



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

December 09, 2017

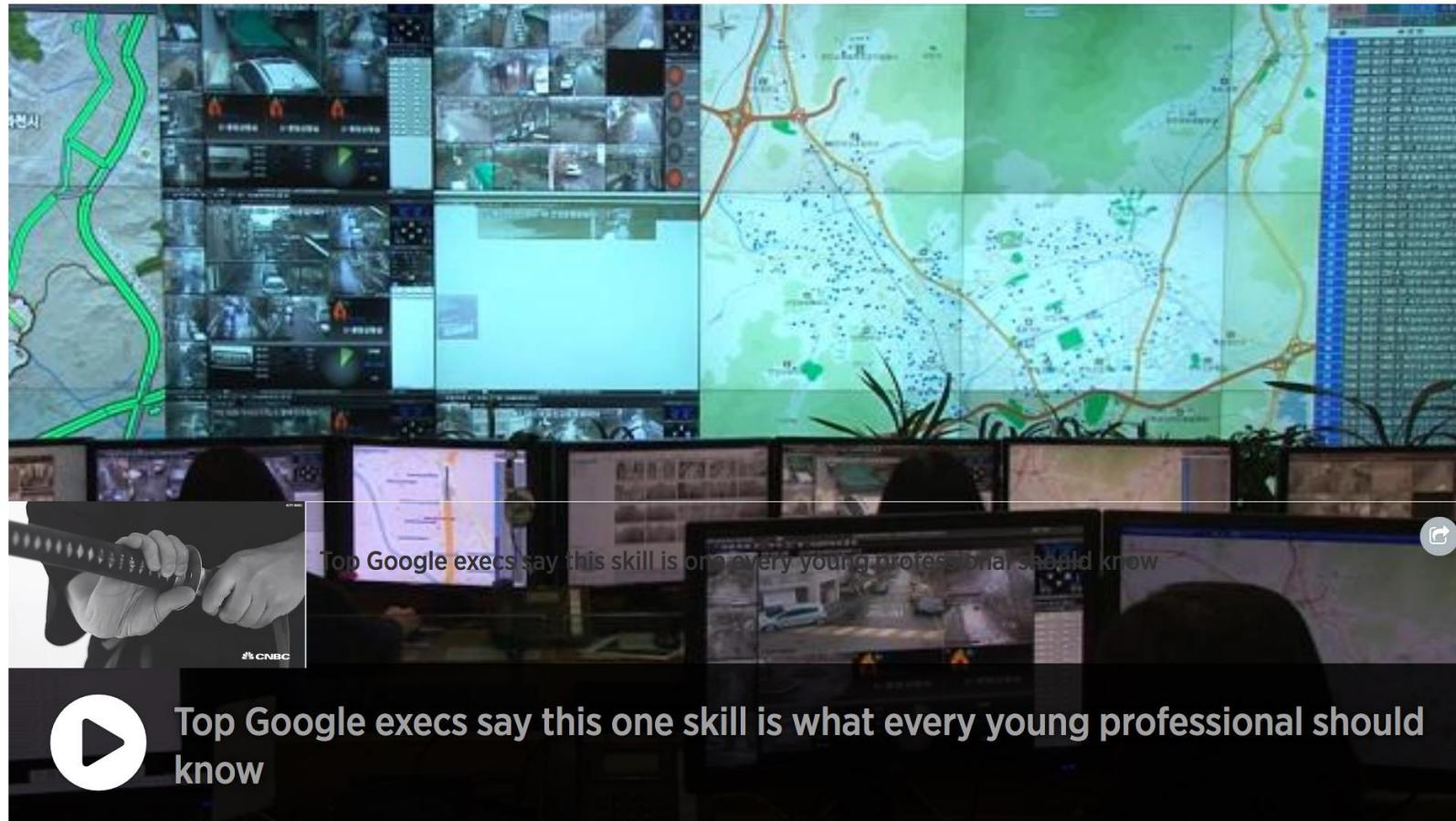
CSE 7315C

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



Source: <http://www.cnbc.com/2017/03/31/google-executives-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

**Digital
THE ELECTION
CENTRE**

**DIGITAL
MEDIA IN**

**NEW AGE
ELECTIONS**

NEW AGE POLL TOOLS: DATA, TECHNOLOGY

PUN (117 / 117)
Majority-59
Poll of Exit Polls



Akali +
BJP

10



CONG

54



AAP +

52



#ResultsWithNDTV

RIES

RAJNATH ON U.S. HATE CRIMES

**NDTV
24X7
in association with
AMITY
UNIVERSITY**

**INTERNATIONAL SCHOOL OF ENGINEERING
INSOFE**

CROSSWIND

14 KMPH

14 CMS



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 %

LIAR LIAR
PANTS ON FIRE!



Problem #1: Data Gathering

Schedule Reference	Parliamentary Constituency			
	SI.	PC No.	PC Name	Type
Schedule no:	7	1	Daman & Diu	GEN
No of PCs going to poll	1			
Issue of Notification:		02 Apr 14 (Wed)		
Last Date for filing Nominations:		09 Apr 14 (Wed)		
Scrutiny of Nominations:		10 Apr 14 (Thu)		
Last date for withdrawal of Candidature:		12 Apr 14 (Sat)		
Date of Poll	30 Apr 14 (Wed)			
Counting of Votes:	16 May 14 (Fri)			
Date before which the election shall be completed		28 May 14 (Wed)		

Source: http://eci.nic.in/eci_main1/GE2014/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).

CESE 7315C

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

Last accessed: April 27, 2014

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 #2: Data Understanding

The ratios reflect the ratios of registered voters.



State/UT	Registered Voters			Voted in 2014 General Elections		
	Male	Female	% Female	Male	Female	% Female
Puducherry	432048	469309	52.07	351360	388657	52.52
Kerala	11734258	12592391	51.76	8678185	9297708	51.72
Manipur	871431	902894	50.89	685427	727210	51.48
Mizoram	346219	355951	50.69	216167	217034	50.1
Daman & Diu	57011	54816	49.02	42378	44855	51.42
Meghalaya	777639	789602	50.38	524774	553284	51.32
Goa	528308	532469	50.2	395766	421234	51.56
Arunachal Pradesh	379627	379760	50.01	289291	307665	51.54

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/articledetailpage.aspx?id=474880>;

Last accessed: April 27, 2014

CSE 7315C



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 #2: Data Analysis/Interpretation

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

State/UT	Male			Female			Male	Female	Female-Male			
	2006-08	2011-13	2014	2006-08	2011-13	2014			2006-08	2011-13	2014	Sparklines
Puducherry	84.48	83.97	81.32	86.29	86.97	82.81			1.81	3	1.49	
Kerala	73.17	75.08	73.96	71.08	74.78	73.84			-2.09	-0.3	-0.12	
Manipur	85.88	76.94	78.66	86.82	81.36	80.54			0.94	4.42	1.88	
Mizoram	78.77	80.3	62.44	81.24	82.2	60.97			2.47	1.9	-1.47	
Daman & Diu				74.33			81.83					7.5
Meghalaya	88.62	85.17	67.48	89.36	88.44	70.07			0.74	3.27	2.59	
Goa	69.7	78.86	74.91	70.3	84.57	79.11			0.6	5.71	4.2	
Arunachal Pradesh				76.2			81.02					4.82

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

CSE 7315C

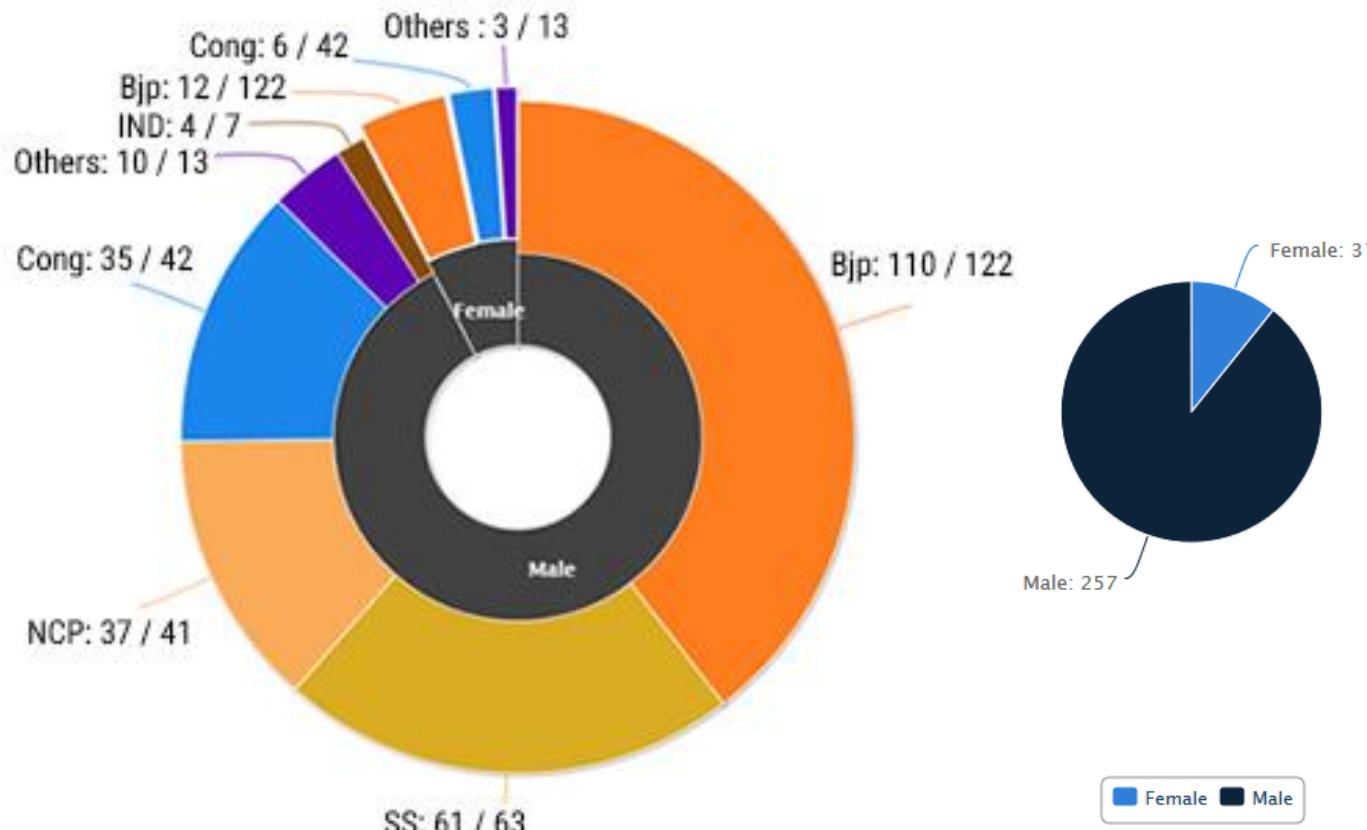
Source: <http://epaper.deccanchronicle.com/articledetailpage.aspx?id=474880>;

