Normal distribution (W5.1)

Notebook:	INIAD Statistics		
Created:	10/25/2018 7:52 PM	**Updated:**	11/13/2018 8:06 PM
Author:	danganhvu1998		

Normal distribution

Normal distribution

- Applied to very various kinds of phenomena in natural and social physics.
- The most important distribution in statistics

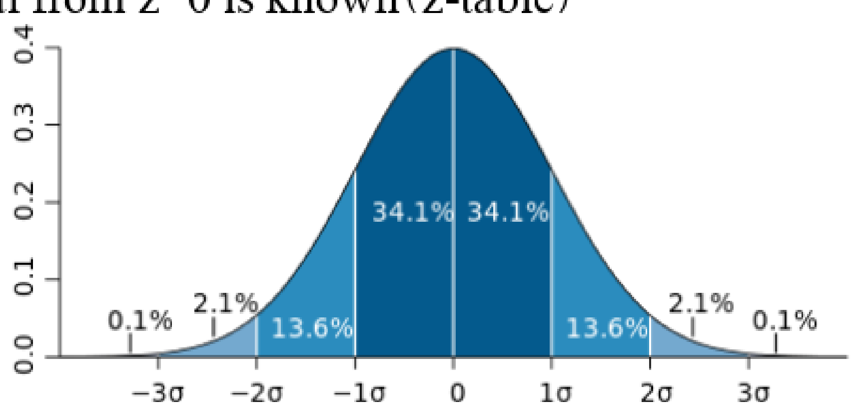
It's density is $$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

  - $E(X) = \mu$
  - $V(X) = \sigma^2$

- Denoted as $N(\mu, \sigma^2)$

Z-distribution

- Symmetric with expected value of z=0.
- You can discuss the peculiarity of data by using the distance from z=0.
- The integral from z=0 is known (z-table)



- Calculate Z

- **Z table**

```python
from scipy.stats import norm

#norm.cdf(A, loc=B, scale=C) -> Ratio that value from -oo -> A if Mean(Expected Value) = B, Standard Deviation = C
print(norm.cdf(25, loc=10, scale=5)) #0.9986501019683699
#Value from -oo->25 is about 99.86% in total of a set with Mean(Expected Value) = 10, Standard Deviation = 5

#norm.ppf(A)*B+C -> Value D that from -oo -> D with Mean(Expected Value) = C, Standard Deviation = C has Ratio = A
print(norm.ppf(0.1)*15+60) #40.77672651683099
#Value from -oo->40.78 is about 10% in total of a set with Mean(Expected Value) = 60, Standard Deviation = 15
```

Z-table

Stands for the integral from z=0. Of course, the integral over the whole real umber should be 1.

For instance, for z=1.00,the corresponding value is 0.3413, the hatched area occupies 34.13%.

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

- **Python**

statistic.py
11/10/2018 11:17 AM, 583 B


```python
from scipy.stats import norm

#norm.cdf(A, loc=B, scale=C) -> Ratio that value from -oo -> A if Mean(Expected Value) = B, Standard Deviation = C
print(norm.cdf(25, loc=10, scale=5)) #0.9986501019683699
#Value from -oo->25 is about 99.86% in total of a set with Mean(Expected Value) = 10, Standard Deviation = 5

#norm.ppf(A)*B+C -> Value D that from -oo -> D with Mean(Expected Value) = C, Standard Deviation = C has Ratio = A
print(norm.ppf(0.1)*15+60) #40.77672651683099
#Value from -oo->40.78 is about 10% in total of a set with Mean(Expected Value) = 60, Standard Deviation = 15
```


Example

- Mean = 10, Standard Deviation = 5

【Exercise1 : Normal distribution】【Answer】

- Let x is subject to N $(10, 5^2)$. Then, find $P(5<X<25)$.

$(5-10)/5=-1$, $(25-10)/5=3$.

Thus,

$P(5<X<25)=P(-1<z<3)$

$=0.3413+0.4987=0.84$

- Mean = 60, Standard Deviation = 10

【Exercise 2: Normal distribution】【Answer】

- Let a r.v. X is subject to N $(60, 10^2)$. Then, find $P(45<X<70)$.

$(45-60)/10=-1.5$, $(70-60)/10=1$.

Thus,

$P(45<X<70)=P(-1.5<z<1)$

$=0.4332+0.3413=0.7745$

- Top 10% so need to find A where 90% value from -oo -> A

【Exercise 3: Normal distribution】【Answer】

- The commute of workers in a certain cooperation is subject to the normal distribution, whose expected value and standard deviation are 60 [minutes] and 15[minutes], respectively.
- Then, find range of the commute of top 10% workers.

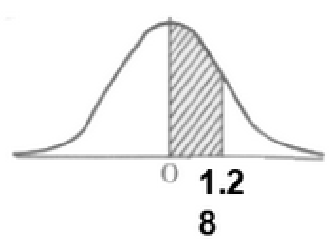
Since $P(0<z<1.28) = 0.3997$,

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

$P(-\infty<z<1.28) \fallingdotseq 0.8997$

Thus, it is at least

$60+15\times1.28=79.2$[minutes]

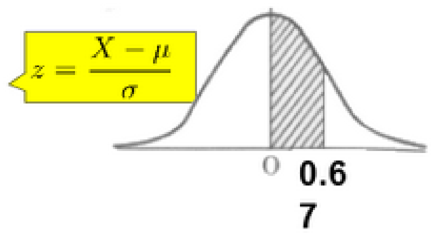


- 50^{th} mean 75%

【Exercise 4: Normal distribution】【Answer】

- The weight of a canned juice is subject to the normal distribution, whose expected value and standard deviation are 150.3[g] and 5[g], respectively. Assume you have 200 canned juice.
- 1) Find the weight of the canned juice of the 50^{th} place.

Since $P(0<z<0.67) = 0.2468$,
$P(-\infty < z < 0.67) \fallingdotseq 0.7468$.
Thus,, $150.3 + 5 \times 0.67 = 153.65$ [g].



-