Lecture 15

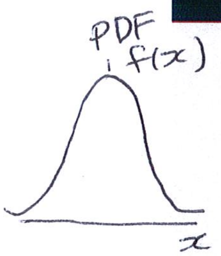
Normal Distribution

Manju M. Johny

STAT 330 - Iowa State University

Normal Distribution

Normal Distribution



Setup: The normal distribution is commonly used to model a wide variety of variables (weight, height, temperature. voltage, etc) due to its *"bell-shaped"* and *symmetric* shape.

If a random variable X follows a *normal distribution*,

$$X \sim N(\mu, \sigma^2)$$

where μ is the mean, and $\sigma^2 > 0$ is the variance

μ "mu"

σ^2 "sigma squared"

- Probability Density Function (pdf)

won't use this directly

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \text{for } -\infty < x < \infty$$

- Expected Value: $E(X) = \mu$
- Variance: $Var(X) = \sigma^2$

Normal PDF

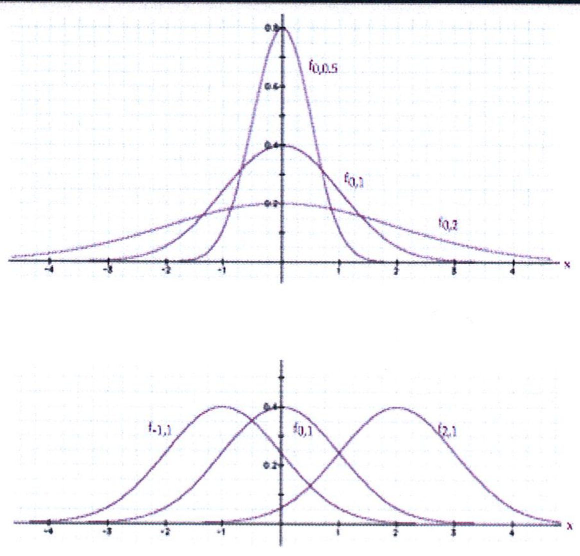


Figure 1: PDFs for normal distribution with various μ and σ^2

μ determines the location of the peak in the $x-$axis,

σ^2 determines the "width" of the bell shape.

Normal CDF

- Cumulative Distribution Function (cdf)

$$F_X(t) = \int_{-\infty}^{t} f(x)dx \qquad \text{(no closed form)}$$

→ The normal cdf does not have a closed form expression.

→ Use cdf table ($z-$table) of *standard normal distribution* (Z-table) $N(\mu = 0, \sigma^2 = 1)$ to obtain probabilities.

→ We need to *standardize* any normal random variable, X, into standard normal random variable, Z.

Standardization of Normal Distribution

Let $X \sim N(\mu, \sigma^2)$. Then,

standard deviation $\sigma = \sqrt{\sigma^2}$

1. $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable
2. $Z \sim N(0, 1)$ (normal distribution with $\mu = 0$, $\sigma^2 = 1$)

Standardization

Example 1: Suppose $X \sim N(20, 100)$. What is the probability that X is less than 23.5?

$\mu = 20$ $\quad \sigma^2 = 100$

To find this probability, we usually ...

- Integrate the PDF (too difficult)
- Plug into CDF (impossible - no closed form for CDF of X)

Instead we standardize X, and obtain probabilities using the standard normal cdf table ($z-$table)

- The standardized R.V is $Z = \frac{X-\mu}{\sigma} = \frac{X-20}{\sqrt{100}} \sim N(0, 1)$
- The standardized observation is $z = \frac{x-\mu}{\sigma} = \frac{23.5-20}{\sqrt{100}} = \cancel{0.23}$ 0.35
- $P(X < 10) = P(Z < 0.35)$ (obtain this from $z-$table)

Standard Normal Distribution

Standard Normal Distribution

Suppose a random variable, X, follows a $N(\mu, \sigma^2)$ distribution.
Then, $Z = \frac{X-\mu}{\sigma}$ follows a *standard normal distribution*

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

$$Z \sim N(0,1)$$

- Probability Density Function (pdf)

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad \text{for } -\infty < z < \infty$$

- Expected Value:

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{E(X)-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0$$

- Variance:

$$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = Var\left(\frac{X}{\sigma}\right) = \frac{1}{\sigma^2} Var(X) = \frac{\sigma^2}{\sigma^2} = 1$$

Standard Normal CDF

- Cumulative Distribution Function (cdf)

$$\Phi(z) = \quad F_Z(t) = \int_{-\infty}^{t} f(z)dz = \Phi(t) \qquad \text{(no closed form)}$$

Since the std. normal CDF is so common, it has its own symbol $\Phi(z)$ instead of $F_Z(t)$

→ Just like the normal cdf, the standard normal cdf does not have a closed form expression.

