Inferential Statistics

Presentation · March 2019

DOI: 10.13140/RG.2.2.26024.62725

1 author:

Tarek Tawfik Amin
Cairo University

INFERENTIAL STATISTICS

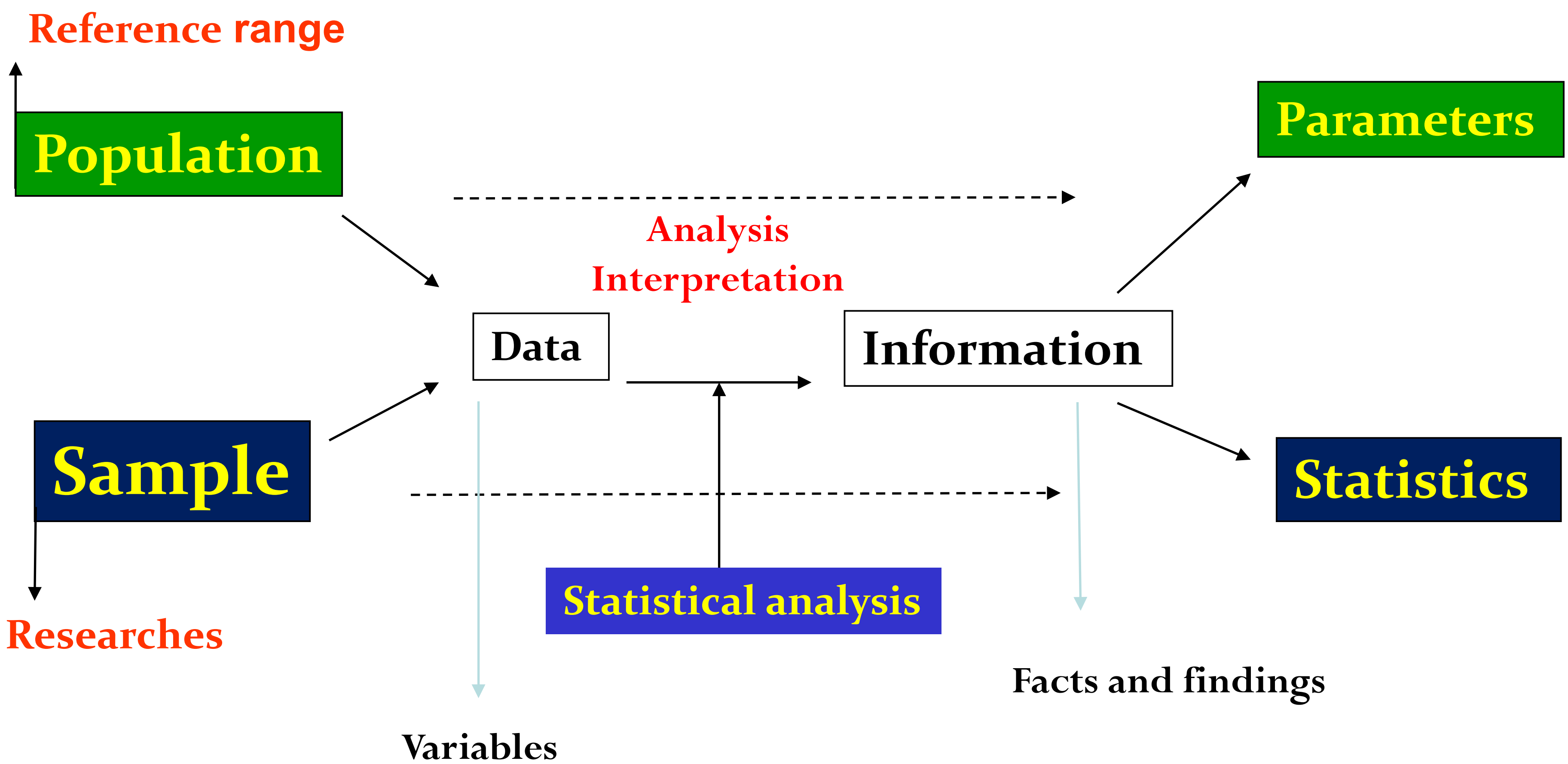
Professor Tarek Tawfik Amin

Faculty of Medicine, Cairo University

dramin55@gmail.com

Introduction

Parameters vs. statistics
Data vs. information



Information: It is the meaningful aggregation and management of data to be evaluated and interpreted for specific use.

Data: are sets of values of one or more variables
(e.g. age, sex, education, received vaccination etc.,)
collected from records, interviews, observations, research, etc.
The data should be managed to provide information.

Statistics [Functions]

Descriptive
- Organize
- Summarize
- Present

Inferential
- Predict
- Estimate
- Generalize

Find a difference / Association

Applications of inferential statistics

- To compare two or more samples to investigate potential differences.(sometimes comparing a sample to the population parameter)
- Studying the relationship between two or more variables.
- To infer from sample group generalizations that can be applied to a wider population.
- Allows detection of large or small differences, in variables or correlations between variables relevant to research question. [1-3]

1. Allison JJ, Calhoun JW, Wall TC, Spettell CM, Fargason CA, Weissman NW, et al. Optimal reporting of health care process measures: inferential statistics as help or hindrance. Manag Care Q 2000;8(4):1e10.
2. Botti M, Endacott R. Clinical research quantitative data collection and analysis. Int Emerg Nurs 2008;16(2):197e204.
3. Gouveia-Oliveira A. Medical decision making. Acta Med Port 1996;9 (10e12):391e6.

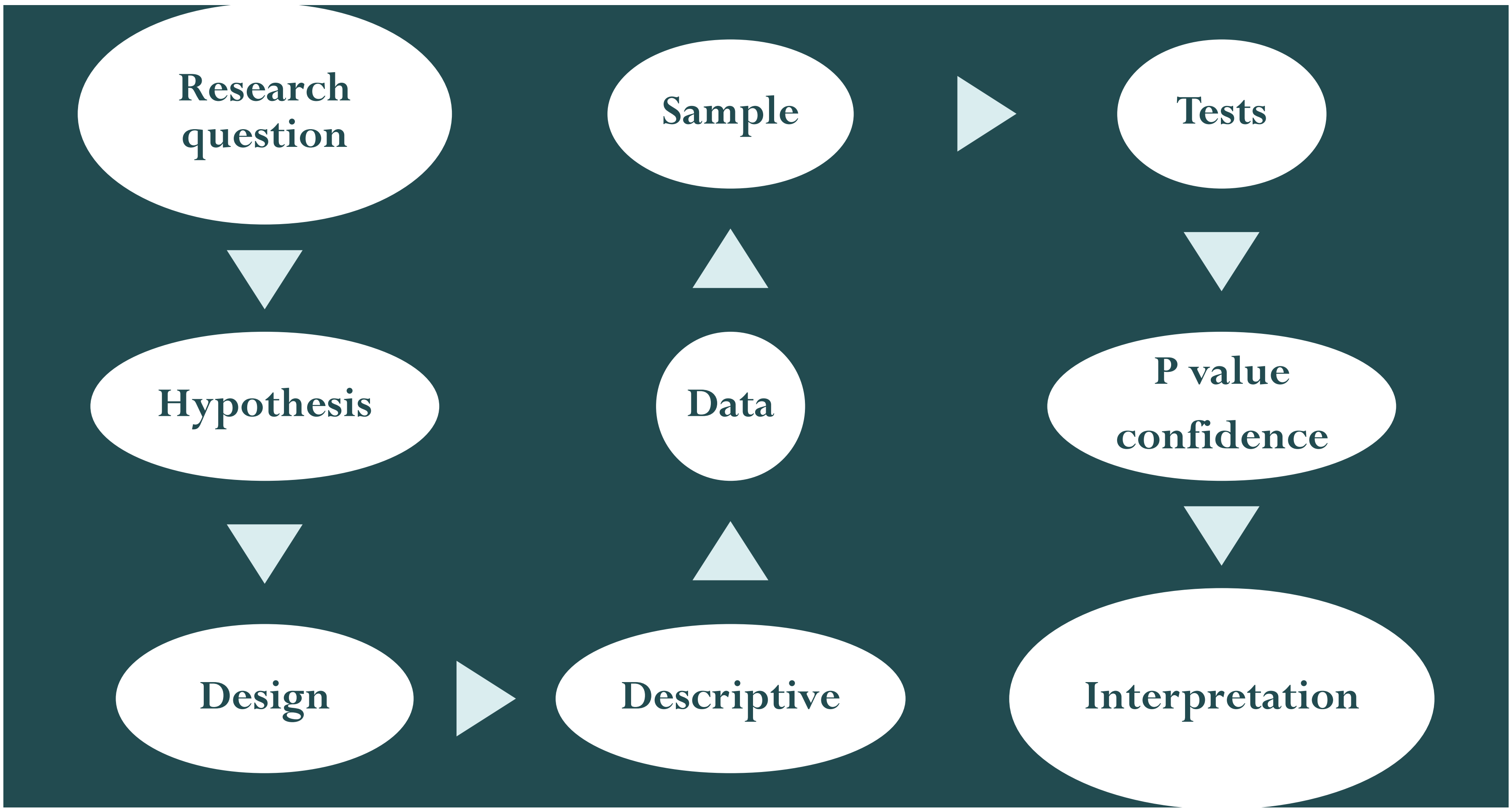
Condition to use inferential statistics

Representative sample = Good estimator

Inferential Statistics

Predicts or estimates characteristics of a population from a knowledge of the characteristics of <u>only a sample</u> of the population

Steps needed



Population and Sample

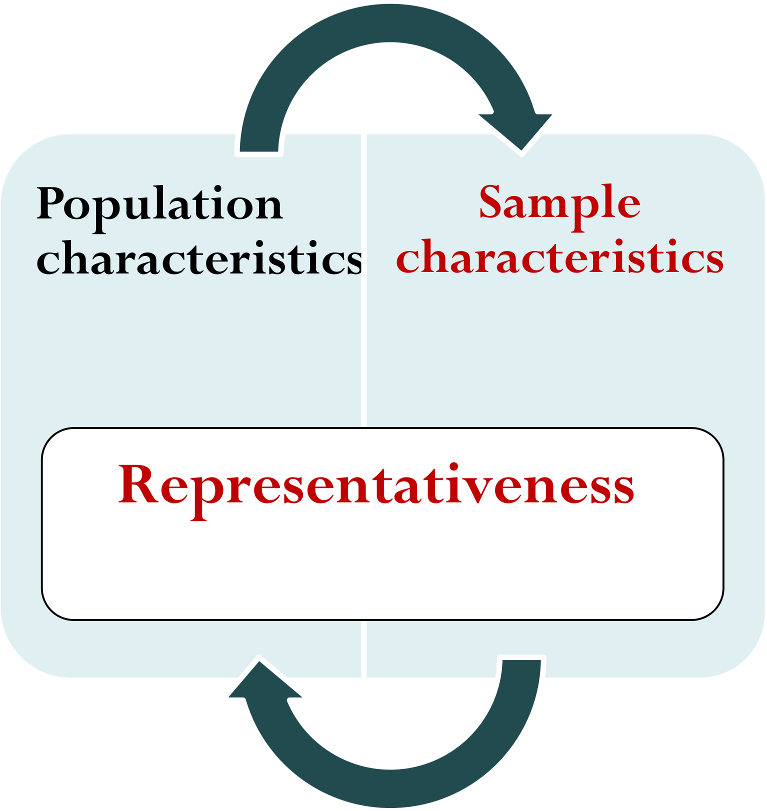
- In research we want to make a (conclusion) about the population.
- Studying the whole population is impossible in terms of money/time/labor.
- Random sampling from the population and infer from the sample data the needed conclusions.
- The task of statistics is to quantify the **uncertainty**

  (is the sample is really representing that population?).

Why sampling instead of whole population?

- 1- Samples are usually cheaper and quicker.
- 2- Impossible to locate all members of a population (*not included in list, unavailable, difficult to reach, unwilling to participate).*
- 3- Research sometimes destroys the units of analysis so that a census would destroy the population *(batteries quality control check).*
- 4- Sometimes sampling is more accurate than census *(inexperienced data collection team during census).*

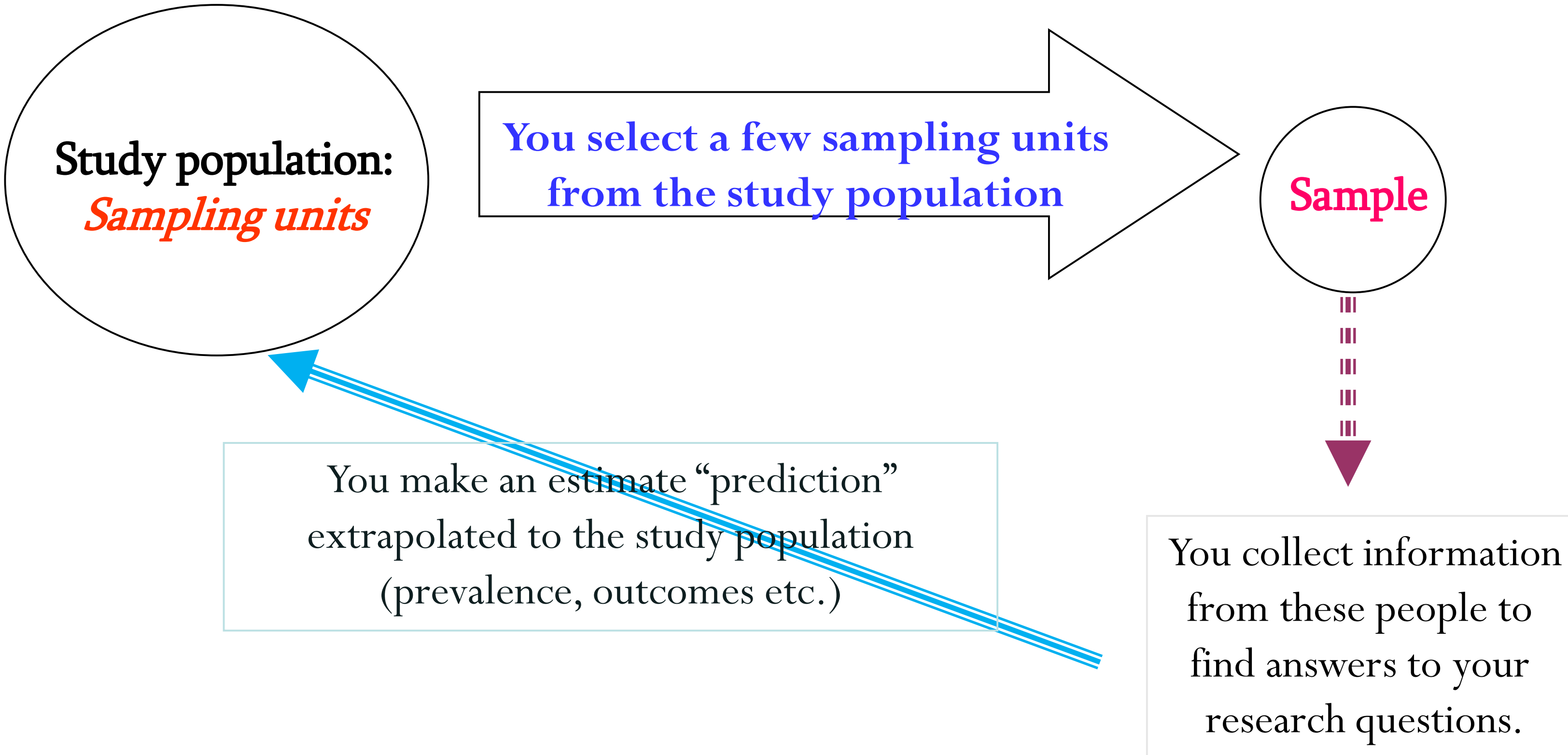
Problem with Sampling



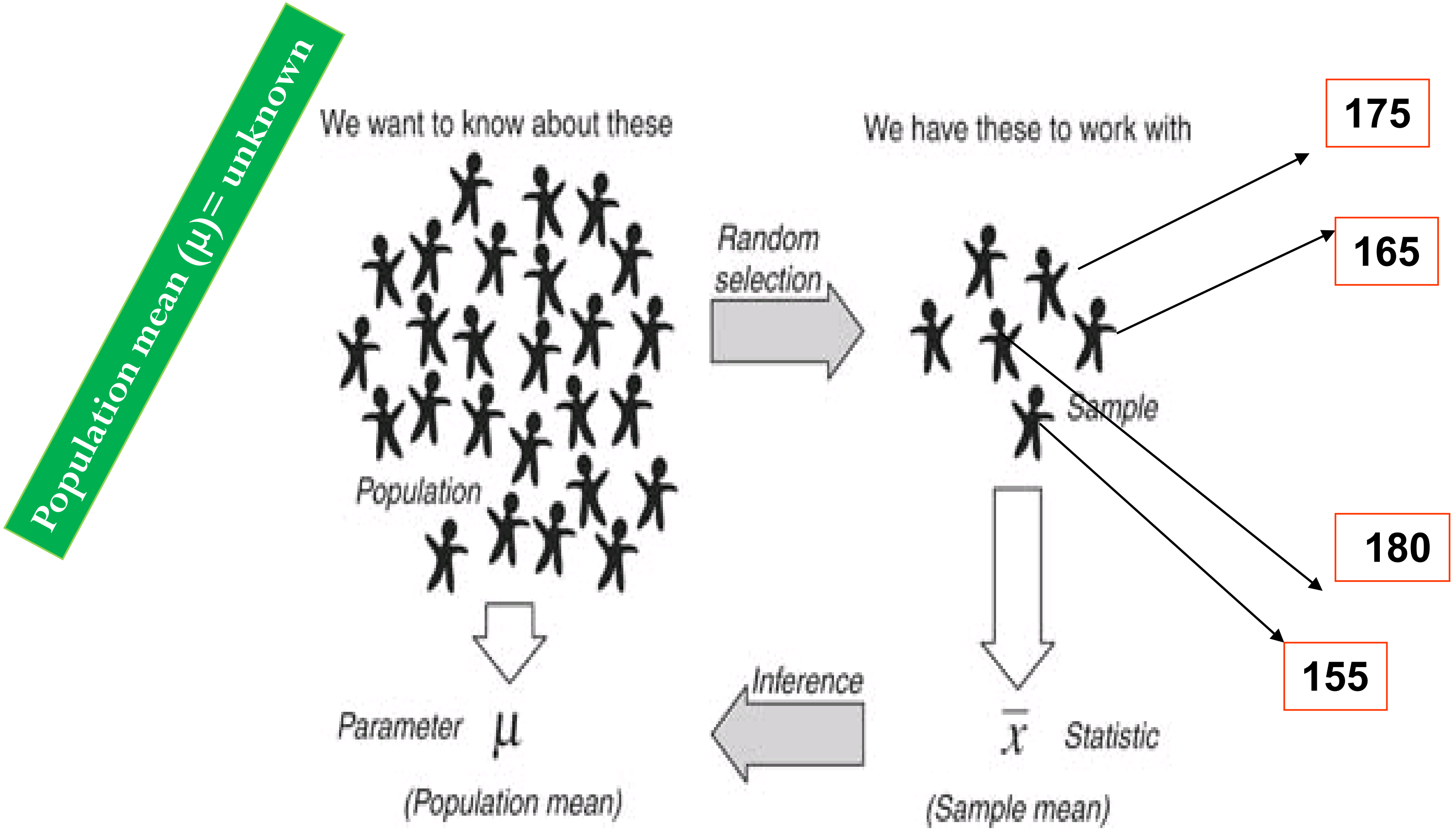
? Sampling Error

Do sample statistics are representative to the population as those concluded if we conduct a complete and accurate census?

