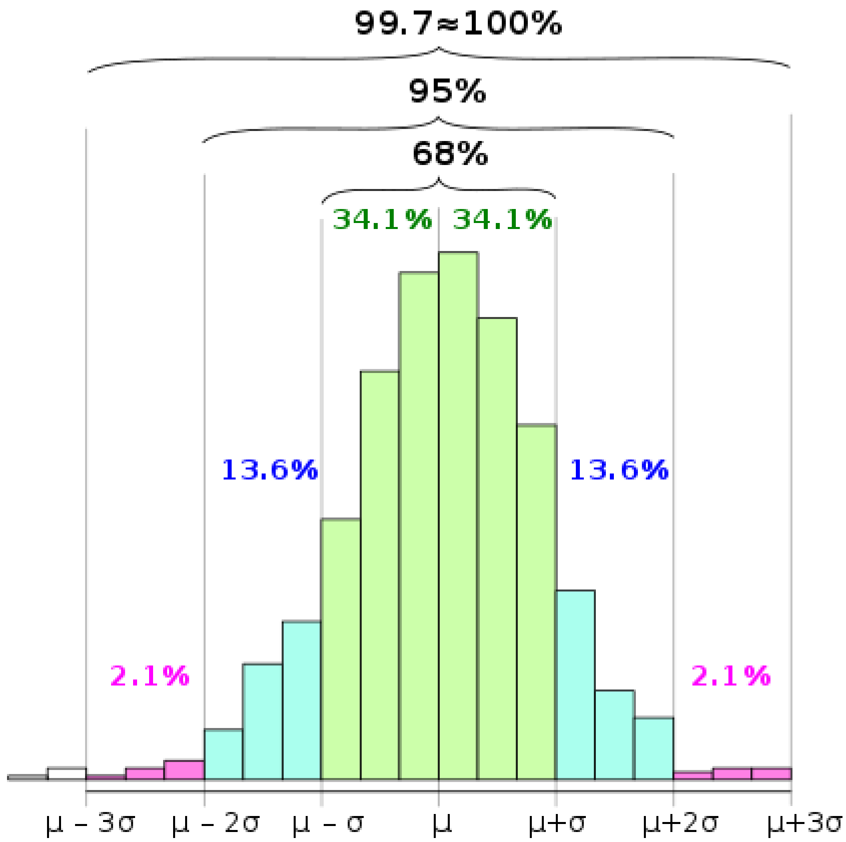
WIKIPEDIA

68–95–99.7 rule

In statistics, the **68–95–99.7 rule**, also known as the **empirical rule**, is a shorthand used to remember the percentage of values that lie within a band around the mean in a normal distribution with a width of two, four and six standard deviations, respectively; more accurately, 68.27%, 95.45% and 99.73% of the values lie within one, two and three standard deviations of the mean, respectively.

In mathematical notation, these facts can be expressed as follows, where X is an observation from a normally distributed random variable, μ is the mean of the distribution, and σ is its standard deviation:

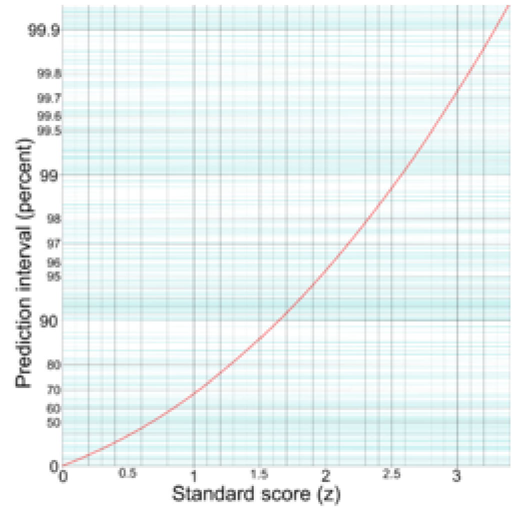


For an approximately normal data set, the values within one standard deviation of the mean account for about 68% of the set; while within two standard deviations account for about 95%; and within three standard deviations account for about 99.7%. Shown percentages are rounded theoretical probabilities intended only to approximate the empirical data derived from a normal population. This rule is generally taught within common core mathematics.

$$\begin{aligned}
\Pr(\mu - 1\sigma \le X \le \mu + 1\sigma) &\approx 0.6827\\
\Pr(\mu - 2\sigma \le X \le \mu + 2\sigma) &\approx 0.9545\\
\Pr(\mu - 3\sigma \le X \le \mu + 3\sigma) &\approx 0.9973
\end{aligned}$$

In the empirical sciences the so-called **three-sigma rule of thumb** expresses a conventional heuristic that nearly all values are taken to lie within three standard deviations of the mean, and thus it is empirically useful to treat 99.7% probability as near certainty.[1] The usefulness of this heuristic depends significantly on the question under consideration. In the social sciences, a result may be considered "significant" if its confidence level is of the order of a two-sigma effect (95%), while in particle physics, there is a convention of a five-sigma effect (99.99994% confidence) being required to qualify as a discovery.