→ The cdf of N(0,1) random variable is denoted by $\Phi(t)$ (or more commonly $\Phi(z)$)

$P(Z \leq t) = F_Z(t) = \Phi(t)$

→ The values of the cdf, $\Phi(z)$, are found in the standard normal table (*z-table*)

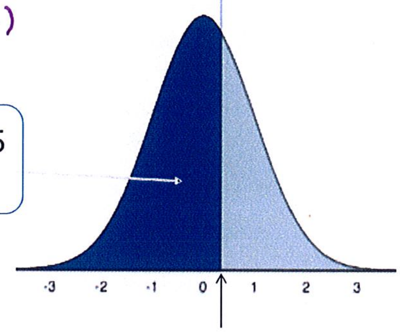
Z-table (Standard Normal Table)

Z-table

standardized observation

- Z-Table gives proportion of normal curve less than a particular z score
  - Gives left-hand area (dark blue shaded region)
  - This is same as the *percentile* value for z
  - Can be referred to as areas, proportions, or percentiles.
  - Denoted P(Z< z) or $P(Z \leq z)$ or $\Phi(z)$

Proportion of area less than z=0.35
Denoted as "P(Z<0.35)"



How to read the Z-table

- z values range from –3.99 to 3.99 on the z-table
- Row – ones and tenths place for z
- Column – hundredths place for z
- P(Z < z) found *inside* z-table

	Second decimal place in z						
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	[illegible]	[illegible]	0.5239
0.1	0.5398	0.5438	0.5478	[illegible]	[illegible]	[illegible]	0.5636
0.2	0.5793	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.6026
0.3	0.6179	[illegible]	[illegible]	[illegible]	0.6331	0.6368	0.6406
0.4	0.6554	[illegible]	[illegible]	0.6664	0.6700	0.6736	0.6772

Left-hand area/probability
P(Z < z)

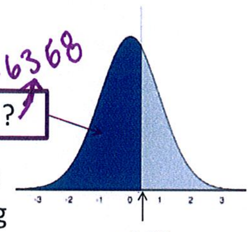
or $P(Z \leq z)$

How to read the Z-table

$P(Z < 0.35)$

0.6368

$\boxed{P(Z<0.35) = ?}$



z=0.35

- Look up z = 0.35 in the margins on z-table
- Percentile/left-hand area is corresponding value inside z-table

	Second decimal place in z						
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	$\boxed{0.6368}$	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772

$\boxed{P(Z < 0.35)}$

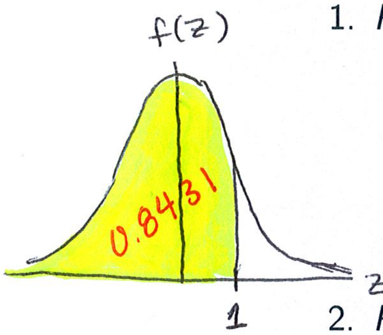
$= P(Z \leq 0.35)$

$= 0.6368$

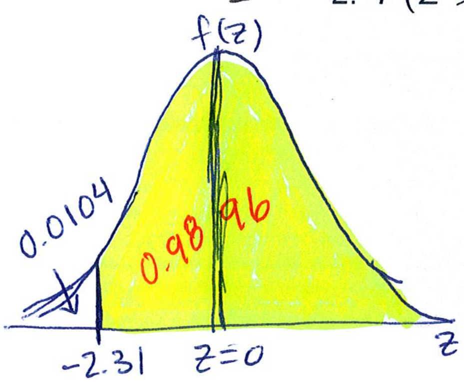
Examples

Z-table Practice

Suppose $Z \sim N(0,1)$



1. $P(Z < 1) = P(Z < 1.00) = \Phi(1.00) = 0.8431$

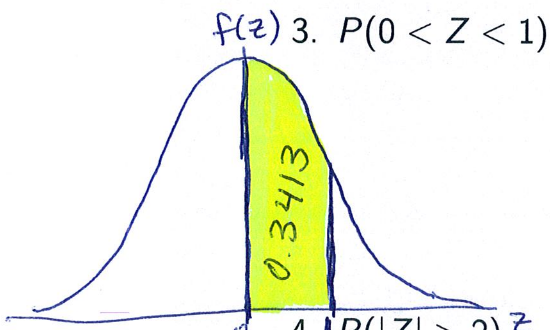


2. $P(Z > -2.31) = 1 - P(Z \le -2.31)$
$= 1 - \Phi(-2.31)$
$= 1 - 0.0104$
$= 0.9896$