Last accessed: April 27, 2014

Problem #4: Data Presentation

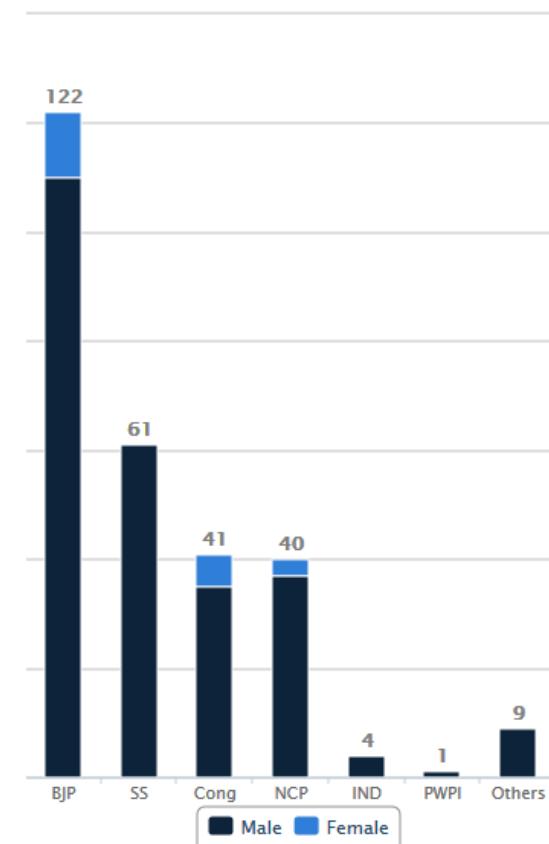
Maharashtra: Gender Break-up

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 Chief Functionary Mr Vinod

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 counseling. 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 attributed to only enforcement or improvement in rule compliance.



Source: <http://epaper.deccanchronicle.com/articledetailpage.aspx?id=6941170>
Last accessed: December 09, 2016

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

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-

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.

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

Last accessed: April 27, 2014

CSE
7315C



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 within 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.

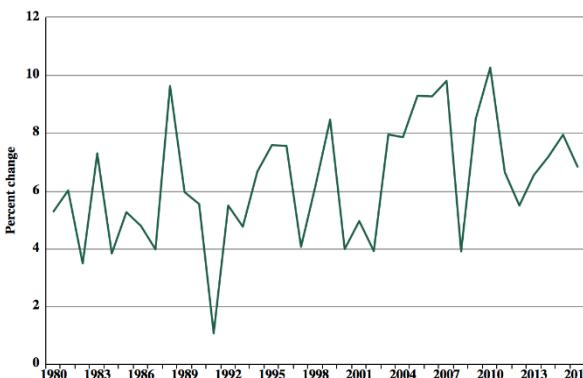
Before this data revision, it was a dominant view that India had seen worst growth in 2013-14.

Some economists said that the latest GDP numbers released on Friday does not reflect the ground reality and people will now have to adjust that 7 per cent plus growth under the new methodology is not the same what it was based on older formula.



While the GDP has grown at a fast pace the corporate earnings are dismal, industrial activity is weak, private investment is yet to start and bank credit uptake is slow. Rating agency, India ratings, principal economist and director (public finance) Sunil Kumar Sinha said that the people need to recognise that methodology to calculate GDP has undergone a change and it is no longer estimated at factor cost but at market cost.

HIGHLIGHTS



Date	Value	Change, %
2016	6.8	-13.94 %
2015	7.9	10.52 %
2014	7.2	9.77 %
2013	6.5	19.26 %
2012	5.5	-17.38 %
2011	6.6	-35.30 %
2010	10.3	20.99 %
2009	8.5	117.94 %
2008	3.9	-60.30 %
2007	9.8	5.80 %
2006	9.3	-0.25 %
2005	9.3	

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. — PTI

ROOM F UP IF O

Mumbai, J
Room for m
cuts in Ind
open up if n
rains are bet
expected or if
ernment take
to prevent foo
from rising
global oil
remain low, s
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, reported the *International Business Times*.

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

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.



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, consistent with the



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.

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" 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

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.

CSE
7315C

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

Last accessed: March 13, 2015



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

CSE 7315C



<u>TEST NAME</u>	<u>RESULT</u>	<u>BIOLOGICAL REFERENCE RANGE</u>
GLUCOSE - SERUM / PLASMA (FASTING) (Hexokinase)	86	Non-Diabetic : <= 100 Impaired : 101 - 125 Diabetic : >= 126 Male : 16 - 63 U/L Female : 14 - 59 U/L
ALT(SGPT) - SERUM / PLASMA (UV,Using Pyridoxal Phosphate:IFCC)	95 *	Desirable : < 200 Borderline high : 200 High Risk : >=240
CHOLESTEROL - SERUM / PLASMA (Cholesterol Oxidase / Esterase)	176.00	

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 MHA, 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 overdrive, drawing up contingency plans till district level to deal with a drought, worried at a spiralling 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.

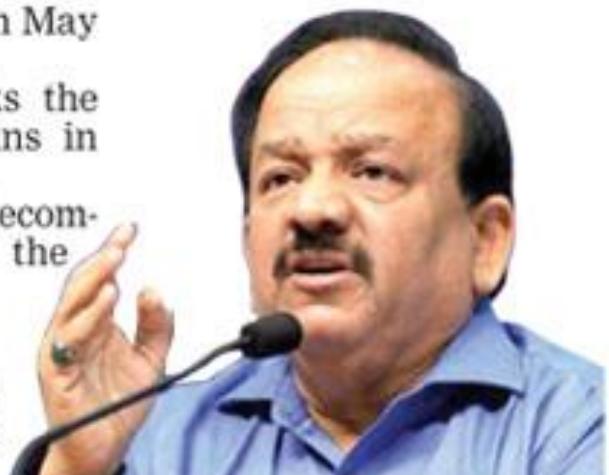
"Conditions are becoming favourable for the

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



CSE 7315C

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.

Providing some more



File photo showing citizens getting drenched due to heavy showers.

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,

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

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

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



■ NARROWER nasal passages help to increase the moisture content of air and warm it — a bonus for those in higher latitudes

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

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

Pvt research plays down risk from mobiles: Study

DurgeshNandan.Jha
@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 underestimate the risk. Government funded studies show in-

PHONE TROUBLE

- 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 Kameshwar Prasad, head of neurology at AIIMS, who is lead author of the study.

► Brain tumour risk, P 12

HUGE CONTROVERSY

CSO asked to explain GDP growth estimates

■ Data has been overestimated post note ban, say experts

PAWAN BALI | DC
NEW DELHI, MARCH 15

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

Central Statistics Office (CSO) in its advance estimate has pegged GDP growth for 2016-17 at 7.1 per cent and it 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.

"The committee would 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," said the committee in its

report submitted in the Parliament.

Many experts 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 arguing that the note ban would have short-term impact on the Indian economy.

The committee would 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



national accounts with 2011-12 as the base year 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.

CSE 7315C



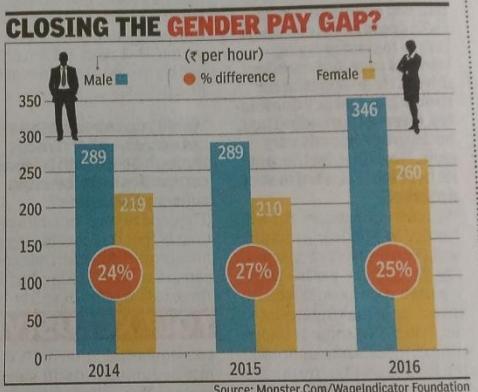
Women earned 25% less than men in 2016: Study

TIMES NEWS NETWORK

Mumbai: The average gender pay gap in India may have narrowed by two percentage points since last year, but it still stands at a yawning 25%. According to Monster Salary Index (MSI) data from 2016, men earned a median gross hourly salary of Rs 346, while women earned only Rs 260 in 2016. While it is an improvement over last year's over 27%, the pay gap is a notch higher than 24% in 2014. In the manufacturing sector, the average gender pay gap is as high as 30%.

Sanjay Modi, managing director (APAC & Middle East), Monster.com, said, "In India, the gender pay gap story holds true and the overall gap across India Inc is at 25%. This primarily is a manifestation of the underlying diversity challenges that organisations currently face."

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



men feel that gender parity need not be a top priority for their organisations. On the other hand, almost 69% expressed that even if gender parity is a priority, the management "does not walk the talk", indicating the need for India Inc to step up and implement pragmatic policies to bridge the gap.
"Unexamined conventions of women's commit-

Half of all pregnant women are anaemic

►Continued from P 1

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



HUNGER CRISIS: Overall, 53% of women and 23% of men in the 15-49 age group were anaemic

those in the south, Haryana and Gujarat have slightly better numbers but are still at unacceptable levels. In Tamil Nadu, 51% children are anaemic while in Kerala it is over one-third. In many states, stunting has declined but the share of severely wasted children has increased. These are clear signs of an endemic crisis of hunger in the country that policy makers don't appear to be addressing.

How hand sanitisers can harm children

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 alcohol 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 sanitisers occurred in children aged 6-12 years. During 2011-2014, 70,669 hand sanitiser exposures in 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

'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

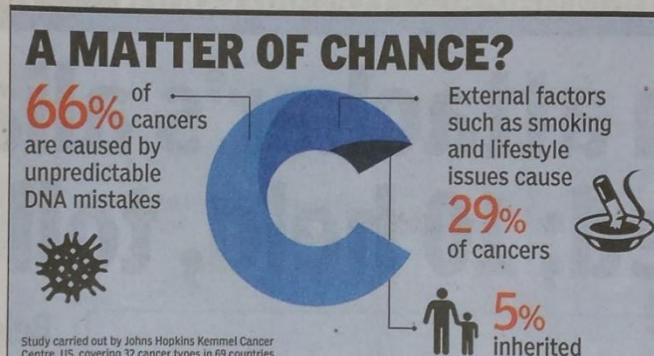
Why does cancer strike some people and not others? New research shows that random changes or 'mistakes' in DNA when cells are dividing cause nearly two-third of all cancers in humans. These changes are neither caused by external factors like smoking or exposure to harmful chemicals, nor by hereditary 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.

Lung cancer is most likely to be caused by environmental factors, mostly smoking. About 65% of all the mutations are due to



smoke and 35% due to DNA copying errors. Inherited factors have negligible role. The study involved a statistical analysis of cancer data from 69 countries including India, representing 4.8 billion people, 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 reviewed journal Science on March 24.

