

Inspire...Educate...Transform.

Foundations of Statistics and Probability for Data Science

Basic Statistical Concepts, Central Tendencies and Measures of Variability, Probability Basics

Dr. Sridhar Pappu

Executive VP – Academics, INSOFE

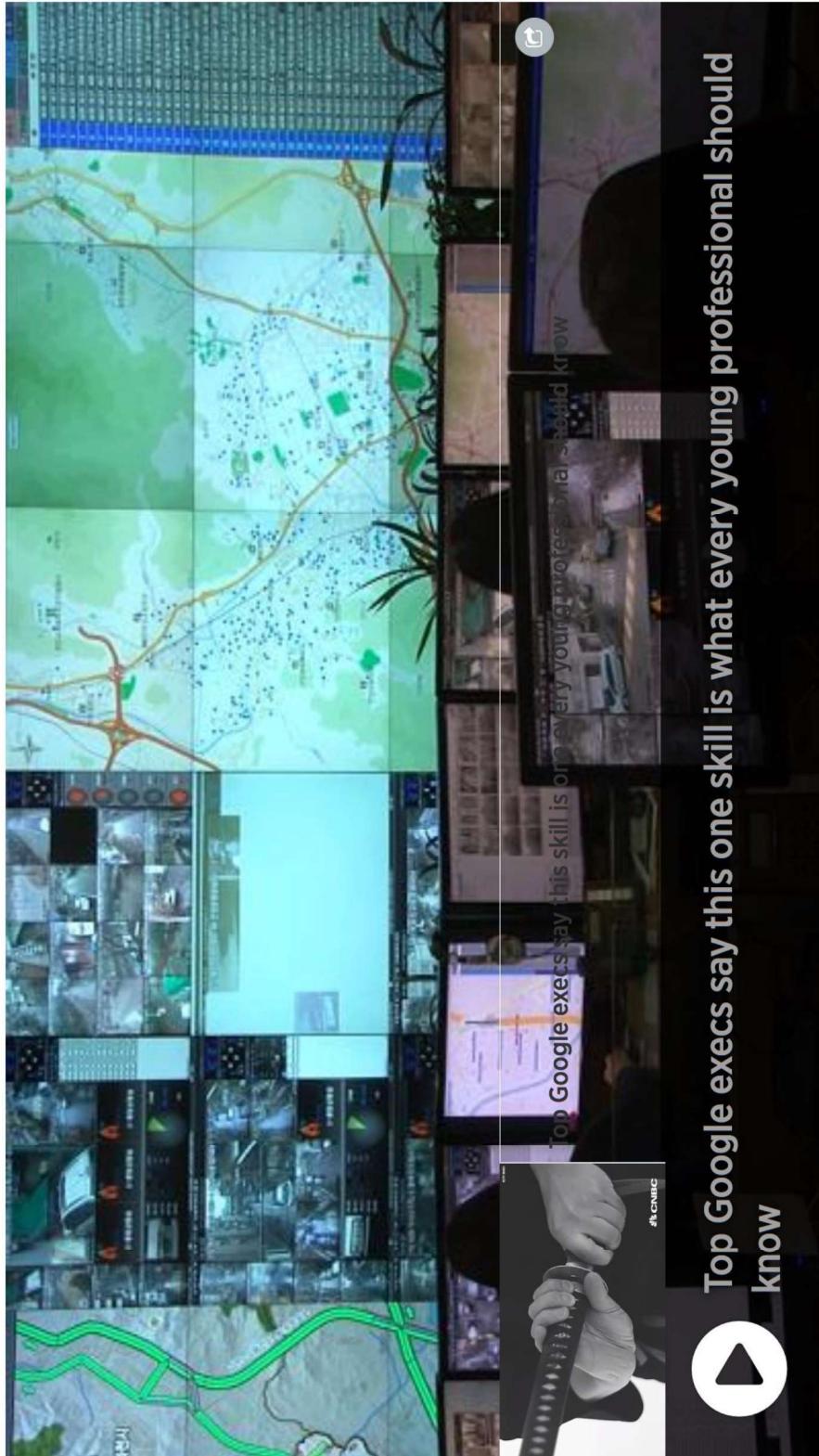
October 07, 2017



MOTIVATION

Google billionaire Eric Schmidt says this is the skill employers will look for in the future

Marguerite Ward | @forwardist | 1:16 PM ET Fri, 31 March 2017



Top Google execs say this one skill is what every young professional should know



Source: <http://www.cnbc.com/2017/03/31/google-execs-agree-on-the-skill-employers-will-look-for-in-the-future.html>

Last accessed: October 06, 2017

"I think a basic understanding of data analytics is incredibly important for this next generation of young people," Schmidt tells CNBC. "That's the world you're going into."

"By data analytics," the executive chairman says, "I mean a basic knowledge of how statistics works, a basic knowledge of how people make conclusions over big data."

Source: <http://www.cnbc.com/2017/03/31/google-execs-agree-on-the-skill-employers-will-look-for-in-the-future.html>

Last accessed: October 06, 2017

The best place for students to learn Applied Engineering



THE ELECTION
CENTRE
DIGITAL
MEDIA IN
MEDIA

**NEW AGE
ELECTIONS**





Kent
Sa
Shudh

CROSSWIND
14 KMPH

14 CMS



SC



*There are three kinds of lies:
lies, damned lies, and
statistics.*

- Mark Twain / Benjamin Disraeli

"DATA IS THE SWORD OF THE 21ST CENTURY, THOSE WHO WIELD IT WELL, THE SAMURAI."

-Jonathan Rosenberg, adviser to Larry Page and former SVP of products at Google

Rosenberg agrees.

"My favorite statement that echoes Eric's," he says, "is 'Data is the sword of the 21st century, those who wield it well, the samurai.'"

The quote comes from an **internal memo** Rosenberg sent to employees in 2009, following the inauguration of President Barack Obama.

"Everyone should be able to defend arguments with data," he writes in the memo. "Information transparency helps people [...] determine who is telling the truth."

GENDER RACE

The sex-ratio of electorates used to be tilted to the male voters, but the trend has started to change. Five of the 13 states along with the three Union Territories which went to polls in the first four phases of LS experienced female electorates outnumbering their male counterparts.

PUDUCHERRY

FEMALE	MALE
52%	48%

KERALA

FEMALE	MALE
51.9%	48.1%

MANIPUR

FEMALE	MALE
51%	49%

MIZORAM

FEMALE	MALE
50.9%	49.1%

DAMAN & DIU

FEMALE	MALE
50.5%	49.5%

MEGHALAYA

FEMALE	MALE
50.4%	49.6%

GOA

FEMALE	MALE
50.1%	49.9%

ARUNACHAL

FEMALE	MALE
50.1%	49.9 %

Problem #1: Data Gathering

LIAR LIAR
PANTS ON FIRE!



Schedule Reference	Parliamentary Constituency		
	Sl.	PC No.	PC Name
7 1 No of PCs going to poll Issue of Notification: Last Date for filing Nominations: Scrutiny of Nominations: Last date for withdrawal of Candidature: Date of Poll Counting of Votes: Date before which the election shall be completed	1	1	Daman & Diu GEN

Source: http://eci.nic.in/eci_main1/GEC2014/Schedule/DD.htm

Last accessed: October 24, 2014

By April 24, when Puducherry went to polls, 6 phases (not 4) were completed, and 19 States and 5 UTs had completed polling (not 13 and 3, respectively; Daman & Diu went to polls on April 30).

Source: <http://epaper.deccanchronicle.com/articledetailpage.aspx?id=474880>

Last accessed: April 27, 2014

The best place for students to learn Applied Engineering

9

<http://www.insofe.edu.in>



CSE 7315c

GENDER RACE

Problem #2: Data Understanding



The ratios reflect the ratios of registered voters.

MANIPUR
FEMALE MALE
51% **49%**

PUDUCHERRY
FEMALE MALE
52% **48%**

KERALA

FEMALE MALE
51.9% **48.1%**

MANIPUR

FEMALE MALE
51% **49%**

MIZORAM

FEMALE MALE
50.9% **49.1%**

DAMAN & DIU

FEMALE MALE
50.5% **49.5%**

MEGHALAYA

FEMALE MALE
50.4% **49.6%**

GOA

FEMALE MALE
50.1% **49.9%**

ARUNACHAL

FEMALE MALE
50.1% **49.9 %**

Voted in 2014 General Elections					
		Registered Voters			
State/UT		Male	Female	Male	Female
Puducherry		432048	469309	52.07	351360
Kerala		11734258	12592391	51.76	8678185
Manipur		871431	902894	50.89	685427
Mizoram		346219	355951	50.69	216167
Daman & Diu		57011	54816	49.02	42378
Meghalaya		777639	789602	50.38	524774
Goa		528308	532469	50.2	395766
Arunachal Pradesh		379627	379760	50.01	289291

Data from <http://pib.nic.in/newsite/PrintRelease.aspx?relid=105116> and <http://pib.nic.in/newsite/efeatures.aspx?relid=104195>.

Source: <http://epaper.deccanchronicle.com/article/detailpage.aspx?id=474880>; Last accessed: April 27, 2014

The best place for students to learn Applied Engineering

10

<http://www.insofe.edu.in>



Problem #2: Data Analysis/Interpretation

The sex-ratio of electorates used to be tilted to the male voters, but the trend has started to change.

GENDER RACE

The sex-ratio of electorates used to be tilted to the male voters, but the trend has started to change. Five of the 13 states along with the three Union Territories which went to polls in the first four phases of LS experienced female electorates outnumbering their male counterparts.

PUDUCHERRY

FEMALE	MALE
48%	

KERALA

FEMALE	MALE
51.9%	

MANIPUR

FEMALE	MALE
51%	

MIZORAM

FEMALE	MALE
50.9%	

DAMAN & DIU

FEMALE	MALE
49.5%	

MEGHALAYA

FEMALE	MALE
50.4%	

GOA

FEMALE	MALE
50.1%	

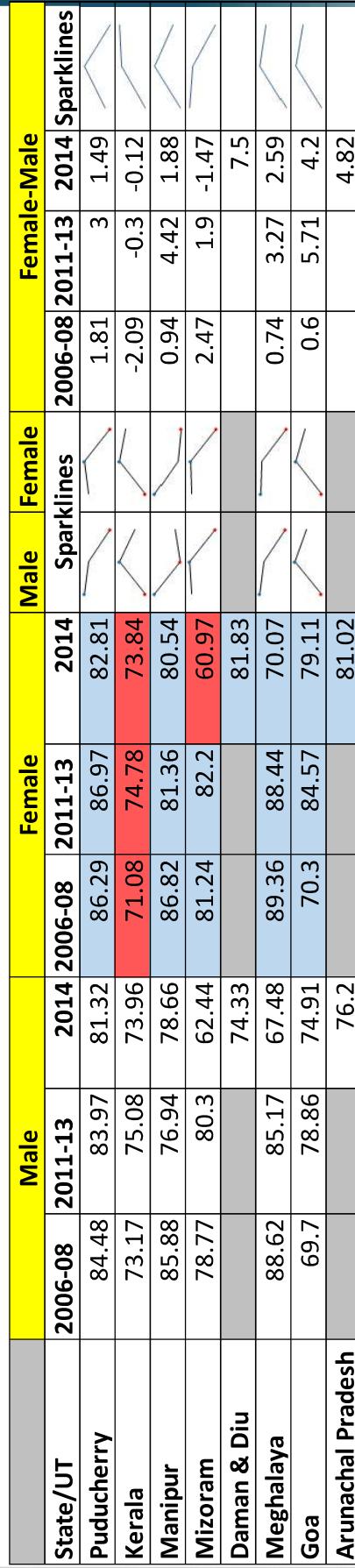
ARUNACHAL

FEMALE	MALE
50.1%	

Source: <http://epaper.deccanchronicle.com/articledetailpage.aspx?id=474880;>

Last accessed: April 27, 2014

<http://www.insofe.edu.in>

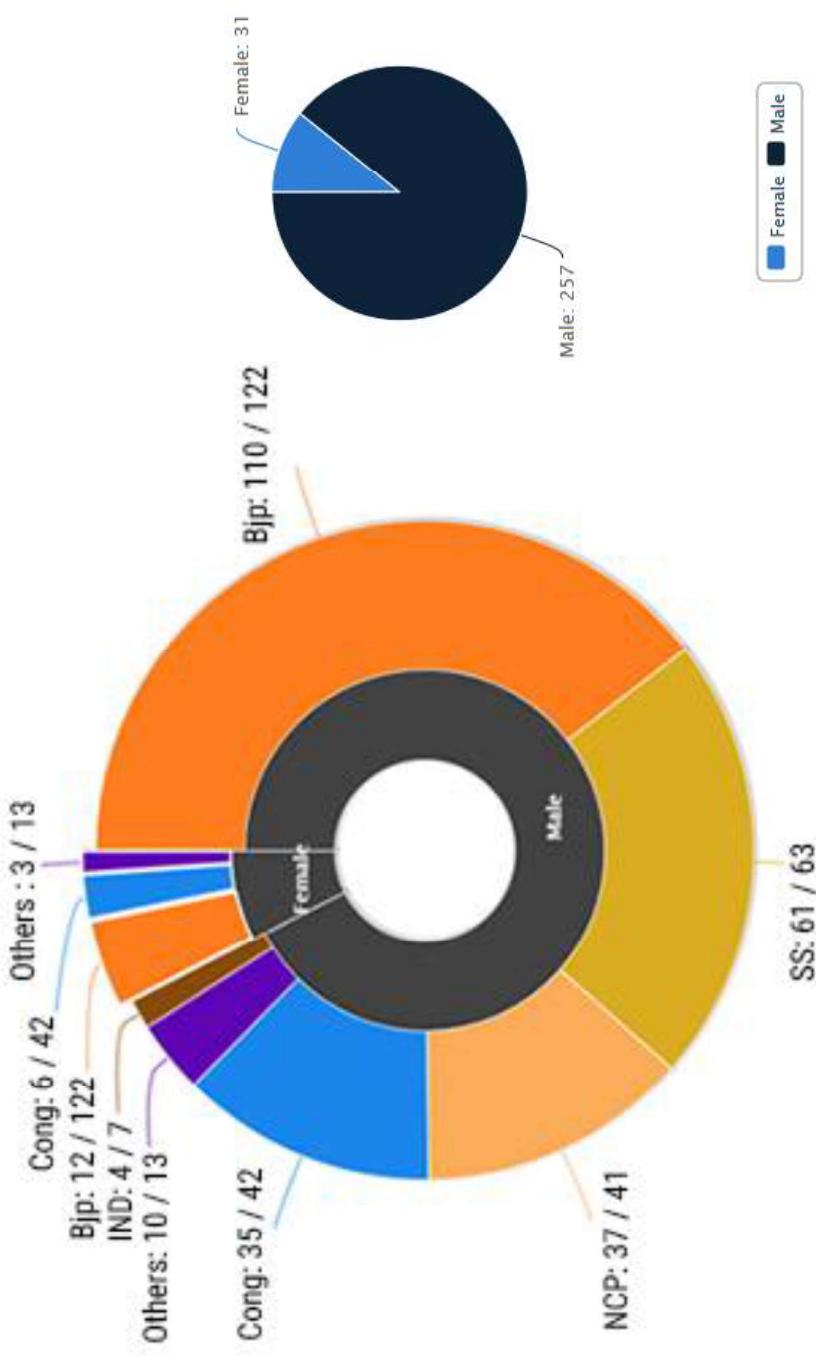


Data from <http://pib.nic.in/newsite/PrintRelease.aspx?relid=105116> and <http://pib.nic.in/newsite/efeatures.aspx?relid=104195>.