The concept of sampling



What would be the mean systolic blood pressure of older subjects (65+) in Egypt?



From our sample we calculate an estimate of the population parameter

The good sample (the good estimator)

Should be:

Unbiased: The mean of sample = population mean

Precise: (narrow dispersion about the mean)

The dispersion in repeated samples is small

??????

The Sampling error

Four individuals A, B, C, D (the whole population):

A = 18 years

B= 20 years

C= 23 years

D= 25 years

Their mean age is = 18+20+23+ 25= 86/4= 21.5 years (population mean μ).

Probability of sampling two individuals: (6 probabilities)

A+B=18+20= 38/2=19.0 years

A+C= 18+23=20.5 years.

A+D=18+25=21.5 years.

B+C=20+23=21.5 years.

B+D=20+25=22.5 years.

C+D=23+25=24.0 years.

Sampling error= population mean-sample mean = ranges from -2.5 to +2.5 years.

Probability of sampling three individuals: (4 probabilities)

A+B+C=18+20+23=20.33 years.

A+B+D=18+20+25=21.00 years.

A+C+D=18+23+25=22.00 years.

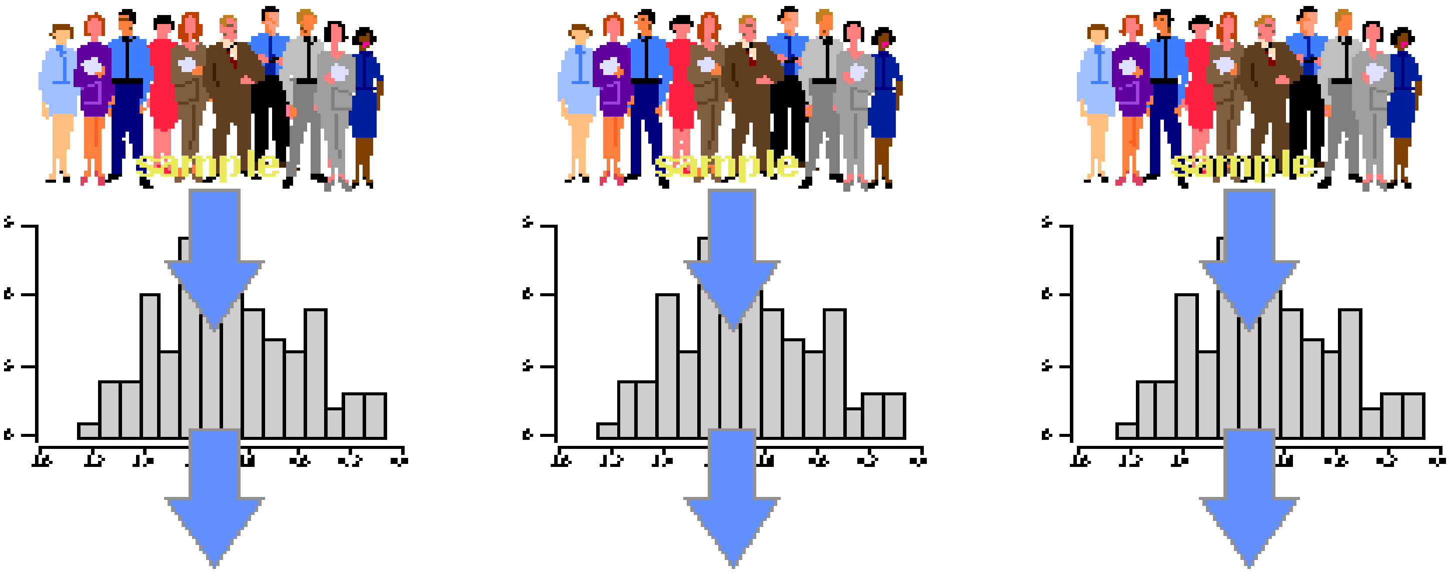
B+C+D=20+23+25=22.67 years.

Sampling Error = ranges from -1.17 to +1.7 years.

If C=32 (instead of 23) years and D=40 (instead of 25) years: sampling of 2= sampling error of -7.00 to +7.00 and in 3= -3.67 to +3.67 years.

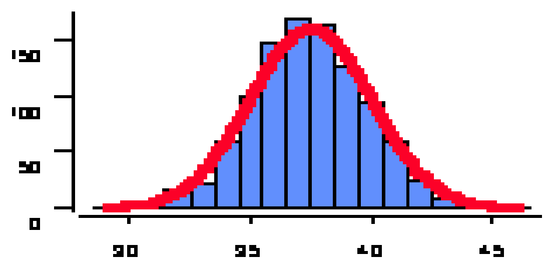
The greater the variability of a given variable the larger the sampling error for a given sample size.

Infinite samples should represents the population it came from (good estimator)



Average **Average** **Average**

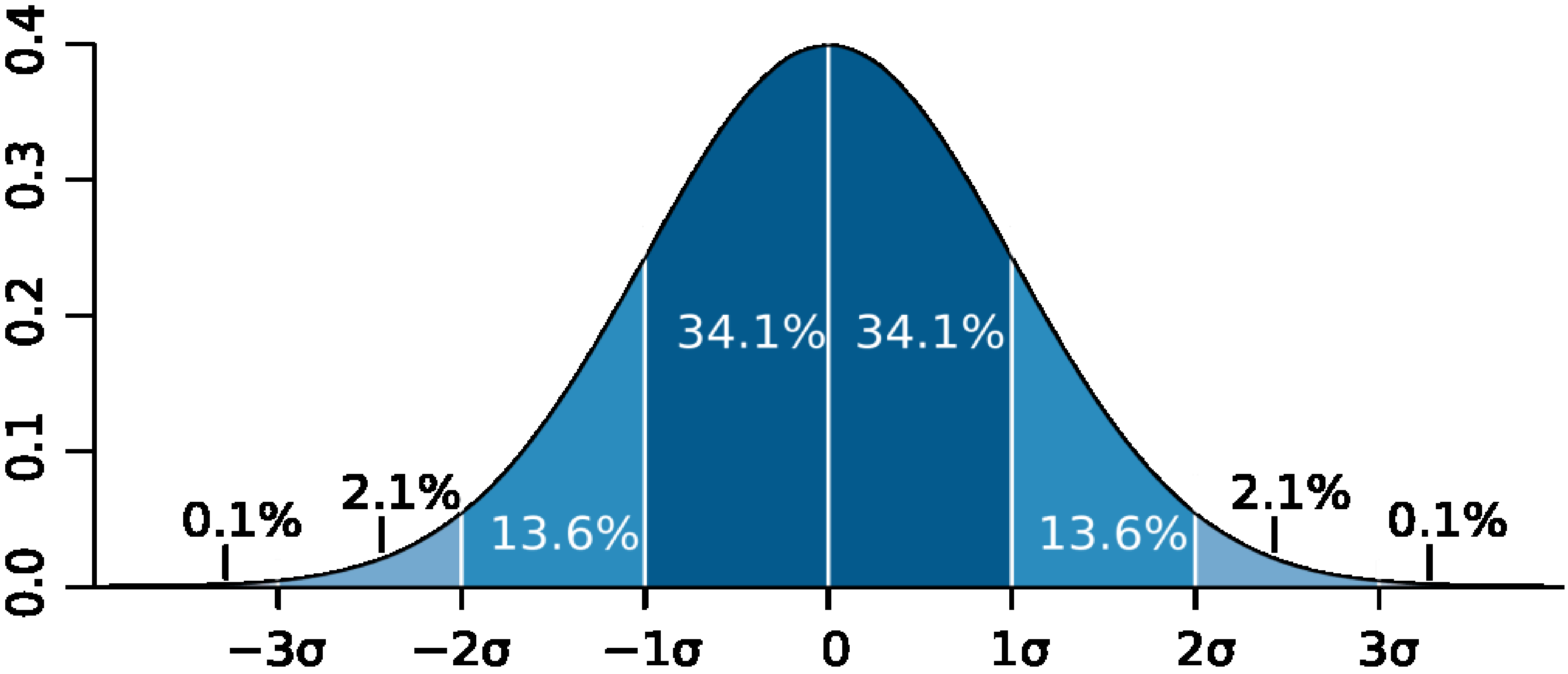
The Sampling Distribution...



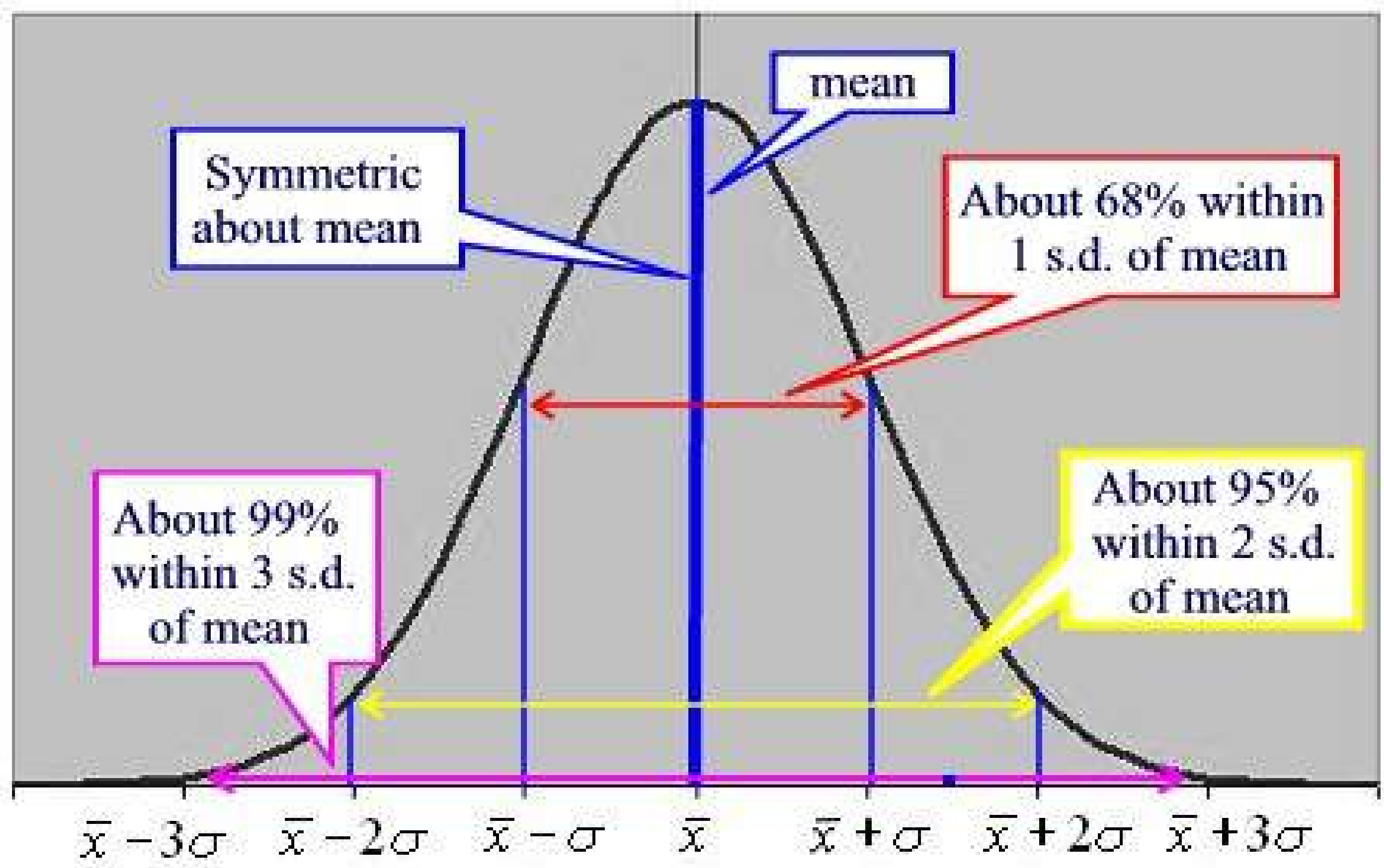
...is the distribution of a statistic across an infinite number of samples

Normal Distribution

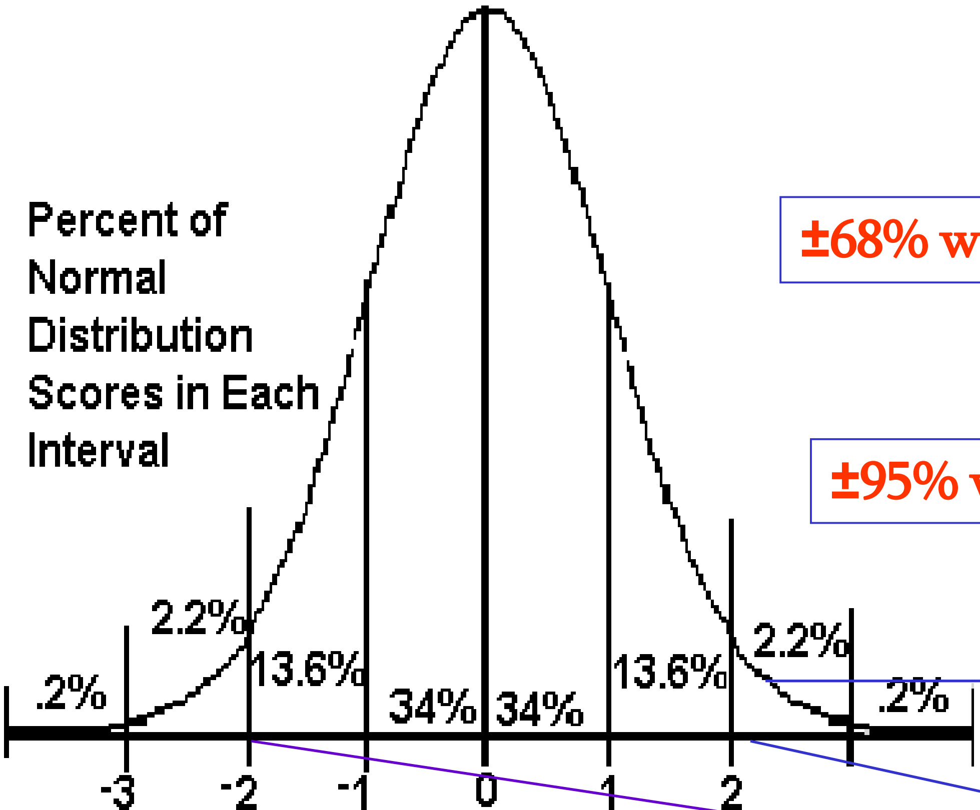
***Normal Distribution*: (Gaussian curve):** **Many human traits, such as intelligence, personality, and attitudes, also, the weight and height, are distributed among the populations in a fairly normal way.**



Characteristics of normal curve



The normal distribution



±68% within between μ ±1 SD (σ)

±95% within between μ ±1.96 SD (σ)

