# Week 3

Tuesday, 3 August 2021 13:52

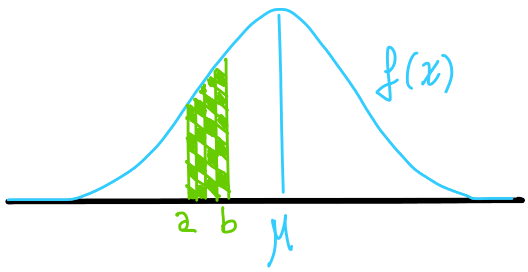
## THE NORMAL DISTRIBUTION (GAUSSIAN)

MEAN = MEDIAN = MODE
PERFECTLY BALANCE AROUND MEAN

THEORETICAL PROBABILITY DISTRIBUTIONS

INFINITE # OF NORMAL CURVES $\forall (\mu, \sigma)$

$$P(a < x < b) = \int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$




CURVE IDENTIFIED ONLY BY 2 PARAMETERS

SAME STRUCTURAL PROPERTIES
WE JUST HAVE TO KNOW $N(0,1)$ STD

ANY NORMAL DISTRIBUTION CAN BE
RESCALED TO THE STANDARD

JUST 1 REFERENCE TABLE

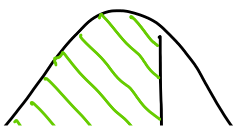
$$\boxed{\begin{aligned} P(|x| < \sigma + \mu) &= 68\% \\ P(|x| < 2\sigma + \mu) &= 95\% \\ P(|x| < 3\sigma + \mu) &= 99.7\% \end{aligned}}$$

PROPORTIONS = PROBABILITY

## STANDARD NORMAL TABLE

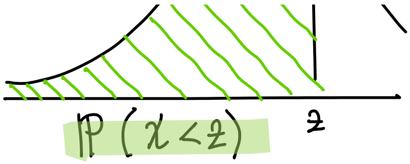
$\mu = 0, \ \sigma = 1$

$N(0, 1)$



CALCULATING USING R

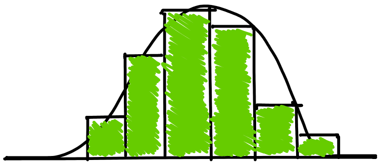
## PNORM (z)




## APPLICATIONS

$\bar{X} = 123.6$
$S = 12.9$
$\hat{m} = 123.0$

$\left.\begin{matrix} \bar{X} = 123.6 \\ S = 12.9 \\ \hat{m} = 123.0 \end{matrix}\right\}$ mmHg



PERCENTILES

$\left.\begin{matrix} q_{0.025} = \bar{X} - 2s = 97.8 \text{ mmHg} \\ q_{0.975} = \bar{X} + 2s = 149.4 \text{ mmHg} \end{matrix}\right\}$ ESTIMATING 95% RANGE

95% OF PEOPLE HAVE THEIR BLOOD PRESSURE BETWEEN:

$x \in [97.8, 149.4]$ mmHg

PREDICTIONS

WE GET $\mu, \delta$ FROM SAMPLE
USING THE RESULTS TO ANALYZE FUTURE SUBJECTS NOT INCLUDED IN FIRST SAMPLE

$\mathbb{P}(X > 130 \text{ mmHg})$ ?

$\mathbb{P}\left(\frac{X - \mathbb{E}(X)}{\sqrt{VAR(X)}} > \frac{X - \mu}{\sigma}\right)$ z-SCORE

$1 - \phi\left(\frac{x - \mu}{\sigma}\right)$

MEASURE OF THE RELATIVE DISTANCE & DIRECTION OF A SINGLE OBSERVATION IN A DATA DISTRIBUTION RELATIVE TO THE MEA

YOU CAN CALCULATE A z-SCORE EVEN IF THE SAMPLE IS NOT NORMAL BUT WILL NOT CORRESPOND DIRECTLY TO PERCENTILES

INVALID & NONSENSICAL RESULTS
WE CANNOT ALWAYS USE THE NORMAL PROPERTIES

| PNORM | QNORM |
|---|---|
| $\phi(x)$ | $z_x$ |

$$\boxed{z_i = \frac{x_i - \mu}{\sigma}}$$

EXAMPLE BMI INDEX

$x = \text{BMI}_{18\text{ y.o.}} \approx N(21.9,\ 3.2^2)$

$x_1 = 26.7$
$x_2 = 23.5$

$z_1, z_2 = ?$

$z_1 = \frac{26.7 - 21.9}{3.2} = 1.5$

$z_2 = \frac{23.5 - 21.9}{3.2} = 0.5$

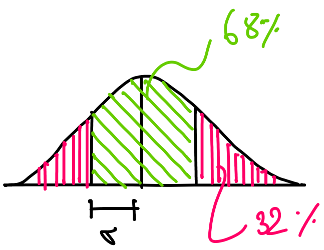
IF DISTRIBUTION SLIGHTLY SKEWED,
CAN WE ESTIMATE THE INTERVAL OF $z_2$ ?

YES $z_{0.025}$, $z_{0.975}$ (NORMAL RANGE)

EXAMPLE 2

$y = \text{BLOOD GLUCOSE LEVELS } \frac{mg}{dL} \sim N(90,\ 13^2)$

$\pm 1$ STD DEVIATION "ABNORMAL"




- NORMAL RANGE ?

MIN $= \mu - 1 \times \delta = 90 - 13 = 77$ mg/dL

MAX $= \mu + 1 \times \delta = 90 + 13 = 103$ mg/dL

NORMAL RANGE $x_{\text{HEALTHY}} \in [77, 103]$

IF NORMAL RANGE OUTSIDE $\pm 2\delta \rightarrow 5\%$ ABNORMAL

MIN $= \mu - 2\delta = 64$

MAX $= \mu + 2\delta = 116$

NORMAL RANGE $x_{\text{HEALTHY}} \in [64, 116]$

EXAMPLE 3

$x = \text{BMI}_{18\text{YO}} \approx N(21.9,\ 3.2^2)$ $\frac{kg}{m^3}$

3 IS $0.75\delta$ BELOW AVERAGE , $X_3 = ?$

$$z_3 = -0.75\delta$$

$$z_3 = \frac{X_3 - \mu}{\sigma} \qquad z_3\sigma = X_3 - \mu$$

$$\boxed{X_3 = z_3\delta + \mu} = 19.5$$

NORMAL RANGE (95%) ?

$$MIN = \mu - 2\delta = 15.5$$

$$MAX = \mu + 2\delta = 28.3$$

$$x_{HEALTHY} \in [15.5, 28.3]$$