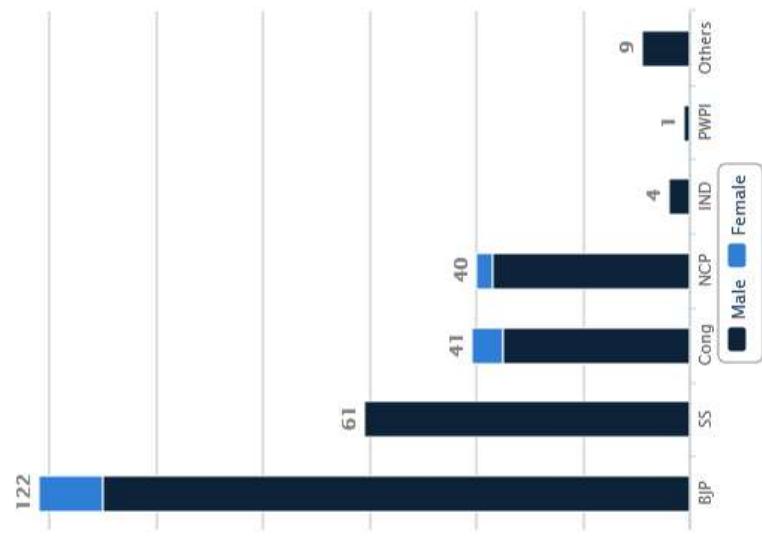
The best place for students to learn Applied Engineering

Problem #4: Data Presentation

Maharashtra: Gender Break-up
Total MLAs: 288*



Maharashtra: Gender Break-up
Total MLAs: 288*



Source: <http://www.ndtv.com/elections/assembly-cabinet/maharashtra>
Last accessed: October 24, 2014

Problem #4: Data Presentation

Cops shift focus, 50% dip in cases

■ Traffic cops chase helmet violations but see cut in cases filed against other culprits

DC CORRESPONDENT
HYDERABAD, DEC. 8

The number of cases booked for triple riding, cell phone driving and signal jumping have come down by over 50 per cent this year compared to previous year.

The drop, however, is neither due to enhanced enforcement by the traffic cops nor because of improved compliance to rules by motorists.

The likely cause of the dip is, among others, is the traffic cops' focus on drives against violation of specific rules like helmet rule violation. Consequently, the number of cases booked for helmet violation jumped from 1.34 lakh in 2015 to over 17 lakh in 2016. During August and September this year, the traffic police was going slow on enforcement due to heavy rains. That was followed in November by demonitisation.

Indian Road Safety Federation Mr Vinod Functionary TS

CASES BOOKED IN 2015 AND IN 2016 UP TO DECEMBER 5

	2015	2016
Triple riding:	73,549	31,704
Cell Phone Driving:	27,342	10,015
Signal Jumping:	51,725	16,105
Not wearing helmet:	1,34,092	17,20,169
Without number plate:	4,934	4351

Kumar Kanumala, said the drop cannot be taken as a benchmark to declare that violation of traffic rules has come down or compliance to rules has increased.

"A scientific study has to be done to check if violations are repeated by motorists, how many are first time violators and if they are given counselling. How the counselling helps needs to be seen," he said.

Social worker TS Gupta said the number of helmet violation cases would go up by 100 per cent if traffic cops were to intensify the drive with many bikers driving without helmets.

"Even the triple riding cases would be in large numbers if traffic cops enforce the rule strictly," he said.

Deputy Commissioner

of Police (Traffic) Mr A V Ranganath admitted

that a fall in number of

cases booked this year

compared to previous

year cannot be attrib-

uted to only enforce-

ment or improvement in

rule compliance.

Source: <http://epaper.deccanchronicle.com/article/detailpage.aspx?id=6941170>

Last accessed: December 09, 2016

The best place for students to learn Applied Engineering

13

<http://www.insofe.edu.in>

9% growth to cut poverty by 36%

DC CORRESPONDENT
 NEW DELHI, APRIL 20

A high growth will be imperative for the new government to pull more people out of poverty, create more jobs and increase the income of households hit by high price rise.

If India is able to see a fast growth of nine per cent then the number of people below poverty line will be reduced to 17.7 crores at the end of fiscal 2019 against 26.9 crores at the end of 2012, according to rating agency Crisil.

However, if India grew at 6.5 per cent then the number of poor during this period will be reduced to 22.6 crores. But in case growth



remained at 5 per cent in the coming years then the number of people below poverty line will be 25.6 crores.

India's working-age population would have swelled by over 85 million to fiscal 2019. Of these, 5.1 crores would be seeking employment, said Crisil.

With 6.5 per cent average GDP growth, non-farm employment over

wage, more-productive industry and service sectors," said Crisil.

Such a dismal situation would not arise if India were to grow at 9 per cent over the next five years as it will result in 5.2 crores non-farm jobs. "Enough non-farm employment would be created to absorb the entire incremental labour force within the industry and service sectors. Indeed, at 9 per cent growth, it would even be possible to pull additional people out of agriculture," said the report. On the other hand, getting stuck in the 5 per cent growth will aggravate India's employment situation as non-farm jobs will then increase by only 2.6 crores.

this period will at best grow by 3.7 crores. This means an additional 14 million will be forced to either depend on low-productivity agriculture or remain unemployed.

"However, much of the increase in farm jobs will be disguised unemployment. That's because, given insufficient job opportunities, labour force will not be able to migrate to the higher-

Source: <http://epaper.deccanchronicle.com/articledetailpage.aspx?id=474956>

Last accessed: April 27, 2014

INDIA BEATS CHINA, GROWS 7.5% IN Q4

Economists doubtful; India Inc sees downside risks

**PAWAN BALI / DC
NEW DELHI, MAY 29**

India's GDP accelerated in January to March 2015 period (fourth quarter of 2014-15) to 7.5 per cent, overtaking China as the fastest growing major economy in the world.

This took the overall GDP growth for FY 2014-15 to 7.3 per cent against 6.9 per cent in the 2013-14.

Finance minister Arun Jaitley said that the economy is clearly on "a recovery path". He said an economy growing at fastest pace in the world cannot be 'fragile' as alleged by the former Prime Minister Manmohan Singh.

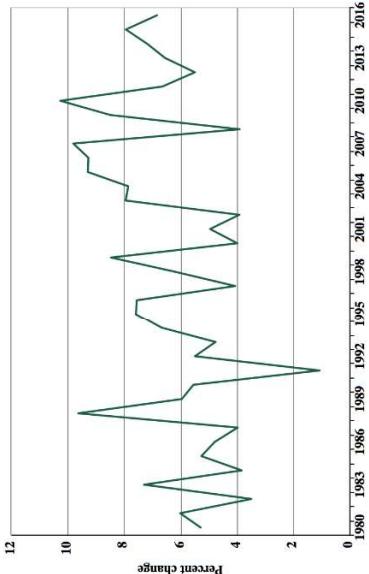
Finance ministry said that those sectors with in control of policy – manufacturing and services – improved substantially while those dependent on factors beyond the policy control such as agriculture (dependent on weather) and exports (on foreign demand), "did less well." However, the high

GDP numbers have come on the back of new methodology which the Central Statistical Organisation (CSO) adopted earlier this year to calculate the GDP. In January, CSO revised India's growth to 6.9 per cent in the 2013-14 from the earlier estimate of 4.7 per cent.

Before this data revision, it was a dominant view that India had seen worst growth in 2013-14.

Some economists said that the latest GDP numbers released on Friday does not reflect the ground reality and people will now have to adjust with under the new methodology is not the same what it was based on older formula.

the same what it was based on older formula.



HIGHLIGHTS

We are not cheerleader,

Mumbai, June 2: Doing some plain-speaking, governor Raghuram Rajan on Tuesday said RBI is not a 'cheerleader'

and he chose to "err a bit" in lowering the rates to push investments as growth on ground may not be as high as what is suggested by the headline GDP numbers.

Wondering why an economy needed rate cut when it was growing at 7.5 per cent, Dr Rajan said there is a "contra-

diction" between the high GDP numbers and the poor corporate earnings along with lack of any visible pick-up in demand.

"In some sense it is a Goldilocks policy, just right given the current situation," he said, as he defended his third 0.25 per cent rate cut this year despite lingering concerns over the below-normal monsoons as well as steadily firming oil prices and their impact on inflation.



ROOM FOR UP IF OIL

Mumbai, J Room for more cuts in India open up if rains are better expected or if government takes to prevent food from rising global oil remain low, says governor Ra

Cholesterol not a threat

■ US to remove high-cholesterol food from 'naughty' list

Washington, May 26: United States officials have finally given the green light for a U-turn on previous warnings on cholesterol, which has been on the "naughty" list of nutrients for nearly 40 years. Health officials have been warning people to stay away from high-cholesterol foods since the 1970s to avoid heart disease and clogged arteries.

However, after a study, eggs, butter, full-fat dairy products, nuts, coconut oil and meat have now been classified as "safe" and have been officially removed from the "nutrients of concern" list.

FOODIES **DELIGHT**

Butter, full-fat dairy products, nuts, coconut oil and meat have now been classified as "safe" and have been officially removed from the "nutrients of concern" list.

longer warn people against eating high-cholesterol foods and will instead focus on sugar as the main substance of dietary concern.

The 70s, 80s and 90s were the 'non fat' years, with the US government warning people to limit the amount of high-cholesterol foods in their diets to avoid heart disease and strokes.

But nutritionists and scientists have long been campaigning for the U-turn, which started with introducing "good cholesterol's" back into the 'safe zone'.

US cardiologist Dr Steven Nissen said: "It's the right decision. We got the dietary guidelines wrong. They've been wrong for decades." Dr Chris Masterjohn added: "When we eat more foods rich in this compound, our bodies make less. If we deprive ourselves of foods high in cholesterol — such as eggs, and butter — our body revs up its cholesterol synthesis." — Agencies



The 70s, 80s and 90s

were the 'non fat' years, with the US government warning people to limit the amount of high-cholesterol foods in their diets.



The US Department of Agriculture, which is responsible for updating the guidelines every five years, stated in its findings for 2015: "Previously, the Dietary Guidelines for Americans recommended that cholesterol intake be limited to no more than 300 mg/day. The 2015 DGAC will not bring forward this recommendation because available evidence shows no appreciable relationship between consumption of dietary cholesterol and serum (blood) cholesterol levels." — Agencies

AHA/ACC (American Heart Association / American College of Cardiology) report. Cholesterol is not a nutrient of concern for overconsumption." The Dietary Guidelines Advisory Committee will, in response, no

TABLE 2: Vitamin D Levels^a

<10 NG/ML	severe deficiency^b
10-24 NG/ML	mild to moderate deficiency^c
25 TO 80 NG/ML	optimal levels
>80 NG/ML	toxicity possible

^a Mayo Medical Laboratories, Mayo Clinic.

^b Could be associated with osteomalacia[os-tee-oh-ma-LASS-ee-ah] (adults) or rickets (children).

^c May be associated with bone loss or osteoporosis.

My level is between 20-30 ng/ml

According to the Vitamin D Council, you're deficient in vitamin D. According to the Endocrine Society, you're insufficient. By the Institute of Medicine's standards, you're getting enough vitamin D. If you're Caucasian American, you're likely to have a vitamin D level in this range.

With a vitamin D level in this range you're less likely to have health problems than when vitamin D is at lower levels. However, your body may still be producing too much parathyroid hormone and not fully absorbing calcium, which can affect your bones. Research shows that levels above 30 ng/ml may be more likely to prevent bone problems such as fractures.

Sources: <http://www.empoweryourhealth.org/issue-1/Low-Vitamin-D-LEVELS-IN-ADULTS>, <https://www.vitamindcouncil.org/further-topics/just-tested-my-vitamin-d-level-what-do-my-results-mean/#>

Last accessed: March 13, 2015

The best place for students to learn Applied Engineering

Although there are few major studies carried out in India to determine the optimum (sufficient) levels of serum vitamin D 25(OH) D to be maintained to prevent chronic ailments, globally there is a consensus that vitamin D deficiency is defined as serum 25(OH) D levels less than 20 ng/ml and insufficiency as serum 25(OH) D less than 30 ng/ml. Whereas, serum 25(OH) D levels above 30 ng/ml is found to be sufficient.

TOI, May 5, 2013

"If your vitamin D level is below 50 or over 100 nanomol per litre, there is a greater connection to deaths," said Peter Schwarz, professor at the University of Copenhagen in Denmark.

The researchers studied the level of vitamin D in 247,574 people and analysed their mortality rate over a seven-year period. In that time 16,645 patients had died.

"We have looked at what caused the death of patients, and when numbers are above 100, it appears that there is an increased risk of dying from a stroke or a coronary," Schwarz added.

"In other words, levels of vitamin D should not be too low, but neither should they be too high. Levels should be somewhere in between 50 and 100 nanomol per litre, and our study indicates that 70 is the most preferable level," Schwartz explained.

That having too much vitamin D in our blood can be bad for our health has never been proven before, and it may have great influence on our future intake of nutritional supplements.

The study appeared in the *Journal of Endocrinology and Metabolism*.

TOI, March 11, 2015

Sources: <http://timesofindia.indiatimes.com/home/science/More-than-80-of-healthy-Indians-are-vitamin-D-deficient-Diabetes-Foundation-of-India/articleshow/19898114.cms>, <http://timesofindia.indiatimes.com/life-style/health-fitness/health/High-vitamin-D-levels-may-lead-to-stroke/articleshow/46527522.cms>

Last accessed: March 13, 2015

TEST NAME	RESULT	BIOLOGICAL REFERENCE RANGE		
GLUCOSE - SERUM / PLASMA (FASTING) (Hexokinase)	86	Non-Diabetic : <= 100 Impaired : 101 - 125 Diabetic : >= 126 Male : 16 - 63 U/L		
ALT(SGPT) - SERUM / PLASMA (UV,Using Pyridoxal Phosphate:IFCC) CHOLESTEROL - SERUM / PLASMA (Cholesterol Oxidase / Esterase)	95 ~ 176.00	Female : 14 - 59 U/L Desirable : < 200 Borderline high : 200 High Risk : >=240		
TEST NAME	TECHNOLOGY	VALUE	UNITS	NORMAL RANGE
ALKALINE PHOSPHATASE	PHOTOMETRY	70	U/l	M:53 to 128 - F:42 to 98
BILIRUBIN - TOTAL	PHOTOMETRY	0.57	mg/dl	0.30 - 1.20
BILIRUBIN -DIRECT	PHOTOMETRY	0.18	mg/dl	0 - 0.30
BILIRUBIN (INDIRECT)	CALCULATED	0.39	mg/dl	0 - 0.9
ASPARTATE AMINOTRANSFERASE (SGOT)	PHOTOMETRY	27	U/l	M: 0 to 37 - F: 0 to 31
ALANINE TRANSAMINASE (SGPT)	PHOTOMETRY	39	U/l	M: 13 to 40 - F: 10 to 28
GAMMA GLUTAMYL TRANSFERASE (GGT)	PHOTOMETRY	24.7	U/l	M: 0 to 55 - F : 0 to 38
PROTEIN - TOTAL	PHOTOMETRY	7.27	gm/dl	5.7 - 8.2

Monsoon to be weak

■ PMO in a crisis mode; asks MIHA, Cab-secy for solutions

DC CORRESPONDENT
NEW DELHI, JUNE 2

Already facing an agrarian crisis, there is bad news in store for India with a "deficient" monsoon predicted this year, raising fears of a drought in parts of the country.

