DATA1001 Topic 4 Flashcards

Normal Model

Abyan Majid

Question	My answer
What type of data can be modeled by a normal curve? What are the parameters of a normal curve?	A continuous variable, usually one with a relatively large set of data points (to satisfy the Central Limit Theorem). The probability density function (PDF) that defines the normal curve requires two parameters: ● Mean ● Standard deviation
How would you formally denote that some random variable X follows a normal distribution of mean μ and standard deviation σ?	$X \sim N(\mu, \sigma^2)$
What is the difference between standard and general normal curves!	A standard normal curve has mean $\mu = 0$ and standard deviation $\sigma = 1$ as parameters. Meanwhile, a general normal curve has any mean and standard deviation as parameters.
How many percent of the data is within the range of: a) 1 SD from the mean in both directions? b) 2 SDs from the mean in both directions? c) 3 SDs from the mean in both directions?	a) 68% b) 95% c) 99.7%
Is it possible to convert a general normal curve to a standard normal curve? If so, how?	Convert all data points into a Z-score (standard unit). $Z = \frac{x_i - \bar{x}}{SD}$ for indexes $i = 1, \ldots, n$ in the set of data points
Recite the R syntax for getting the area under a normal curve (1) from the left until a data point x, and (2) from x all the way to the right!	(1) pnorm(x, mean, sd, lower.tail = TRUE) (2) pnorm(x, mean, sd, lower.tail = FALSE)
Can a model ever be true? Why or why not?	No, models can never be exactly true. Any model only approximates how the dependent variable grows based on the data points from which the model is derived. A good model does not tell the truth - but it does allow for you to reliably interpolate (or, though more prone to unreliability, extrapolate) from the data.

Identify 4 ways in which we can get an insight of whether it is appropriate to use the normal curve on a given data!	1. Plot a histogram of the data and see whether it's shaped approximately like a bell, and has no excessively long tails on either end. 2. Check the proportions of the data to verify whether it follows the 68%-95%-99.7% rule 3. Plot the QQ plot and see whether the data is roughly linear (if it is, then it is likely normal). 4. Conduct the shapiro test. If the p-value is greater than a significance level of 0.05 (by convention), then a normal curve is likely to fit the data.
Define measurement error in terms of chance error, and bias!	Measurement error is the difference between a measured value and the actual value due to chance and bias. A chance error is a measurement error caused due to chance. Bias is a systematic error that causes a constant number to be added or subtracted from a measurement deliberately or accidentally. In a more compact sense, $Measurement = exact\ value + chance\ error + bias$
How to minimize chance error?	Replicate the measurement under the same exact conditions.
What is the convention for identifying outliers in measurements in order to assume normality?	Any data point that is less than or greater than 3 SD from the mean should be considered an outlier.