Human bodies grow by constant division of cells, starting from the first cell formed by fusion of the male sperm with the female egg. Every time a cell divides into two, the genetic code carrying DNA is copied. What the scientists are saying is that mistakes occur in this copying process that accumulate over time and ultimately cause cancer. "These copying mistakes are a potent source of cancer mutations that historically have been scientifically undervalued, and this new work provides the first estimate of the fraction of mutations caused by these mistakes," said the paper's lead author Cristian Tomasetti.

The researchers studied all 32 cancer types and estimated that

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.

CSE 7315C



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

CSE 7315C



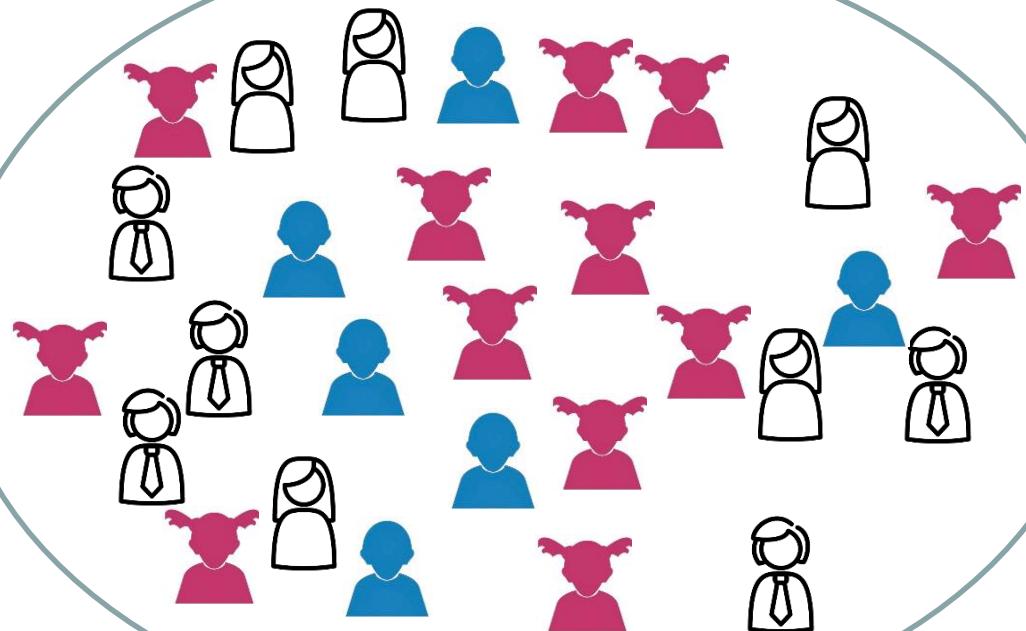
BASIC STATISTICAL TERMINOLOGY

CSE 7315C



Population and Sample

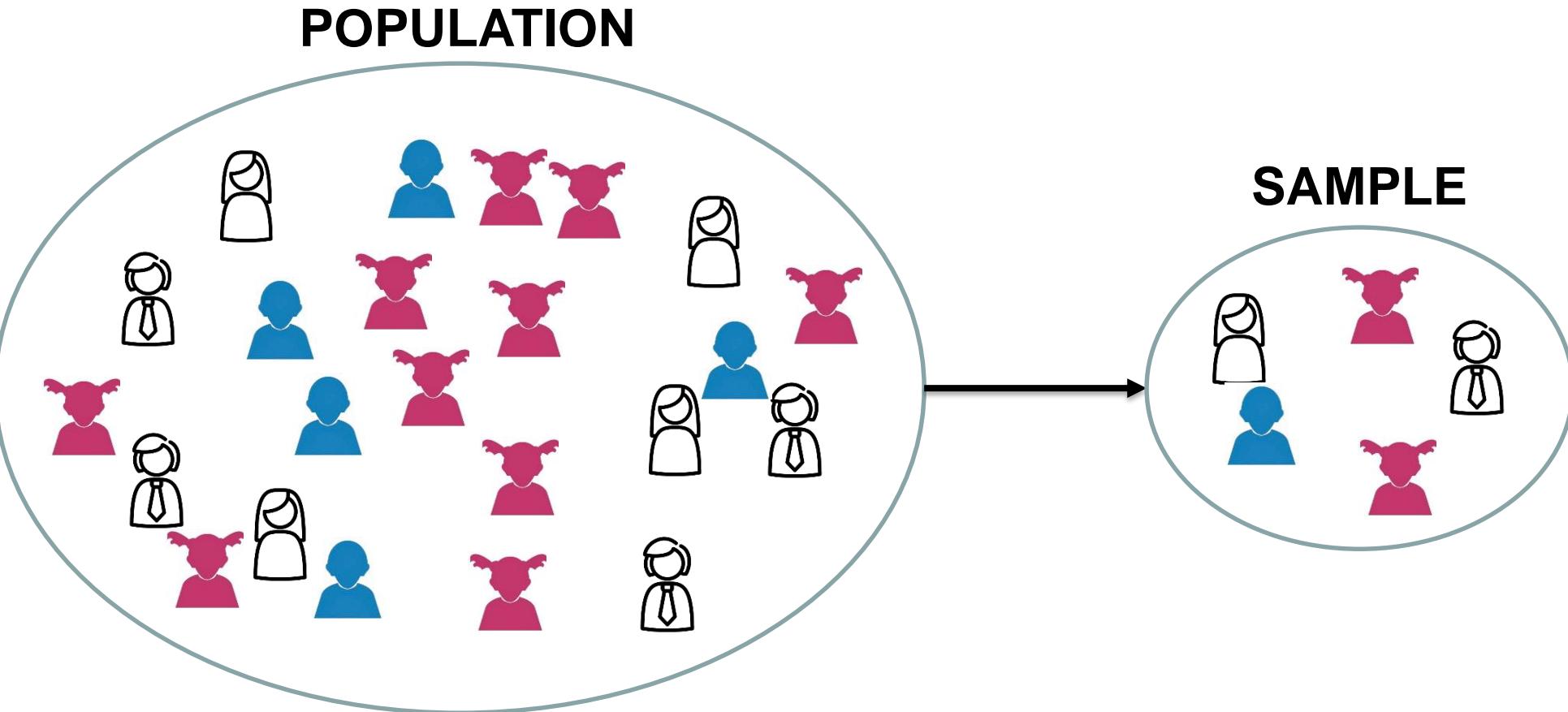
POPULATION



CSE 7315C



Population and Sample



CSE 7315C





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.
<u>Chelseae Ramos</u>	5'2"	Madison, WI	Jr.
<u>Sydney Shelton</u>	4'10"	Scottsville, KY	Jr.
Ashley Wettstain	5'0"	Owensboro, KY	Fr.
<u>Madison Yee</u>	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	<u>Jennifer O'Neill</u>	PG	SR-3L	5-6	Bronx, N.Y./Saint Michael Academy
2	<u>Ivana Jakubcova</u>	C	JR-JC	6-6	Bratislava, Slovakia//Murray State College
3	<u>Janee Thompson</u>	PG	JR-2L	5-7	Chicago, Ill./Whitney Young
5	<u>Kywin Goodin-Rogers</u>	F	SO-HS	6-1	Lebanon, Ky./Marion Co.
12	<u>Jelleah Sidney</u>	F/C	SR-2L	6-2	Queens Village, N.Y./Saint Michael Academy/Chipola JC
13	<u>Bria Goss</u>	G	SR-3L	5-10	Indianapolis, Ind./Ben Davis
15	<u>Linnae Harper</u>	G	SO-1L	5-8	Chicago, Ill./Whitney Young
24	<u>Jaycee Coe</u>	G	FR-HS	5-11	Gainesboro, Tenn./Jackson Co.
25	<u>Makayla Epps</u>	G	SO-1L	5-10	Lebanon, Ky./Marion Co.
35	<u>Alexis Jennings</u>	F/C	FR-HS	6-2	Madison, Ala./Sparkman
45	<u>Alyssa Rice</u>	C	FR-HS	6-3	Reynoldsburg, Ohio/Reynoldsburg
50	<u>Azia Bishop</u>	F/C	SR-3L	6-3	Toledo, Ohio/Start

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

Last accessed: October 7, 2014

Source: <http://www.dailymail.co.uk/news/article-2742468/Tall-small-s-basketball-Ladies-Kentucky-Wildcats-team-tower-cheerleaders.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.

Census and Survey

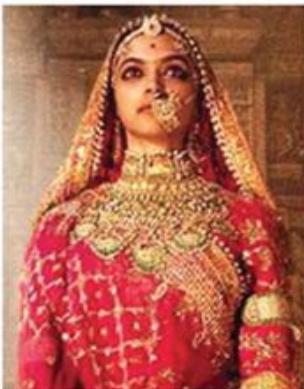
‘Majority of people interested in ‘Padmavati’ despite dispute

ANUSHA PUPPALA | DC
HYDERABAD, DEC. 7

A recent study reflects that the controversy over the film ‘Padmavati’ is driven by caste considerations led by caste groups. The study by Chrome Data Analytics was conducted in Delhi, Mumbai, Kolkata, Chennai, Bengaluru and Hyderabad, among adult respondents. The study focussed on freedom of expression.

So how many people are interested in watching ‘Padmavati’ the film after all the acrimony over it? Research found that 81 per cent of the respondents in the metros plan to watch the movie despite the controversies surrounding it.

The survey seeks to analyse public insight around the controversy, their views and if artistic freedom should be bereft of religious sentiments. The sample-size constituted 53 per cent of women and 47 per cent men. About 29 per cent of the people said that film makers should not think about religious senti-



Research found that 81 per cent of the respondents in metros plan to watch ‘Padmavati’ despite controversies. —DC

ments or cultural sensitivity while making a movie. However, 71 per cent respondents felt freedom of expression should be respectful of religious sentiments.

The film was initially scheduled for release on December 1, but that has now been voluntarily deferred by the producers. Director of the film Sanjay Leela Bhansali is again

planning to postpone the release date for multiple reasons.

Amidst of all the controversies, BJP MLA from Goshamahal, T Raja Singh, had even requested Telangana CM K. Chandrasekhar Rao to ban the screening of the film till the film-maker makes changes protestors have asked for. States of Rajasthan, MP and UP have already imposed a ban on the film. In a letter to Mr Rao, Mr Singh said the film hurts the sentiments of Hindus by manipulating history of Rani Padmavati. MLA Raja Singh said, “Rani Padmavatiji was from the Rajput community, and Rajputs treat her as a goddess and respect her immensely.”

On the other side of the divide, Diana Monteiro, director and counseling psychologist at Hyderabad Academy of Psychology says, “Everyone must get to express themselves and films aren’t real life. We must realise that before we get offended by films.”

Source: <http://epaper.deccanchronicle.com/articledetailpage.aspx?id=9564134>

Last accessed: December 08, 2017

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

I AM INDEBTED TO
MY FATHER FOR
LIVING, BUT
TO MY
TEACHER FOR
LIVING WELL.