The Prime Minister's Office has gone into over-drive, drawing up contingency plans till district level to deal with a drought, worried at a spillover effect on all sectors.

The Cabinet Secretary and the MHA disaster management division have been tasked with putting in place detailed plans and out-of-the-box solutions to tackle crop losses.

Earth sciences minister Harsh Vardhan, giving the revised forecast for the 2015 monsoons, said on Tuesday: "I have to say with a heavy heart that as per our revised forecast, India will receive 88 per cent of rainfall of the Long Period Average, plus or

minus four per cent."

The southwest monsoon is also delayed, and may now hit the Kerala coast on June 5.

Though the normal date for onset of monsoon over Kerala is June 1, the Indian Meteorological Department (IMD) had predicted that it will hit the southern state on May 30 this year.

June 1 also marks the official onset of rains in the country.

"Conditions are becoming favourable for the

onset of southwest monsoon over Kerala around June 5," said an IMD statement.

Last year too, monsoon was delayed and hit the Kerala coast on June 6.

■ Another report inside



I have to say with a heavy heart that as per our revised forecast, India will receive 88% of rainfall of the Long Period Average, plus or minus four per cent. **HARSH VARDHAN, EARTH SCIENCES MINISTER**



Skymet rejects IMD forecast

■ **Claims India will get good rainfall; Centre offers subsidy to farmers**

New Delhi, June 5: Rejecting projections for a drought by India's meteorological office, the country's only private weather forecaster Skymet said there will be a good monsoon season this year as the Indian Ocean Dipole (IOD) phenomenon counters an El Nino weather event.

The outlook by Skymet should help allay concerns of farmers who have already been hit hard by unseasonal rains earlier this year and have been on edge ever since the government forecast what could be India's first drought in six years. Skymet, which says its predictions have been correct for the past three years, forecast rains at 102 per cent of the long-term average over the June-September monsoon season, versus Indian Meteorological Department's 88 per cent. "The El Nino threat remains, we do not deny

that," Skymet's chief meteorologist, G.P. Sharma, said. "But the positive IOD effect could bail us out."

Partly due to the emergence of an El Nino, or the rise of sea-surface temperatures in the Pacific that lead to dry weather in Asia, India's monsoon rains this year arrived five days later than expected on Friday.

Mr Sharma is, however, hopeful that a "positive IOD" this year will slightly offset the impact of an El Nino. The IOD phenomenon is characterised by higher sea-surface temperatures in the Indian Ocean.

A positive IOD creates a barrier in the eastern Indian Ocean and all the southwesterly winds blow towards the Indian sub-continent, causing rains there but leading to droughts in parts of Indonesia and Australia, according to Skymet.

"News will bring relief to farmers who are already facing unseasonal rains



File photo showing citizens getting drenched due to heavy showers.

HAPPY SHOWERS

■ Skymet said there will be a good monsoon season this year as the Indian Ocean Dipole phenomenon counters an El Nino weather event.

■ The IOD is characterised by higher surface temperature in the Indian Ocean.

■ A positive IOD creates a barrier in the eastern Indian Ocean and all the southwesterly winds blow towards the Indian sub-continent, causing rains there but leading to droughts in parts of Indonesia and Australia, according to Skymet.

■ News will bring relief to farmers who are already facing unseasonal rains

rural development, food and fertiliser ministries, among others were also present in the meet.

"We will offer subsidy on diesel, power and seeds to farmers like last year to deal with drought-like situation," Mr Singh said. The minister said the government is fully prepared to face drought-like situation. — Agencies

relief to farmers, the government on Friday announced it will offer subsidy on diesel, power and seeds to farmers in case of deficient monsoon affecting crops.

Agriculture minister Radha Mohan Singh on Friday held a meeting with senior officials of IMD. Officials from power, water resource,

CSO asked to explain GDP growth estimates

■ Data has been overestimated post note ban, say experts

RAVAN BALUDU

NEW DELHI, March 15

The Parliamentary standing committee on finance headed by Dr M Venkateswaran Moily has sought an explanation from CSO over its GDP estimate for 2016-17, which it said is considered by experts as "over-estimation" in view of demonetisation.

Central Statistics Office (CSO) in its advance estimates has pegged GDP growth for 2016-17 at 7.1 per cent and, if kept unchanged while announcing Q3 growth figures, last month, CSO had said that despite demonetisation, India's GDP grew by 7 per cent between October to December 2016 (Q3 2016-17) to retain the title of the world's fastest growing major economy.

Pvt research plays down risk from mobiles: Study

PHONE TROUBLE

DurgeshNandanJha

@timesgroup.com

New Delhi: Is radiation from mobile phones harmful? Multiple studies globally have not conclusively reached an answer. But an analysis by AIIMS of all research on the subject has found an interesting pattern — government funded studies show increased risk of brain tumour

on long-term exposure to mobile phone radiation while industry-funded research tends to underestimate the risk. "We found that industry-funded studies are not of good quality and tend to under-estimate the risk. Government funded studies show in-

▲ 1.33 times higher risk of brain tumour due to mobile use, according to average of all studies

▲ Mixed-funded (govt, industry, mobile makers) research says risk much lower, at 1.05 times

▲ Studies funded by phone industry and mixed sources have low quality score of 5.6

▲ Score of govt-funded research higher at 7.8

creased risk of brain tumour on long-term exposure," said Dr Kamleshwar Prasad, head of neurology at AIIMS, who is lead author of the study.

► Brain tumour risk, P 12

was the advance 7.1% GDP estimate for FY17. The CSO kept its estimate unchanged while announcing the Q3 growth figures for 2016-17, which has been considered by independent experts as a possible over-estimation, particularly in the backdrop of demonetisation

— PARLIAMENTARY STANDING COMMITTEE ON FINANCE



national accounts with national accounts with base year 2011-12 as the base year methodology from 2015 has raised more questions than answers. "The GDP data does not seem to reflect the momentum of economic activities across the spectrum. It is, therefore, imperative that a more realistic computation method be adopted with a view to enhancing the credibility of official statistics," said the report.

The committee also raised question over CSO methodology from 2015 for computing GDP data. In its report, the committee noted that in currency of old ₹500 and ₹1,000 on November 8 will hit the economy. The Reserve Bank of India and other agencies like IMF and OECD had lowered GDP projections for 2016-17 replacing the previous series with the new series of the note ban would have a short-term impact on the Indian economy.

The committee had warned that demonetisation of old currency of ₹500 and ₹1,000 on November 8 will hit the economy. The Reserve Bank of India and other agencies like IMF and OECD had lowered GDP projections for 2016-17 replacing the previous series with the new series of the note ban would have a short-term impact on the Indian economy.

The committee had also like to be apprised about the rationale/process/assumptions made and adopted by the CSO in their recent GDP advance estimates for 2016-17, which has been considered by independent experts as a possible over-estimation, particularly in the backdrop of demonetisation

— PARLIAMENTARY STANDING COMMITTEE ON FINANCE



CLIMATE HAS A NOSE FOR SHAPES

THE SHAPE OF OUR NOSES WAS FORMED BY A LONG PROCESS OF ADAPTATION TO CLIMATE, A STUDY SAYS

■ WIDER noses are more common in warm-humid climates

■ NARROWER noses are more common in cold-dry climates

ADAPTING MECHANISM

Narrower nostrils alter the airflow so the mucous-covered inside of the nose can humidify and warm the air

People with narrower nostrils therefore fared better in colder climates

This lead to a gradual decrease in nose width in populations living far away from the equator

EVOLUTION OF THE NOSE

The width of the nostrils and the base of the nose measurements differ across populations, researchers found

This indicates a role for natural selection in the evolution of nose shape.

The width of the nostrils is strongly correlated with temperature and absolute humidity

How hand sanitisers can harm children

Half of all pregnant women are anaemic



►Continued from P1

The WHO defines wasting as low weight for height, stunting as low height for age, and underweight as low weight for age.

The survey also found that just over half of all pregnant women were anaemic. This would automatically translate into their newborn being weak. Overall, 53% of women and 23% of men in the 15-49 age group were anaemic.

There is wide variation among states. The data for UP has not been released in view of the ongoing polls, according to Balram Paswan, professor at Mumbai-based International Institute for Population Sciences which was the nodal agency for the survey done for the health ministry. But poorer states like Bihar, Madhya Pradesh, Jharkhand, Assam, Rajasthan and Chhattisgarh have higher than national average rates on all markers.

More advanced states like

CLOSING THE GENDER PAY GAP?



Source: Monster.com/WageIndicator Foundation

men feel that gender parity need not be a top priority for their organisations. On the other hand, almost 68% expressed that even if gender parity is a priority, themanagement "does not walk the talk," indicating the need for India Inc to step up and implement pragmatic policies to bridge the gap. "Uneexamined conventions of women's commitment

to work, distractions of family commitments, societal perception of women who work for long hours, etc, are the challenges that constrain women's progress. Sensitisation towards gender diversity and a step towards examining our socialised preconceptions would help break barriers to a more fair and inclusive work environment," said Modi.

According to the survey which aimed at understanding the working women of India and their workplace concerns, nearly 44% wo-

New York: Scientists have warned that hand sanitisers might do more harm than good. They have found these alcohol-based, scented products might tempt young kids to swallow the substance — leading to stomach pain, nausea, apnea and even coma.

Researchers from US Centres for Disease Control (CDC) and Prevention have identified serious consequences, including apnea, acidosis and coma in young children who swallowed alcohol-based hand sanitisers. To characterise paediatric hand sanitiser exposures, researchers from US

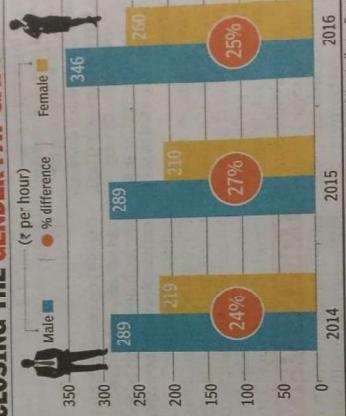
Centres for Disease Control and Prevention analysed data reported by poison centres among children aged 12 years during 2011 to 2014.

The study found majority of intentional exposures to alcohol hand sanitiser occurred in children aged 6-12 years. During 2011-2014, 70,669

children aged 12 years were reported, of which 65,293 (92%) were alcohol exposures and 5,376 (8%) were non-alcohol exposures. These data also indicate that, among older children, exposures occur less frequently during the summer months. *PTI*

Women earned 25% less than men in 2016: Study

CLOSING THE GENDER PAY GAP?



Source: Monster.com/WageIndicator Foundation

men feel that gender parity need not be a top priority for their organisations. On the other hand, almost 68% expressed that even if gender parity is a priority, themanagement "does not walk the talk," indicating the need for India Inc to step up and implement pragmatic policies to bridge the gap. "Uneexamined conventions of women's commitment

to work, distractions of family commitments, societal perception of women who work for long hours, etc, are the challenges that constrain women's progress. Sensitisation towards gender diversity and a step towards examining our socialised preconceptions would help break barriers to a more fair and inclusive work environment," said Modi.

According to the survey which aimed at understanding the working women of India and their workplace concerns, nearly 44% wo-

'Cancer may strike due to bad luck, not lifestyle'

Random Changes In DNA During Cell Division Cause Nearly Two-Third Of All Cancers In Humans, Finds Study

Subodh Varma@timesgroup.com

Why does cancer strike some people and not others? New research shows that random changes or 'mistakes' in DNA when cells are dividing cause nearly two-third of all cancers in humans. These changes are neither caused by external factors like smoking or exposure to harmful chemicals, nor by hereditary factors. They are chance events occurring at the molecular level. In other words, cancer can strike anybody.

This upends prevailing wisdom that cancer is mostly a lifestyle related disease caused by external or environmental factors like smoking, harmful chemicals

and conditions like obesity. While all these are valid and important risk increasing factors, random chance may be the real driver, if one goes by this new research. Different types of cancers have different origins, according to the study. For example, in pancreatic cancers, 77% are due to random DNA copying errors, 18% to environmental factors, such as smoking, and the remaining 5% to heredity. In other cancer types, such as those of the prostate, brain or bone, more than 95% of the mutations are due to random copying errors.

smoke and 35% due to DNA copy-

ing errors. Inherited factors have

negligible role. The study involved

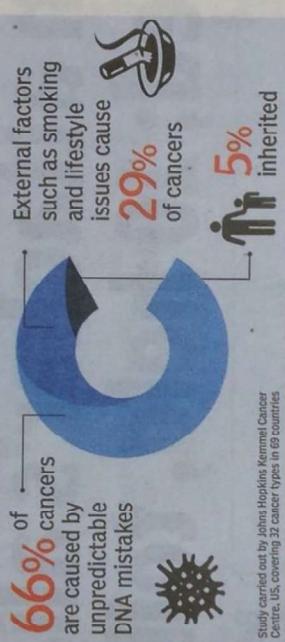
more than half of the world's

population. It was done by scientists from Johns Hopkins Kimmel Cancer Center at Baltimore, US,

and published in the peer review

journal Science on March 24.

A MATTER OF CHANCE?



Study carried out by Johns Hopkins Kimmel Cancer Centre, US, covering 32 cancer types in 69 countries

66% of cancer mutations result from copying errors, 29% can be attributed to lifestyle or environmental factors, and the remaining 5% are inherited. They found a strong correlation between cancer incidence and normal cell divisions among 17 cancer types, regardless of the countries environment or stage of economic development. This means that lifestyle factors like smoking or exposure to toxic chemicals are also very important factors causing nearly a third of cancers.

"We need to continue to encourage people to avoid environmental agents and lifestyles that increase their risk of developing cancer mutations," co-author Bert Vogelstein emphasised.

Why Study Statistics?

Statistics are part of your daily life and are all around you.

Why Study Statistics?