> 3 SDs Definite outliers

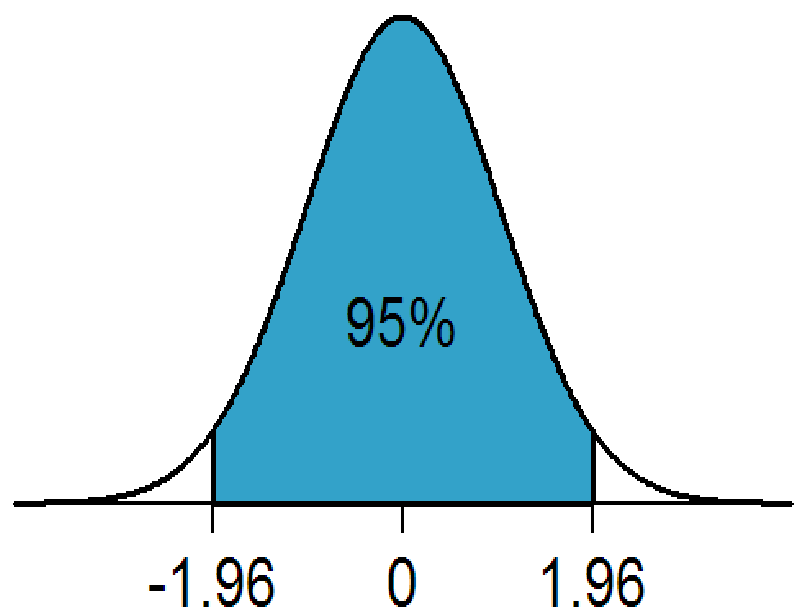
> 2SDs Possible outliers

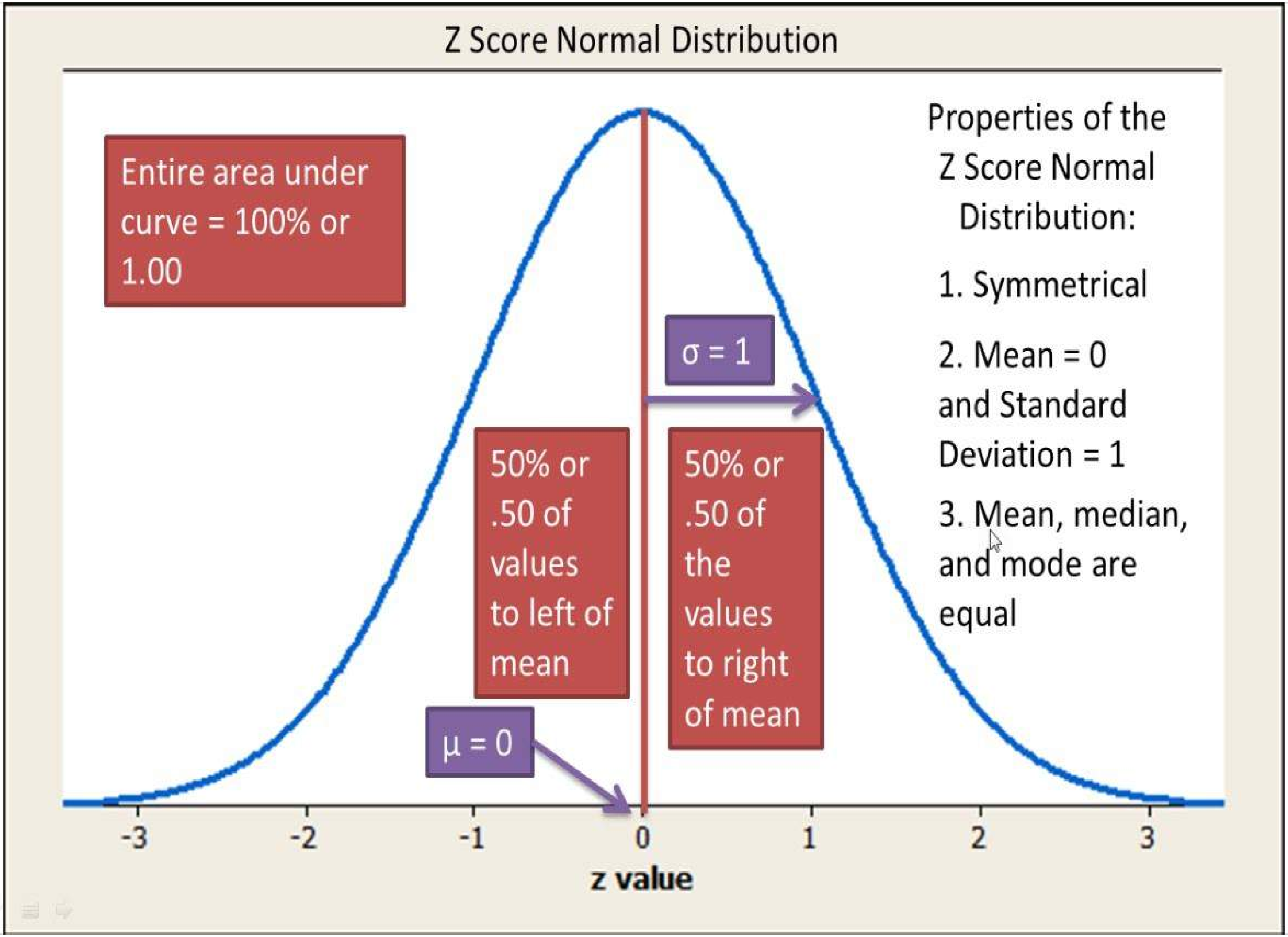
The Standard Normal Distribution

- **Standard Normal Curve (Z-Distribution)**
- Has a mean of "0" and a standard deviation of "1"
- The area under the whole distribution is 1.
- Essential to understanding P value and Alpha.

To locate any point under the curve $Z = X – u \ / \ s$

- Where: X is the variable to be standardized, u is the population mean and s is the population standard



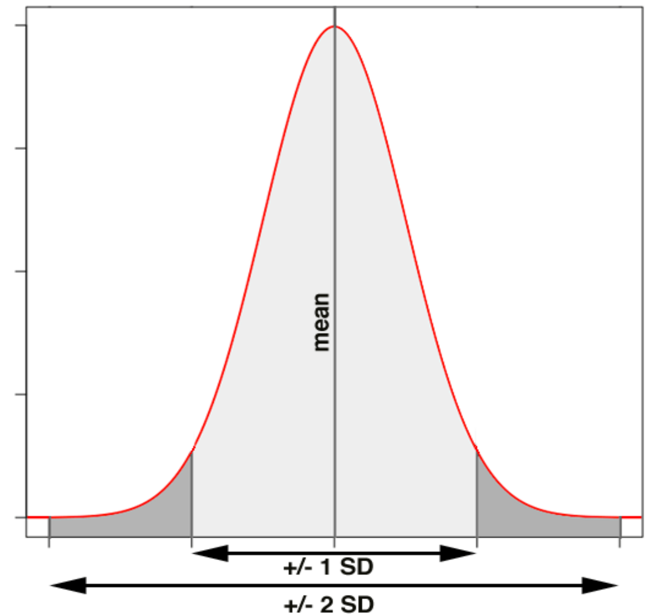

Z Score Normal Distribution
Entire area under curve = 100% or 1.00
σ = 1
50% or .50 of values to left of mean
50% or .50 of the values to right of mean
μ = 0
Properties of the Z Score Normal Distribution:
1. Symmetrical
2. Mean = 0 and Standard Deviation = 1
3. Mean, median, and mode are equal
-3
-2
-1
0
1
2
3
z value

Areas under the standard normal curve.

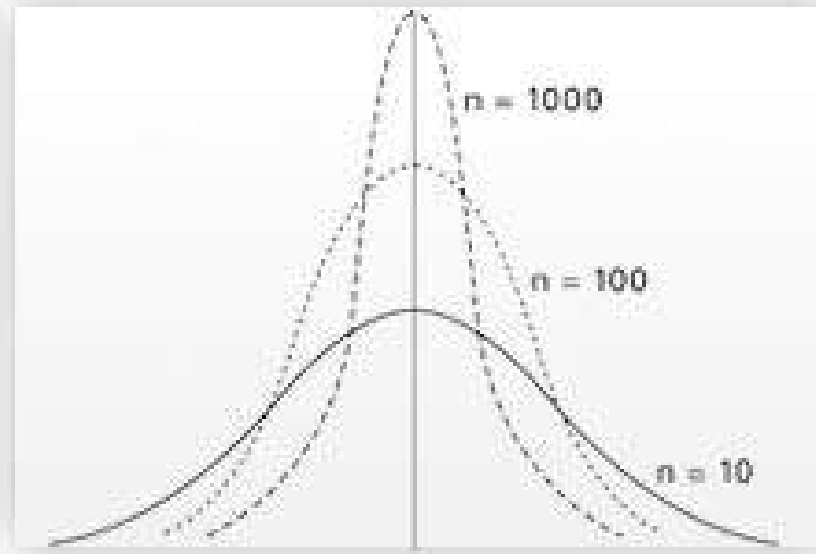
Z score	Area under curve between both points (Around the mean)	Beyond both points (two tails)	Beyond one point (one tail)
±0.1	0.080	0.920	0.4600
±0.2	0.159	0.841	0.4205
±0.3	0.236	0.764	0.3820
±0.4	0.311	0.689	0.3445
±0.5	0.383	0.617	0.3085
±0.6	0.451	0.549	0.2745
±0.7	0.516	0.484	0.2420
±0.8	0.576	0.424	0.2120
±0.9	0.632	0.368	0.1840
±1	0.683	0.317	0.1585
±1.1	0.729	0.271	0.1355
±1.2	0.770	0.230	0.1150
±1.3	0.806	0.194	0.0970
±1.4	0.838	0.162	0.0810
±1.5	0.866	0.134	0.0670
±1.6	0.890	0.110	0.0550
±1.645	0.900	0.100	0.0500
±1.7	0.911	0.089	0.0445
±1.8	0.928	0.072	0.0360
±1.9	0.943	0.057	0.0290
1.96	0.950	0.050	0.0250
±2	0.954	0.046	0.0230
±2.1	0.964	0.036	0.0180
±2.2	0.972	0.028	0.0140
±2.3	0.979	0.021	0.0105
±2.4	0.984	0.010	0.0100
±3	0.996	0.004	0.0020

Areas under the standard normal curve.

Z	Area under curve between both points (around the mean)	Beyond both points (two tails)	Beyond one point (one tail)
±0.1	0.080	0.920	0.4600
±0.2	0.159	0.841	0.4205
±0.3	0.236	0.764	0.3820
±0.4	0.311	0.689	0.3445
±0.5	0.383	0.617	0.3085
±0.6	0.451	0.549	0.2745
±0.7	0.516	0.484	0.2420
±0.8	0.576	0.424	0.2120
±0.9	0.632	0.368	0.1840
±1	0.683	0.317	0.1585
±1.1	0.729	0.271	0.1355
±1.2	0.770	0.230	0.1150
±1.3	0.806	0.194	0.0970
±1.4	0.838	0.162	0.0810
±1.5	0.866	0.134	0.0670
±1.6	0.890	0.110	0.0550
±1.645	0.900	0.100	0.0500
±1.7	0.911	0.089	0.0445
±1.8	0.928	0.072	0.0360
±1.9	0.943	0.057	0.0290
1.96	0.950	0.050	0.0250
±2	0.954	0.046	0.0230
±2.1	0.964	0.036	0.0180
±2.2	0.972	0.028	0.0140
±2.3	0.979	0.021	0.0105
±2.4	0.984	0.010	0.0100
±2.576	0.99	0.004	0.0020



c= level of confidence	Z_c= Z critical values (under normal curve)
90%	1.645
95%	1.960
99%	2.578

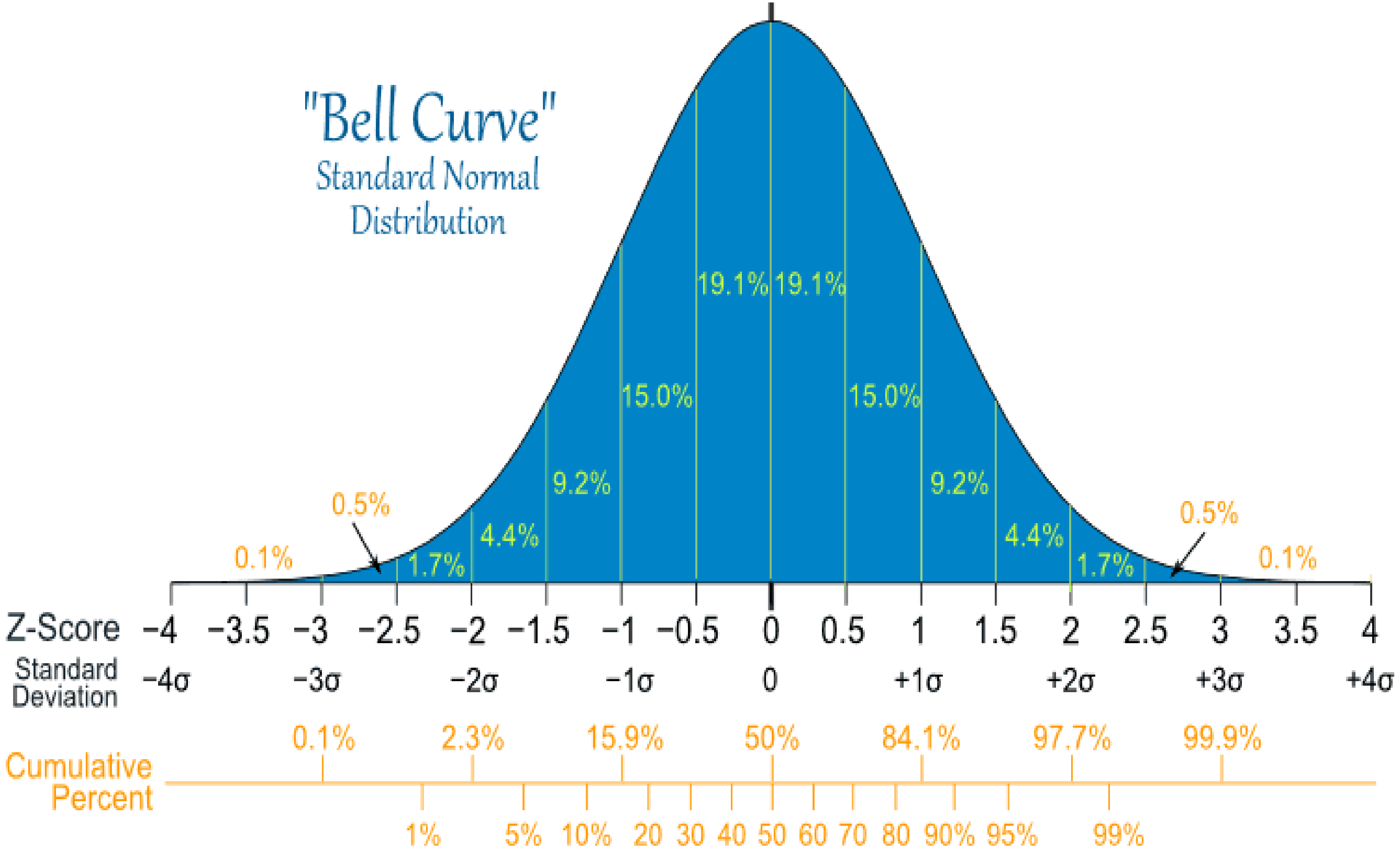


Confidence Interval Chart Example

$$\overline{\overline{x}} \pm Z_c\left(\frac{\sigma}{\sqrt{n}}\right)$$

C.I= Mean of the sample ±Z critical scores (SEM)
SEM= SD/√n

See explanation later



Normal distribution in Z score and percentiles
The concept of sampling distribution

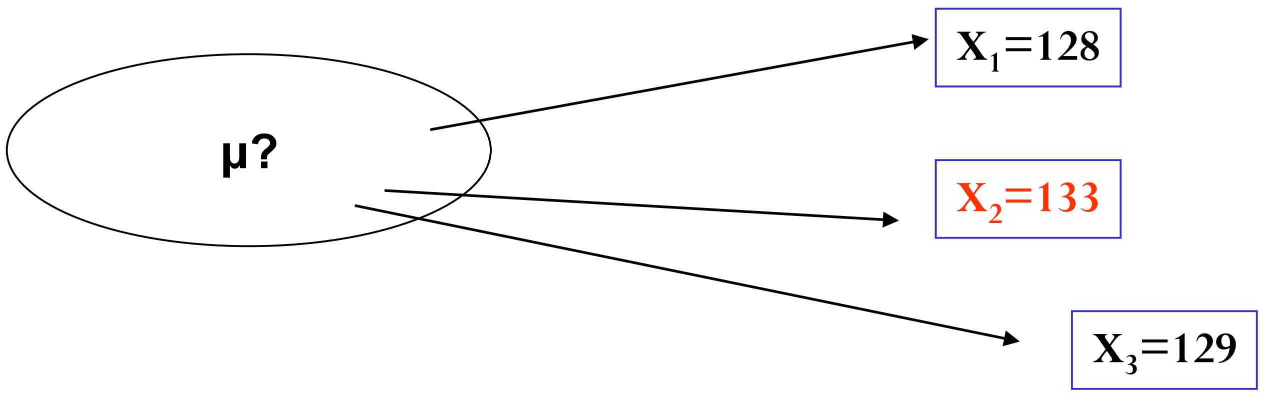
- **Sampling distributions – are theoretical distributions developed by mathematicians to organize statistical outcomes from various sample sizes**

Rule of thumb with caution

If we do not have a normal distribution of data? Is it possible to draw conclusions about areas under the curve, as we did with the Z distribution? It depends -----Rule of thumb, n=30.

Estimation population parameters from a random sample

Random sample for estimating a population mean



From the information in the sample, we will estimate the unknown population mean ($\overline{X}$ is an estimator for μ).

What could have happened if we had another random sample?
What is the measure of variation of sample means?

The Sampling Distribution of a Sample Statistics

- Let's assume that we want to survey a community of 400, the age of them were recorded and having the following parameters

$$\mu = 35 \text{ years}$$

$$\sigma = 13 \text{ years}$$

- ≈ **Let's assume, however, that we do not survey all 400, instead we randomly select 120 people and ask them about their ages and calculate the mean age.**
- ≈ **Then, we put them back into the community and randomly select another 120 residents (may include members of the first sample).**

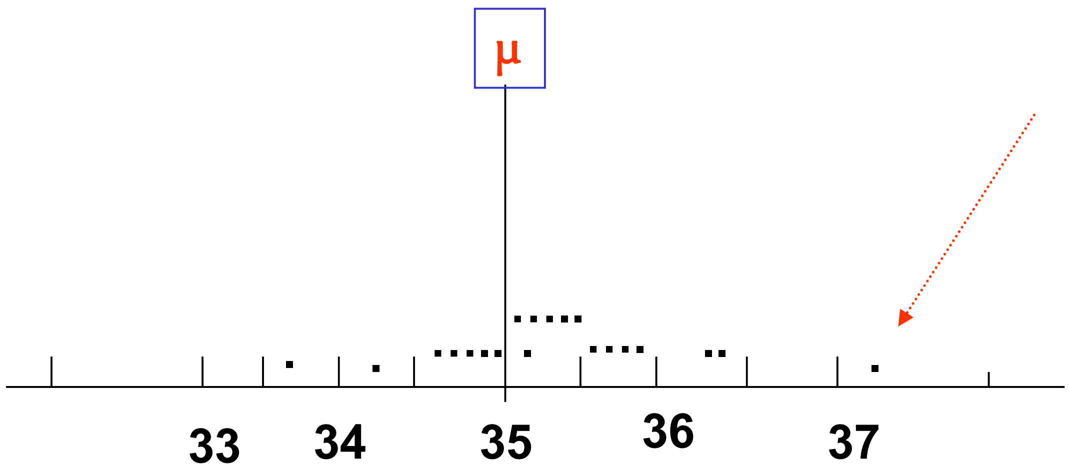
- ≈ **We did this over and over and each time we calculate the mean age.**

SPSS Demo.