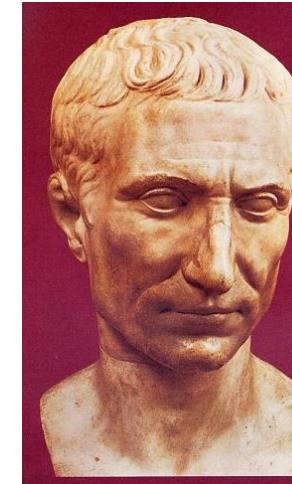
Alexander the Great
www.quote-coyote.com

Greek – Population Parameter

Mean – μ

Variance – σ^2

Standard Deviation - σ



*"What we wish,
we readily
believe, and
what we
ourselves think,
we imagine
others think
also."*

Julius Caesar

Roman – Sample Statistic

Mean – \bar{x}

Variance – s^2

Standard Deviation - s

CSE 7315C



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.searo.who.int/PDS_DOCS/B4793.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* 54.12 (2011): 3022-7. NCBI. Web. March 2013.

Source:

http://www.arogyaworld.org/wp-content/uploads/2010/10/ArogyaWorld_IndiaDiabetes_FactSheets_CGI2013_web.pdf

Last accessed: November 25, 2015

Variables and Data

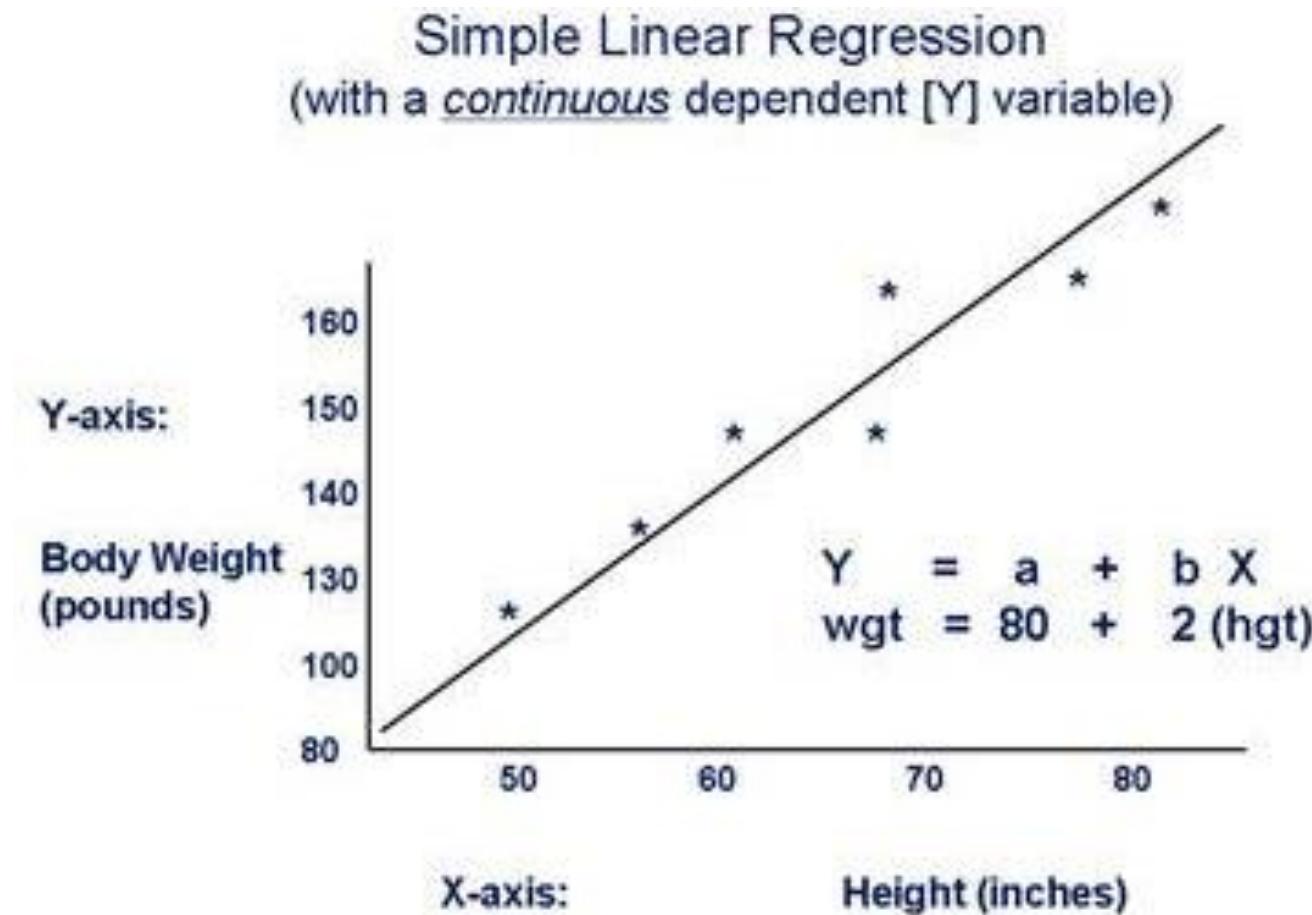
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
58	management	married	tertiary	no	2143	yes	no	unknown	5	may	261	1	-1	0	unknown	no
44	technician	single	secondary	no	29	yes	no	unknown	5	may	151	1	-1	0	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	yes	unknown	5	may	181	1	-1	0	unknown	no

CSE 7315C



Variables – Dependent and Independent

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



CSE 7315C



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

CSE 7315C



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.

CSE 7315C



Numeric Data (Quantitative)

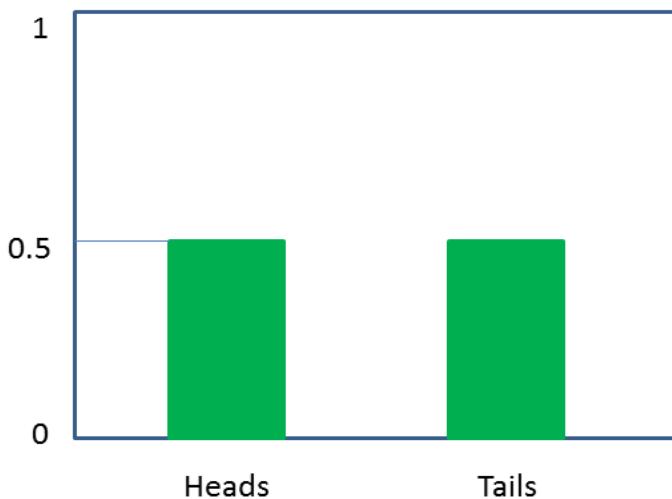
Examples

- 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

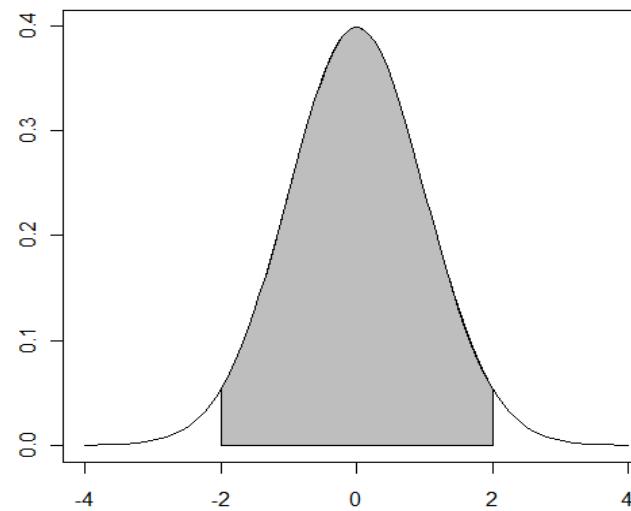
CSE 7315C



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

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

CSE 7315C

Source: <http://www.salaryexplorer.com/salary-survey.php?loc=17&loctype=1>

Last accessed: May 17, 2016



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 1,200 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 public sector companies show a totally different picture with their chiefs getting salaries of just about 3-4 times of their median employee remunerations.

■ 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-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 00:45 IST

Ramco HRIS Solutions

Cloud Based HRIS Solution With Role Based Workspaces. Ask For Demo! www.ramcos.com/HRIS

Ads by Google



RELATED NEWS

- RIL rejigs investment portfolio, sells blue chip stocks
- Indra Nooyi: A test of RIL's core competence
- Ambani company gets loan restructuring package
- RIL margins to rise to six-year high: CLSA
- Refining drives RIL's Q4 performance

Top 10 Mutual Funds
Buy Mutual Funds online. Select from 5000+ schemes. Open Free A/c now.
www.myuniverse.co.in/ZipSip

Regus™ Office Solutions
No Hidden Costs & Flexible Options. Workspaces To Suit. Get A Quote Now.
www.regus.co.in

Ads by Google

CHEQUES & BALANCES

Head honcho	Ratio to median employee salary in the company
Y C Deveshwar, executive chairman, ITC	439
Navin Agarwal, chairman, Vedanta	293
Mukesh Ambani, chairman, RIL	205
Aditya Puri, MD, HDFC Bank	117
Vishal Sikka, CEO, Infosys	116
Chanda Kochhar, CEO, ICICI Bank	97
Sanjiv Mehta, CEO, HUL	93
Azim Premji, CMD, Wipro	89
Shikha Sharma, MD & CEO, Axis Bank	74
Deepak Parekh, chairman, HDFC	19

■ Ambani has kept his salary capped at Rs 15 crore for seven years now, while the median remuneration of employees increased by 3.71 per cent to Rs 7.29 lakh during 2014-15. The total remuneration of key managerial 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 median employee remuneration. The overall key managerial personnel remuneration rose 20 per cent. Deveshwar's gross remuneration in 2014-15 stood at over Rs 15 crore, but net pay was lower at Rs 7.3 crore.

