Examples

CIND 123

Q1

- A computer crashes once every 2 days on average. What is the probability of there being 2 crashes in one week?

- What is Lambda?

Solution

- average rate given: lambda = 0.5 crashes/day. Hence, Poisson.
- (lambda t) = (0.5 per day * 7 days) = 3.5/week and n = 2.
- P(2) = (3.5)^2/2! * exp (-3.5) = 0.185

- dpois(2, lambda=3.5)

Q2

- Components are packed in boxes of 20. The probability of a component being defective is 0.1.
- What is the probability of a box containing 2 defective components?
- What is the probability of a box containing 11 non-defective components?
- $P(12 <= X <= 15)$ $p=0.9$

Solution

- What is the probability of not being defected?
- What is the probability of being defected?

- dbinom(2, size=20, prob=.1)
- dbinom(11,20,0.9)

- $\sum_{n=12}^{15} dbinom(n, 20,0.9)$ => n= 12:15, sum(dbinom(n,20,0.9))
- pbinom(15,20,0.9) – pbinom(11,20,0.9)

Q3

- Find the height of the probability distribution curve the student arriving in 28.5 minutes, mean=30, sd=5. What is z score for the t=28.5 minutes? What is t if z=3?

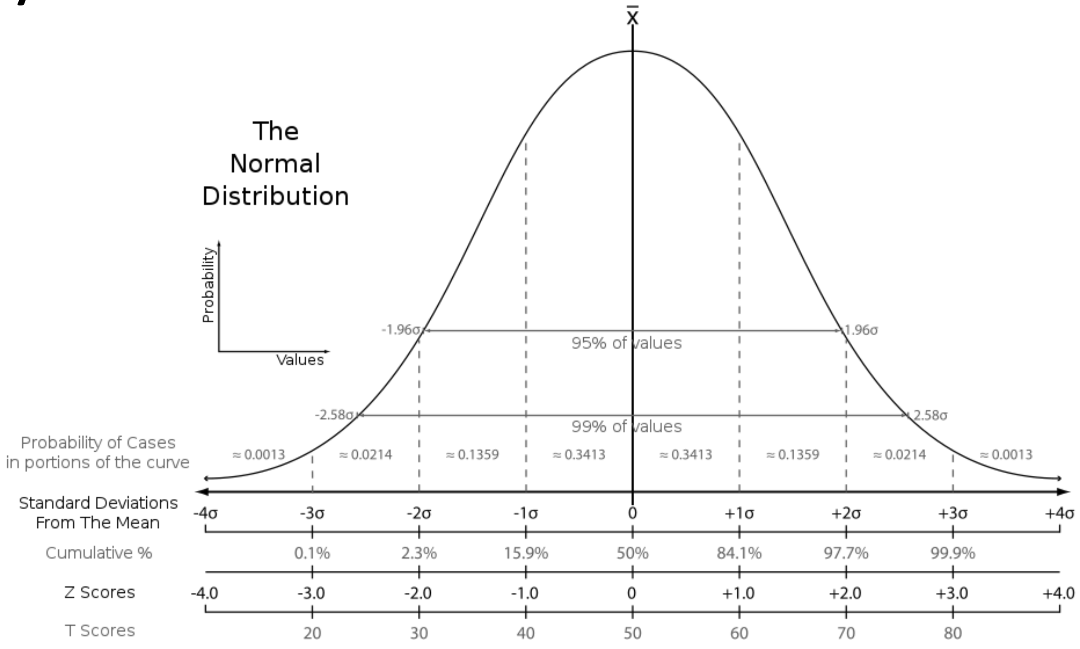
Solution

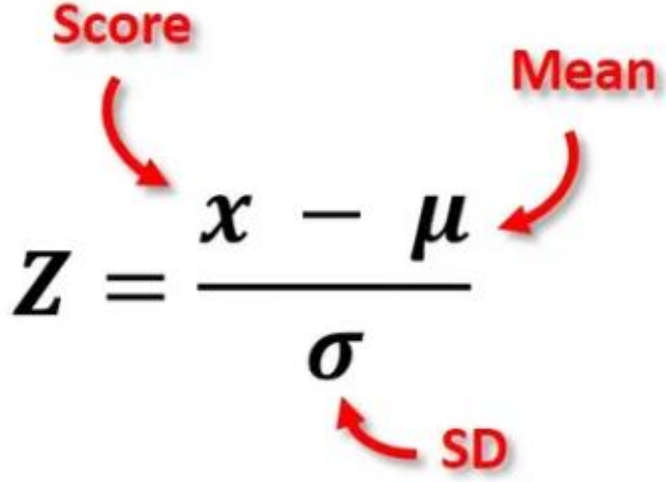
- `dnorm(28.5, mean=30, sd= 5)`

- `z=(x-mean)/sd=(28.5-30)/5=-0.3`
- `z=3, t=(z*sd)+mean` → `t=45 minutes`

Q4

- The monthly utility bills in a city are normally distributed with a mean value of 70 CAD, and variance of 64 CAD, what is the z score of 80 CAD utility bill.