The "three-sigma rule of thumb" is related to a result also known as the **three-sigma rule,** which states that even for non-normally distributed variables, at least 88.8% of cases should fall within properly calculated three-sigma intervals. It follows from Chebyshev's Inequality. For unimodal distributions the probability of being within the interval is at least 95%. There may be certain assumptions for a distribution that force this probability to be at least 98%.[2]



Prediction interval (on the y-axis) given from the standard score (on the x-axis). The y-axis is logarithmically scaled (but the values on it are not modified).

Contents

Cumulative distribution function

These numerical values "68%, 95%, 99.7%" come from the cumulative distribution function of the normal distribution.

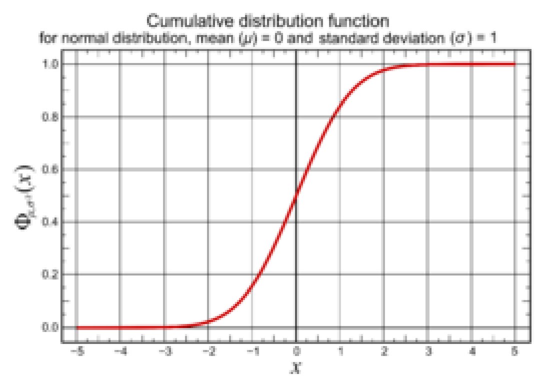


Diagram showing the cumulative distribution function for the normal distribution with mean (μ) 0 and variance (σ^2) 1.

The prediction interval for any standard score z corresponds numerically to $(1-(1-\Phi_{\mu,\sigma^2}(z))\cdot 2)$.

For example, $\Phi(2) \approx 0.9772$, or $\Pr(X \le \mu + 2\sigma) \approx 0.9772$, corresponding to a prediction interval of $(1 - (1 - 0.97725)\cdot 2) = 0.9545 = 95.45\%$. Note that this is not a symmetrical interval – this is merely the probability that an observation is less than $\mu + 2\sigma$. To compute the probability that an observation is within two standard deviations of the mean (small differences due to rounding):

$$\Pr(\mu - 2\sigma \le X \le \mu + 2\sigma) = \Phi(2) - \Phi(-2) \approx 0.9772 - (1 - 0.9772) \approx 0.9545$$

This is related to confidence interval as used in statistics: $\bar{X} \pm 2\frac{\sigma}{\sqrt{n}}$ is approximately a 95% confidence interval when $\bar{X}$ is the average of a sample of size n.

Normality tests

The "68–95–99.7 rule" is often used to quickly get a rough probability estimate of something, given its standard deviation, if the population is assumed to be normal. It is also used as a simple test for outliers if the population is assumed normal, and as a normality test if the population is potentially not normal.

To pass from a sample to a number of standard deviations, one first computes the deviation, either the error or residual depending on whether one knows the population mean or only estimates it. The next step is standardizing (dividing by the population standard deviation), if the population parameters are known, or studentizing (dividing by an estimate of the standard deviation), if the parameters are unknown and only estimated.

To use as a test for outliers or a normality test, one computes the size of deviations in terms of standard deviations, and compares this to expected frequency. Given a sample set, one can compute the studentized residuals and compare these to the expected frequency: points that fall more than 3 standard deviations from the norm are likely outliers (unless the sample size is significantly large, by which point one expects a sample this extreme), and if there are many points more than 3 standard deviations from the norm, one likely has reason to question the assumed normality of the distribution. This holds ever more strongly for moves of 4 or more standard deviations.

One can compute more precisely, approximating the number of extreme moves of a given magnitude or greater by a Poisson distribution, but simply, if one has multiple 4 standard deviation moves in a sample of size 1,000, one has strong reason to consider these outliers or question the assumed normality of the distribution.

For example, a 6σ event corresponds to a chance of about two parts per billion. For illustration, if events are taken to occur daily, this would correspond to an event expected every 1.4 million years. This gives a simple normality test: if one witnesses a 6σ in daily data and significantly fewer than 1 million years have passed, then a normal distribution most likely does not provide a good model for the magnitude or frequency of large deviations in this respect.