■ Premji saw his pay decline by 53 per cent to Rs 4.78 crore, while median employee remuneration rose by 9.5 per cent. Wipro CEO T K Kurian got a package that was 170 times the

The Central Tendencies

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

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

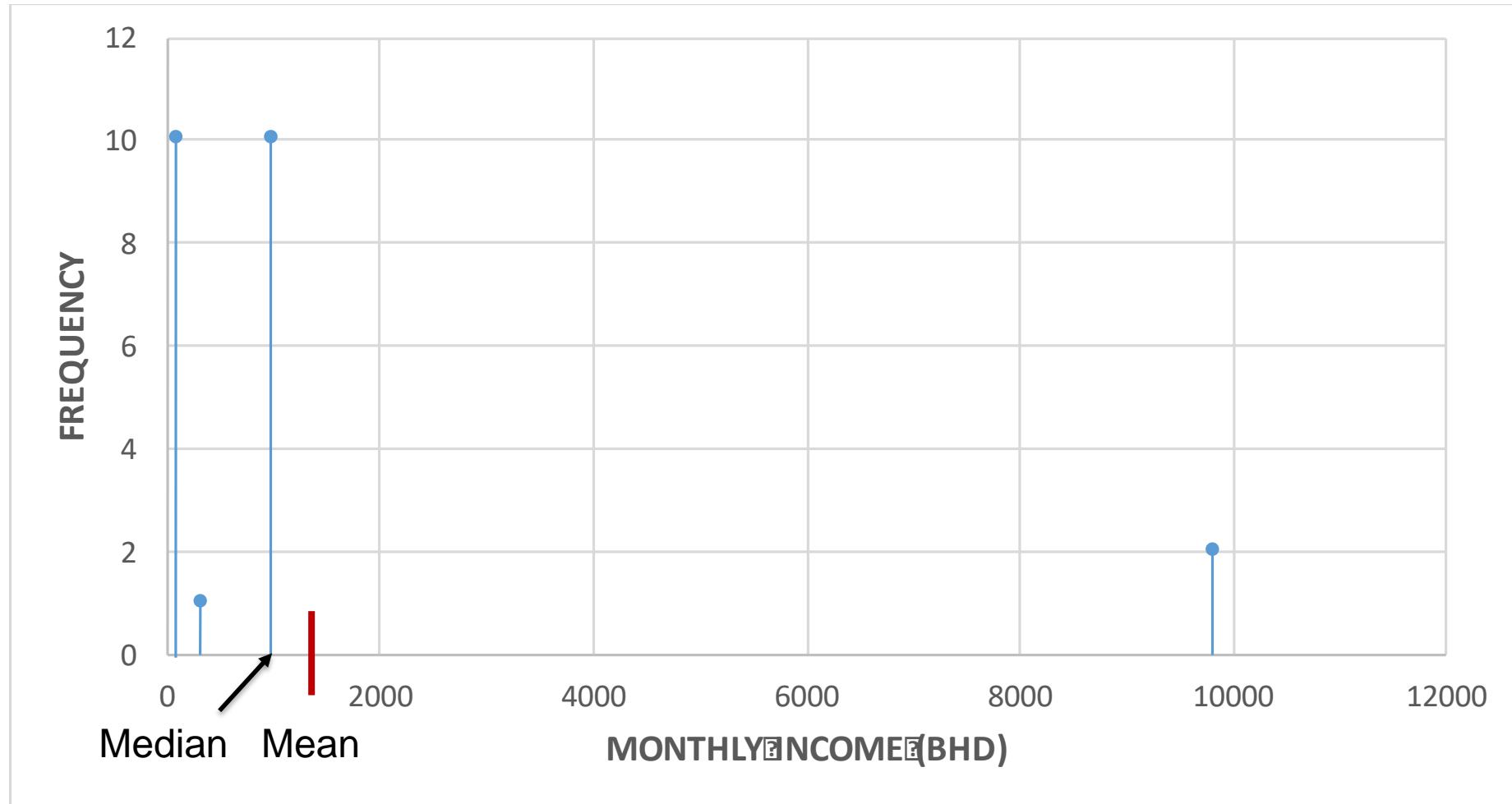
CSE 7315C

Source: http://www.business-standard.com/article/companies/ambani-gets-205-times-ril-s-median-pay-115070500340_1.html
Last accessed: July 7, 2015



The Central Tendencies - Excel

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



CSE 7315C



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?

CSE 7315C



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

Range = Max - 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

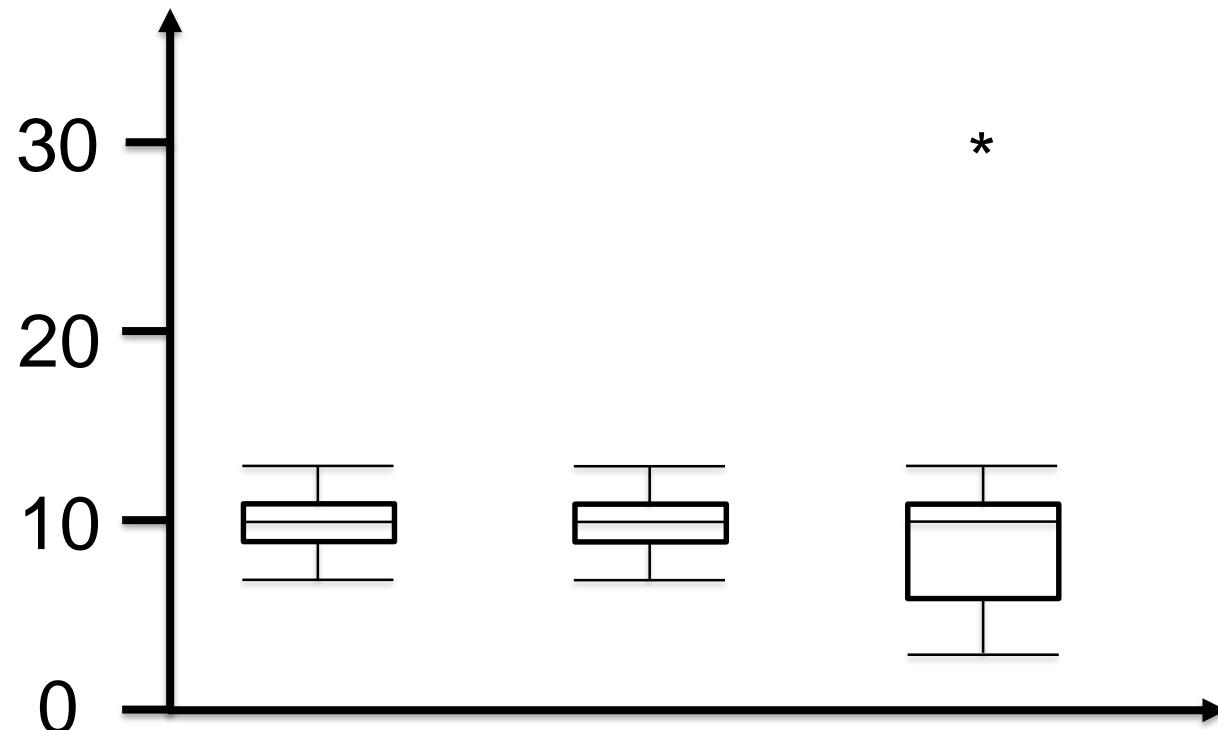
Upper quartile (75th percentile, Q3) = $\frac{3*(n+1)}{4}$ th

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

Measuring Variability and Spread

Exclude outliers scientifically – Quartiles

Box and whisker diagram or Box plot



CSE 7315C

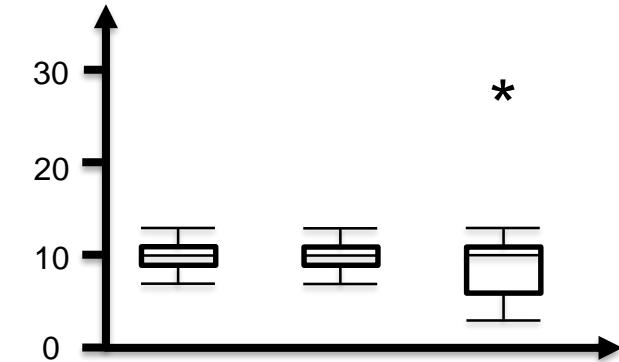


Measuring Variability and Spread

Exclude outliers scientifically – Quartiles

Box and whisker diagram or Box plot

Tukey fences



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	$1.5 \times \text{H-Spread}$ ($1.5 \times \text{IQR}$)	3	3	7.5
Upper Inner Fence	Upper Hinge + 1 Step (75th percentile + $1.5 \times \text{IQR}$)	14	14	18.5
Lower Inner Fence	Lower Hinge - 1 Step (25th percentile - $1.5 \times \text{IQR}$)	6	6	-1.5
Upper Outer Fence	Upper Hinge + 2 Steps (75th percentile + $3 \times \text{IQR}$)	17	17	26
Lower Outer Fence	Lower Hinge - 2 Steps (25th percentile - $3 \times \text{IQR}$)	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			30

CSE 7315C



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/-0YwIRjLMWr0/T4DqOwVClgI/AAAAAAAAGg/Yjf-ttkQLSg/s1600/fishy.jpg>