The results will be like those in the following slide.

Sample Number	Sample mean
1	34.7
2	35.9
3	35.5
4	34.7
5	34.5
6	34.4
7	35.7
8	34.6
9	37.4
10	35.3
11	34.1
12	35.5
13	34.9
14	36.2
15	35.6
16	35.0
17	35.1
18	36.4
19	35.6
20	33.6
SD of the means	13.37

Distribution of 20 random sample means (n=20)



All the results are clustered around the population value (35 years), with a few scores a bit further out and one extreme score of 37.4 years (random variation=1/20=5%).

Those 400 people have age range from 2 to 69 years, while the <u>means</u> of the samples have a very narrow range of value of about 4 years and 10 samples coincide with the population mean (35 years).

In case of infinite sampling

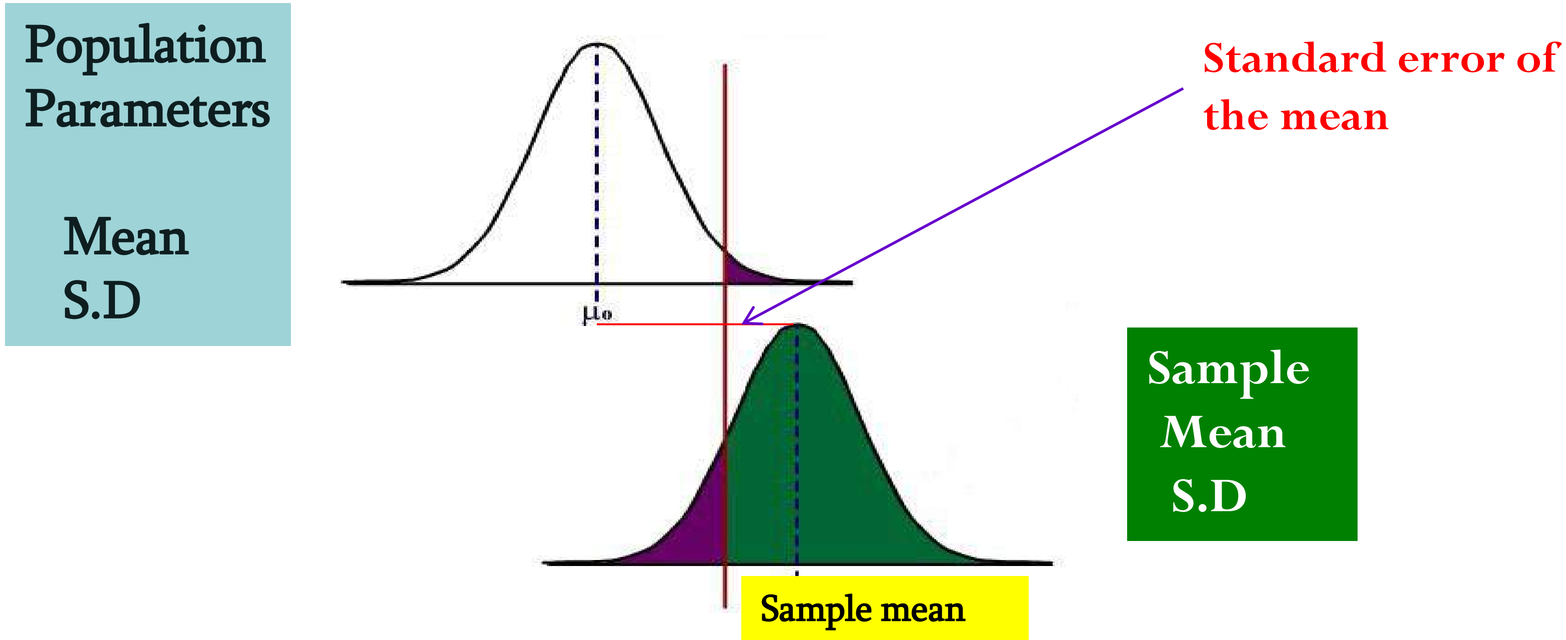
Most of the samples will cluster around the population parameters (population mean) with occasional sample result falling relatively further to one side or the other of the distribution ***(this called the sampling distribution of sample means).***

I- The mean of the sampling distribution is equal to the population mean,

The average of the averages (μχ) will be the same as the population mean.

II- The standard deviation of the sample means = the standard error (of the means)

SE= $\sigma/\sqrt{n}$, (σ= population SD).



The degree the sample statistics are deviating /different from the population parameters.

The term error indicates the fact that due to sampling error, each sample mean is likely to deviate some what from true population mean.

III- The distribution of the sample means is Normal if the population distribution is Normal.

IV- If the population distribution is Not Normal,

The distribution of the sample means is almost Normal when n is large *(Central Limit Theorem)*.

Central Limit Theorem

If you draw a sample from a population and calculate its mean, how close have you come to knowing the mean of the population?

How your mean expresses and close enough to the population mean?

When many samples are drawn from a population,

I- The means of these samples tend to be distributed normally and

II- The mean of the means is very close to the actual mean of the population.

├ **The *Z*-score can be calculated for these means** ┤

To calculate we need to know the S.D of the means,

$(\sigma/\sqrt{n})$ this new S.D of the means is called the

<u>standard error of the mean.</u>

Standard Error of the Means (SEM)

The formula for *SE= SD/√n*.

The formula indicates that we are estimating the SE given the SD of a sample of size *n*.

For a sample of 100 and S.D of 40 the SE= 40 /√100 = 4.

For a sample of 1000 and S.D of 40 the SE= 40 / √1000 = 1.26.

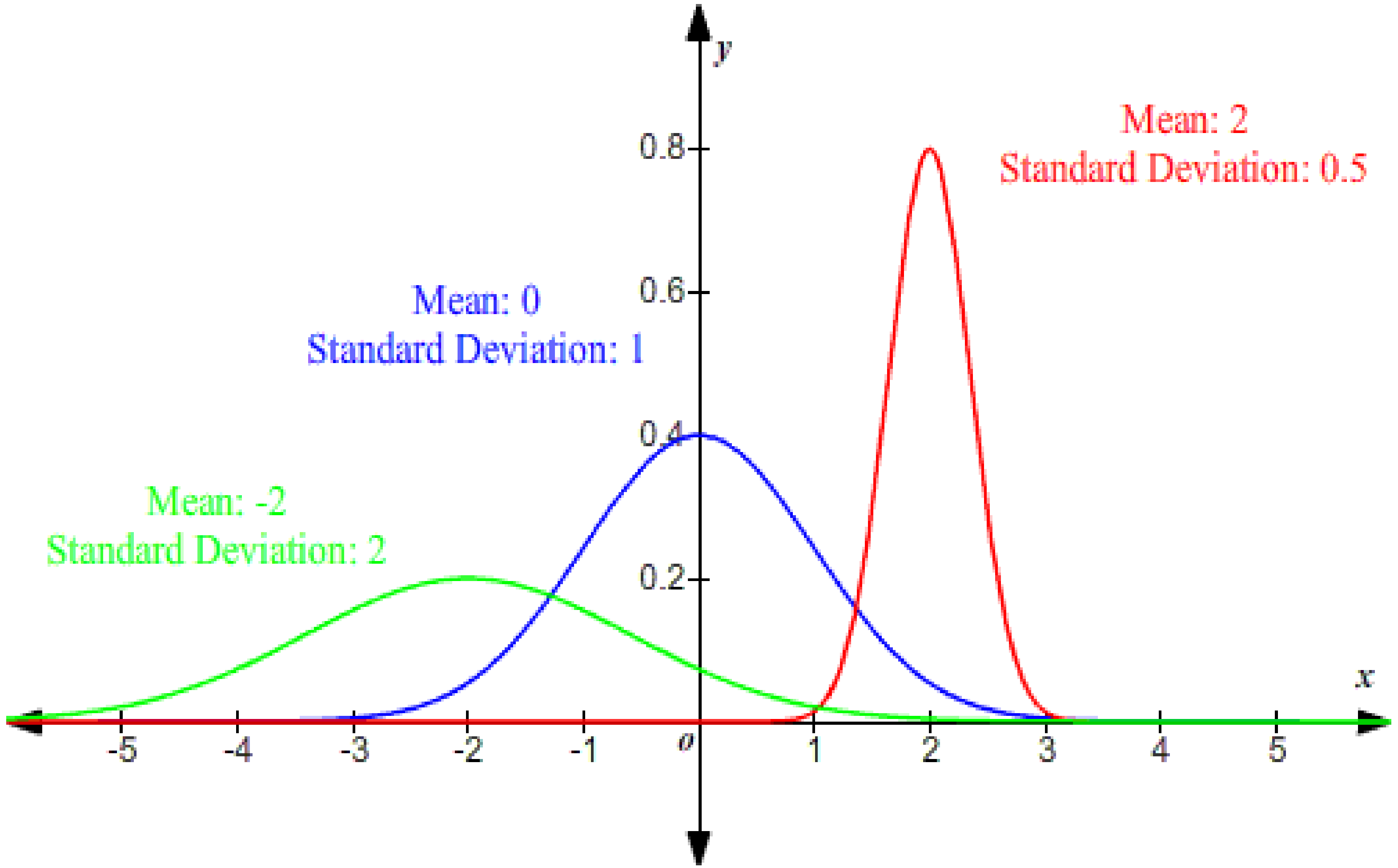
Standard Error of the Means

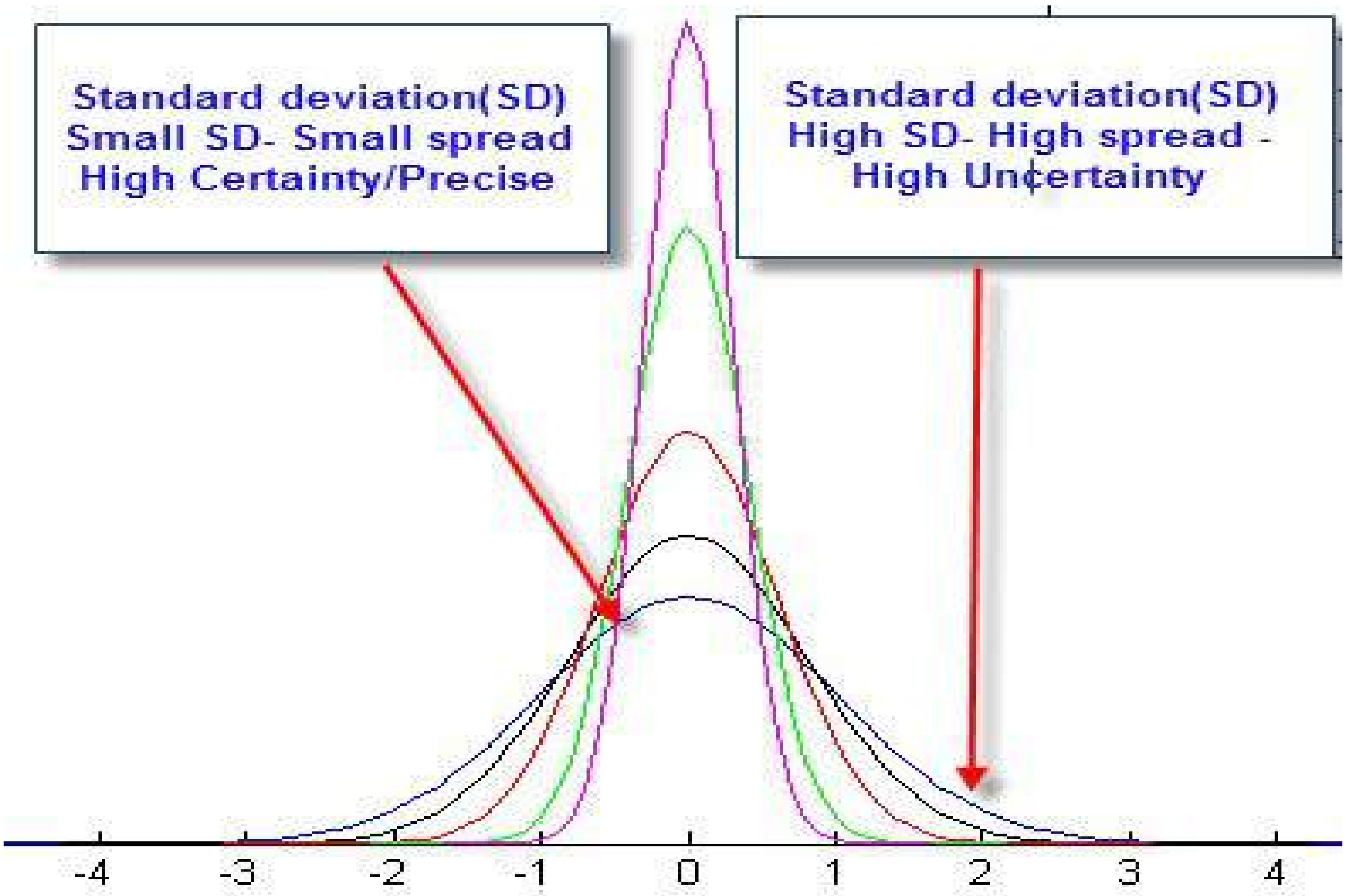
Two factors influence the SE, sample size and S.D of the sample:

<u>Sample size</u> has greater impact as it is used a denominator.

For a sample of 100 and S.D of 20 the SE = 20 / √100 = 2.
For a sample of 100 and S.D of 40 the SE = 40 / √100 = 4.

If there is more variability within a sample the greater the SE.


y
x
0.8
0.6
0.4
0.2
-5
-4
-3
-2
-1
o
1
2
3
4
5
Mean: 2
Standard Deviation: 0.5
Mean: 0
Standard Deviation: 1
Mean: -2
Standard Deviation: 2


Standard deviation(SD)
Small SD- Small spread
High Certainty/Precise
Standard deviation(SD)
High SD- High spread -
High Uncertainty
-4
-3
-2
-1
0
1
2
3
4

Conclusion of the central limit theorem

The distribution of samples (means, medians, variances, and most other statistical measures) approaches a normal distribution as the sample size, *n*, increases.

SPSS Demo.

Standard Deviation *Vs*. Standard Error

The value of σ measures the standard deviation in the population and is based on measurements of individuals "tell us how much variability can be expected among individuals".

SEM (the standard deviation of the means) tell us how much variability can be found among means of the samples.

SEM pertains to means of repeated samples not to individuals.

Estimation and Confidence

Random Error= Chance= Confidence Intervals

Estimation

In research we usually have a single sample and we need to estimate the population value from such sample.

The population parameters are unknown, and we assume that the sample actually fall within a certain region of the sampling distribution

- ***Our sample is one of the 95 % of all possible samples that fall within ±1.96 standard error from the population mean (Confidence Level), so our estimate of the population value is valid.***

Confidence Intervals (CI)

Confidence interval gives an *estimated range of values which is likely to include an unknown population parameter*, the estimated range being calculated from a given set of sample data .

c= level of confidence	Z c= Z critical values (under normal curve)
90%	1.645
95%	1.960
99%	2.578

$$\overline{\overline{\overline{\chi}}} \pm Z_c \left(\frac{\sigma}{\sqrt{n}} \right)$$

C.I= Mean of the sample ±Z critical scores (SEM)
SEM= SD/√n

Confidence Intervals

- Provides a range that is highly likely (often 95% or 99%) to contain the true population parameter that is being estimated.
- The narrower the interval the more informative is the result.
- Calculated using the estimate (sample mean) and its standard error (SEM).

CI for μ
Systolic blood pressure in 287 diabetic patients

Descriptives

			Statistic	Std. Error
syst. blood pressure at start	Mean		151.20	1.319
	90% Confidence Interval for Mean	Lower Bound	149.02	
		Upper Bound	153.38	
	5% Trimmed Mean		150.30	
	Median		150.00	
	Variance		483.880	
	Std. Deviation		21.997	
	Minimum		100	
	Maximum		220	
	Range		120	

90% C.I= 151.20±1.65 (21.997/√287)
C.I=149.02-153.38 mmHg

Descriptives

			Statistic	Std. Error
syst. blood pressure at start	Mean		155.06	3.064
	90% Confidence Interval for Mean	Lower Bound	149.92	
		Upper Bound	160.20	
	5% Trimmed Mean		154.72	
	Median		151.20	
	Variance		460.033	
	Std. Deviation		21.448	
	Minimum		115	
	Maximum		205	
	Range		90	
	Interquartile Range		30.00	
	Skewness		.263	.340
	Kurtosis		-.506	.668

Random sample of 50 out of 287

Descriptives

			Statistic	Std. Error
syst. blood pressure at start	Mean		151.20	1.319
	95% Confidence Interval for Mean	Lower Bound	148.60	
		Upper Bound	153.80	
	5% Trimmed Mean		150.30	
	Median		150.00	
	Variance		483.880	
	Std. Deviation		21.997	
	Minimum		100	
	Maximum		220	
	Range		120	
	Interquartile Range		30.00	
	Skewness		.540	.146

Descriptives

			Statistic	Std. Error
syst. blood pressure at start	Mean		155.06	3.064
	95% Confidence Interval for Mean	Lower Bound	148.90	
		Upper Bound	161.22	
	5% Trimmed Mean		154.72	
	Median		151.20	
	Variance		460.033	
	Std. Deviation		21.448	
	Minimum		115	
	Maximum		205	
	Range		90	
	Interquartile Range		30.00	
	Skewness		.263	.340
	Kurtosis		-.506	.668

95% C.I = 151.20±1.96(21.997/√287)

C.I=148.60-153.80 mmHg

Random Sample of 50 out of 287

Descriptives

			Statistic	Std. Error
syst. blood pressure at start	Mean		151.20	1.319
	99% Confidence Interval for Mean	Lower Bound	147.78	
		Upper Bound	154.62	
	5% Trimmed Mean		150.30	
	Median		150.00	
	Variance		483.880	
	Std. Deviation		21.997	
	Minimum		100	
	Maximum		220	
	Range		120	

99% C.I=151.20±2.58 (21.997/√287)
C.I=147.78-154.62 mmHg

Descriptives

			Statistic	Std. Error
syst. blood pressure at start	Mean		155.06	3.064
	99% Confidence Interval for Mean	Lower Bound	146.84	
		Upper Bound	163.28	
	5% Trimmed Mean		154.72	
	Median		151.20	
	Variance		460.033	
	Std. Deviation		21.448	
	Minimum		115	
	Maximum		205	
	Range		90	
	Interquartile Range		30.00	
	Skewness		.263	.340
	Kurtosis		-.506	.668

Random sample of 50 out of 287

After calculation of the previous examples

90% C.I= $151.20 \pm 1.65(21.997/\sqrt{287})$
C.I=149.02-153.38 mmHg

95% C.I=$151.20 \pm 1.96(21.997/\sqrt{287})$
C.I=148.60-153.80 mmHg

99% C.I=$151.20 \pm 2.58(21.997/\sqrt{287})$
C.I=147.78-154.62 mmHg

What does this mean? It means that if the same population is sampled on numerous occasions and interval estimates are made on each occasion, the resulting intervals would bracket the true population parameter (ranged) in approximately 90, 95 and 99 % of the cases.