Sampling

- Consider a large population with a mean of 100 and a standard deviation of 40. A random sample of size is 25 taken from this population. What is the standard error of the sampling distribution of sample mean? What is the mean value of the sample?

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}}$$

Q5

Suppose IQ's are normally distributed with a mean of 100 and a standard deviation of 15.

What percentage of people have an IQ between 110 and 125?

$P(110<X<125)$

Solution

If x is a normally distributed random variable, with mean $= \mu$ and standard deviation $= \sigma$, then

$$P(x < x_{\max}) = \texttt{pnorm(x}_{\texttt{max}}\texttt{, mean = } \mu\texttt{, sd = } \sigma\texttt{, lower.tail=TRUE)}$$

$$P(x > x_{\min}) = \texttt{pnorm(x}_{\texttt{min}}\texttt{, mean = } \mu\texttt{, sd = } \sigma\texttt{, lower.tail=FALSE)}$$

$$P(x_{\min} < x < x_{\max}) = \texttt{pnorm(x}_{\texttt{max}}\texttt{, mean = } \mu\texttt{, sd = } \sigma\texttt{, lower.tail=TRUE)} \\ \texttt{- pnorm(x}_{\texttt{min}}\texttt{, mean = } \mu\texttt{, sd = } \sigma\texttt{, lower.tail=TRUE)}$$

```
pnorm(125, mean = 100, sd = 15, lower.tail=TRUE)
- pnorm(110, mean = 100, sd = 15, lower.tail=TRUE)
```

$= \mathbf{0.2047}$ or about 20%

Q6

- A sample of 10 scores are selected from a normally distributed population with mean 100 and standard deviation of 5.
- What is the probability that the sample mean is between 99 and 101?

Solution

- pnorm(101, mean = 100,sd = (5/sqrt(10))) - pnorm(99, mean = 100,sd = (5/sqrt(10)))