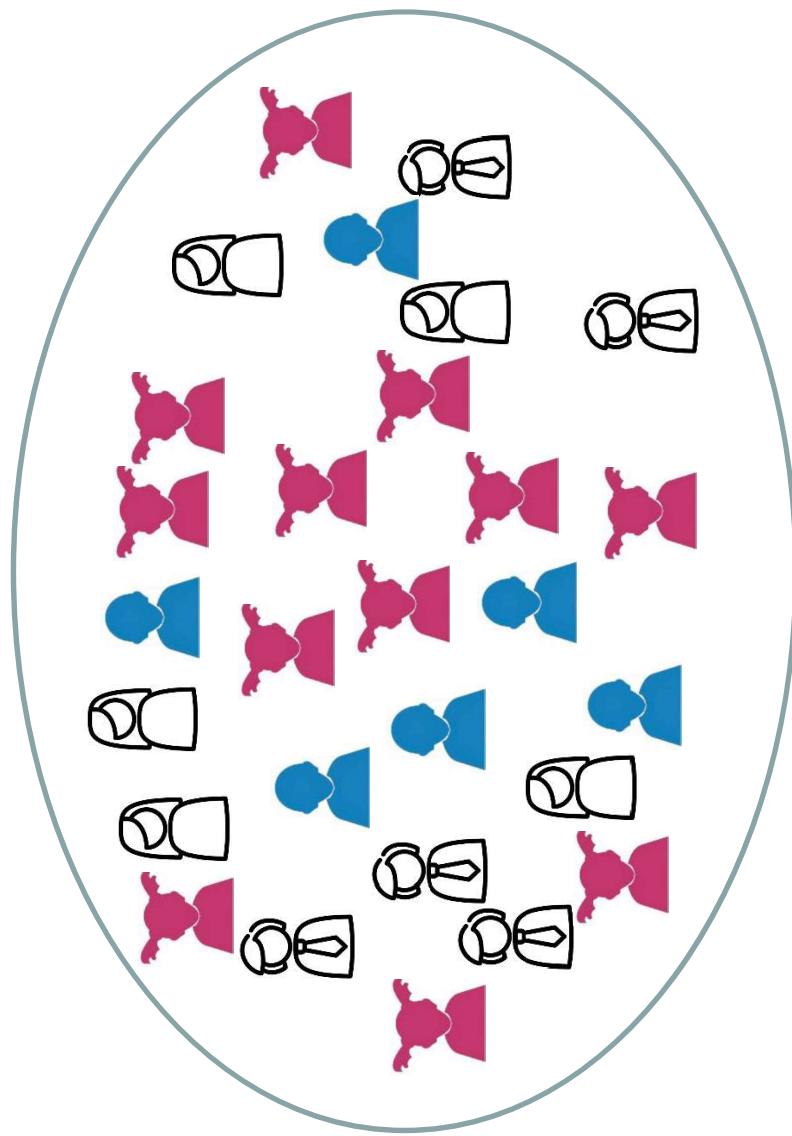
Statistics don't lie but Statisticians will in any of the following situations:

- Data Gathering
- Data Understanding
- Data Analysis/Interpretation
- Data Presentation

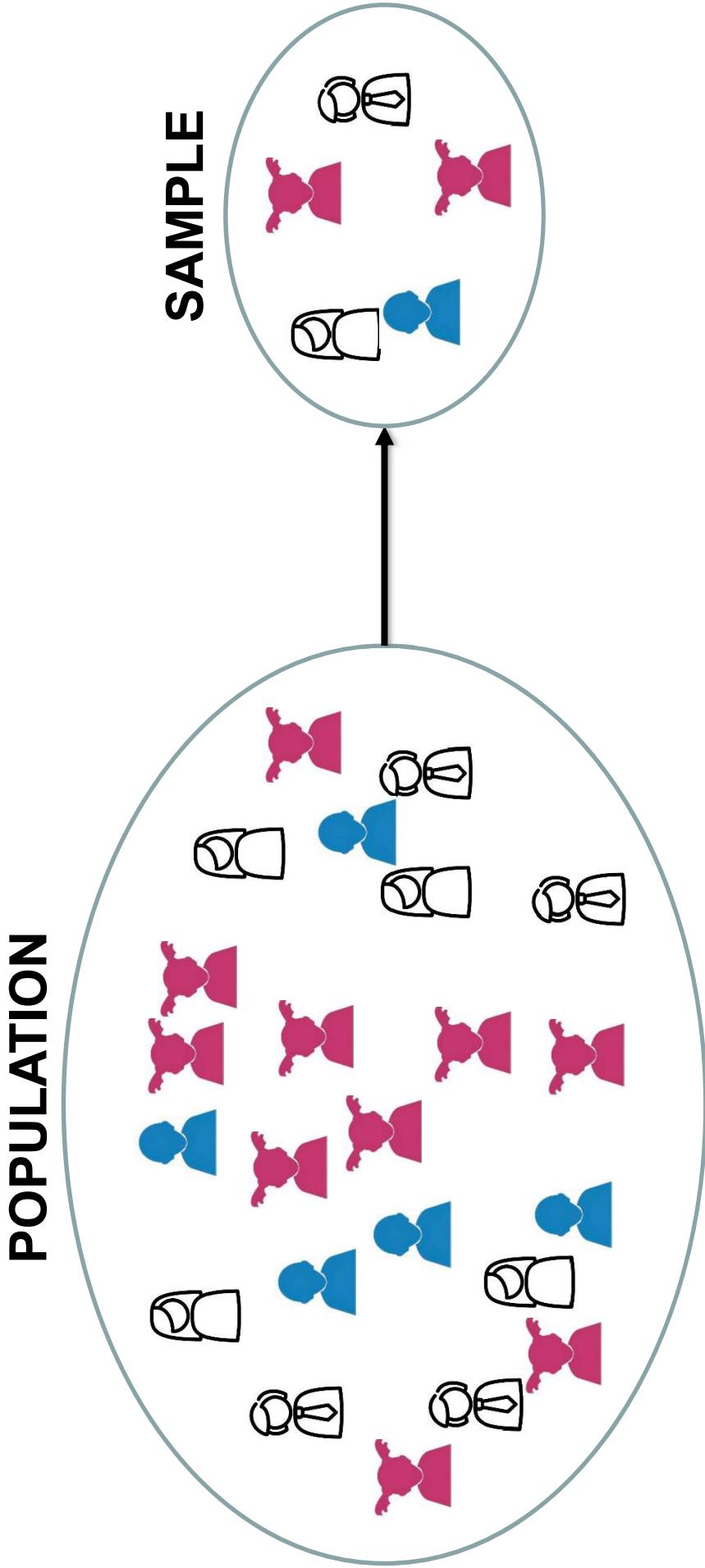
BASIC STATISTICAL TERMINOLOGY

Population and Sample

POPULATION



Population and Sample



Name	Ht.	Hometown	Class
Cheyanne Bustle	5'0"	Prestonburg, KY	Fr.
Jaclyn Fyffe	5'3"	Richmond, KY	Fr.
Brooke Gibbs	4'11"	Pineville, KY	So.
Michelle Malavasi	4'10"	Heredia, Costa Rica	So.
Madison Mullin	5'2"	Georgetown, KY	Fr.
Dallas Pringle	5'2"	Reno, NV	Fr.
Chelsee Ramos	5'2"	Madison, WI	Jr.
Sydney Shelton	4'10"	Scottsville, KY	Jr.
Ashley Wettstein	5'0"	Owensboro, KY	Fr.
Madison Yee	5'2"	San Marcos, CA	So.

Source: <http://www.ukathletics.com/trads/cheer-roster.html>

Last accessed: October 7, 2014

No.	Name	Pos.	Cl./Exp.	Ht.	Hometown/High School/Last College
0	Jennifer O'Neill	PG	SR-3L	5-6	Bronx, N.Y./Saint Michael Academy
2	Ivana Jakubcová	C	JR-JC	6-6	Bratislava, Slovakia/Murray State College
3	Janeen Thompson	PG	JR-2L	5-7	Chicago, Ill./Whitney Young
5	Kwin Goodin-Rogers	F	SO-HS	6-1	Lebanon, Ky./Marion Co.
12	Jelleah Sidney	F/C	SR-2L	6-2	Queens Village, N.Y./Saint Michael Academy/Chipola JC
13	Bria Goss	G	SR-3L	5-10	Indianapolis, Ind./Ben Davis
15	Linnae Harper	G	SO-1L	5-8	Chicago, Ill./Whitney Young
24	Jaycee Coe	G	FR-HS	5-11	Gainesboro, Tenn./Jackson Co.
25	Makayla Epps	G	SO-1L	5-10	Lebanon, Ky./Marion Co.
35	Alexis Jennings	F/C	FR-HS	6-2	Madison, Ala./Sparkman
45	Alyssa Rice	C	FR-HS	6-3	Reynoldsburg, Ohio/Reynoldsburg
50	Azia Bishop	F/C	SR-3L	6-3	Toledo, Ohio/Start

Source: <http://www.ukathletics.com/sports/w-basketball/mtt/ktv-w-baskbl-mtt.html>

Last accessed: October 7, 2014



Census and Survey

Census: Gathering data from the whole **population** of interest.

For example, elections, 10-year census, etc.

Survey: Gathering data from the **sample** in order to make conclusions about the population.

For example, opinion polls, quality control checks in manufacturing units, etc.

Parameter and Statistic

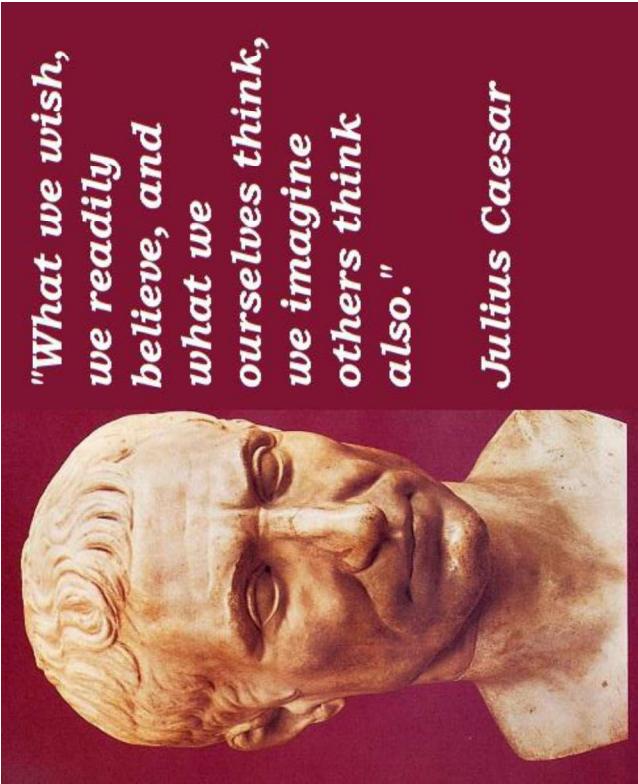
Parameter: A descriptive measure of the **population**.

For example, population mean, population variance, population standard deviation, etc.

Statistic: A descriptive measure of the **sample**.

For example, sample mean, sample variance, sample standard deviation, etc.

Parameter and Statistic



Greek – Population Parameter

Mean – μ

Variance – σ^2

Standard Deviation - σ

Roman – Sample Statistic

Mean – \bar{x}

Variance – s^2

Standard Deviation - s

Descriptive and Inferential Statistics

- Descriptive Statistics – Data gathered about a group to reach conclusions about the same group.
- Inferential Statistics – Data gathered from a sample and the statistics generated to reach conclusions about the population from which the sample is taken. Also known as Inductive Statistics.

1 Diabetes is a huge problem in India.

The prevalence of diabetes increased tenfold, from 1.2% to 12.1%, between 1971 and 2000.

Noncommunicable Diseases in the Southeast Asia Region, Situation and Response, World Health Organization, 2011.
http://apps.searWHO/int/PDS_DOCS/IB4793.pdf

It is estimated that 61.3 million people aged 20-79 years live with diabetes in India (2011 estimates). This number is expected to increase to 101.2 million by 2030.

David R Whiting, et al. IDF Diabetes Atlas: Global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Research and Clinical Practice, Volume 94, Issue 3, December 2011, Pages 311-321, <http://www.sciencedirect.com/science/article/pii/S0168822711005912>

And, 77.2 million people in India are said to have pre-diabetes.

Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, Unnikrishnan R, et al. "Prevalence of diabetes and prediabetes (impaired fasting glucose and/or impaired glucose tolerance) in urban and rural India: phase I results of the Indian Council of Medical Research-India Diabetes (ICMR-INDIAB) study." Diabetologia 54(2) (2011): 302-27. NCBI. Web. March 2013.

Source:
http://www.aryyaworld.org/wp-content/uploads/2010/10/ArogyaWorld_IndiaDiabetes_FactSheets_CGI2013_web.pdf
Last accessed: November 25, 2015

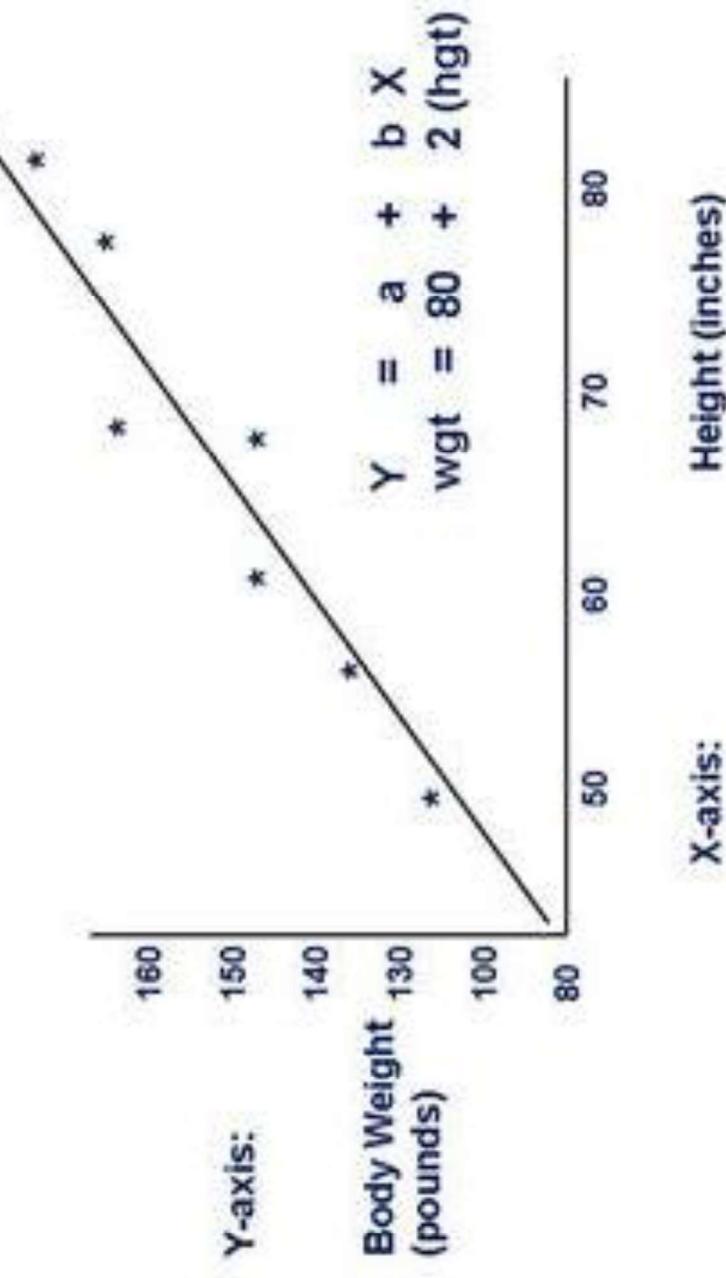
Variables and Data

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	unknown	no
33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5	may	76	1	-1	0	unknown	no
47	blue-collar	married	unknown	no	1506	yes	no	unknown	5	may	92	1	-1	0	unknown	no
33	unknown	single	unknown	no	1	no	no	unknown	5	may	198	1	-1	0	unknown	no
35	management	married	tertiary	no	231	yes	no	unknown	5	may	139	1	-1	0	unknown	no
28	management	single	tertiary	no	447	yes	yes	unknown	5	may	217	1	-1	0	unknown	no
42	entrepreneur	divorced	tertiary	yes	2	yes	no	unknown	5	may	380	1	-1	0	unknown	no
58	retired	married	primary	no	121	yes	no	unknown	5	may	50	1	-1	0	unknown	no
43	technician	single	secondary	no	593	yes	no	unknown	5	may	55	1	-1	0	unknown	no
41	admin.	divorced	secondary	no	270	yes	no	unknown	5	may	222	1	-1	0	unknown	no
29	admin.	single	secondary	no	390	yes	no	unknown	5	may	137	1	-1	0	unknown	no
53	technician	married	secondary	no	6	yes	no	unknown	5	may	517	1	-1	0	unknown	no
58	technician	married	unknown	no	71	yes	no	unknown	5	may	71	1	-1	0	unknown	no
57	services	married	secondary	no	162	yes	no	unknown	5	may	174	1	-1	0	unknown	no
51	retired	married	primary	no	229	yes	no	unknown	5	may	353	1	-1	0	unknown	no
45	admin.	single	unknown	no	13	yes	no	unknown	5	may	98	1	-1	0	unknown	no
57	blue-collar	married	primary	no	52	yes	no	unknown	5	may	38	1	-1	0	unknown	no
60	retired	married	primary	no	60	yes	no	unknown	5	may	219	1	-1	0	unknown	no
33	services	married	secondary	no	0	yes	no	unknown	5	may	54	1	-1	0	unknown	no
28	blue-collar	married	secondary	no	723	yes	yes	unknown	5	may	262	1	-1	0	unknown	no
56	management	married	tertiary	no	779	yes	no	unknown	5	may	164	1	-1	0	unknown	no
32	blue-collar	single	primary	no	23	yes	yes	unknown	5	may	160	1	-1	0	unknown	no
25	services	married	secondary	no	50	yes	no	unknown	5	may	342	1	-1	0	unknown	no
40	retired	married	primary	no	0	yes	no	unknown	5	may	181	1	-1	0	unknown	no

Variables – Dependent and Independent

- Dependent variables on y-axis and Independent on x-axis.
- Dependent variable also called Target variable or Class variable.