Sample distribution of a proportion: Confidence intervals

$$\mu_p = \pi$$

$$SE(p) = \sqrt{\frac{p(1-p)}{n}}$$

$$p = K / n$$

$$CI_p = p \pm 1.96(SE)$$

Z critical score equal 95%

Smokers among diabetics

Sample=400

Smokers=40

P=40/400=0.1

SE $_{(p)}$ = √0.1-0.9/400=0.015

CI $_{p}$ 95%= 0.1±1.96(0.015)

[0.07-0.13]

for % it is the same SE=1.5% C.I=[7-13]

95% CI for the difference between two means (μ_1-μ_2)

Smoke	n	Mean SBP	SE (mean)
No	214	153.1	1.50
Yes	64	144.8	2.62
Difference		8.3	

$$\overline{\chi_1} - \overline{\chi_2} \pm 1.96 * SE(\overline{\chi_1} - \overline{\chi_2})$$

$$SE = \sqrt{(SE(\overline{\chi_1}))^2} + (SE(\overline{\chi_2}))^2$$

C.I= 2.4 to 14.2

95% CI for percentage

Smoke (n)	% died	SE
No (212)	28.8	3.11
Yes (64)	23.4	5.30

Difference= 5.4%

$$P_{ns} - P_s \pm 1.96 * SE(P_{ns} - P_s)$$

$$SE = \sqrt{\left[\frac{P_1 \times (100 - p1)}{n1} + \frac{P_2 \times (100 - p2)}{n2}\right]}$$

95% C.I=-6.7% to 17.4%

95% CI for RR and OR

Use available software

http://www.medcalc.org/calc/odds_ratio.php

http://www.medcalc.org/calc/relative_risk.php

xls.vl.academicdirect.org/applied statistics/.../Calculator

Confidence Intervals: conclusion

- **A 95% confidence interval is the estimated range of values within which it is 95% possible or likely that the precise or true population effect lies.** [4]

- Confidence intervals are a pivotal tool in evidence-based practice, because they allow results to be extrapolated into the relevant population.

4. Bolton J. Clinicians and the “S-Word”. Clin Chiropractic 2006;9:88e91.

Interpretation of Confidence Intervals

- **By using the 95 % confidence level, we are prepared to be wrong only five times in every 100 samples (1 in 20).**

- **The risk of not including the population parameter in our interval estimate is only 5 %.**

- **This probability of error is known as alpha level (ά) which is simply one minus the confidence level expressed as proportion,**

1- 0.95 = 0.05.

Changing the Confidence Level

A random sample of 200 nurses is taken and each nurse asked his or her annual income in whole dollars. These 200 nurses have an average income of $ 35,000, SD of $ 5000.

The 90 % CI will be : (34,415 to 35,585).

The interval width = 1170 $.

The 95 % CI = 35,000± 1.96 (5000/ √200) = 35,000 ± 695.

(34,305 to 35,695 $) = this is the estimate of the average income of all nurses with a 95 % confidence.

The interval width = 35,695-34,305 = $ 1390.

The 99 % CI = 35,000 ± 2.58 (5000 / √200) = 35,000 ± 915

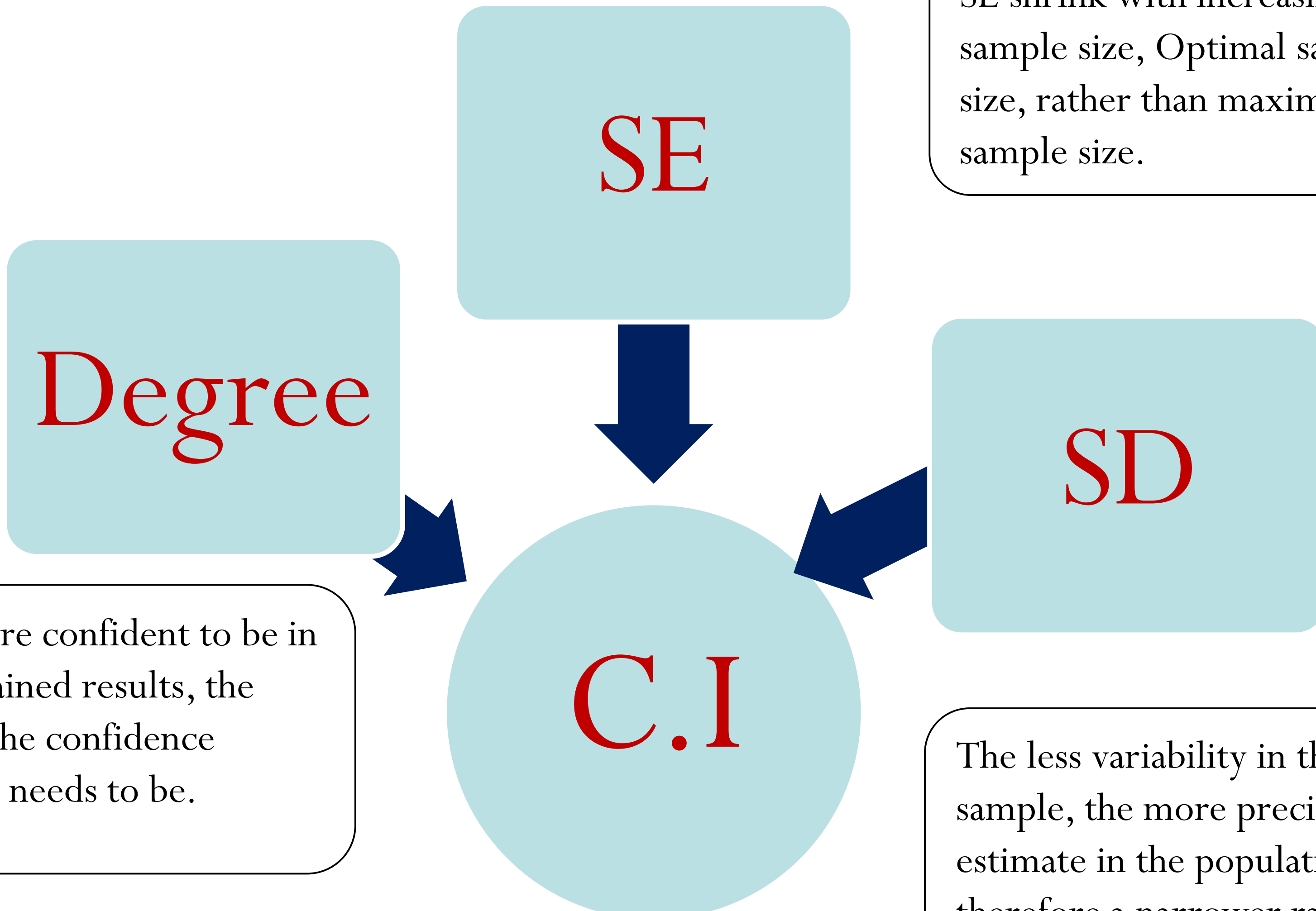
(34,085 to 35,915) = the interval width is 1830.

SPSS Demo

- If a 99% confidence interval is desired then the range will have to be wider, to cover the extra data that needs to be covered over and above the arbitrary 95%, to ensure that it is possible to be more confident that the average for the population (the population mean) lies within it.

- Conversely, if a 90% confidence interval is considered sufficient then the range of data required will be narrower, and hence the required sample size will be smaller.

Elements affecting the confidence intervals



SE shrink with increasing sample size, Optimal sample size, rather than maximal sample size.

The more confident to be in the obtained results, the higher the confidence interval needs to be.

The less variability in the sample, the more precise the estimate in the population and therefore a narrower range.

Hypothesis Testing

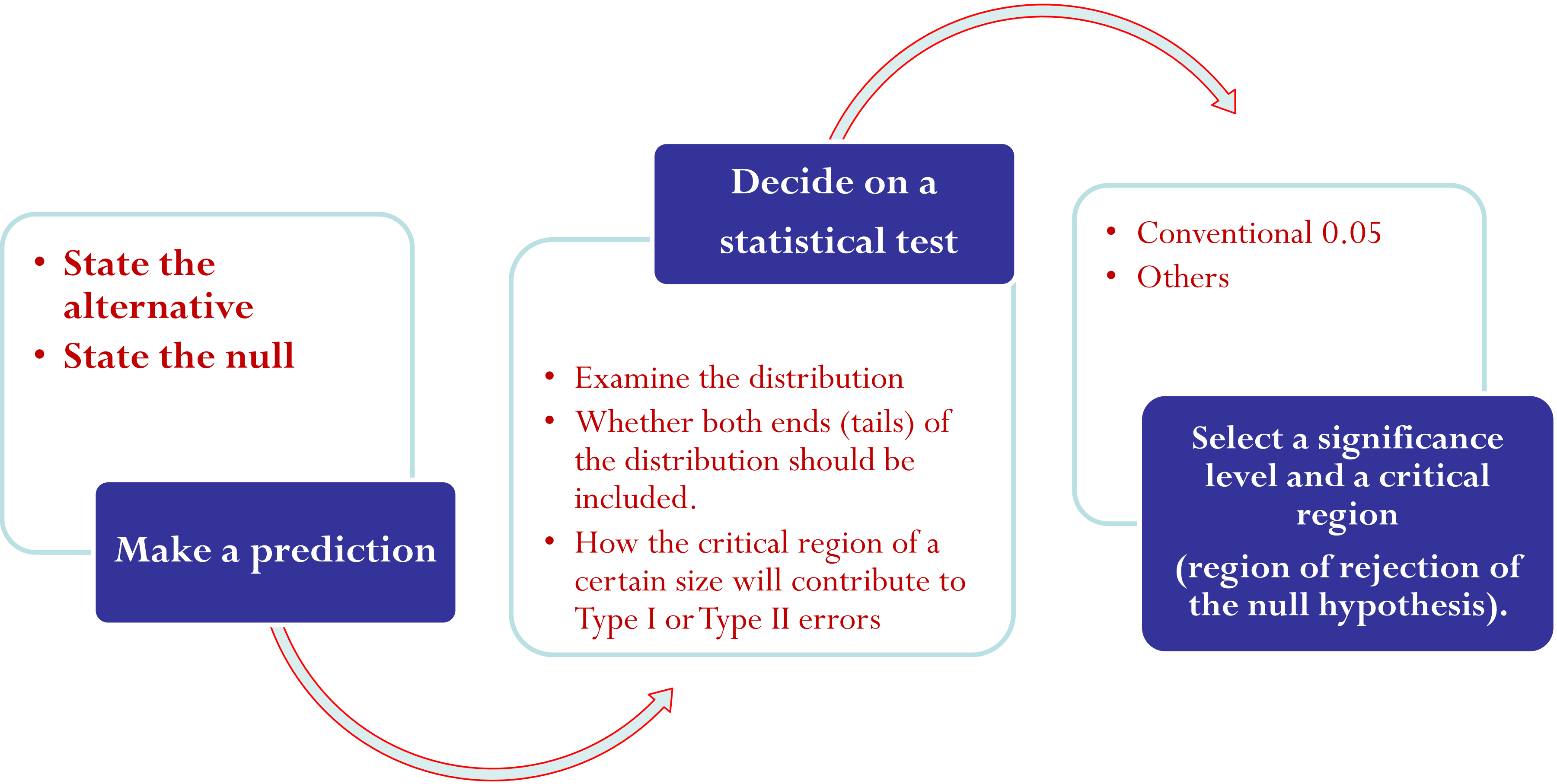
Sampling Distribution.

- **Sampling distributions – are theoretical distributions developed by mathematicians to organize statistical outcomes from various sample sizes**

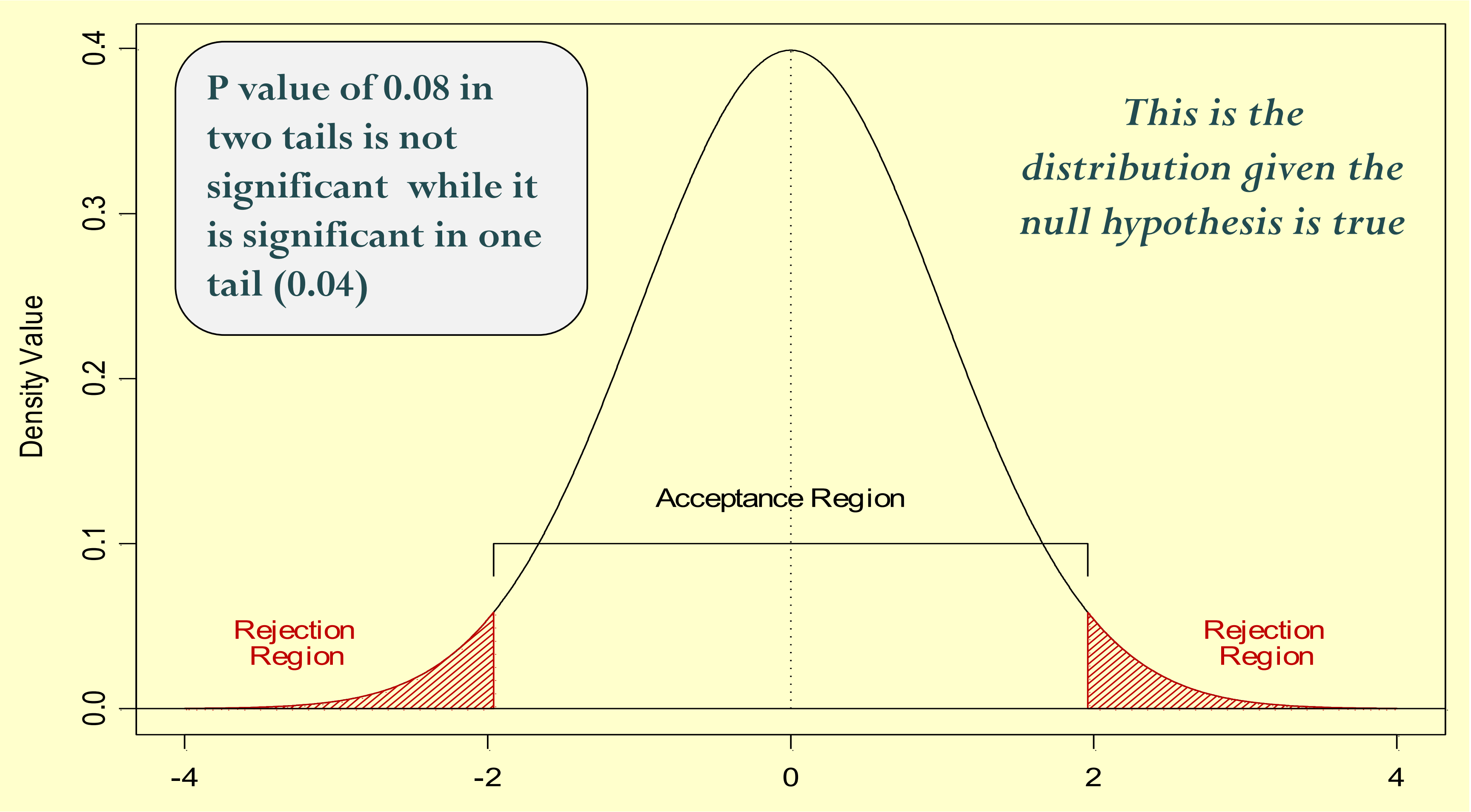
So we can determine the probability of (of an event/outcome) happening by chance in the population from which the sample was drawn.

They allow us to know the probability of occurrence of any value in the distribution.