In The Black Swan, Nassim Nicholas Taleb gives the example of risk models according to which the Black Monday crash would correspond to a 36-σ event: the occurrence of such an event should instantly suggest that the model is flawed, i.e. that the process under consideration is not satisfactorily modelled by a normal distribution. Refined models should then be considered, e.g. by the introduction of stochastic volatility. In such discussions it is important to be aware of problem of the gambler's fallacy, which states that a single observation of a rare event does not contradict that the event is in fact rare. It is the observation of a plurality of purportedly rare events that increasingly undermines the hypothesis that they are rare, i.e. the validity of the assumed model. A proper modelling of this process of gradual loss of confidence in a hypothesis would involve the designation of prior probability not just to the hypothesis itself but to all possible alternative hypotheses. For this reason, statistical hypothesis testing works not so much by confirming a hypothesis considered to be likely, but by refuting hypotheses considered unlikely.

Table of numerical values

Because of the exponential tails of the normal distribution, odds of higher deviations decrease very quickly. From the rules for normally distributed data for a daily event:

Range	Expected fraction of population inside range	Approximate expected frequency outside range	Approximate frequency for daily event
μ ± 0.5σ	0.382 924 922 548 026	2 in 3	Four or five times a week
μ ± σ	0.682 689 492 137 086	1 in 3	Twice a week
μ ± 1.5σ	0.866 385 597 462 284	1 in 7	Weekly
μ ± 2σ	0.954 499 736 103 642	1 in 22	Every three weeks
μ ± 2.5σ	0.987 580 669 348 448	1 in 81	Quarterly
μ ± 3σ	0.997 300 203 936 740	1 in 370	Yearly
μ ± 3.5σ	0.999 534 741 841 929	1 in 2149	Every six years
μ ± 4σ	0.999 936 657 516 334	1 in 15 787	Every 43 years (twice in a lifetime)
μ ± 4.5σ	0.999 993 204 653 751	1 in 147 160	Every 403 years (once in the modern era)
μ ± 5σ	0.999 999 426 696 856	1 in 1 744 278	Every 4776 years (once in recorded history)
μ ± 5.5σ	0.999 999 962 020 875	1 in 26 330 254	Every 72 090 years (thrice in history of modern humankind)
μ ± 6σ	0.999 999 998 026 825	1 in 506 797 346	Every 1.38 million years (twice in history of humankind)
μ ± 6.5σ	0.999 999 999 919 680	1 in 12 450 197 393	Every 34 million years (twice since the extinction of dinosaurs)
μ ± 7σ	0.999 999 999 997 440	1 in 390 682 215 445	Every 1.07 billion years (four times in history of Earth)
μ ± xσ	$\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)$	1 in $\frac{1}{1-\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)}$	Every $\frac{1}{1-\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)}$ days

See also

- p-value
- Six Sigma#Sigma levels
- Standard score
- t-statistic

References

1. this usage of "three-sigma rule" entered common usage in the 2000s, e.g. cited in Schaum's Outline of Business Statistics. McGraw Hill Professional. 2003. p. 359, and in Grafarend, Erik W. (2006). Linear and Nonlinear Models: Fixed Effects, Random Effects, and Mixed Models. Walter de Gruyter. p. 553.
2. See:
   - Wheeler, D. J.; Chambers, D. S. (1992). Understanding Statistical Process Control (https://books.google.com/books?id=XvMJAQAAMAAJ). SPC Press.
   - Czitrom, Veronica; Spagon, Patrick D. (1997). Statistical Case Studies for Industrial Process Improvement (https://books.google.com/books?id=gEpkYxwDbvgC&pg=PA342). SIAM. p. 342.
   - Pukelsheim, F. (1994). "The Three Sigma Rule". American Statistician. **48**: 88–91. JSTOR 2684253 (https://www.jstor.org/stable/2684253).

External links

- "The Normal Distribution (http://www-stat.stanford.edu/~naras/jsm/NormalDensity/NormalDensity.html)" by Balasubramanian Narasimhan
- "Calculate percentage proportion within x sigmas (http://www.wolframalpha.com/input/?i=erf%28x%2Fsqrt%282%29%29) at WolframAlpha

Retrieved from "https://en.wikipedia.org/w/index.php?title=68–95–99.7_rule&oldid=891319550"

This page was last edited on 7 April 2019, at 04:42 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.