.

Notes

The **mean** and **standard deviation** of the distribution of X are

$$\mu_x \;=\; n\,p \;=\; 100\left(\frac{3}{4}\right) \;=\; \mathbf{75}$$

and

$$\sigma_x \;=\; \sqrt{n\,p\,(1-p)} \;=\; \sqrt{100\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)} \;=\; \mathbf{4.3}$$

Notes

Because n **is large**, the **normal approximation to the binomial** gives

$$\begin{aligned} P(X \leq 70) &\approx \phi\left(\frac{70+0.5-np}{\sqrt{np(1-p)}}\right) \\ &= \phi\left(\frac{70.5-75}{4.3}\right) \\ &= \phi(-1.05) \\ &= \mathbf{0.1469}. \end{aligned}$$

(Note that the exact binomial probability is 0.1495).

Notes

- In practice, n **is large** enough for the ***normal approximation to the binomial*** to be valid as long as

$$np \geq 10 \quad \textbf{and} \quad n(1-p) \geq 10.$$

Notes

Notes

Notes

Notes