Simple Linear Regression
 (with a continuous dependent [Y] variable)



Data – Numeric and Categorical



18 kg



27 kg



Sources: <http://banglanews24.com/en/files/2013August/SM/Gold-sm20130830024804.jpg>,
<http://myoor.com/wp-content/uploads/2014/01/gold.jpg> and <http://im.rediff.com/cricket/2014/feb/01india1.jpg>

Last accessed: November 22, 2014

Categorical Data (Qualitative)

Nominal

Examples

- Employee ID
- Gender
- Religion
- Ethnicity
- Pin codes
- Place of birth
- Aadhaar numbers

Ordinal

Examples

- Mutual fund risk ratings
- Fortune 50 rankings
- Movie ratings

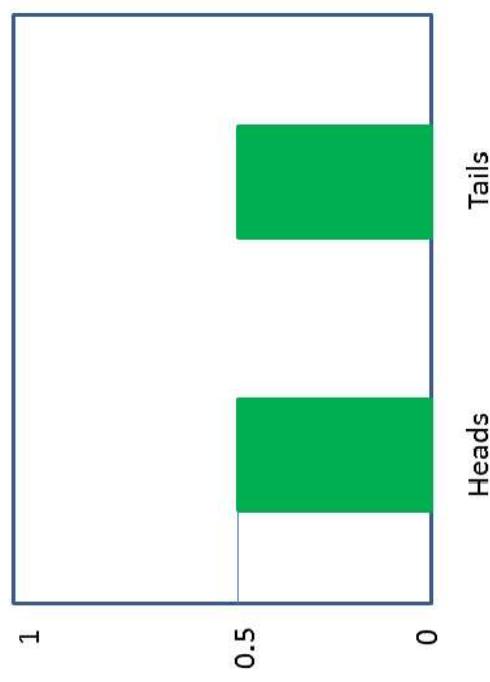
While there is an order, difference between consecutive levels are not always equal.

Numeric Data (Quantitative)

Examples

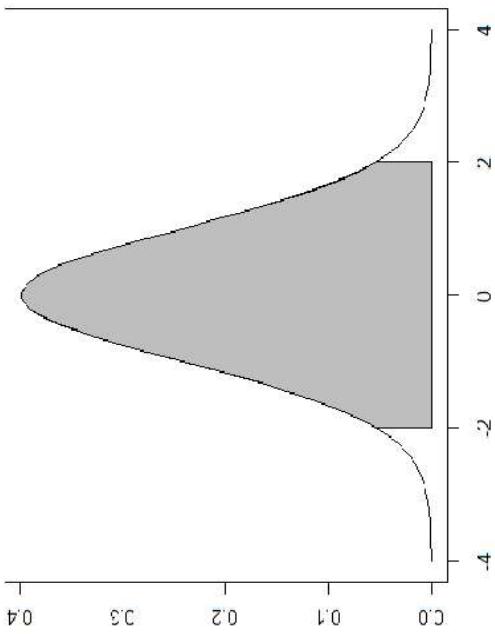
- Height
- Weight
- Time
- Volume
- Number of iPads sold
- Number of complaints received at the call centre
- Number of employees
- Percentage return on a stock
- Rupee change in stock price

Discrete and Continuous



Countable

Measurable



Discrete or Continuous?

Time between customer arrivals at a retail outlet	Continuous
Sampling 100 voters in an exit poll and determining how many voted for the winning candidate	Discrete
Lengths of newly designed automobiles	Continuous
No. of customers arriving at a retail outlet during a five-minute period	Discrete
No. of defects in a batch of 50 items	Discrete

DESCRIBING DATA THROUGH STATISTICS

The Central Tendencies - Excel

Average and Median Monthly Salary Comparison in Bahrain



Salary (BHD)	100	345	1000	9833
Frequency, f	10	1	10	2

$$\text{Mean, } \mu = \frac{\sum x}{n} = \frac{\sum f x}{\sum f} = \frac{100X10+345X1+1000X10+9833X2}{10+1+10+2} = 1348$$

The Central Tendencies

SALARY GAP

A PAY GAP BETWEEN CEOS AND OTHER EMPLOYEES HAS COME TO THE FORE, WITH FIRMS DOLING OUT SALARY PACKAGES OF UP TO 1200 TIMES OF THEIR MEDIAN EMPLOYEE REMUNERATIONS.

■ The public sector companies show a totally different picture with their chiefs getting salaries of just about 3-4 times of their median employee remunerations.

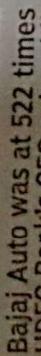


MARKET OVERVIEW

■ Wipro (down from 260 times to 259 times), Infosys (283 times)

■ Dr Reddy's Lab (from 312 times to 233 times)

■ Hero MotoCorp (from 755 times to 731 times).



Bajaj Auto was at 522 times

HDFC Bank's CEO salary hike saw it rise from 179 to 187 times.

■ The median employee remuneration fell or remained almost same during the last fiscal, while the ratio of the top executive's pay to the median employee remuneration remained at astronomically high levels of hundreds of times in many cases.

■ The median employee remuneration fell or remained almost same during the last fiscal, while the ratio of the top executive's pay to the median employee remuneration remained at astronomically high levels of hundreds of times in many cases.

■ The median employee remuneration fell or remained almost same during the last fiscal, while the ratio of the top executive's pay to the median employee remuneration remained at astronomically high levels of hundreds of times in many cases.

■ The median employee remuneration fell or remained almost same during the last fiscal, while the ratio of the top executive's pay to the median employee remuneration remained at astronomically high levels of hundreds of times in many cases.

RIL chairman Mukesh Ambani gets 205 times company's median salary

This ratio stands at 439 times in case of ITC Executive Chairman Y C Deveshwar

Press Trust of India | New Delhi | July 6, 2015 Last Updated at 02:42 IST

Ramcon HRIS Solutions Case Study Press Solution With Case Study Prodigy Case For Oracle insafe.com/HRIS

Access To Strategic

Information Technology (IT) Major Wipro Chairman And Managing Director Atul Puri And At 39 Times For HDFC Chairman Deepak Parekh For 2014-15.

However, HDFC Banks Managing Director (MD) Atul Puri got a remuneration package that was 117 times of the median employees' pay, while for IOCL, Bank Chief Executive Officer (CEO) Chandra Kochhar it was 97 times and Aditya 74 times for Axis Banks MD and CEO Shikha Sharma.

IT giant TCS CEO Vishal Sikka got 116 times of median employees' pay. The same ratio for HUL's CEO Sanjiv Mehta was 93 times, but much higher at 263 times for Vedanta Chairman Naveen Agarwal.

United firms have begun disclosing these ratios and other compensation such as salary ratios for top management, personnel and average staff members, for the first time pursuant to the new Companies Act and SEBI's Listed Corporate Governance Code coming into force.

While a majority of the companies are still in the process of disclosing such details, the discussions made so far by top companies show a wide variance in these ratios. There is also a more difference between the pay increases for top management personnel and average staff in many cases.

Ambani has kept his salary capped at Rs 1.5 crore for seven years now, while the median remuneration of employees increased by 3.71 per cent to Rs 2.78 lakh during 2014-15.

The total remuneration of key managers personnel in fact, dipped by 1.93 per cent to Rs 73.28 crore.

Deveshwar's remuneration rose by 24 per cent during the year, against an increase of 14 per cent in the company's mean employees' remuneration. The overall net managerial personnel remuneration rose 20 per cent. Deveshwar's gross remuneration in 2014-15 stood at over Rs 15 crore, while median employee's pay at Rs 1.3 crore.

Parekh saw his pay declined by 33 per cent to Rs 4.78 crore, while median employee's remuneration rose by 3.5 per cent. Wipro CEO T V Kurian got a package that was 170 times the

median employees' pay.

The ratio stands much lower at 99 times in case of Information Technology (IT) major Wipro Chairman And Managing Director Atul Puri And At 39 Times For HDFC Chairman Deepak Parekh For 2014-15.

However, HDFC Banks Managing Director (MD) Atul Puri got a remuneration package that was 117 times of the median employees' pay, while for IOCL, Bank Chief Executive Officer (CEO) Chandra Kochhar it was 97 times and Aditya 74 times for Axis Banks MD and CEO Shikha Sharma.

IT giant TCS CEO Vishal Sikka got 116 times of median employees' pay. The same ratio for HUL's CEO Sanjiv Mehta was 93 times, but much higher at 263 times for Vedanta Chairman Naveen Agarwal.

United firms have begun disclosing these ratios and other compensation such as salary ratios for top management, personnel and average staff members, for the first time pursuant to the new Companies Act and SEBI's Listed Corporate Governance Code coming into force.

While a majority of the companies are still in the process of disclosing such details, the discussions made so far by top companies show a wide variance in these ratios. There is also a more difference between the pay increases for top management personnel and average staff in many cases.

Ambani has kept his salary capped at Rs 1.5 crore for seven years now, while the median remuneration of employees increased by 3.71 per cent to Rs 2.78 lakh during 2014-15.

The total remuneration of key managers personnel in fact, dipped by 1.93 per cent to Rs 73.28 crore.

Deveshwar's remuneration

rose by 24 per cent during the year, against an increase of 14 per cent in the company's mean employees' remuneration. The overall net managerial personnel remuneration rose 20 per cent. Deveshwar's gross remuneration in 2014-15 stood at over Rs 15 crore, while median employee's pay at Rs 1.3 crore.

Parekh saw his pay declined by 33 per cent to Rs 4.78 crore, while median employee's remuneration rose by 3.5 per cent. Wipro CEO T V Kurian got a package that was 170 times the

CHEQUES & BALANCES

Ratio to median employee salary in the company Head Honchos

	Y Chaudhury, executive chairman, ITC	Nitin Agarwal, chairman, Vedanta	Mukesh Ambani, chairman, Reli.	Aditya Puri, MD, HUL	Vishal Sikha, ED, Infosys	Chand Kochhar, ED, IOCL	Sanjiv Mehta, ED, HUL	Adm Parekh, MD, Wipro	Shikha Sharma, MD & ED, Axis Bank	Deepak Parekh, chairman, HDFC
Ratio to median employee salary in the company	439	293	205	117	116	97	93	89	74	19
Median employees' pay	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh
Median employees' pay	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh
Median employees' pay	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh	Rs 2.78 lakh

Bajaj Auto was at 522 times

HDFC Bank's CEO salary hike saw it rise from 179 to 187 times.

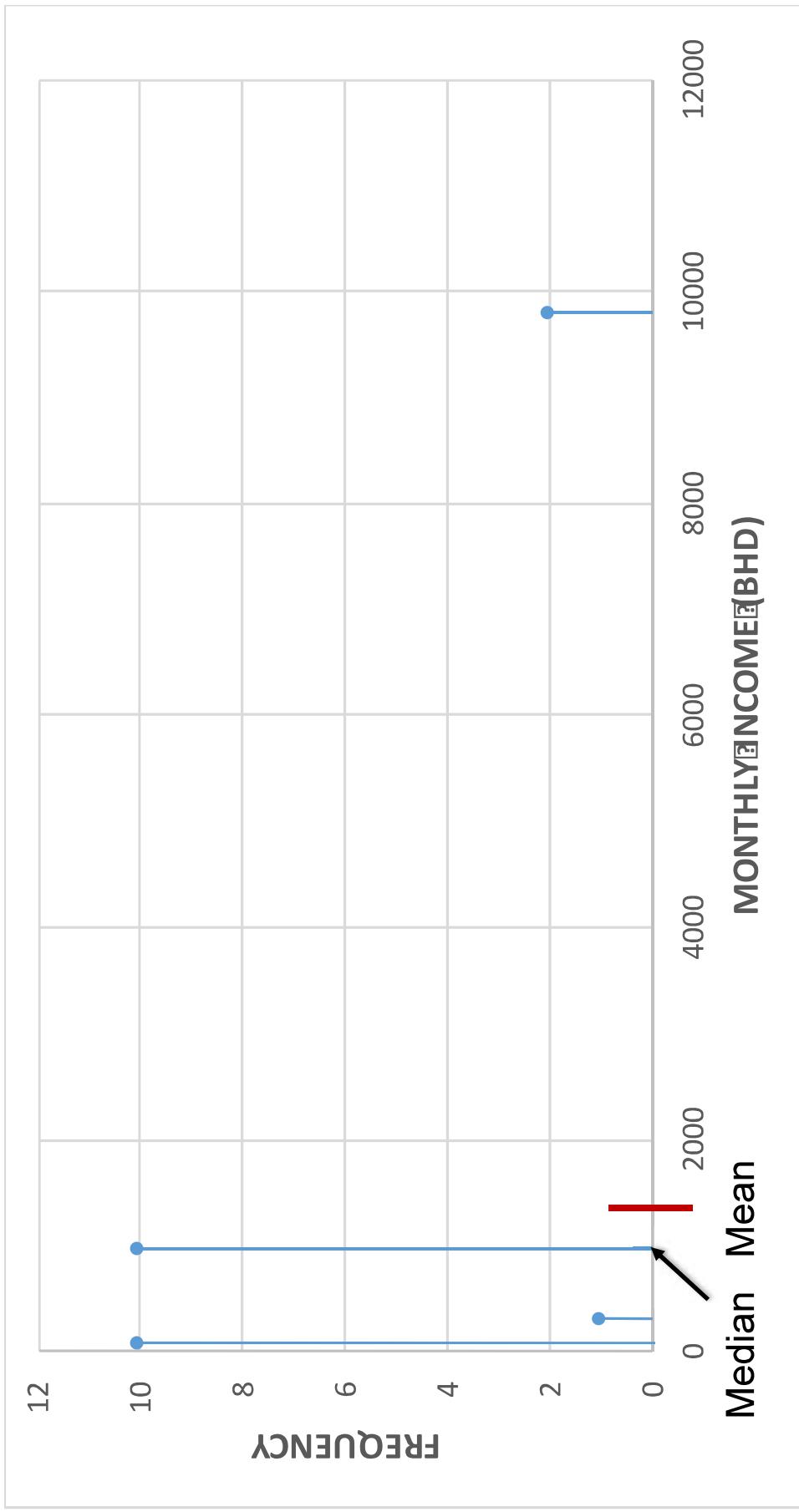
The Central Tendencies

Median: Arrange data in increasing order and find the mid-point
 $\frac{(n+1)}{2}$.