Hypothesis testing: steps



Hypothesis Test



Hypothesis testing: steps

Computing the test statistic

- Test statistic is not a mean, *SD* or any form of descriptive data. It is a number that compared with a set of results predicted by the sampling distribution

Compare

- Test statistic to the sampling distribution and make a decision about the null hypothesis – reject it if the statistic falls in the region of rejection

Consider power of the test

- The probability of detecting a significant difference – parametric tests are more powerful

Statistics software will do all the calculation and comparison for you

TABLE 8—2. CRITICAL VALUES OF *t*

	Test	Level of Significance				
df	2-Tailed α:	.10	.05	.02	.01	.001
1		6.31	12.71	31.82	63.66	636.62
2		2.92	4.30	6.97	9.93	31.60
3		2.35	3.18	4.54	5.84	12.94
4		2.13	2.78	3.75	4.60	8.61
5		2.02	2.57	3.37	4.03	6.86
6		1.94	2.45	3.14	3.71	5.96
7		1.90	2.37	3.00	3.45	5.41
8		1.86	2.31	2.90	3.36	5.04
9		1.83	2.26	2.82	3.25	4.78
10		1.81	2.23	2.76	3.17	4.58
11		1.80	2.20	2.72	3.11	4.44
12		1.78	2.18	2.68	3.06	4.32
13		1.77	2.16	2.65	3.01	4.22
14		1.76	2.15	2.62	2.98	4.14
15		1.75	2.13	2.60	2.95	4.07
16		1.75	2.12	2.58	2.92	4.02
17		1.74	2.11	2.57	2.90	3.97
18		1.73	2.10	2.55	2.88	3.92
19		1.73	2.09	2.54	2.86	3.88
20		1.73	2.09	2.53	2.85	3.85
21		1.72	2.08	2.52	2.83	3.82
22		1.72	2.07	2.51	2.82	3.79
23		1.71	2.07	2.50	2.81	3.77
24		1.71	2.06	2.49	2.80	3.75
25		1.71	2.06	2.49	2.79	3.73
26		1.71	2.06	2.48	2.78	3.71
27		1.71	2.05	2.47	2.77	3.69
28		1.70	2.05	2.47	2.76	3.67
29		1.70	2.05	2.46	2.76	3.66
30		1.70	2.04	2.46	2.75	3.65
40		1.68	2.02	2.42	2.70	3.55
60		1.67	2.00	2.39	2.66	3.46
120		1.66	1.98	2.36	2.62	3.73
∞		1.65	1.96	2.33	2.58	3.29
df	1-Tailed α:	.05	.025	.01	.005	.0005

Prediction: the outcomes

Test statistics of the sample decides that the null hypothesis is	The truth in the population : null hypothesis is **True**	The truth in the population : null hypothesis is **False**
True (null accepted)	Correct decision	False negative False acceptance Type II error (sample size) Power problem
False (null rejected)	False positive False rejection Type I error (internal validity)	Correct decision

Possible outcomes of epidemiological study

Decision based on sample	REALITY: H_0 True *(No association)*	REALITY: H_1 True *(Yes association)*
Do not reject H_0 *(not statistically significant)*	**Correct decision**	**Type II (beta error)**
Reject H_0 *(statistically significant)*	**Type I (alpha error)**	**Correct decision**

Possible outcomes of any epidemiologic study

Decision based on sample	**REALITY** **H_0 True** *(No association)*	**H_1 True** *(Yes association)*
Do not reject H_0 *(not statistically significant)*	**Correct decision**	**Failing to find a difference when one exists** ***Type II*** ***(beta error)***
Reject H_0 *(statistically significant)*	**Finding a difference when there is none** ***Type I*** ***(alpha error)***	**Correct decision**

Possible outcomes of intervention study

REALITY

Decision based on sample	H_0 True (No association)	H_1 True (Yes association)
Do not reject H_0 *(not statistically significant)*	**Correct decision**	**Effective drug off the market (trial result is ineffective while actually it is)** *Type II* *(beta error)*
Reject H_0 *(statistically significant)*	Useless drug on the market (effective while actually not) *Type I* *(alpha error)*	**Correct decision**

Type I and Type II errors

Type I

α is the probability of committing type I error.

Set at 0.05 (or 5%)

[redacted]

Type II

β is the probability of committing type II error.

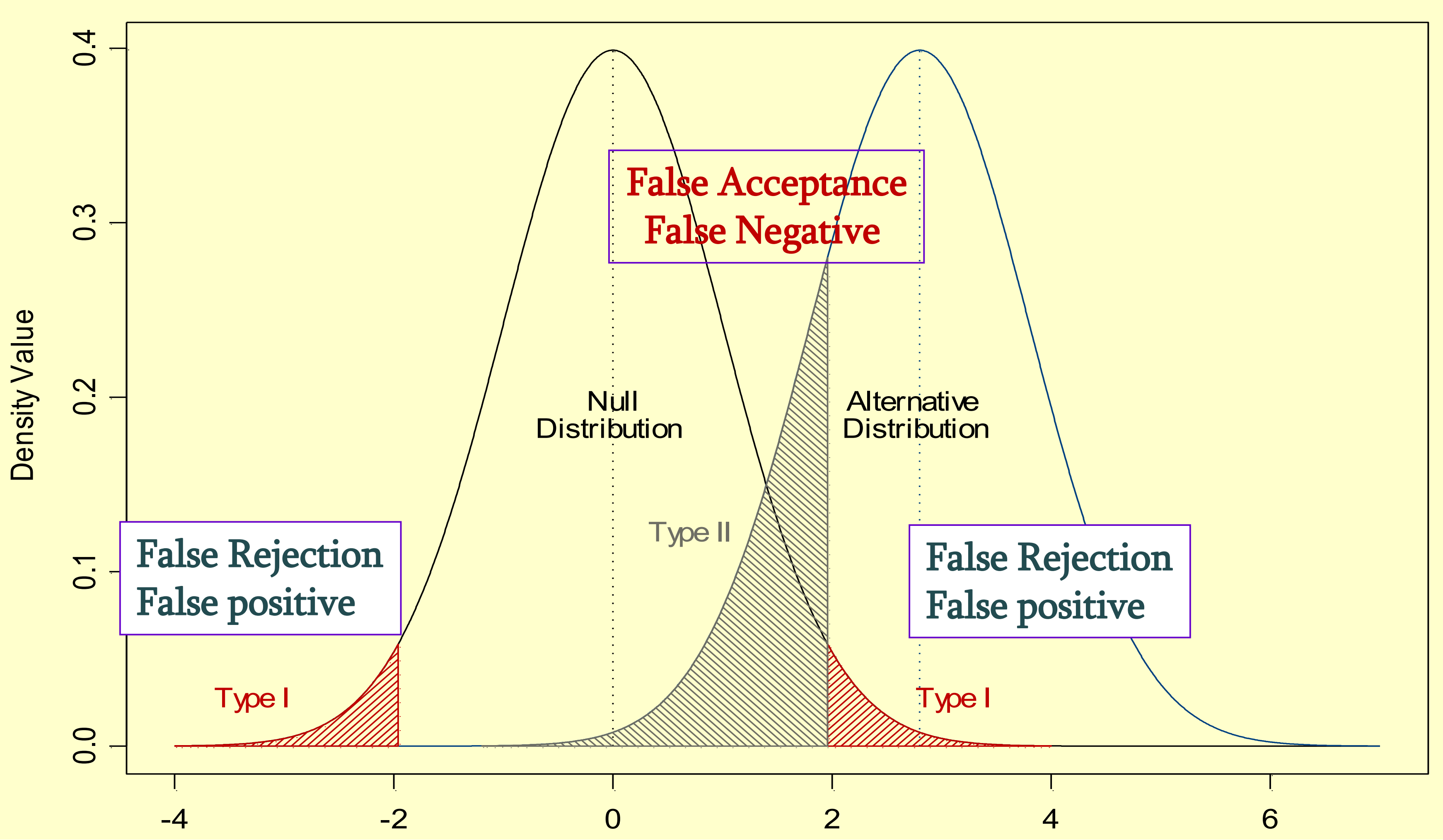
Set at 20% or 0.20

Concluding no difference, when it is actually false and there is a difference

Failure to find a difference

POWER

Type I and Type II Error



Alpha (Type I)

- Probability of concluding a difference between groups when there really is no.
- Statistically significant if probability of (alpha) is less than 5%.
- Error of rejecting null hypothesis if it is really true – or saying something is significant when it is not.

.

- With this error, a drug that does not work could get to market.
- Referred to as a False positive.

Beta (Type II)

- Probability of concluding no difference between treatment groups when there really is.
- Error of accepting null and concluding no difference, when it is actually false and there is a difference.
  - Failing to recognize a real difference. False Negative.
  - A potentially life saving drug off the market.

The Alpha-Fetoprotein (AFP) test has both Type I and Type II error possibilities.

This test screens the mother's blood during pregnancy for AFP and determines risk.
Abnormally high or low levels may indicate Down syndrome.

H_0: patient is healthy

H_a: patient is unhealthy

Error Type I (False positive or False Rejection) is: Test wrongly indicates that patient has a Down syndrome, which means that pregnancy must be aborted for no reason.
Error Type II (False negative or False Acceptance) is: Test is negative and the child will be born with multiple anomalies

P-value (The level of statistical significance)

- **A value of P <0.05 means that the probability that the result is due to chance is less than 1 in 20 and is the same as alpha < 0.05.**

- **The smaller the P-value, the greater your confidence in the statistical result.**

- ***Alpha does not change whereas P values are dependent on the actual value of the statistic in question.***

Difference between an alpha level and a P-value.

- In research, a relatively conservative standard is set to claim that a discovery is made *(of some phenomenon or some research question).*
- The standard is the alpha level, usually set of .05.
- Assuming the null hypothesis is true, e.g. we reject the null only if the observed data are so unusual *(would occurred by chance at most 5 % of the time).*
- The smaller the alpha, the more stringent the test *(the more unlikely it is to find a statistically significant result).*

Difference between an alpha level and a P-value.

- Once alpha level been set, statistic (like *t or Z*) is computed.

Alpha sets the standard for statistical significance, (yes or no) – whether or not we can reject the null hypothesis.

- Each statistic has an associated probability value called a P-value, or ***the likelihood of an observed statistic occurring due to chance, given the sampling distribution.***

The P-value indicates the actual level of how extreme the data are.

Statistical Power

The ability of a study to detect a significant difference between treatment groups

The probability that a study will have a statistically significant result ($P<0.05$).

Power = 1- beta (the false-negative rate).

By convention, adequate study power is usually set at 0.8 (80%), beta of 0.2 (a false-negative rate of 20%).

Power increases as sample size increases.

The power of a study should be stated in the methods section of a study report.

Statistical Power

- The "confidence" in research results.
  - Power is (1-Type II) error
    - Its 1 – the chance you got it wrong = the probability you got it right.
    - Typically we see Alpha 0.05, Beta 0.20, Power is 80%
    - While alpha = 0.05 is an absolute according to most statistical experts, power is not.
    - Power analysis is used in sample size planning and can be used for hypothesis testing.

Statistical Power

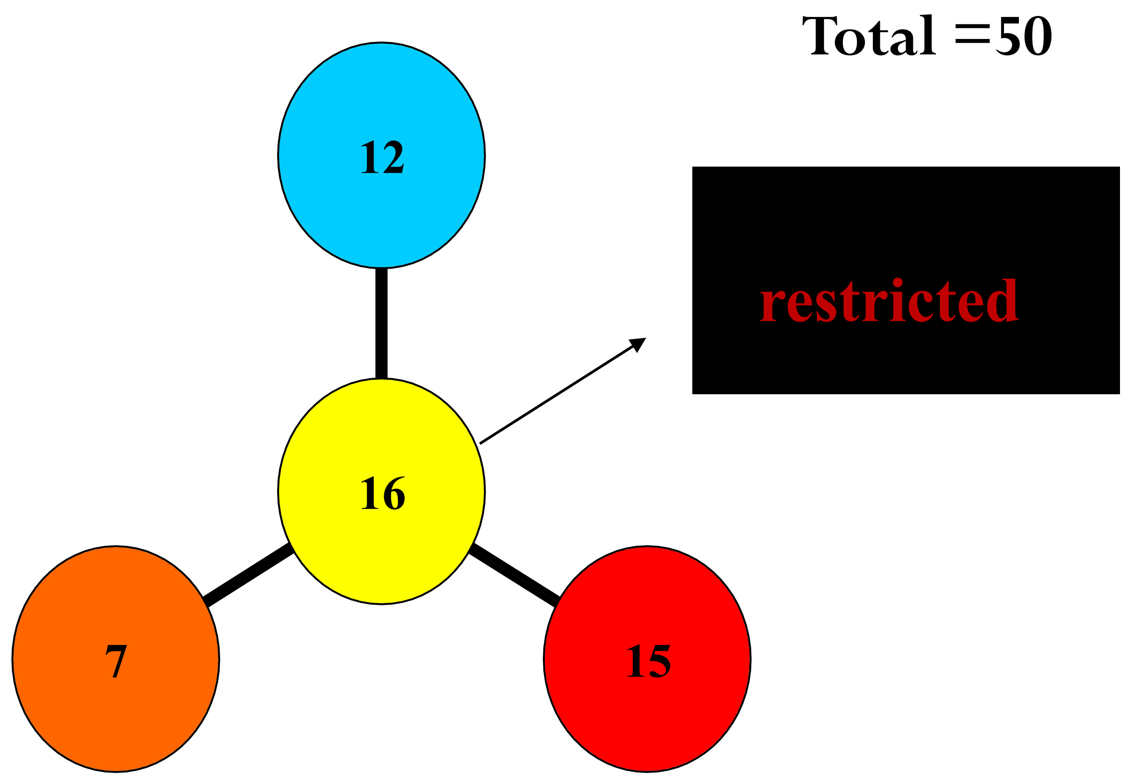
To calculate power all you need is:

1- Desired alpha level

2- Estimate of how big the effect is in the population.

3- Estimate of the variability.

Degree of freedom: concept



$df = \text{n-1}$

Degrees of Freedom

- One sample t-test or paired t-test = N-1
- Independent t-test = N-2
- Chi-square test = (# rows - 1) x (# columns – 1)
- ANOVA :

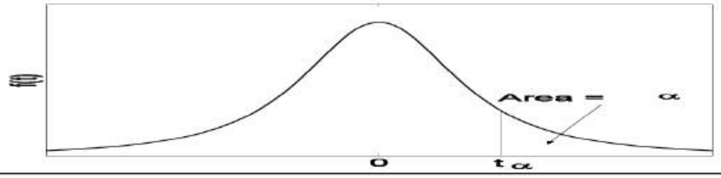
  df between groups = (# levels or groups – 1)

  df within groups = (# subjects - # of levels)
- Correlations = N-2

t-distribution

Table B

Critical Values of the Student *t* Distribution



d.f.	α						
	0.100	**0.050**	**0.025**	**0.010**	**0.005**	**0.001**	**0.0005**
1	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088	636.6193
2	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271	31.5991
3	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.9240
4	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
40	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510
60	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602
120	1.2886	1.6577	1.9799	2.3578	2.6174	3.1595	3.3735
∞	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905

Levels of Measurement

- Four levels or scales of measurement, Each is classified according to certain characteristics.
- Data fall in the first level are limited to certain statistical tests.
- Choices of statistical tests (and power of the tests) increase as the levels go up.

Nominal

Weakest – numbers used to classify a class into mutually exclusive subclasses

Test hypotheses by Chi-square test.

Ordinal

Shows relationships among classes, Allows ranking

Test hypotheses using non-parametric statistics of order and ranking.

Interval

Distance between two numbers is of a known size.

Numbers have absolute values and interval between each number is equal.

Variable is continuous, no true zero –All parametric tests can be used

Ratio

Like interval scale but it has a true zero point as its origin. Time, length and weight are ratio scales when used alone, but not as a characteristic of a person.
Arithmetic, all parametric tests and geometric means can be used with ratio data.

Tests of Significance

Tests

Non-parametric

Parametric

Parametric: pre-requisites

Normality

Create a frequency histogram of each sample.

Normal probability plots

Tests for normality Kolmogorov-Smirnov test, Lilliefors test and chi-square goodness-of-fit test.

Equal variances

Uniform variance or homogeneity.

Are the standard deviations for the samples similar?

Several tests that can be used

F-test and the Levene test.

Non-parametric Tests

Binary, ordinal or nominal data,

+

Interval ratio data that is not normally distributed or, does not exhibit equality of variance.

→

Testing is carried out on ranked data, even when interval data is analyzed, lose information about the magnitude of differences within the scale data.

Less powerful than parametric testing, with the reduced power increasing the chance of a type II error.

Choosing a statistical test

Level of the data

Binary, nominal, ordinal, interval/ratio.

Number of groups in the investigation

The data collected from independent groups/ dependent measurements

Distribution of the data

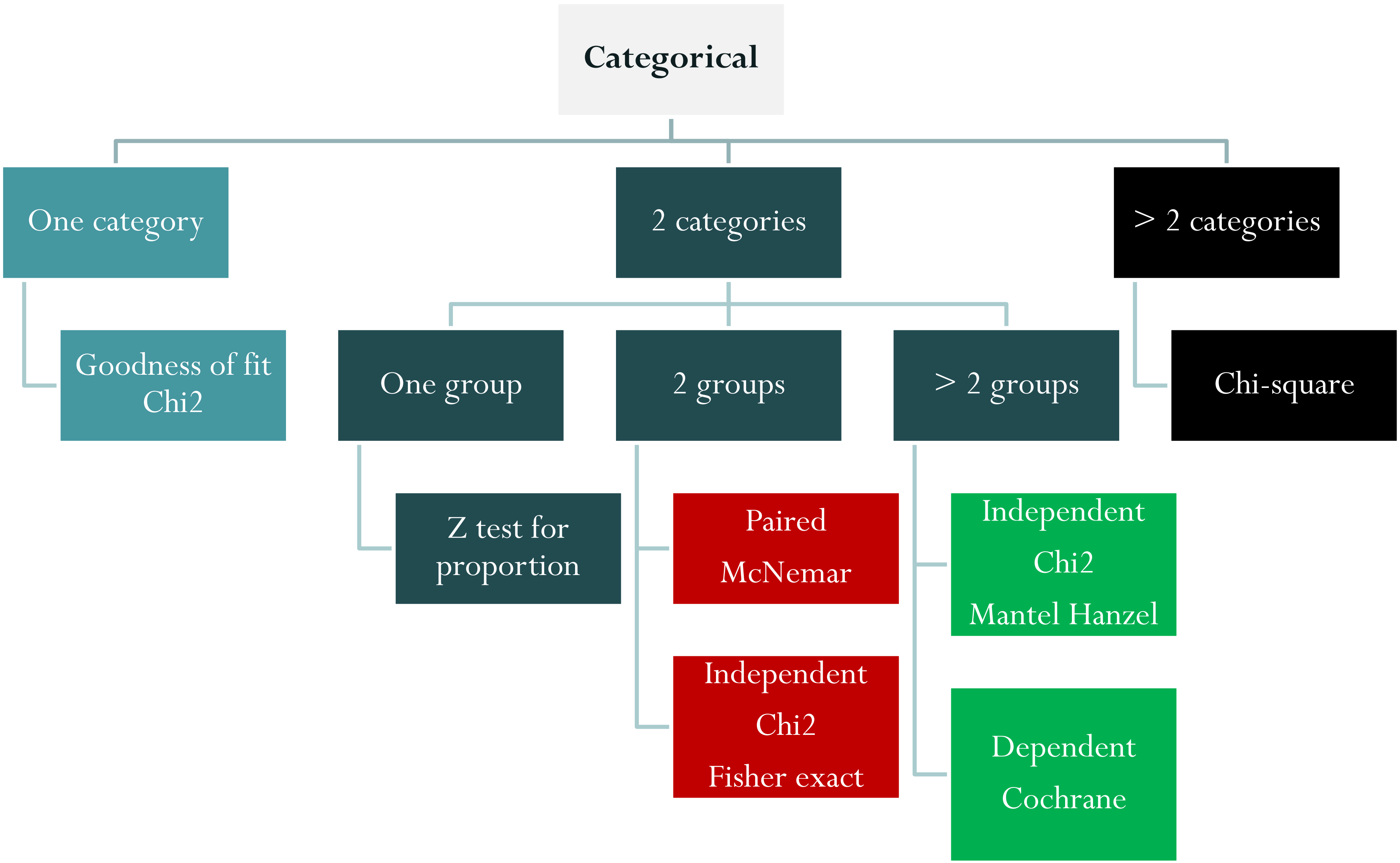
Test designed to investigate a correlation or difference