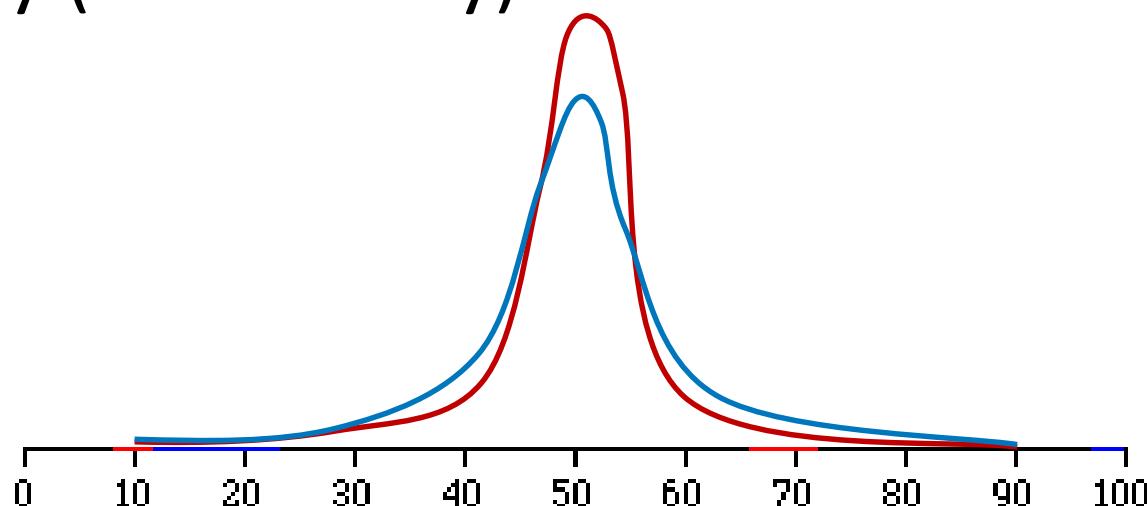
Last accessed: November 01, 2014

CSE 7315C



Measuring Variability and Spread

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



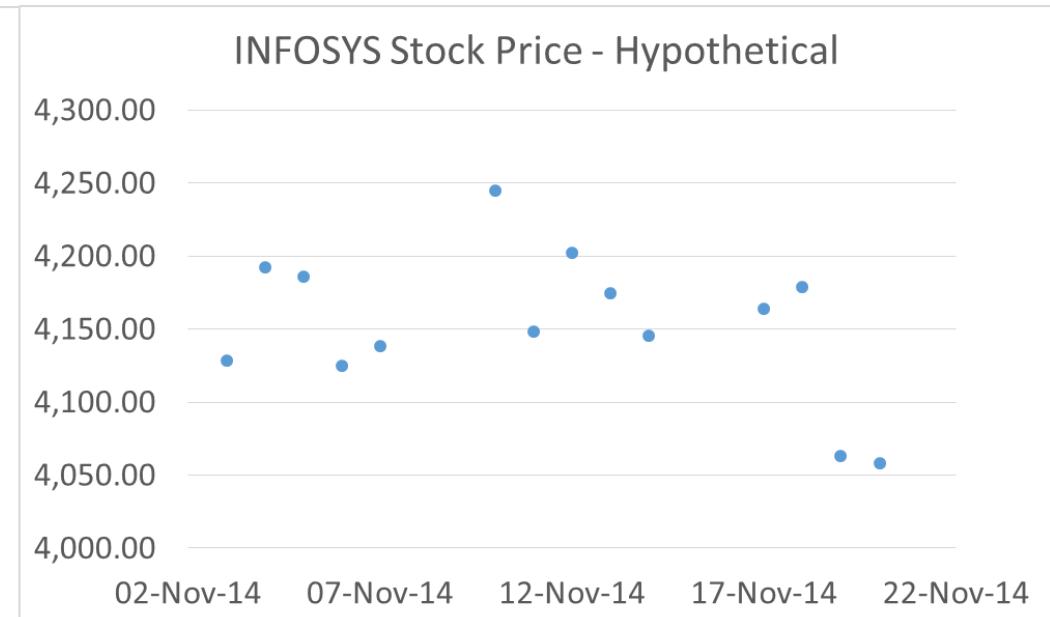
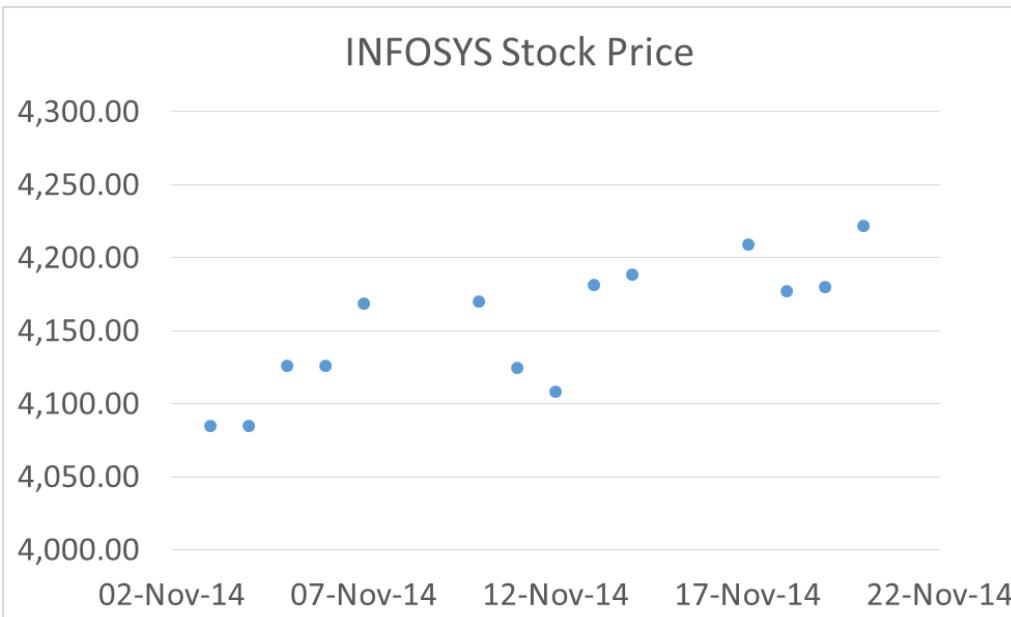
Average distance from the mean?

3 3 6 7 7 10 10 11 13 30

CSE 7315C

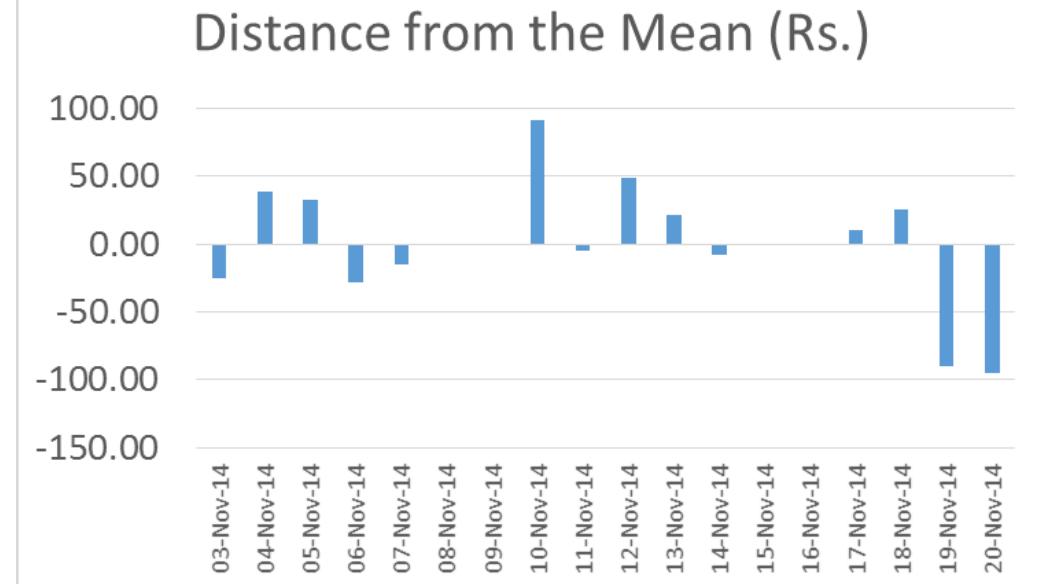
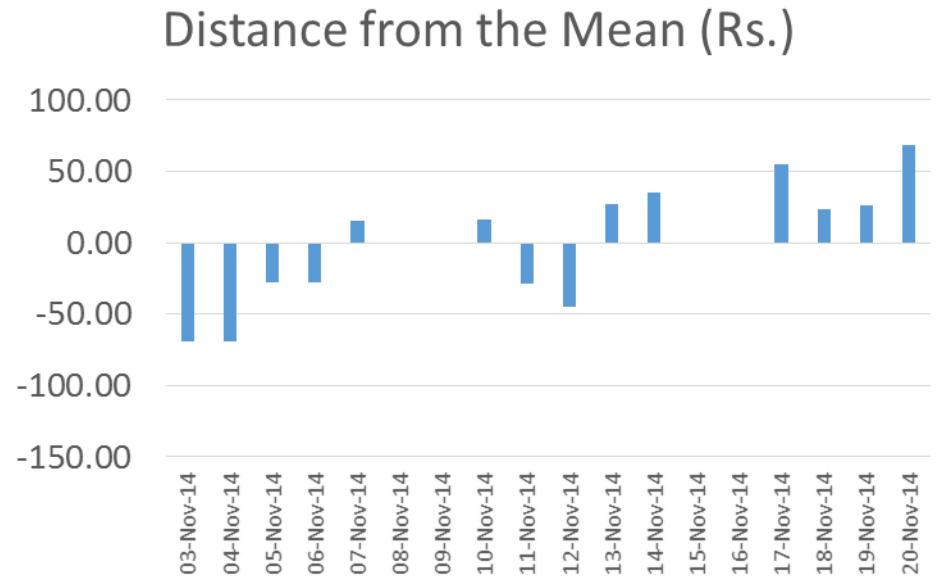


Measures of Spread – Mean Distance, Mean Absolute Deviation or Standard Deviation - Excel



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

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.

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

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

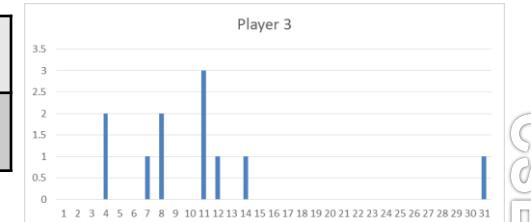
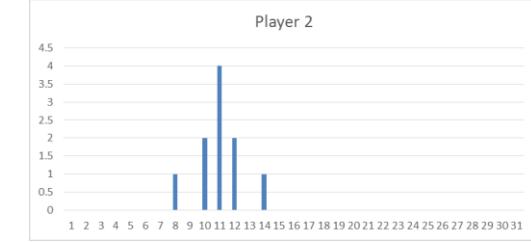
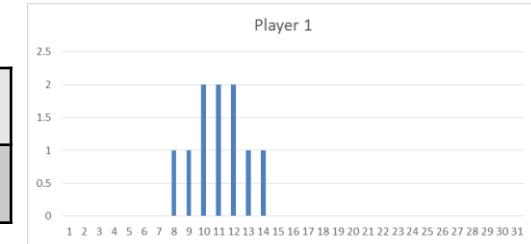
Measuring Variability and Spread

Calculate standard deviation for each player.

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



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.



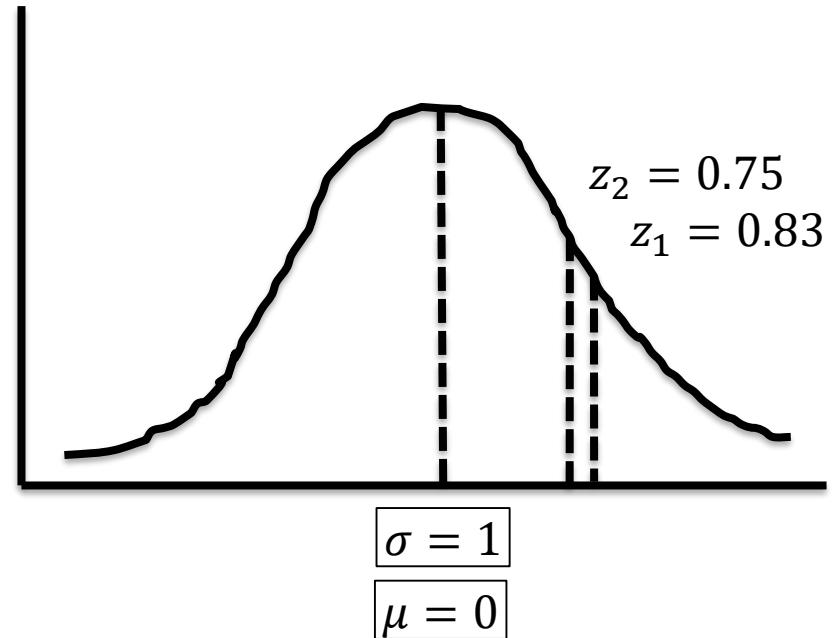
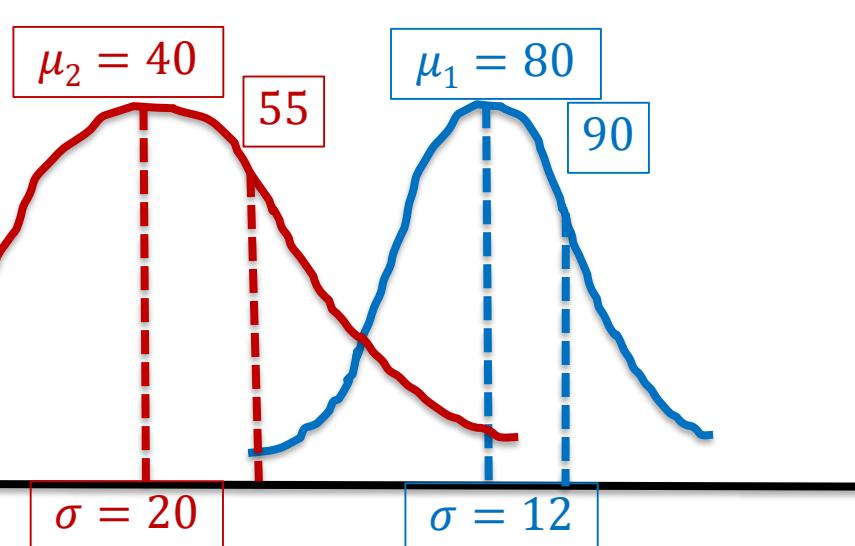
Measuring Variability and Spread