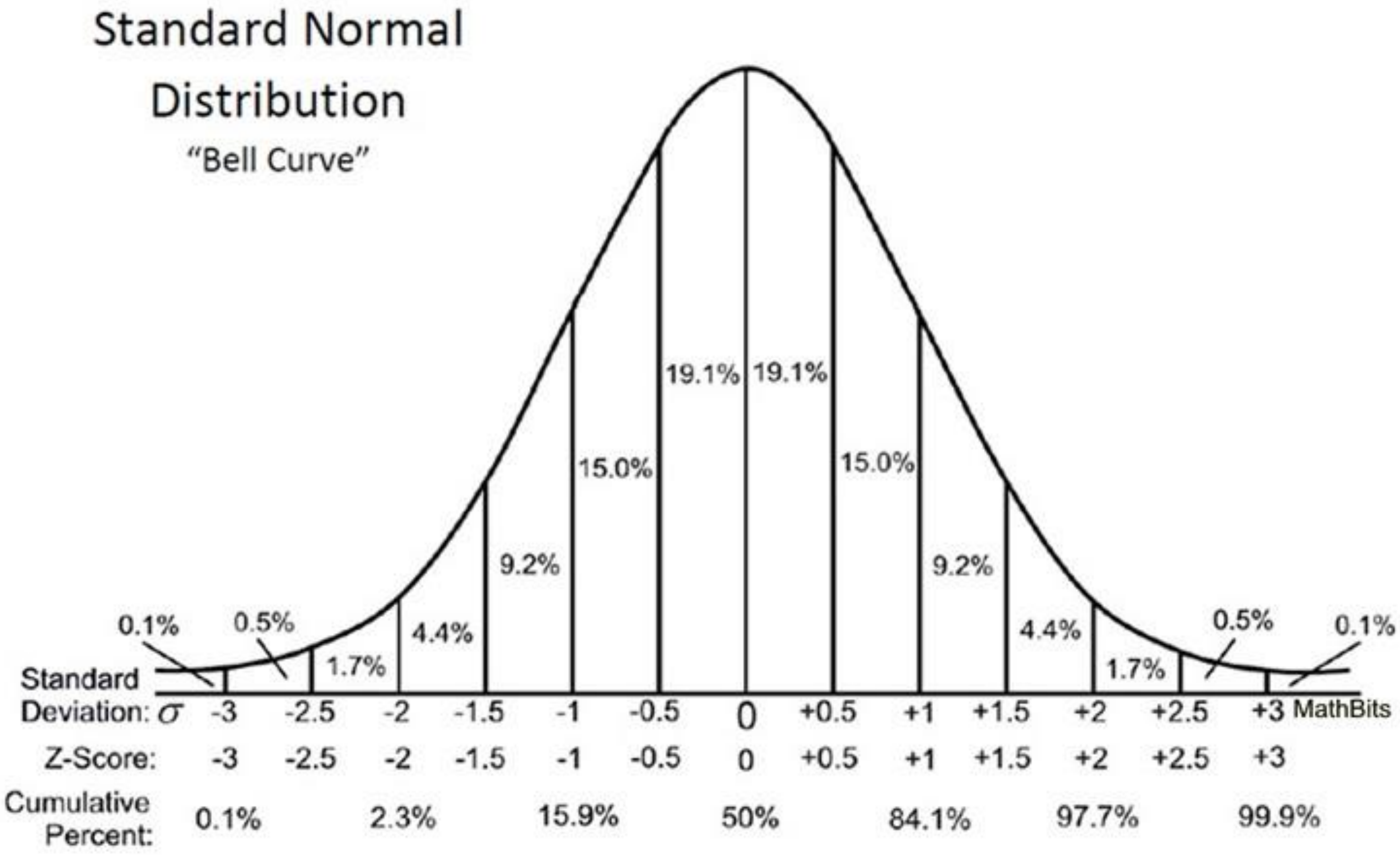
Q7

- How do you distinguish binomial and poisson distribution?

Answer

- The binomial distribution counts discrete occurrences among discrete trials.
- The poisson distribution counts discrete occurrences among a continuous domain.
- Ideally speaking, the poisson should only be used when success could occur at any point in a domain. Such as, for example, cars on a road over a period of time, or random knots in a string over a length, etc. We are talking about infinitely many infinitesimally small trials, each having at most one success.

Q8



$P(0<X<1.5)=?$

What is the z score of 85%?

What are the z scores of the area that is covered 38%?

Answer

- $P(0<X<1.5)$= 19.1 + 15+ 9.2
- Or
- pnorm(1.5) – pnorm(0)
- qnorm(0.85)
- 38%/2=19% --- 50%-19%=31% and 50%+19%=69%
- Lower z value = qnorm(0.31) = - 0.4958503
- Upper z value = qnorm(0.69) = 0.4958503

Q9

- The regression line between the math test (x <- independent variable) and calculus grade (y <- dependent) of the randomly selected students are as follows;

$$\hat{y}=40.78 + 0.76x$$

- What is residual value of a student who’s math test score is 52 and calculus grade is 75?

Answer

- y= 40.78 + 0.76 * (52) = 80.3 → predicted value
- Residual = Actual – predicted
- Residual = 75-80.3= -5.3

Q10

- What are the corresponding relationship for following correlation values?

- 1. r= 0.82
- 2. r= 0.1
- 3. r= -0.96
- 4. r= - 0.22

Answer

- 1. $r = 0.82$ → Strong positive
- 2. $r = 0.1$ → weak positive
- 3. $r = -0.96$ → Strong negative
- 4. $r = -\ 0.22$ → weak negative

Example

Simple Output

```
Call:
lm(formula = Sales ~ Spend, data = dataset)

Residuals:
   Min     1Q Median     3Q    Max 
 -3385  -2097    258   1726   3034 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1383.4714  1255.2404   1.102    0.296    
Spend         10.6222     0.1625  65.378 1.71e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2313 on 10 degrees of freedom
Multiple R-squared:  0.9977,	Adjusted R-squared:  0.9974 
F-statistic:  4274 on 1 and 10 DF,  p-value: 1.707e-14
```

Multiple Regression Output

```
Call:
lm(formula = Sales ~ Spend + Month, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max 
-1793.73 -1558.33    -1.73  1374.19  1911.58 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -567.6098  1041.8836  -0.545  0.59913    
Spend         10.3825     0.1328  78.159 4.65e-14 ***
Month        541.3736   158.1660   3.423  0.00759 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1607 on 9 degrees of freedom
Multiple R-squared:  0.999,	Adjusted R-squared:  0.9988 
F-statistic:  4433 on 2 and 9 DF,  p-value: 3.368e-14
```