Tests of Significance

Type of data	Type of sample(s)	Statistical tests
Nominal	One or paired two independent samples	McNemar's test Chi-Square, Odds ratio
Ordinal	One or paired two independent samples	Stuart test Chi-Square
Interval	One or paired two independent samples	Wilcoxon test Mann Whitney U-test
Ratio	One or paired two independent samples	Wilcoxon test (non-normal) Paired t-test (normal) Mann Whitney U-test (non-normal) Unpaired t-test (normal)

Numerical data

- **One group**
  - One sample t test
  - Sign test
- **2 groups**
  - **Paired**
    - **t-test**
    - **Sign test**
    - **Wilcoxon sign rank test**
  - Independent
    - Unpaired t test
    - Mann Whitney
- **> 2 groups**
  - **Dependent**
    - Repeated measure ANOVA
    - Friedman test
  - Independent
    - One way ANOVA
    - Kruskall Wallis test


Categorical
One category
2 categories
> 2 categories
Goodness of fit
Chi2
One group
2 groups
> 2 groups
Chi-square
Z test for proportion
Paired
McNemar
Independent
Chi2
Fisher exact
Independent
Chi2
Mantel Hanzel
Dependent
Cochrane

Tests of Significance

- **Parametric tests of significance:**

- *At least 30 observations,*
- *Population assumed to be normally distributed,*
- *Variables are at least in an interval scale.*

Tests of Significance

- Non-parametric tests of significance:

- *Small numbers,*
- *Can't assume a normal distribution, or*
- *Measurement not interval*
- *Using scores for measurements*

Testing for Normality

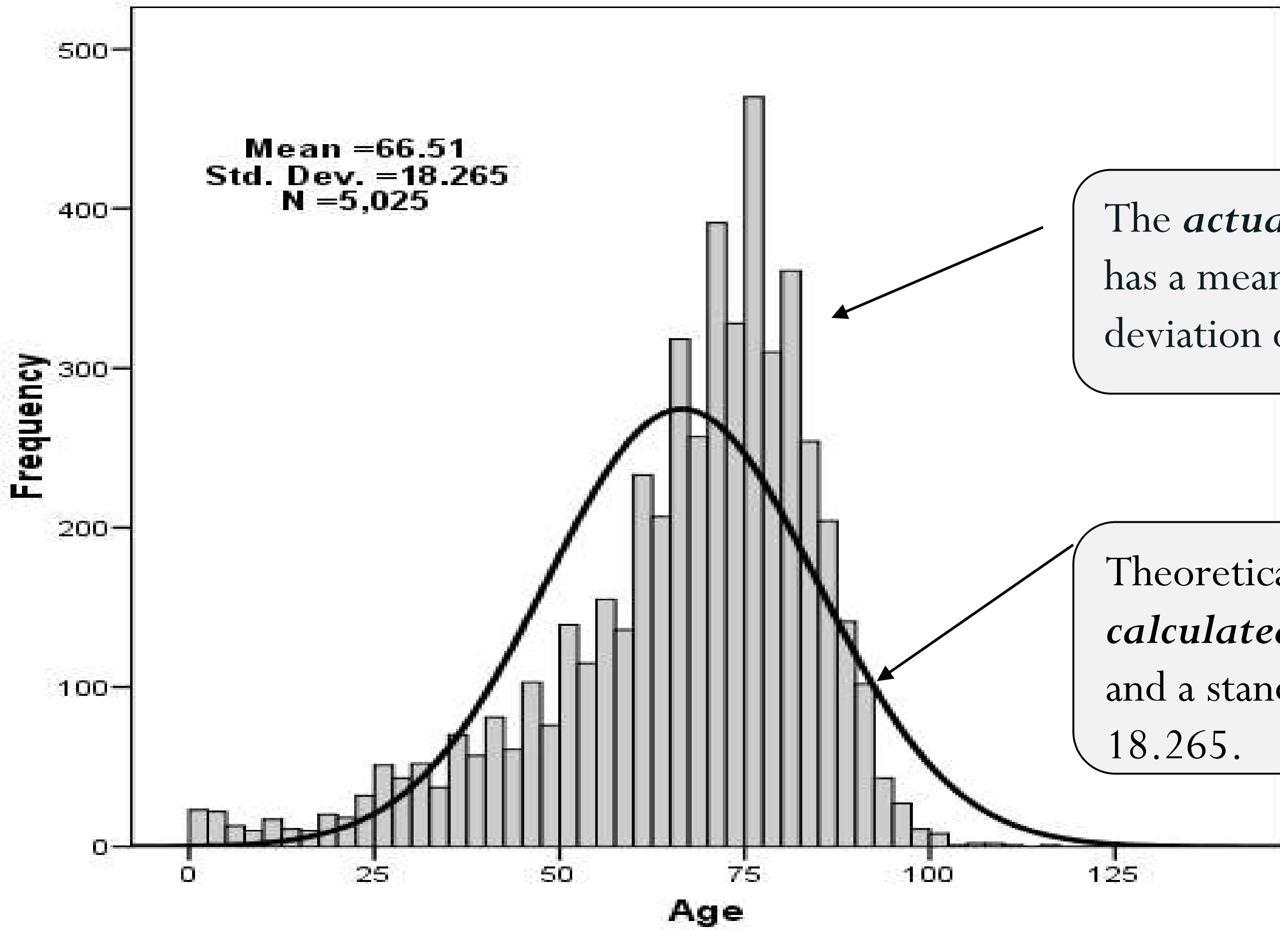
When is non-normality a problem?

- When the sample size is small (< 50).
- Highly skewed data create problems.
- Highly leptokurtic data are problematic, but not as much as skewed data.
- Serious concern when there is "activity" in the tails of the data set.

  Outliers are a problem.

  "Clumps" of data in the tails are worse.

Testing for normality



The ***actual data*** distribution that has a mean of 66.51 and a standard deviation of 18.265.

Theoretical normal distribution ***calculated from*** a mean of 66.51 and a standard deviation of 18.265.

Are the actual data statistically different than the computed normal curve?

Methods

Graphical

- Q-Q probability plots
- Cumulative frequency (P-P) plots

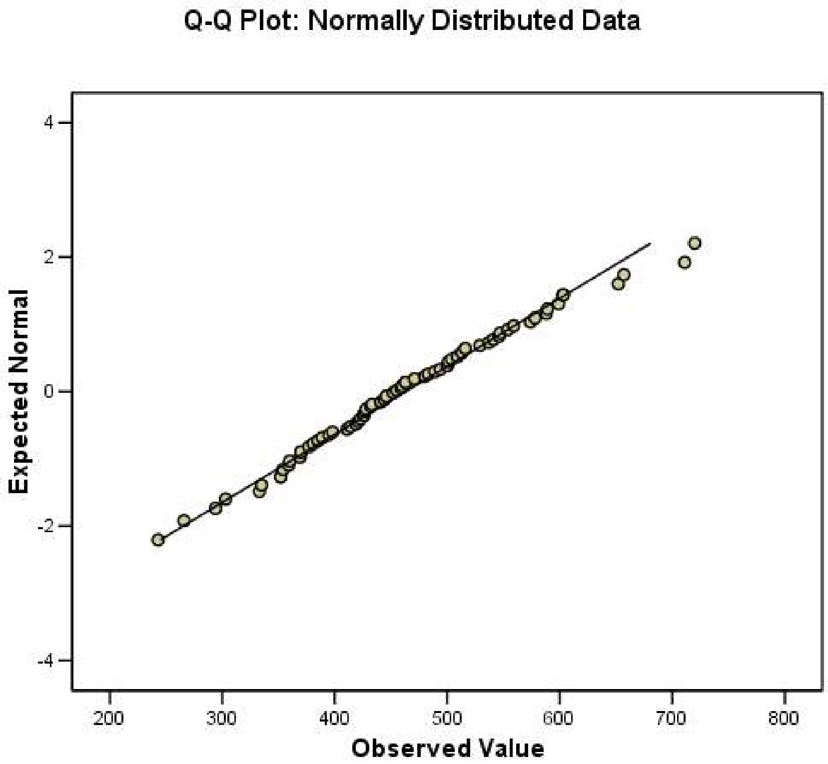
Statistical

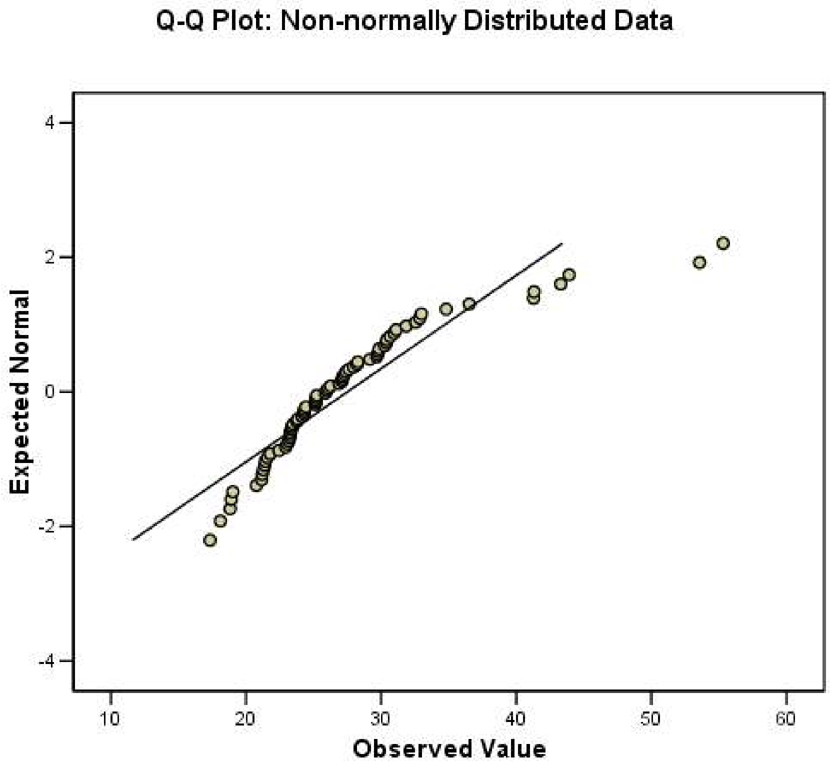
- W/S test
  Jarque-Bera test
  D'Agostino test

  Not available on SPSS

- Shapiro-Wilks test
  Kolmogorov-Smirnov test

Q-Q plots display the observed values against normally distributed data (represented by the line).

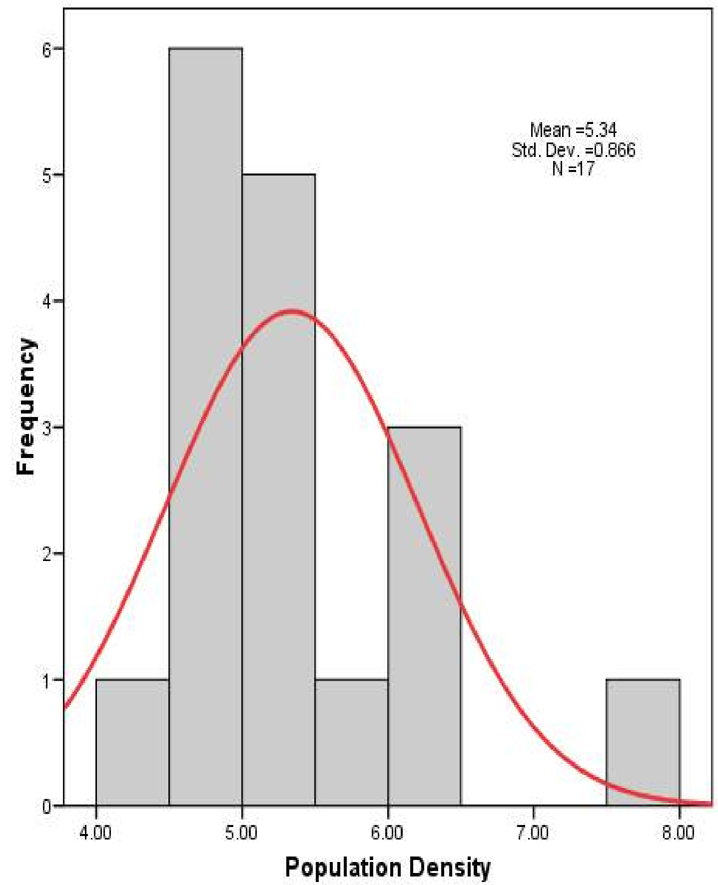




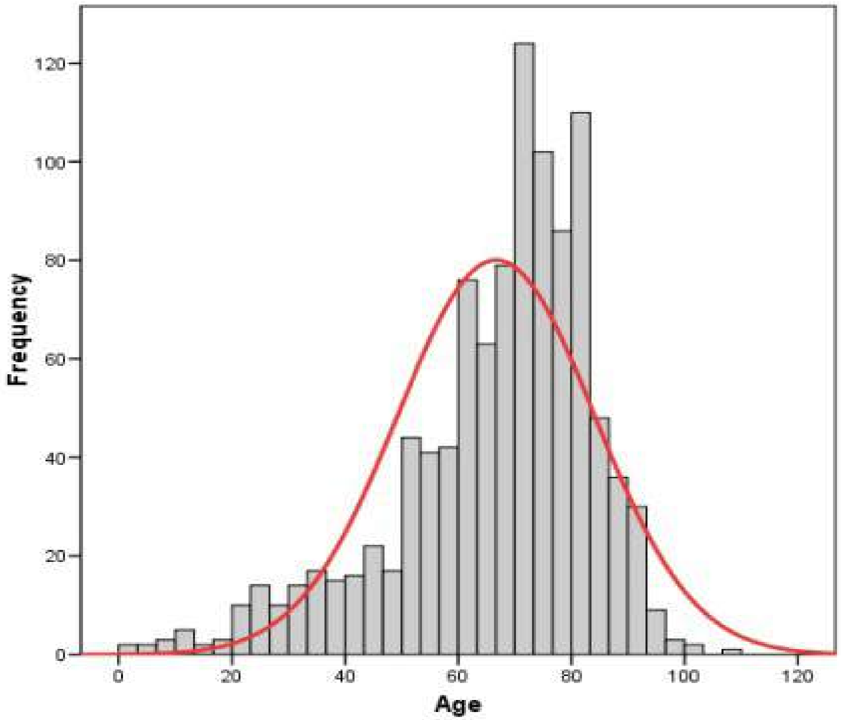
Normally distributed data fall along the line.

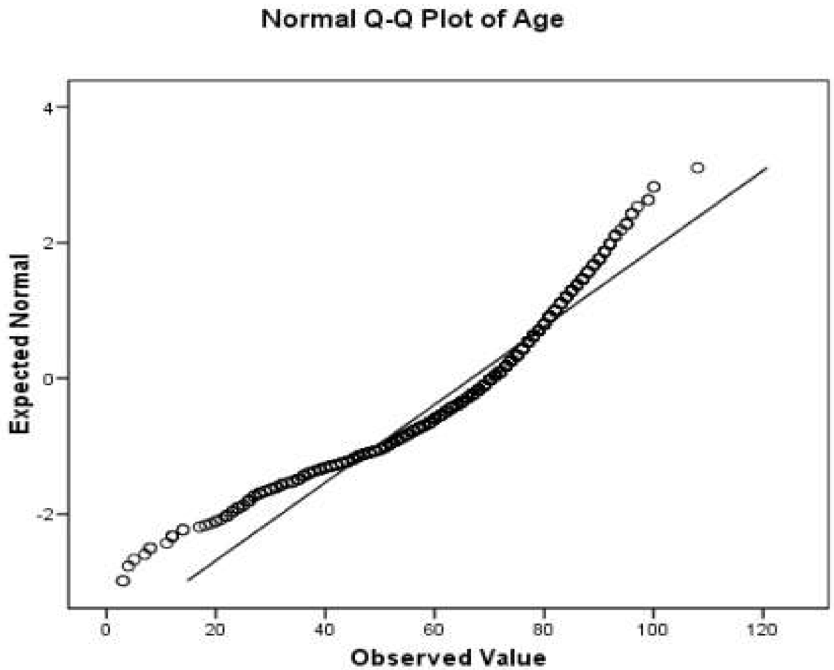
Graphical methods: disadvantages

- Typically not very useful when the sample size is small.
- This histogram of the data depicts that the data do not 'look' normal, but they are not statistically different than normal.