100,100,100,100,100,100,100,
345,**1000**,1000,1000,1000,1000,
1000,1000,1000,9833,9833

The Central Tendencies - Excel

Salary (BHD)	100	345	1000	9833
Frequency, f	10	1	10	2



The Central Tendencies

Salary (BHD)	100	345	1000	9833
Frequency, f	10	1	10	2

What is the Mode – the most frequently occurring data point?

The Central Tendencies

Mean and Median need not be in the dataset but Mode has to be in it.

Mode is also the only average that works with categorical data.

The Central Tendencies

The management of Good Heart Inc. wants to give all its employees a raise. They are unable to decide if they should give a straight Rs 2000 to everyone or to increase salaries by 10% across the board. The mean salary is Rs 50,000, the median is Rs 20,000 and the mode is Rs 10,000.

How do these central tendencies change in both cases?

Measuring Variability and Spread

Basketball coach Statson is in a dilemma choosing between 3 players all having the same average scores.

Points scored per game	7	8	9	10	11	12	13
Frequency, f	1	1	2	2	2	1	1

Points scored per game	7	9	10	11	13
Frequency, f	1	2	4	2	1

Points scored per game	3	6	7	10	11	13	30
Frequency, f	2	1	2	3	1	1	1

Mean = Median = Mode = 10 for all 3.

Measuring Variability and Spread

$$\text{Range} = \text{Max} - \text{Min}$$

Points scored per game	7	8	9	10	11	12	13
Frequency, f	1	1	2	2	2	1	1

Points scored per game	7	9	10	11	13
Frequency, f	1	2	4	2	1

Points scored per game	3	6	7	10	11	13	30
Frequency, f	2	1	2	3	1	1	1

Measuring Variability and Spread

Exclude outliers scientifically – Quartiles

Points scored per game	3	6	7	10	11	13	30
Frequency, f	2	1	2	3	1	1	1

3 3 6 7 7 10 10 10 11 13 30

Lower quartile (25th percentile, Q₁) = $\frac{(n+1)}{4}$ th

Middle quartile = Median = $\frac{2*(n+1)}{4}$ th

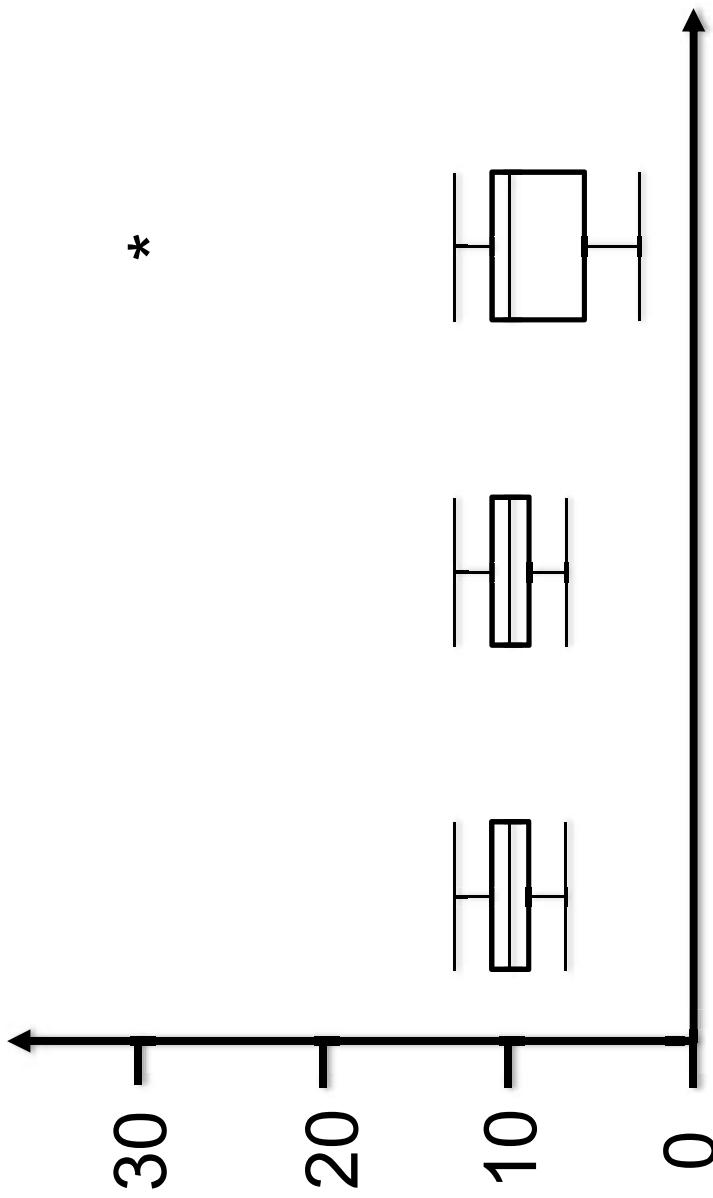
Upper quartile (75th percentile, Q₃) = $\frac{3*(n+1)}{4}$ th

Interquartile range, IQR = Q₃-Q₁ (central 50% of data)

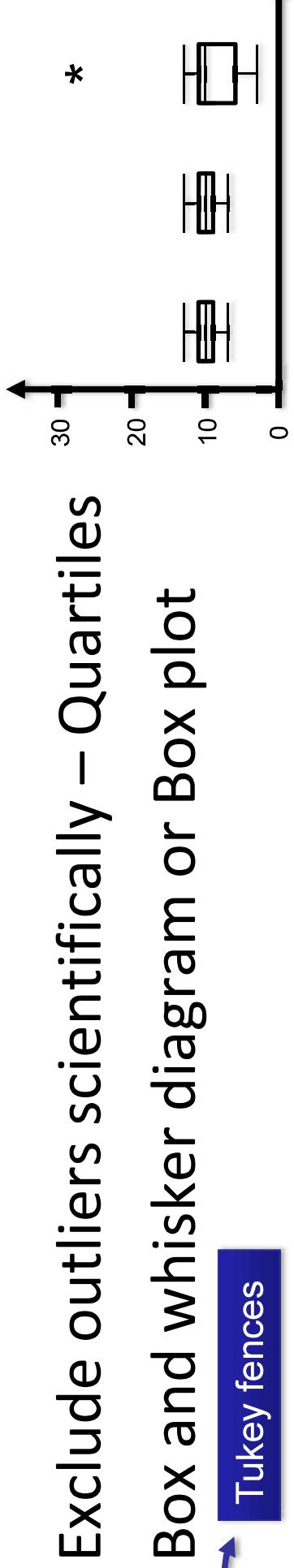
Measuring Variability and Spread

Exclude outliers scientifically – Quartiles

Box and whisker diagram or Box plot



Measuring Variability and Spread



Name	Formula	Player 1	Player 2	Player 3
Upper Hinge	75th Percentile	11	11	11
Lower Hinge	25th Percentile	9	9	6
H-Spread	Upper Hinge - Lower Hinge (IQR)	2	2	5
Step	$1.5 \times \text{H-Spread} (1.5 * \text{IQR})$	3	3	7.5
Upper Inner Fence	Upper Hinge + 1 Step ($75\text{th percentile} + 1.5 * \text{IQR}$)	14	14	18.5
Lower Inner Fence	Lower Hinge - 1 Step ($25\text{th percentile} - 1.5 * \text{IQR}$)	6	6	-1.5
Upper Outer Fence	Upper Hinge + 2 Steps ($75\text{th percentile} + 3 * \text{IQR}$)	17	17	26
Lower Outer Fence	Lower Hinge - 2 Steps ($25\text{th percentile} - 3 * \text{IQR}$)	3	3	-9
Upper Adjacent	Largest value below Upper Inner Fence	13	13	13
Lower Adjacent	Smallest value above Lower Inner Fence	7	7	3
Outside Value (Outliers)	A value beyond an Inner Fence but not beyond an Outer Fence			
Far Out Value (Extreme Values)	A value beyond an Outer Fence			

Outlier detection – Excel and Box Plot Steps

Hadlum vs Hadlum case



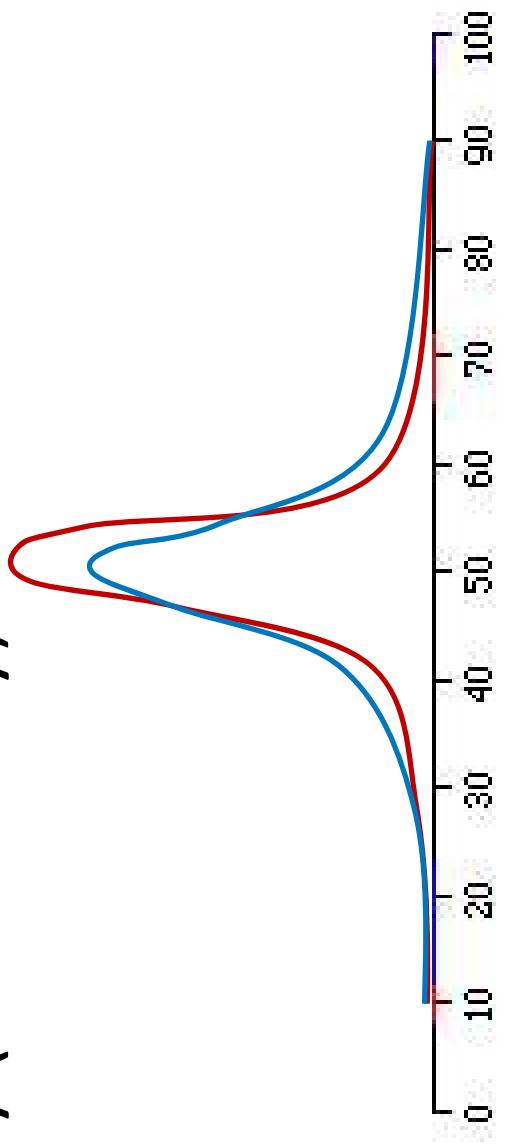
Source: <http://www.alphamom.com/legacy/pregnancy-calendar/week36.jpg>
Last accessed: November 01, 2014



Source: <http://3.bp.blogspot.com/-OYwIRiLMWr0/T4DqQwVCigIAAAAAAAAgg/Yf-ttkQLSq/s1600/fishy.jpg>
Last accessed: November 01, 2014

Measuring Variability and Spread

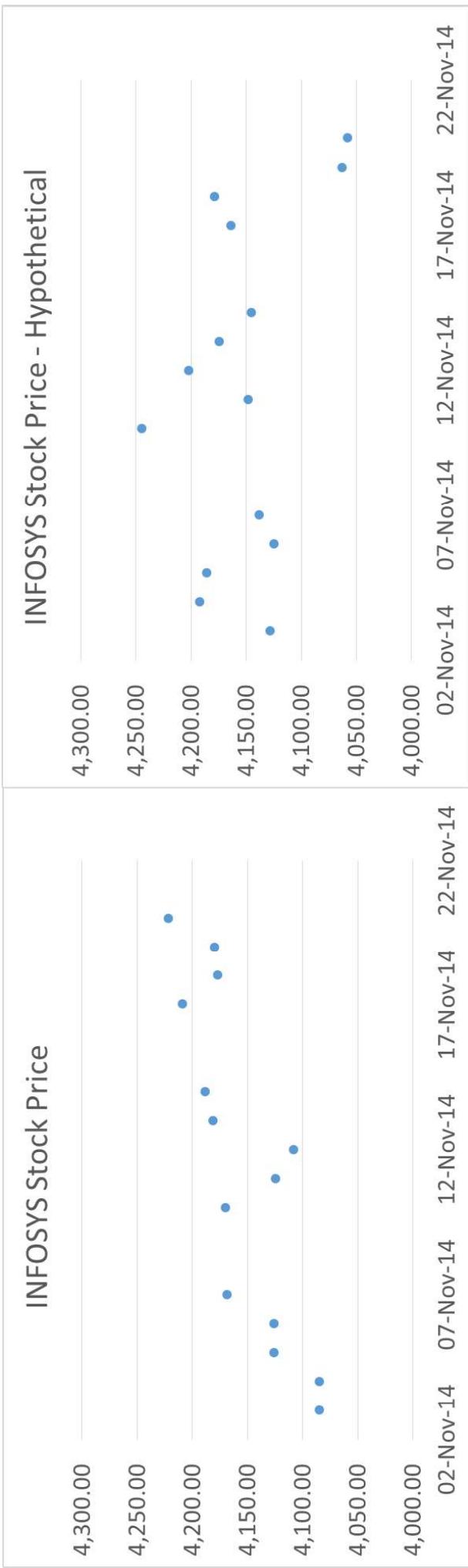
Range and IQR give the spread but still do not describe variability (consistency).



Average distance from the mean?

3 3 6 7 7 10 10 10 11 13 30

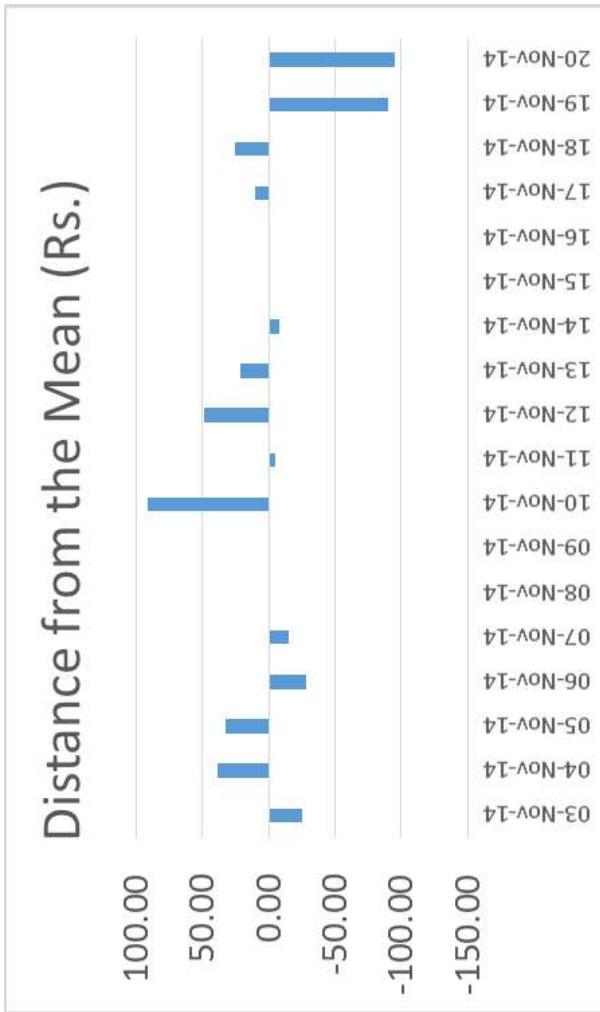
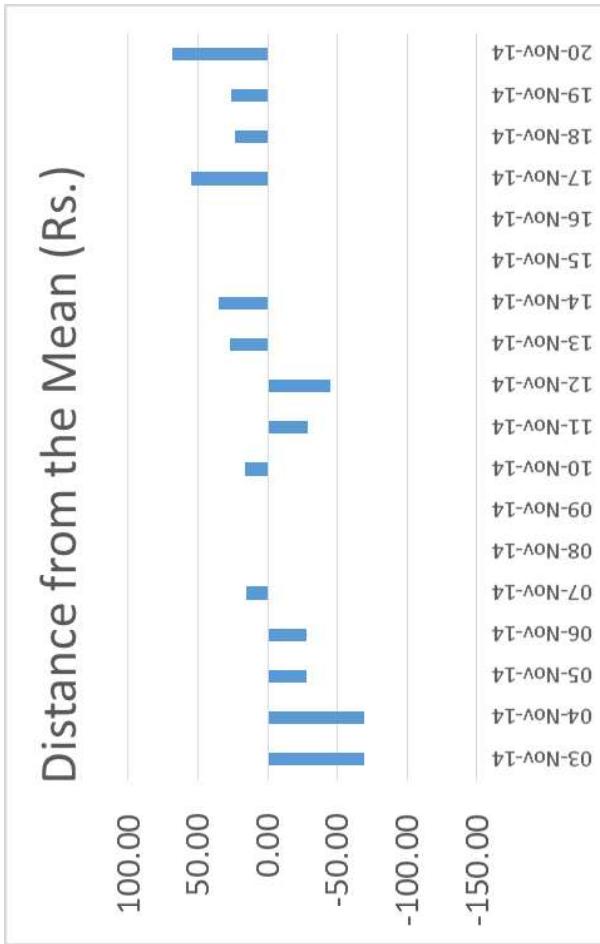
Measures of Spread – Mean Distance, Mean Absolute Deviation or Standard Deviation - Excel



Data Source: <https://in.finance.yahoo.com/q/hp?s=INFY.BO>

The best place for students to learn Applied Engineering

Measures of Spread – Mean Distance, Mean Absolute Deviation or Standard Deviation - Excel



- Mean Distance in both cases = 0
- Mean Absolute Deviation in both cases = 38.17
- Std Dev is 42.54 in the first case and 48.80 in the second.

