Normal Distribution

Module 6

Statistics 251: Statistical Methods

Updated 2021

Normal

The normal distribution is a very important continuous distribution since many things naturally have a normal distribution. It is widely used and widely abused.

IQ scores, often real-estate prices, and even grades mostly follow a normal distribution (I do not grade "on the curve" because I would rather that most get As and Bs).

We will investigate the normal distribution, standard normal distribution, and applications thereof.

Terms

The normal distribution has two parameters: mean (μ) and standard devaition (σ).

$$\text{Shorthand notation: } X \sim N(\mu, \sigma)$$

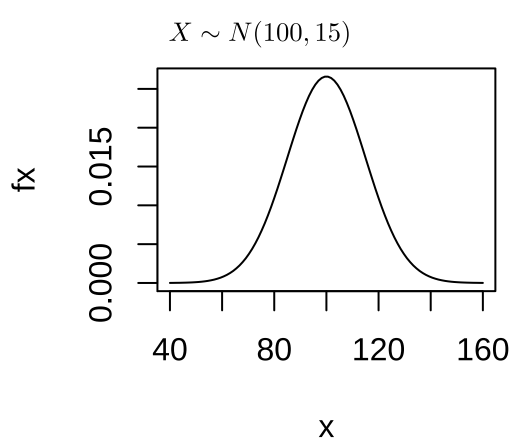
The probability density function (pdf) for a normal distribution is:

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

To use this, you would have to use calculus but we will not have to, there is another (easier) way to calculate probabilities.

Normal Graph

With mean $\mu = 100$ and standard deviation $\sigma = 15$, thus



Standard Normal Distribution

The standard normal distribution is a normal distribution of standardized values, called z-scores. The z-scores are measured in units of the standard deviation. It is always centered at 0 and always has a standard deivation of 1, thus $Z \sim N(0, 1)$.

Z-score logistics I

If $X \sim N(\mu, \sigma)$ then

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

The z-score tells you how many standard deviations you are in relation to the mean. If $z = -1$, that is one standard deviation *below* the mean, if $z = 1$, that is one standard deviation *above* the mean. Values of x that are smaller than the mean will have negative z-scores, x values that are larger than the mean will have positive z-scores. If x is equal to the mean, then it has a z-score of 0. Z-scores allow comparison of data that are on different scales

Z-score logistics II

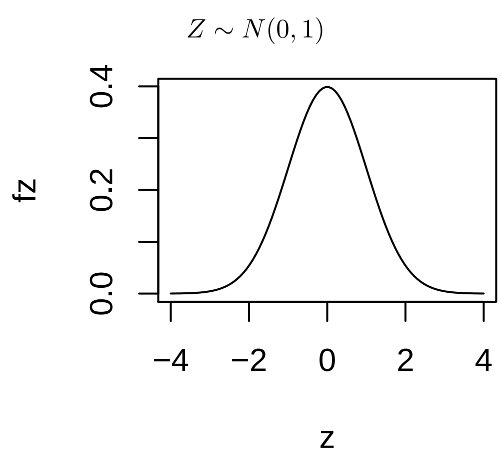
The Empirical Rule is derived from the normal distribution:

68% of observations are within the interval $\overline{X} \pm 1s$
95% of observations are within the interval $\overline{X} \pm 2s$
99.7% of observations are within the interval $\overline{X} \pm 3s$

One of the examples from the Standard Normal examples will show the area between -1 and 1

Standard Normal Graph

With mean $\mu = 0$ and standard deviation $\sigma = 1$, thus



Standard Normal Example

With z-scores:

$P(Z < 1)$
$P(Z > 1)$
$P(Z < -1)$
$P(Z > -1)$

$P(-1 < Z < 1)$
z-score for top 1%
z-score for $Q1$
z-score for $Q3$

Normal Example I

Suppose that $X \sim N(68, 3)$ for golf scores of a high school team. Find:

$P(X < 65)$
$P(X < 68)$
$P(X > 70)$
$P(60 < X < 69)$
Top 5% of golf scores
IQR of golf scores

Normal Example II (bahaha)

What is normal?

Spongeboob

Spongeboob II

Standard Normal Solutions

With z-scores:

$P(Z < 1) = 0.841345$

$P(Z > 1) = 0.158655$

$P(Z < -1) = 0.158655$

$P(Z > -1) = 0.841345$

$P(-1 < Z < 1) = 0.682689$

(Empirical Rule derivation)

$P(-2 < Z < 2) = 0.9545$
$P(-3 < Z < 3) = 0.9973$

Normal Solutions I part 1

$P(X < 65) = P\left(Z < \frac{65-68}{3}\right) = P(Z < -1) = 0.158655$

$P(X < 68) = 0.5$ (because the mean is 68 and z will be 0, which is in the center)

$P(X > 70) = P\left(Z < \frac{70-68}{3}\right) = P(Z > 0.67)$

$= 1 - P(Z < 0.67) = 0.251429$

$P(60 < X < 69) = P\left(\frac{60-68}{3} < Z < \frac{69-68}{3}\right)$

$= P(-2.67 < Z < 0.33) = 0.625507$

Normal Solutions I part 2

Top 5% of golf scores: find z with tail area to the *right* of 0.05 (same as bottom 95%)
$z = 1.644854$ and solve for X where $x = z\sigma + \mu \Rightarrow$
$x = (1.644854)(3) + 68 = 72.934561$

IQR of golf scores: find $z_{0.25}$ and $z_{0.75}$, and solve for the two values of X, then calculate the IQR.

$z_{0.25} = -0.67449$ and $z_{0.75} = 0.67449$.

Now $x_{Q1} = (-0.67449)(3) + 68 = 65.976531$ and
$x_{Q3} = (0.67449)(3) + 68 = 70.023469$

$IQR = 4.046939$

Normal Solutions II

There is no solution for a normal Spongebob... he will be crazy forever... forever... forever...