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

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

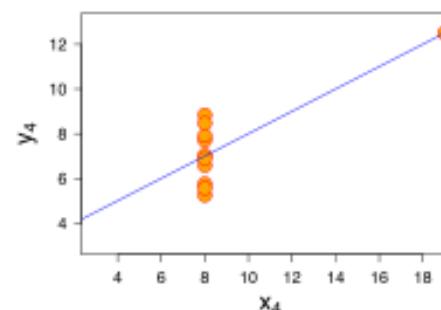
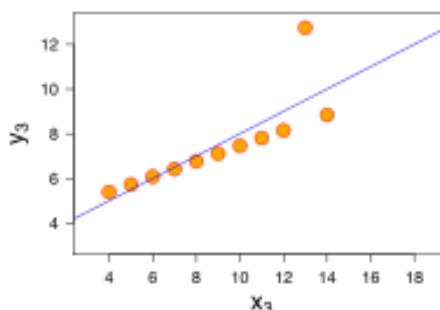
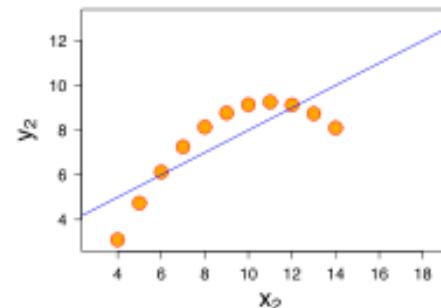
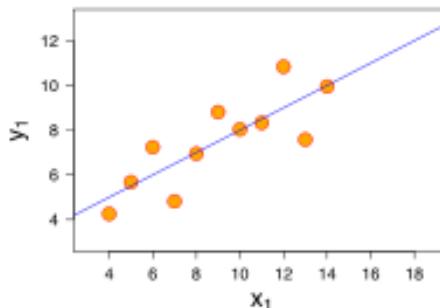
Standard score, $z = \frac{x-\mu}{\sigma}$, # of stdevs from the mean



Measuring Variability and Spread

Anscombe's quartet								
I		II		III		IV		
x	y	x	y	x	y	x	y	
10	8.04	10	9.1	10	7.46	8	6.6	
8	6.95	8	8.1	8	6.77	8	5.8	
13	7.58	13	8.7	13	12.7	8	7.7	
9	8.81	9	8.8	9	7.11	8	8.8	
11	8.33	11	9.3	11	7.81	8	8.5	
14	9.96	14	8.1	14	8.84	8	7	
6	7.24	6	6.1	6	6.08	8	5.3	
4	4.26	4	3.1	4	5.39	19	13	
12	10.8	12	9.1	12	8.15	8	5.6	
7	4.82	7	7.3	7	6.42	8	7.9	
5	5.68	5	4.7	5	5.73	8	6.9	

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)



CSE 7315C

PROBABILITY BASICS





Sholay

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?

CSE 7315C



Probability vs Statistics



Probability is the basis of inferential statistics.

CSE 7315C



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	Probability of dying between ages x to $x+1$	Number surviving to age 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
	$q(x)$	$l(x)$	$d(x)$	$L(x)$	$T(x)$	$e(x)$
0-1	0.006865	100,000	687	99,394	7,743,016	77.4
1-2	0.000469	99,313	47	99,290	7,643,622	77.0
2-3	0.000337	99,267	33	99,250	7,544,332	76.0
3-4	0.000254	99,233	25	99,221	7,445,082	75.0
4-5	0.000194	99,208	19	99,199	7,345,861	74.0
5-6	0.000177	99,189	18	99,180	7,246,663	73.1
6-7	0.000160	99,171	16	99,163	7,147,482	72.1

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

CSE 7315C



Probability - Applications

Gaming industry – Establish charges and payoffs

HR – Does a company have biased hiring policies?

Manufacturing/Aerospace – Prevent major breakdowns

CSE 7315C



Assigning Probabilities

Classical Method – *A priori* or Theoretical

Probability can be determined prior to conducting any experiment.

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

Example: Tossing of a fair die



CSE 7315C



Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

Probability can be determined post conducting a thought experiment.

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

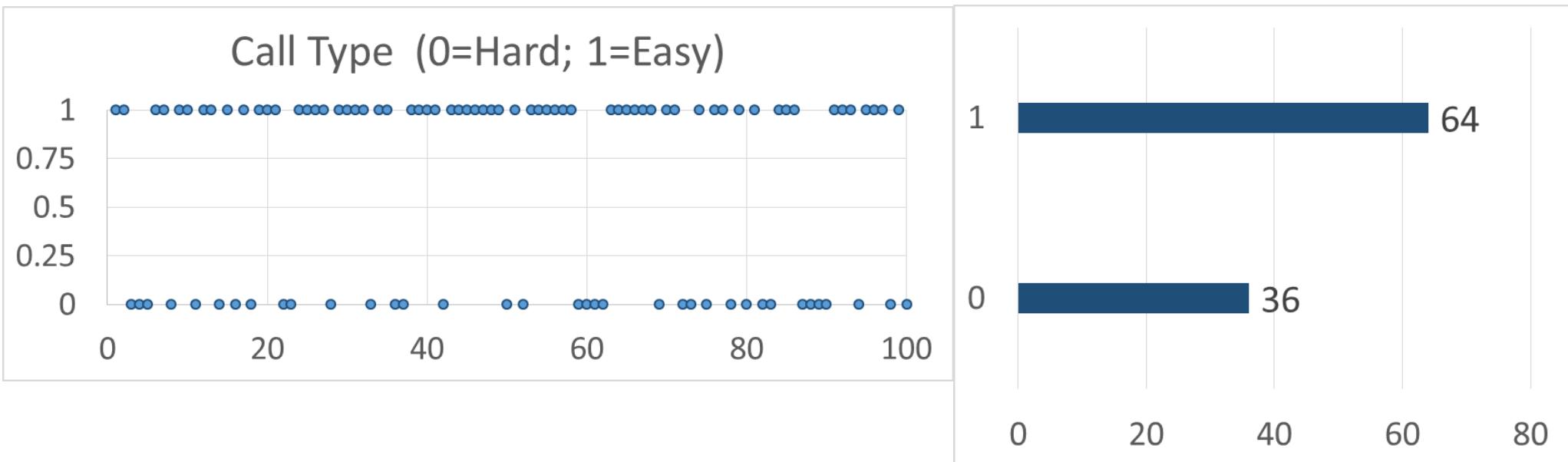
Example: Tossing of a weighted die...well!, even a fair die. The larger the number of experiments, the better the approximation.

This is the most used method in statistical inference.

Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

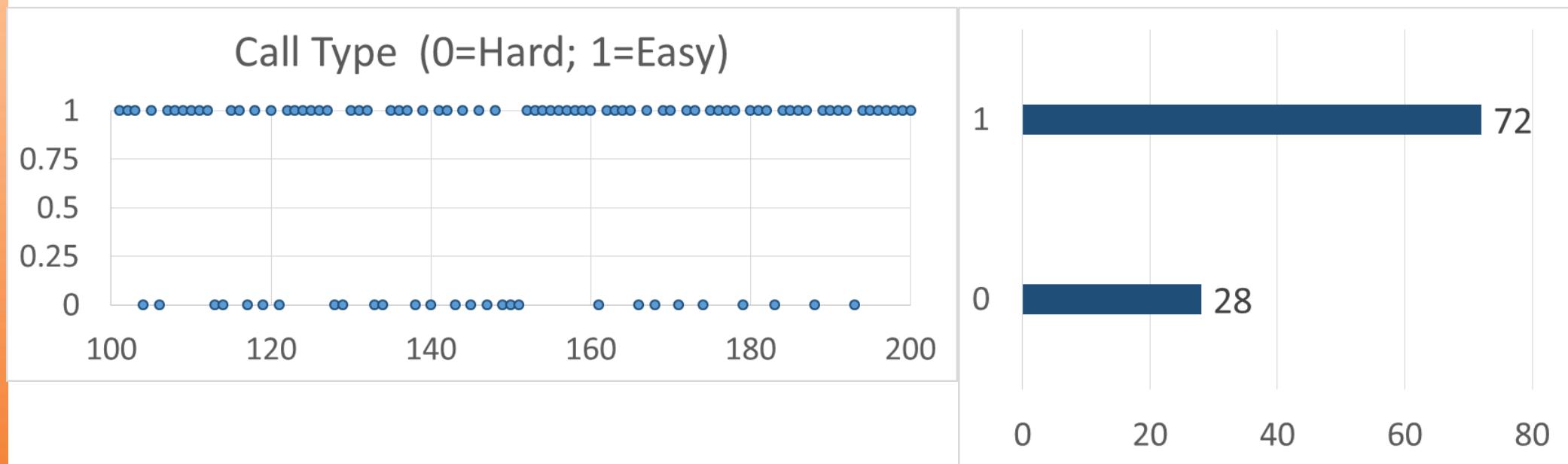
100 calls handled by an agent at a call centre