Measuring Variability and Spread

$$\text{Variance} = \frac{\sum(x - \mu)^2}{n} = \frac{\sum x^2}{n} - \mu^2 \text{ (Derive)}$$

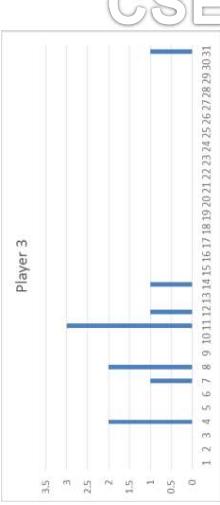
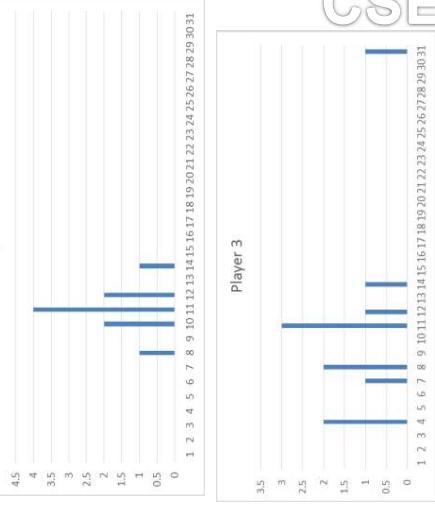
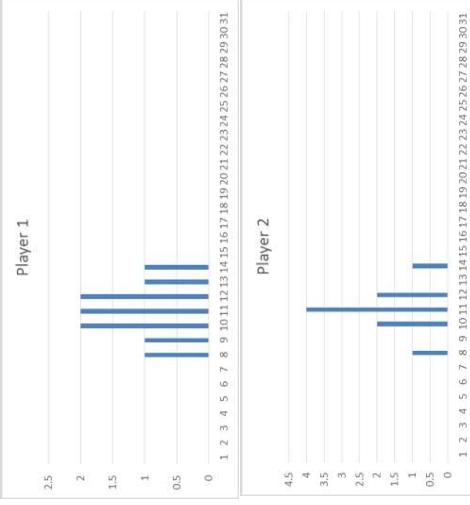
3 3 6 7 7 10 10 10 11 13 30

Units are squared, which is not intuitive.

Standard Deviation, $\sigma = \sqrt{\text{Variance}}$

Measuring Variability and Spread

Calculate standard deviation for each player.



1.73, 1.48, 7.02

Player 3 is the least reliable.

Measuring Variability and Spread

What happens to Standard Deviation if Good Heart Inc. gave all employees a Rs 2000 raise?

What happens to Standard Deviation if Good Heart Inc. gave all employees a 10% raise?

No change.

Increases by 1.1 times.

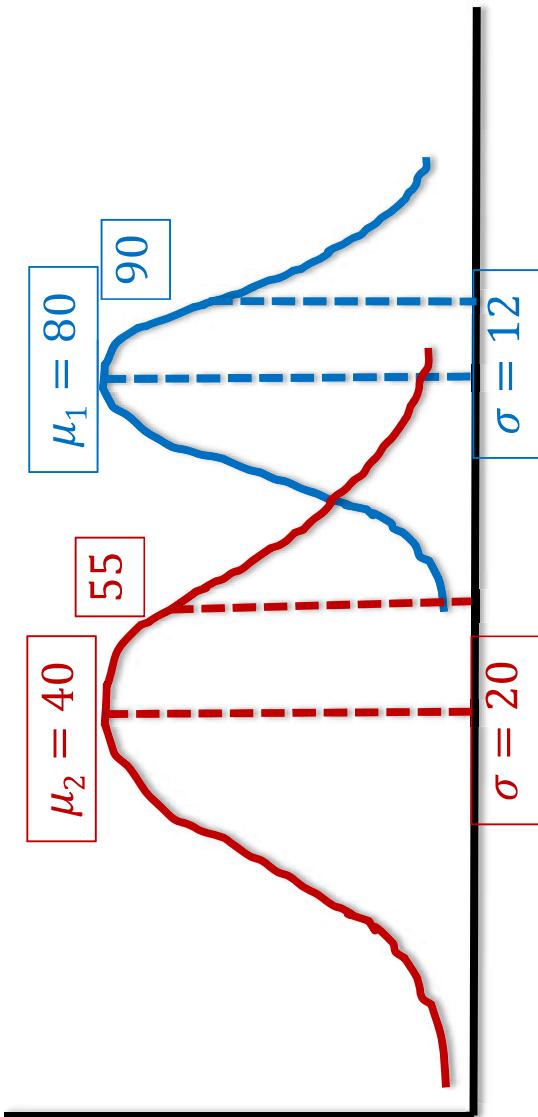
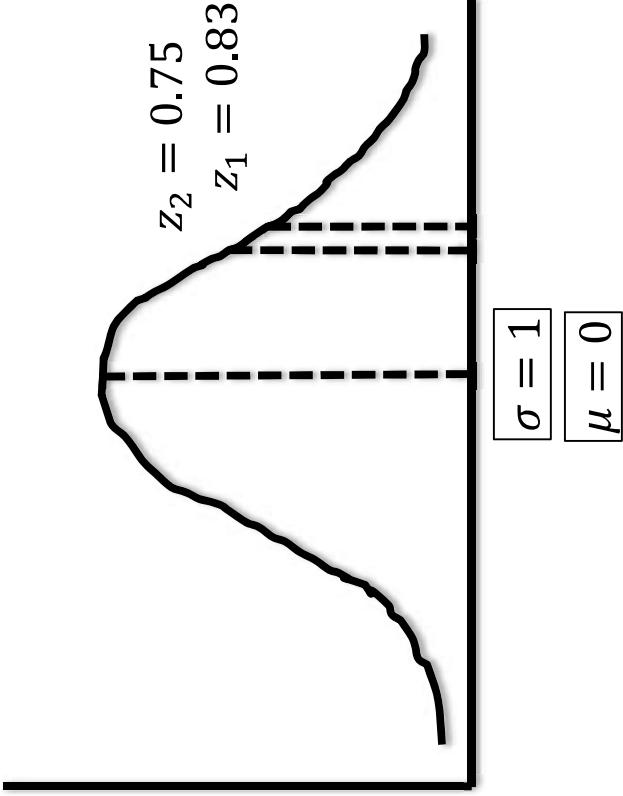
Measuring Variability and Spread

Imagine 2 players with different abilities: one has an average of 80% with 12% StdDev and the other 40% with 20% StdDev.

In a particular practice session, the first one scores 90% of the time and the second 55%. Who did best against their PERSONAL track record?

Measuring Variability and Spread

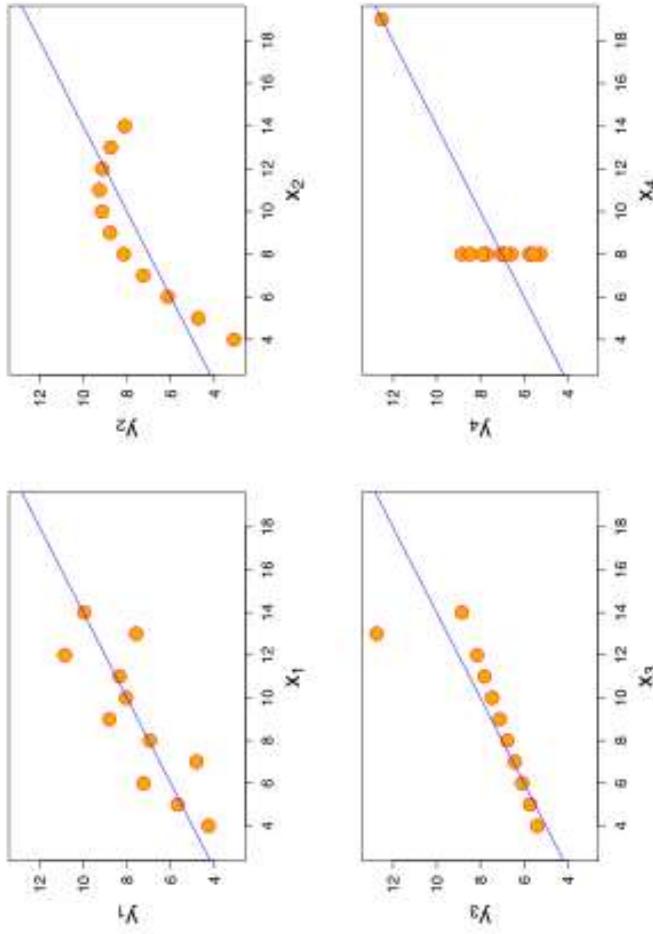
Standard score, $Z = \frac{x - \mu}{\sigma}$, # of std devs from the mean



Measuring Variability and Spread

Anscombe's quartet					
I	II	III	IV		
x	y	x	y	x	y
10	8.04	10	9.1	10	7.46
8	6.95	8	8.1	8	6.77
13	7.58	13	8.7	13	12.7
9	8.81	9	8.8	9	7.11
11	8.33	11	9.3	11	7.81
14	9.96	14	8.1	14	8.84
6	7.24	6	6.1	6	6.08
4	4.26	4	3.1	4	5.39
12	10.8	12	9.1	12	8.15
7	4.82	7	7.3	7	6.42
5	5.68	5	4.7	5	5.73

Property	Value
Mean of x in each case	9 (exact)
Sample variance of x in each case	11 (exact)
Mean of y in each case	7.50 (to 2 decimal places)
Sample variance of y in each case	4.122 or 4.127 (to 3 decimal places)
Correlation between x and y in each case	0.816 (to 3 decimal places)
Linear regression line in each case	$y = 3.00 + 0.500x$ (to 2 and 3 decimal places, respectively)



PROBABILITY BASICS

**Sholay**

The best place for students to learn Applied Engineering

66

<http://www.insofe.edu.in>

Probability vs Statistics

- Probability – Predict the likelihood of a future event
- Statistics – Analyze the past events
- Probability – What will happen in a given ideal world?
- Statistics – How ideal is the world?

Probability vs Statistics

Probability is the basis of inferential statistics.



Probability - Applications

8 National Vital Statistics Reports, Vol. 54, No. 14, April 19, 2006

Table 1. Life table for the total population: United States, 2003

Age	$q(x)$	$l(x)$	$d(x)$	Number dying between ages x to $x+1$	Person-years lived between ages x to $x+1$	Total number of person years lived above age x	Expectation of life at age x
						$\bar{l}(x)$	
0-1	0.006965	100,000	687	99,394	99,394	7,743,016	77.4
1-2	0.000469	99,313	47	99,290	99,290	7,643,622	77.0
2-3	0.000337	99,267	33	99,250	99,250	7,544,332	76.0
3-4	0.000254	99,233	25	99,221	99,221	7,445,082	75.0
4-5	0.000194	99,208	19	99,199	99,199	7,345,861	74.0
5-6	0.000177	99,189	18	99,180	99,180	7,246,663	73.1
6-7	0.000160	99,171	16	99,163	99,163	7,147,482	72.1

Insurance industry uses probabilities in actuarial tables for setting premiums and coverages.

Probability - Applications

- Gaming industry – Establish charges and payoffs
- HR – Does a company have biased hiring policies?
- Manufacturing/Aerospace – Prevent major breakdowns

Assigning Probabilities

Classical Method – *A priori* or Theoretical

Probability can be determined prior to conducting any experiment.

$$P(E) = \frac{\# \text{ of outcomes in which the event occurs}}{\text{total possible # of outcomes}}$$

Example: Tossing of a fair die



Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

Probability can be determined post conducting a thought experiment.

$$P(E) = \frac{\text{\# of times an event occurred}}{\text{total \# of opportunities for the event to have occurred}}$$

Example: Tossing of a weighted die...well!, even a fair die.

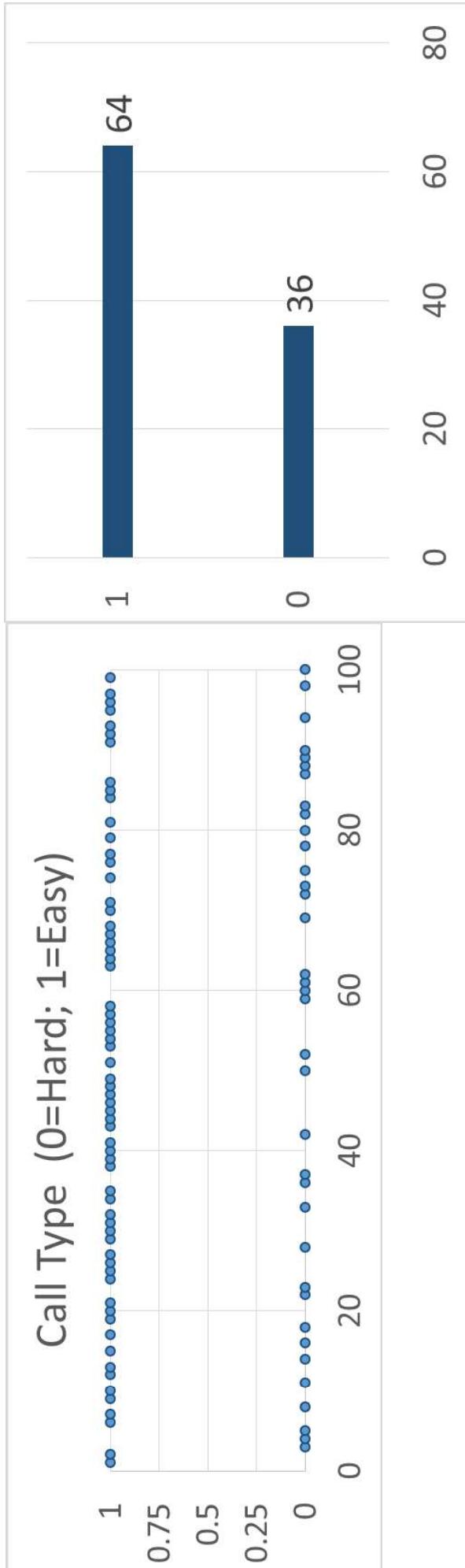
The larger the number of experiments, the better the approximation.

This is the most used method in statistical inference.

Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

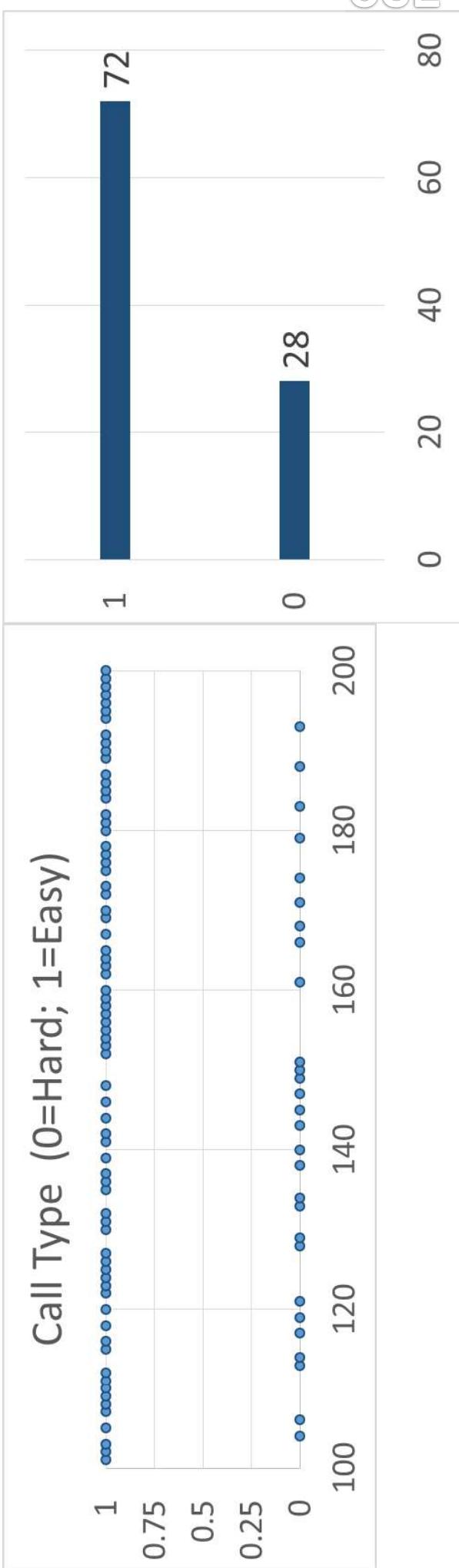
100 calls handled by an agent at a call centre



Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

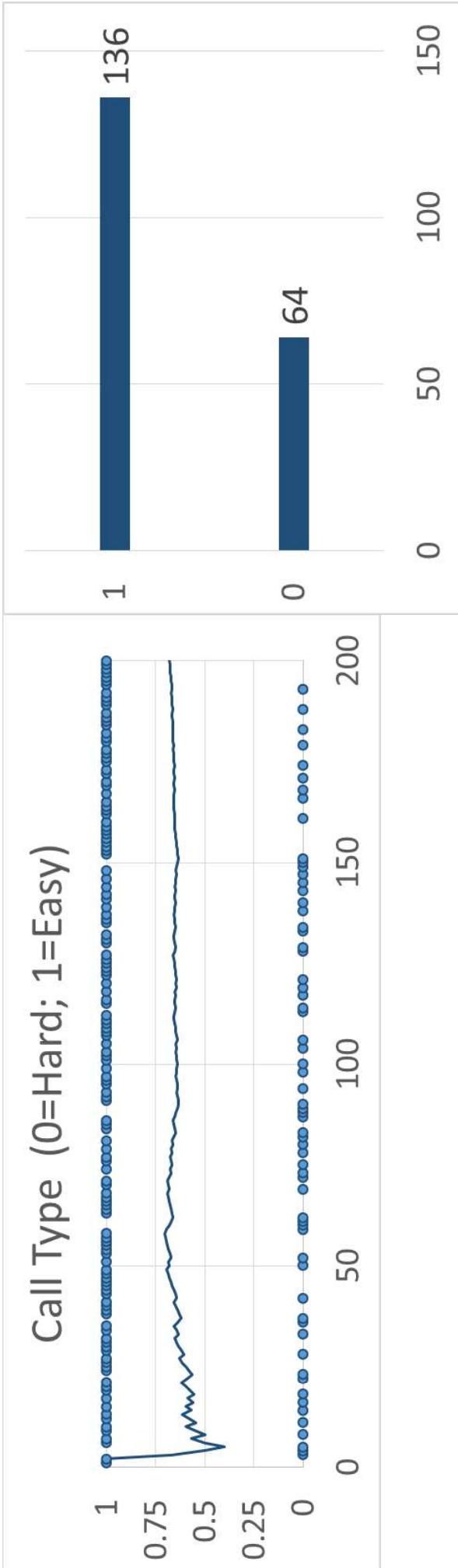
Next 100 calls handled by an agent at a call centre



Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

Averages over the long run



$$P(\text{easy}) = 0.7$$

Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

Probability of having a monthly income of 1000 BHD is

$$10/23 = 0.43$$

INCOME(BHD)	FREQUENCY
100	10
345	1
1000	10
9833	2

Assigning Probabilities

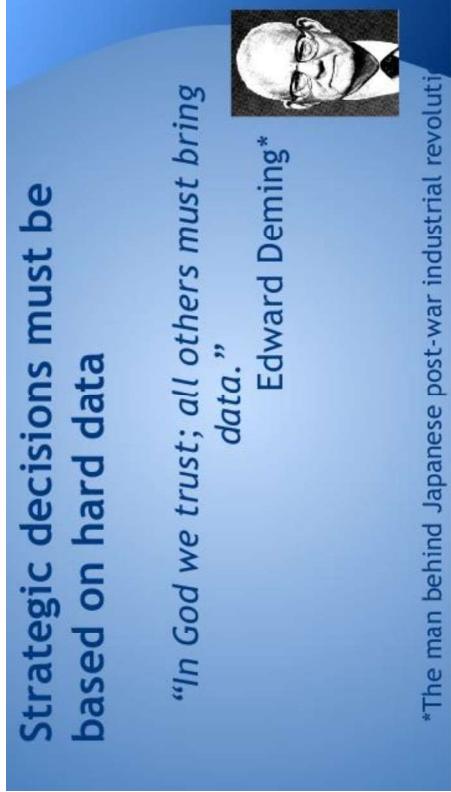
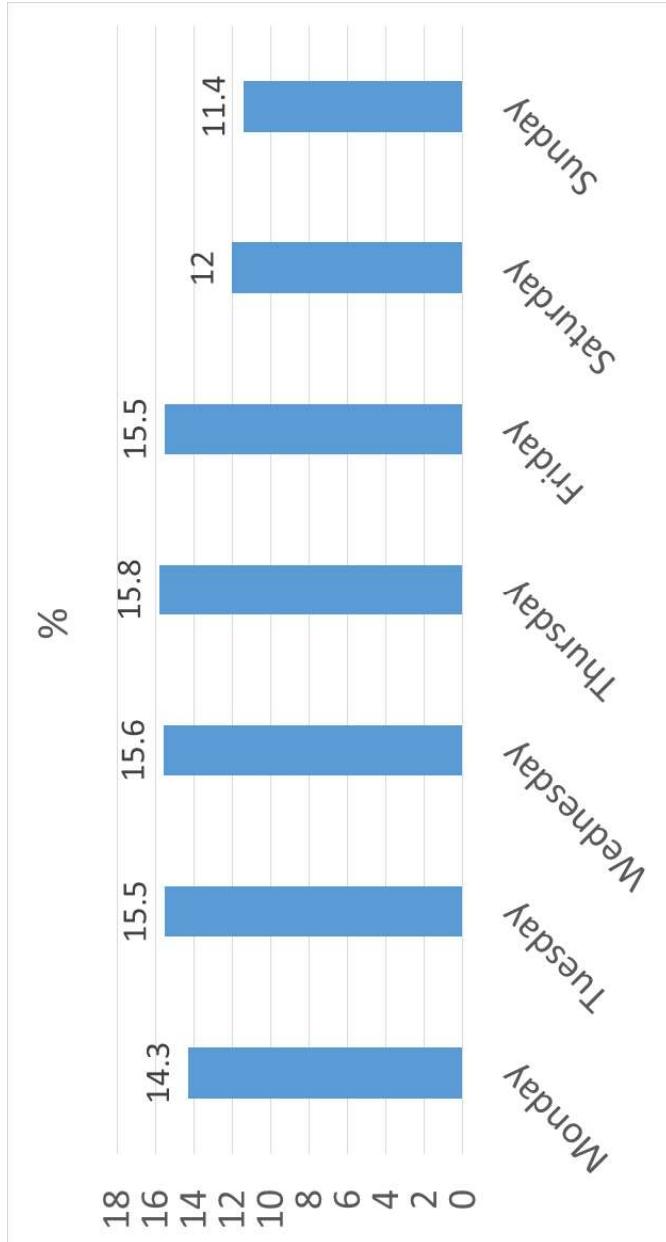
Subjective Method

Based on feelings, insights, knowledge, etc. of a person.

What is the probability of rain tomorrow?

Assigning Probabilities

What is the probability of a baby being born on a Sunday?

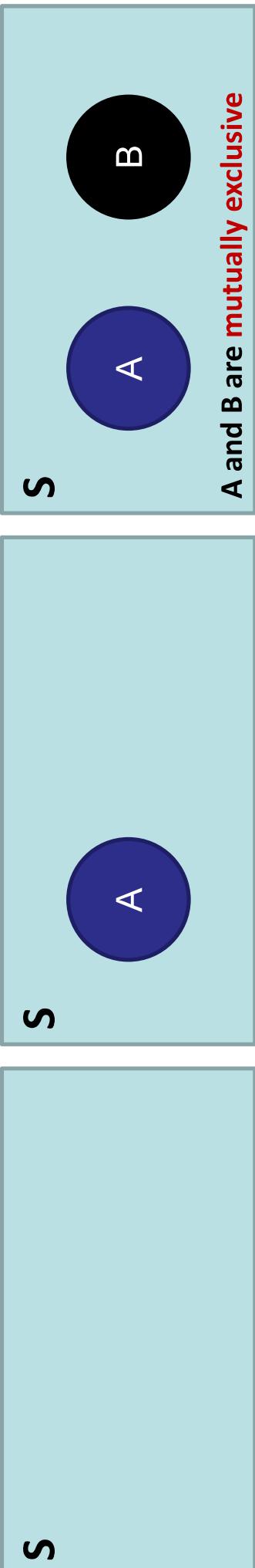


Data from "Risks of Stillbirth and Early Neonatal Death by Day of Week", by Zhong-Cheng Luo, Shiliang Liu, Russell Wilkins, and Michael S. Kramer, for the Fetal and Infant Health Study Group of the Canadian Perinatal Surveillance System. Data of 3,239,972 births in Canada between 1985 and 1998. The reported percentages do not add up to 100% due to rounding.

Probability - Terminology

- Sample Space – Set of all possible outcomes, denoted S .
- Event – A subset of the sample space.

Probability - Rules

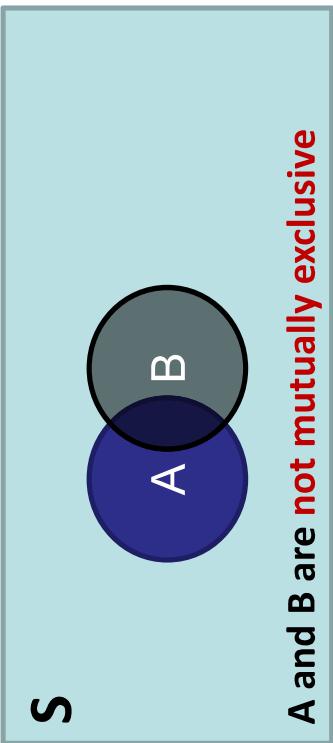


$$\begin{aligned}P(S) &= 1 & 0 \leq P(A) \leq 1 & P(A \text{ or } B) \\&&&= P(A) + P(B)\end{aligned}$$

Area of the rectangle denotes sample space, and since probability is associated with area, it cannot be negative.

Mutually Exclusive – If event A happens, event B cannot.

Probability - Rules



$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Example

Event A – Customers who default on loans

Event B – Customers who are High Net Worth Individuals

Probability - Rules

Independent Events – Outcome of event B is not dependent on the outcome of event A.

Probability of customer B defaulting on the loan is not dependent on default (or otherwise) by customer A.

$$P(A \text{ and } B) = P(A) * P(B)$$

If the probability of getting an *easy* call is 0.7, what is the probability that the next 3 calls will be *easy*?

$$P(easy_1 \text{ and } easy_2 \text{ and } easy_3) = 0.7^3 = 0.343$$

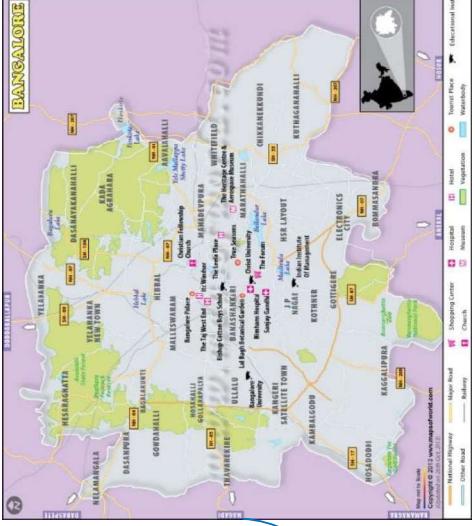
Probability - Question

A basketball team is down by 2 points with only a few seconds remaining in the game. Given that:

- Chance of making a 2-point shot to tie the game = 50%
- Chance of winning in overtime = 50%
- Chance of making a 3-point shot to win the game = 30%

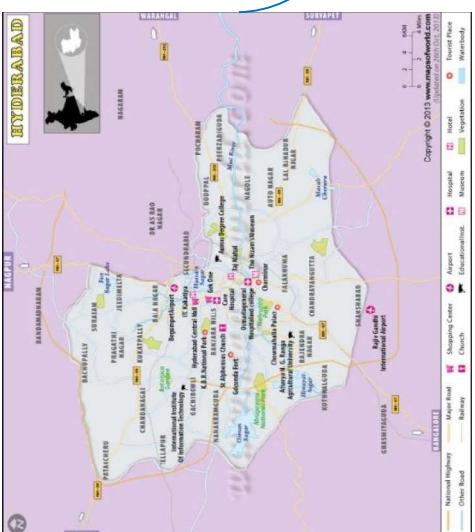


What should the coach do: go for 2-point or 3-point shot?
What are the assumptions, if any?



INSOFE

Inspire...Educate...Transform.



HYDERABAD

2nd Floor, Jyothi Imperial, Vamsiram Builders, Old Mumbai Highway, Gachibowli, Hyderabad - 500 032
+91-9701685511 (Individuals)
+91-9618483483 (Corporates)

BENGALURU

L77, 15th Cross Road, 3A Main Road, Sector 6,
HSR Layout, Bengaluru – 102
+91-9502334561 (Individuals)
+91-9502799088 (Corporates)

Social Media

Web: <http://www.insofe.edu.in>

Facebook: <https://www.facebook.com/insofe>

Twitter: <https://twitter.com/Insofeedu>

YouTube: <http://www.youtube.com/InsofeVideos>

SlideShare: <http://www.slideshare.net/INSOFE>

LinkedIn: <http://www.linkedin.com/company/international-school-of-engineering>

This presentation may contain references to findings of various reports available in the public domain. INSOFE makes no representation as to their accuracy or that the organization subscribes to those findings.