Right-hand area = 1 − left-hand area

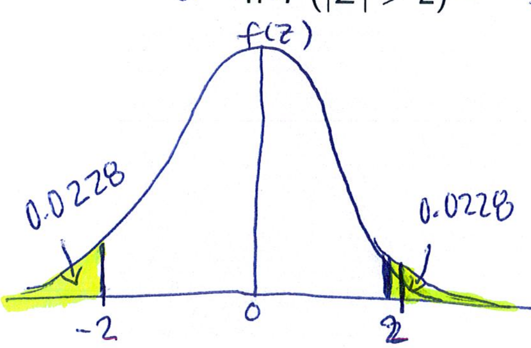
Z-table Practice

Suppose $Z \sim N(0,1)$



3. $P(0 < Z < 1) = P(Z < 1) - P(Z \le 0)$
$= \Phi(1) - \Phi(0)$
$= 0.8413 - 0.5$
$= 0.3413$

[Bigger left-hand area − smaller area]



4. $P(|Z| > 2) = P(Z < -2) + P(Z > 2)$
$= 2P(Z < -2)$
$= 2\Phi(-2)$
$= 2(0.0228)$
$= 0.0456$

[Find the left-hand area and multiply it by 2.]

Normal Distribution Example

$\mu = 1 \quad \sigma^2 = 2$

Suppose $X \sim N(1, 2)$, and we want to find $P(1 < X < 2)$.

- First, standardize X into std. normal R.V Z

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 1}{\sqrt{2}} \sim N(0, 1)$$

- Standardize our observations

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{1 - 1}{\sqrt{2}} = 0$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{2 - 1}{\sqrt{2}} = 0.71$$

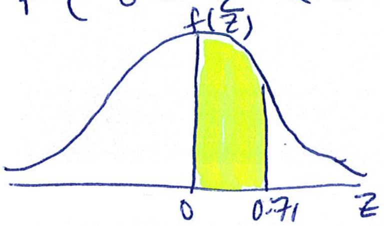
round to 2 decimal places b/c that's as far as my z-table z's go

$$P(0 < Z < 0.71) = P(Z < 0.71) - P(Z \leq 0)$$

$$= \Phi(0.71) - \Phi(0)$$

$$= 0.7611 - 0.5$$

$$= 0.2611$$



Normal Distribution Example

or do it all in one sweep

$$P(1 < X < 2)$$

$$= P\left(\frac{1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{2 - \mu}{\sigma}\right)$$

$$= P\left(\frac{1 - 1}{\sqrt{2}} < Z < \frac{2 - 1}{\sqrt{2}}\right)$$

$$= P(0 < Z < 0.71)$$

$$= P(Z < 0.71) - P(Z \leq 0)$$

$$= \Phi(0.71) - \Phi(0)$$

$$= 0.7611 - 0.5$$

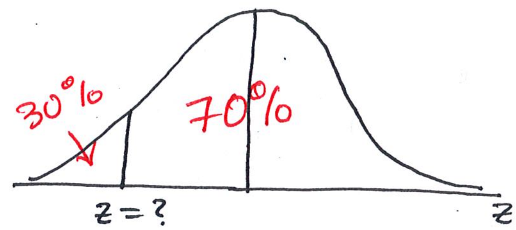
$$= 0.2611$$

<u>Example</u>: Start w/ probability and work backwards to get observation.

Suppose weight of apples follows a normal distribution w/ mean of 60g and variance of 25 g.
70% of apples weigh more than what weight?

$$X = \text{apple weight}$$
$$X \sim N(60, 25)$$

If 70% of apples weigh <u>more</u> than some weight, then 30% weigh <u>less</u> than this weight.



The z-table gives CDF values (left-hand areas). Look up 0.30 (or closest value) inside z-table, and obtain corresponding z-value in margins.

Closest area to 0.30 insize z-table is 0.3015, which has corresponding z-value of $z = -0.52$.

$$\Rightarrow P(Z \le \underbrace{-0.52}_{z}) = 0.3015$$

We need to "reverse standardize" to go from z back to X (apple weight).

$$z = \frac{x - \mu}{\sigma}$$ ← standardization formula

$$\Rightarrow x = \mu + z\sigma$$ ← "reverse" standardization formula.

$$\Rightarrow x = 60 + (-0.52)(5) = \boxed{57.4\text{g}}$$

⇒ 30% of apples weigh less than 57.4 g.
⇒ 70% of apples weigh <u>more</u> than 57.4g.