Assigning Probabilities

Empirical Method – *A posteriori* or Frequentist

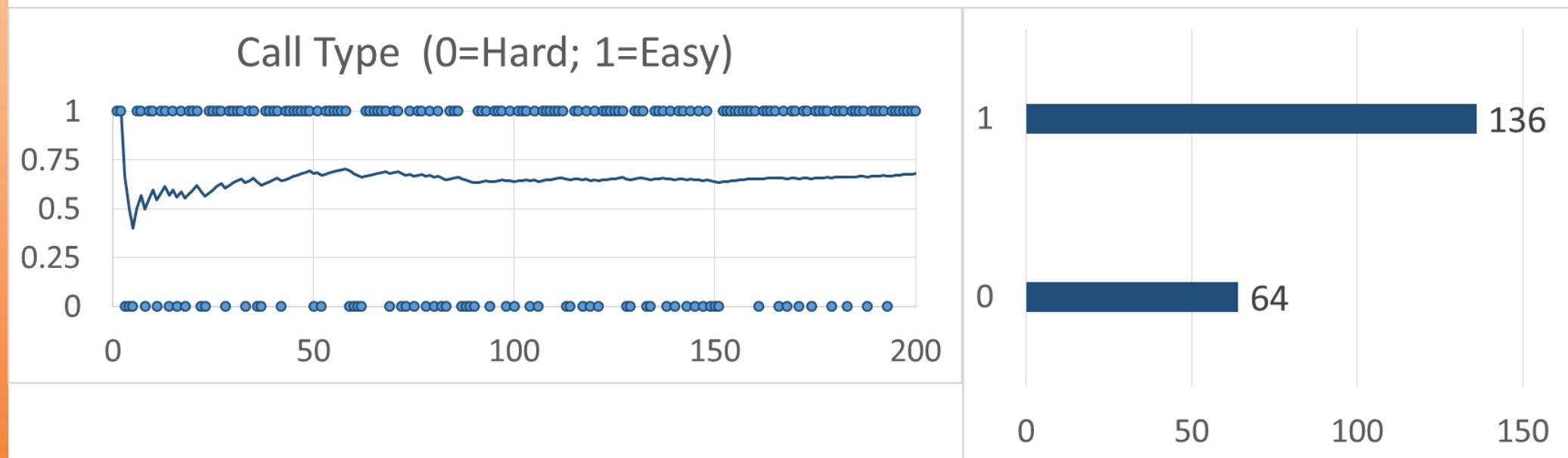
Next 100 calls handled by an agent at a call centre



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

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

What is the probability of having a monthly income of 1000 BHD?

$$10/23 = 0.43$$

CSE 7315C



Assigning Probabilities

Subjective Method

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

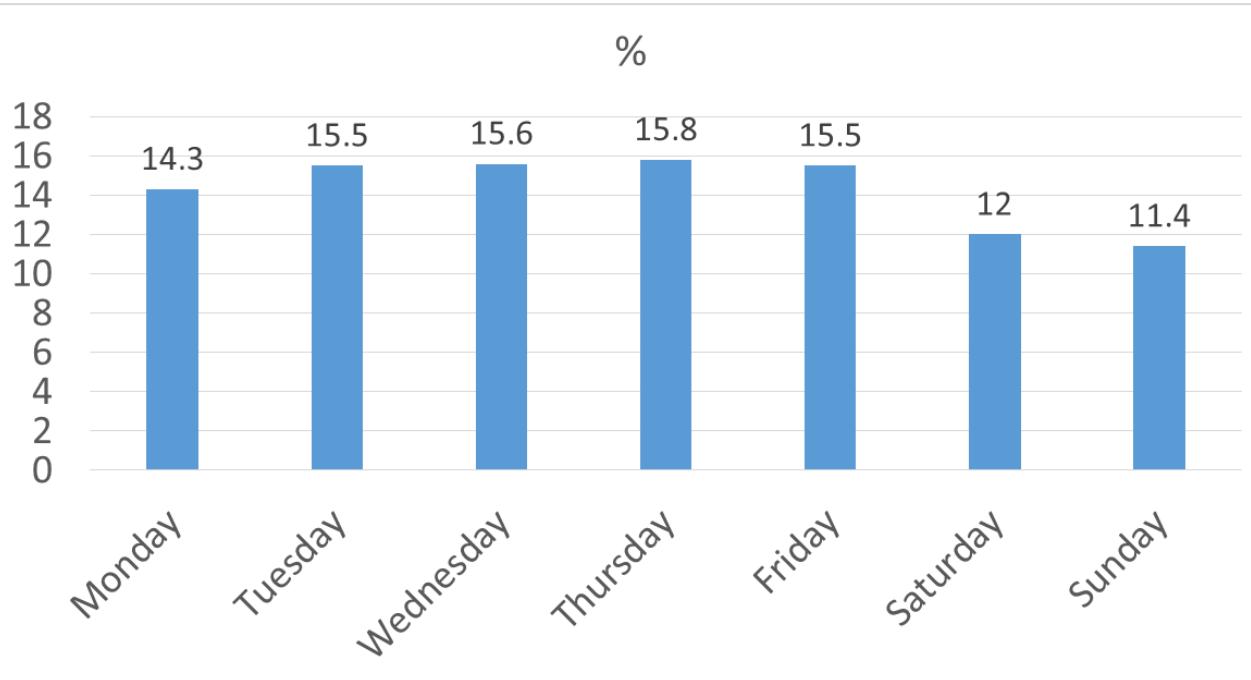
What is the probability of rain tomorrow?

CSE 7315C



Assigning Probabilities

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



Strategic decisions must be based on hard data

"In God we trust; all others must bring data."

Edward Deming*



*The man behind Japanese post-war industrial revolution

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.

CSE 7315C



Probability - Terminology

Sample Space – Set of all possible outcomes, denoted S.

Event – A subset of the sample space.

CSE 7315C



Probability - Rules

S

S

A

S

A

B

A and B are **mutually exclusive**

$$P(S) = 1$$

$$0 \leq P(A) \leq 1$$

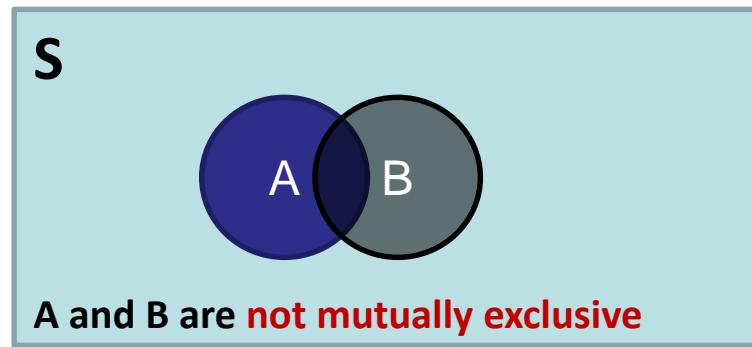
$$\begin{aligned}P(A \text{ or } B) \\= P(A) + P(B)\end{aligned}$$

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?



CSE 7315c



Probability - Types

Contingency table summarizing 2 variables, *Loan Default* and *Age*:

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	10,503	27,368	259	38,130
	Yes	3,586	4,851	120	8,557
	Total	14,089	32,219	379	46,687

Probability - Types

Convert it into probabilities:

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

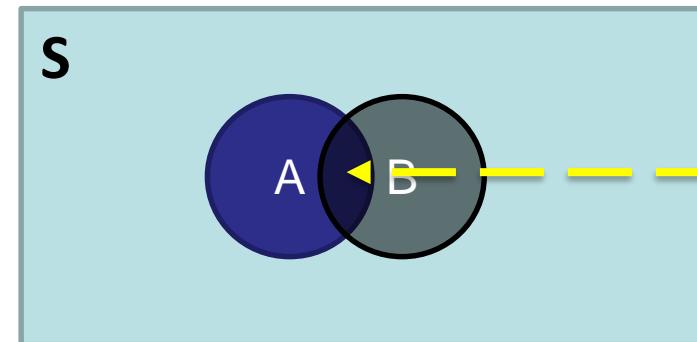
Probability - Types

Joint Probability

		Age			
		Young	Middle-aged	Old	
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Probability describing a combination of attributes.

$$P(\text{Yes and Young}) = 0.077$$

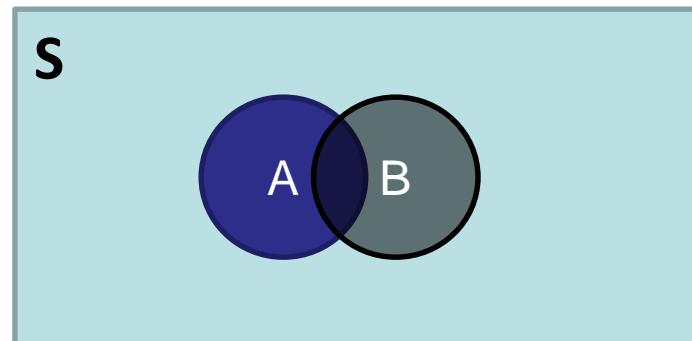


Probability - Types

Union Probability

		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

$$P(\text{Yes or Young}) = P(\text{Yes}) + P(\text{Young}) - P(\text{Yes and Young}) = \\ 0.184 + 0.302 - 0.077 = 0.409$$



Probability - Types

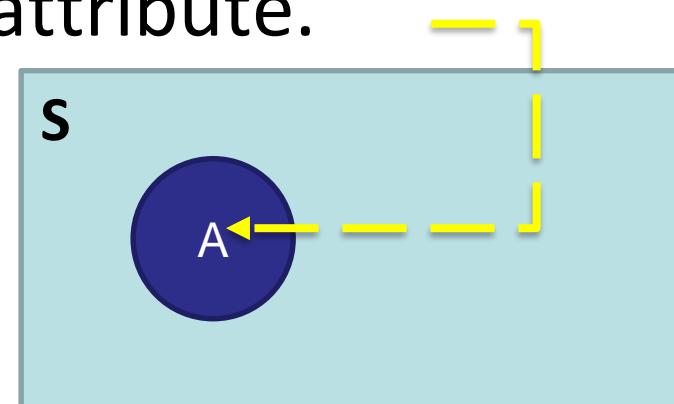
Marginal Probability

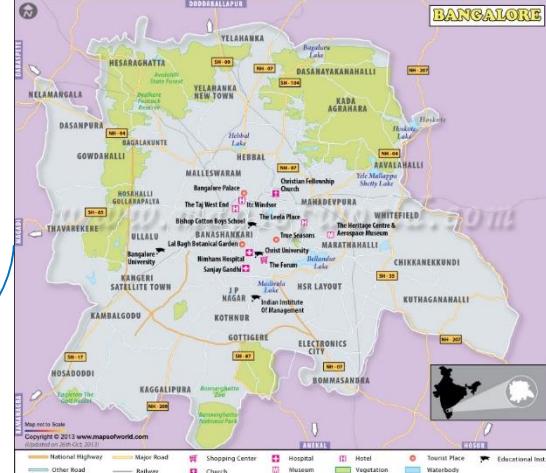
		Age			
		Young	Middle-aged	Old	Total
Loan Default	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Probability describing a single attribute.

$$P(\text{No}) = 0.816$$

$$P(\text{Old}) = 0.008$$





HYDERABAD

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

BENGALURU

Floors 1-3, L77, 15th Cross Road, 3A Main Road, Sector 6, HSR Layout, Bengaluru – 560 102
 +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.