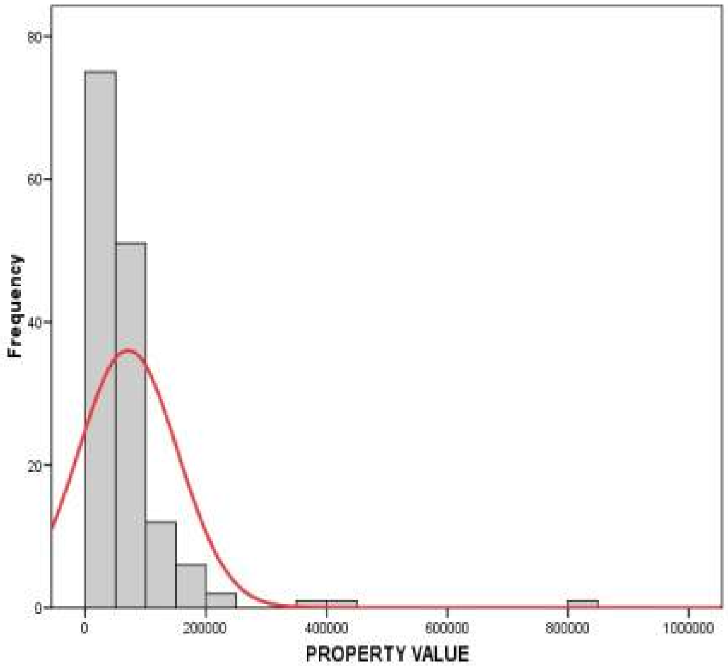


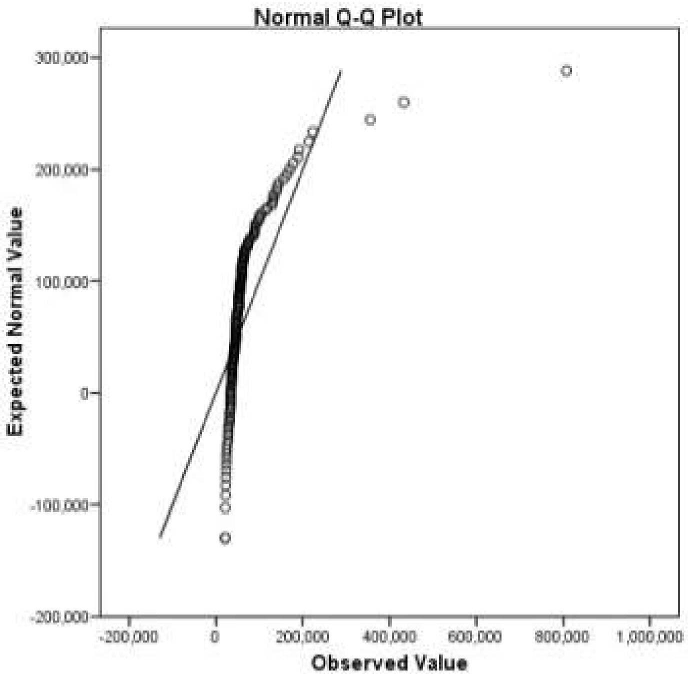
(Kolmogorov, P=0.221)



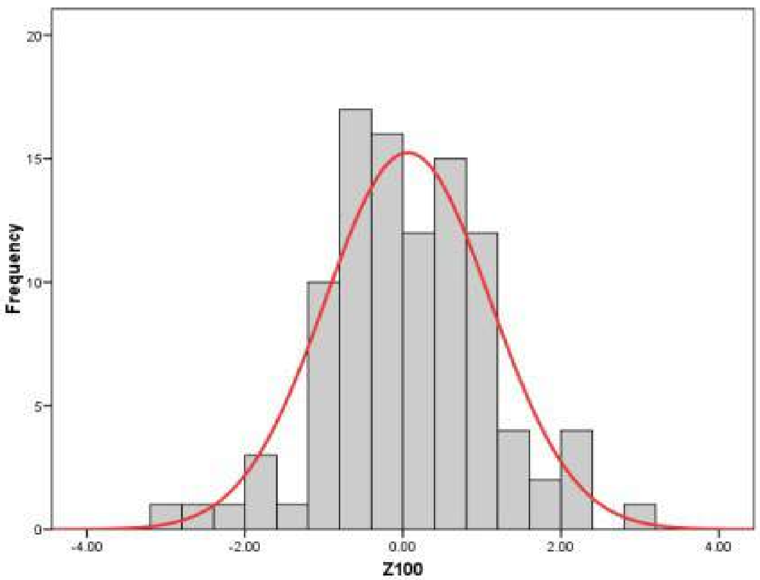


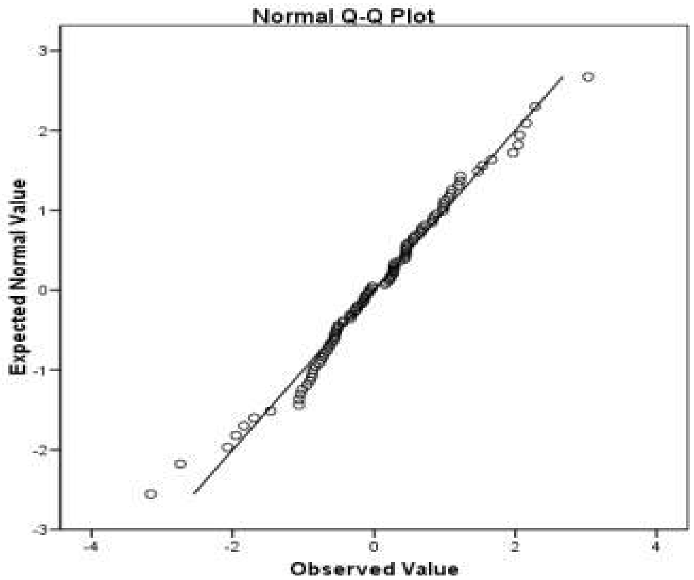
Df=1100, Kolmogorov/Shapiro, P=0.000





Df=149, Kolmogorov/Shapiro, P=0.000





Df=100, Kolmogorov/Shapiro, P=0.331

Statistical tests for Normality

- More precise since actual probabilities are calculated.
- Calculate the probability that the sample was drawn from a normal population.
- Hypotheses used are:

- *Ho: The sample data are not significantly different than a normal population.*

- *Ha: The sample data are significantly different than a normal population.*

Statistical Methods

- Typically, interested in finding a difference between groups, 'looking' for small probabilities.
- Testing normality, not 'looking' for a difference.
- The data set to be NO DIFFERENT than normal.
- Accept the null hypothesis.
- When testing for normality:
- *Probabilities > 0.05 mean the data are normal.*
- *Probabilities < 0.05 mean the data are NOT normal.*

SPSS Demo.

- *Remember that LARGE probabilities denote normally distributed data.*

Important

- ***Important*: As the sample size *increases*, normality parameters becomes *MORE* restrictive and it becomes harder to declare that the data are normally distributed.

- ***For very large data sets, normality testing becomes less important.***

Simple Tests for Normality: W/S Test for Normality

- A fairly simple test requires only sample standard deviation and data range.
- Based on the q statistic, which is the 'studentized' (meaning t distribution) range, or the range expressed in standard deviation units.
- Tests kurtosis. where *q*is the test statistic, *w*is the range of the data and the standard deviation.*swq*

W/S test

- Uses a critical range.

- IF the calculated value falls WITHIN the range, then accept Ho.

- ***IF the calculated value falls outside the range then reject Ho.***

- Since using a critical range, it is difficult to determine a probability range for the results, simply state alpha level.
- *The sample data set is not significantly different than normal (W/S4.16, $p > 0.05$)*

Critical Values of q for the W/S Normality Test

Taken from Kanji, 1994 Table 14

Columns *a* denote the lower boundaries or the left-sided critical values.
Columns *b* denote the upper boundaries or the right-sided critical values.

n	Level of significance α											
	0.000		0.005		0.01		0.025		0.05		0.10	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*	*a*	*b*
3	1.732	2.000	1.735	2.000	1.737	2.000	1.745	2.000	1.758	1.999	1.782	1.997
4	1.732	2.449	1.82	2.447	1.87	2.445	1.93	2.439	1.98	2.429	2.04	2.409
5	1.826	2.828	1.98	2.813	2.02	2.803	2.09	2.782	2.15	2.753	2.22	2.712
6	1.826	3.162	2.11	3.115	2.15	3.095	2.22	3.056	2.28	3.012	2.37	2.949
7	1.871	3.464	2.22	3.369	2.26	3.338	2.33	3.282	2.40	3.222	2.49	3.143
8	1.871	3.742	2.31	3.585	2.35	3.543	2.43	3.471	2.50	3.399	2.59	3.308
9	1.897	4.000	2.39	3.772	2.44	3.720	2.51	3.634	2.59	3.552	2.68	3.449
10	1.897	4.243	2.46	3.935	2.51	3.875	2.59	3.777	2.67	3.685	2.76	3.57
11	1.915	4.472	2.53	4.079	2.58	4.012	2.66	3.903	2.74	3.80	2.84	3.68
12	1.915	4.690	2.59	4.208	2.64	4.134	2.72	4.02	2.80	3.91	2.90	3.78
13	1.927	4.899	2.64	4.325	2.70	4.244	2.78	4.12	2.86	4.00	2.96	3.87
14	1.927	5.099	2.70	4.431	2.75	4.34	2.83	4.21	2.92	4.09	3.02	3.95
15	1.936	5.292	2.74	4.53	2.80	4.44	2.88	4.29	2.97	4.17	3.07	4.02
16	1.936	5.477	2.79	4.62	2.84	4.52	2.93	4.37	3.01	4.24	3.12	4.09
17	1.944	5.657	2.83	4.70	2.88	4.60	2.97	4.44	3.06	4.31	3.17	4.15
18	1.944	5.831	2.87	4.78	2.92	4.67	3.01	4.51	3.10	4.37	3.21	4.21
19	1.949	6.000	2.90	4.85	2.96	4.74	3.05	4.56	3.14	4.43	3.25	4.27
20	1.949	6.164	2.94	4.91	2.99	4.80	3.09	4.63	3.18	4.49	3.29	4.32
25	1.961	6.93	3.09	5.19	3.15	5.06	3.24	4.87	3.34	4.71	3.45	4.53
30	1.966	7.62	3.21	5.40	3.27	5.26	3.37	5.06	3.47	4.89	3.59	4.70
35	1.972	8.25	3.32	5.57	3.38	5.42	3.48	5.21	3.58	5.04	3.70	4.84
40	1.975	8.83	3.41	5.71	3.47	5.56	3.57	5.34	3.67	5.16	3.79	4.96
45	1.978	9.38	3.49	5.83	3.55	5.67	3.66	5.45	3.75	5.26	3.88	5.06
50	1.980	9.90	3.56	5.93	3.62	5.77	3.73	5.54	3.83	5.35	3.95	5.14
55	1.982	10.39	3.62	6.02	3.69	5.86	3.80	5.63	3.90	5.43	4.02	5.22
60	1.983	10.86	3.68	6.10	3.75	5.94	3.86	5.70	3.96	5.51	4.08	5.29
65	1.985	11.31	3.74	6.17	3.80	6.01	3.91	5.77	4.01	5.57	4.14	5.35
70	1.986	11.75	3.79	6.24	3.85	6.07	3.96	5.83	4.06	5.63	4.19	5.41
75	1.987	12.17	3.83	6.30	3.90	6.13	4.01	5.88	4.11	5.68	4.24	5.46
80	1.987	12.57	3.88	6.35	3.94	6.18	4.05	5.93	4.16	5.73	4.28	5.51
85	1.988	12.96	3.92	6.40	3.99	6.23	4.09	5.98	4.20	5.78	4.33	5.56
90	1.989	13.34	3.96	6.45	4.02	6.27	4.13	6.03	4.24	5.82	4.36	5.60
95	1.990	13.71	3.99	6.49	4.06	6.32	4.17	6.07	4.27	5.86	4.40	5.64
100	1.990	14.07	4.03	6.53	4.10	6.36	4.21	6.11	4.31	5.90	4.44	5.68
150	1.993	17.26	4.32	6.82	4.38	6.64	4.48	6.39	4.59	6.18	4.72	5.96
200	1.995	19.95	4.53	7.01	4.59	6.84	4.68	6.60	4.78	6.39	4.90	6.15
500	1.998	31.59	5.06	7.60	5.13	7.42	5.25	7.15	5.47	6.94	5.49	6.72
1000	1.999	44.70	5.50	7.99	5.57	7.80	5.68	7.54	5.79	7.33	5.92	7.11

Source: Sachs, 1972

Jarque–Bera Test

- A goodness-of-fit test of whether sample data have the skewness and kurtosis matching a normal distribution.

D'Agostino Test

A very powerful test for departures from normality.

Based on the D statistic, which gives an upper and lower critical value.

Critical Values for the D'Agostino D Normality Test
Taken from Zar, 1981 Table B.22

n	$\alpha =$ 0.20	0.10	0.05	0.02	0.01
10	0.2632, 0.2835	0.2573, 0.2843	0.2513, 0.2849	0.2436, 0.2855	0.2379, 0.2857
12	0.2653, 0.2841	0.2598, 0.2849	0.2544, 0.2854	0.2473, 0.2859	0.2420, 0.2862
14	0.2669, 0.2846	0.2618, 0.2853	0.2568, 0.2858	0.2503, 0.2862	0.2455, 0.2865
16	0.2681, 0.2848	0.2634, 0.2855	0.2587, 0.2860	0.2527, 0.2865	0.2482, 0.2867
18	0.2690, 0.2850	0.2646, 0.2855	0.2603, 0.2862	0.2547, 0.2866	0.2505, 0.2868
20	0.2699, 0.2852	0.2657, 0.2857	0.2617, 0.2863	0.2564, 0.2867	0.2525, 0.2869
22	0.2705, 0.2853	0.2670, 0.2859	0.2629, 0.2864	0.2579, 0.2869	0.2542, 0.2870
24	0.2711, 0.2853	0.2675, 0.2860	0.2638, 0.2865	0.2591, 0.2870	0.2557, 0.2871
26	0.2717, 0.2854	0.2682, 0.2861	0.2647, 0.2866	0.2603, 0.2870	0.2570, 0.2872
28	0.2721, 0.2854	0.2688, 0.2861	0.2655, 0.2866	0.2612, 0.2870	0.2581, 0.2873
30	0.2725, 0.2854	0.2693, 0.2861	0.2662, 0.2866	0.2622, 0.2871	0.2592, 0.2872
32	0.2729, 0.2854	0.2698, 0.2862	0.2668, 0.2867	0.2630, 0.2871	0.2600, 0.2873
34	0.2732, 0.2854	0.2703, 0.2862	0.2674, 0.2867	0.2636, 0.2871	0.2609, 0.2873
36	0.2735, 0.2854	0.2707, 0.2862	0.2679, 0.2867	0.2643, 0.2871	0.2617, 0.2873
38	0.2738, 0.2854	0.2710, 0.2862	0.2683, 0.2867	0.2649, 0.2871	0.2623, 0.2873
40	0.2740, 0.2854	0.2714, 0.2862	0.2688, 0.2867	0.2655, 0.2871	0.2630, 0.2874
42	0.2743, 0.2854	0.2717, 0.2861	0.2691, 0.2867	0.2659, 0.2871	0.2636, 0.2874
44	0.2745, 0.2854	0.2720, 0.2861	0.2695, 0.2867	0.2664, 0.2871	0.2641, 0.2874
46	0.2747, 0.2854	0.2722, 0.2861	0.2698, 0.2866	0.2668, 0.2871	0.2646, 0.2874
48	0.2749, 0.2854	0.2725, 0.2861	0.2702, 0.2866	0.2672, 0.2871	0.2651, 0.2874
50	0.2751, 0.2853	0.2727, 0.2861	0.2705, 0.2866	0.2676, 0.2871	0.2655, 0.2874
60	0.2757, 0.2852	0.2737, 0.2860	0.2717, 0.2865	0.2692, 0.2870	0.2673, 0.2873
70	0.2763, 0.2851	0.2744, 0.2859	0.2726, 0.2864	0.2708, 0.2869	0.2687, 0.2872
80	0.2768, 0.2850	0.2750, 0.2857	0.2734, 0.2863	0.2713, 0.2868	0.2698, 0.2871
90	0.2771, 0.2849	0.2755, 0.2856	0.2740, 0.2862	0.2721, 0.2866	0.2707, 0.2870
100	0.2774, 0.2849	0.2759, 0.2855	0.2745, 0.2860	0.2727, 0.2865	0.2714, 0.2869
120	0.2779, 0.2847	0.2765, 0.2853	0.2752, 0.2858	0.2737, 0.2863	0.2725, 0.2866
140	0.2782, 0.2846	0.2770, 0.2852	0.2758, 0.2856	0.2744, 0.2862	0.2734, 0.2865
160	0.2785, 0.2845	0.2774, 0.2851	0.2763, 0.2855	0.2750, 0.2860	0.2741, 0.2863
180	0.2787, 0.2844	0.2777, 0.2850	0.2767, 0.2854	0.2755, 0.2859	0.2746, 0.2862
200	0.2789, 0.2843	0.2779, 0.2848	0.2770, 0.2853	0.2759, 0.2857	0.2751, 0.2860
250	0.2793, 0.2841	0.2784, 0.2846	0.2776, 0.2850	0.2767, 0.2855	0.2760, 0.2858
300	0.2796, 0.2840	0.2788, 0.2844	0.2781, 0.2848	0.2772, 0.2853	0.2766, 0.2855
350	0.2798, 0.2839	0.2791, 0.2843	0.2784, 0.2847	0.2776, 0.2851	0.2771, 0.2853
400	0.2799, 0.2838	0.2793, 0.2842	0.2787, 0.2845	0.2780, 0.2849	0.2775, 0.2852
450	0.2801, 0.2837	0.2795, 0.2841	0.2789, 0.2844	0.2782, 0.2848	0.2778, 0.2851
500	0.2802, 0.2836	0.2796, 0.2840	0.2791, 0.2843	0.2785, 0.2847	0.2780, 0.2849
600	0.2804, 0.2835	0.2799, 0.2839	0.2794, 0.2842	0.2788, 0.2845	0.2784, 0.2847
700	0.2805, 0.2834	0.2800, 0.2838	0.2796, 0.2840	0.2791, 0.2844	0.2787, 0.2846
800	0.2806, 0.2833	0.2802, 0.2837	0.2798, 0.2839	0.2793, 0.2842	0.2790, 0.2844
900	0.2807, 0.2833	0.2803, 0.2836	0.2799, 0.2838	0.2795, 0.2841	0.2792, 0.2843
1000	0.2808, 0.2832	0.2804, 0.2835	0.2800, 0.2838	0.2796, 0.2840	0.2793, 0.2842
1250	0.2809, 0.2831	0.2806, 0.2834	0.2803, 0.2836	0.2799, 0.2839	0.2797, 0.2840
1500	0.2810, 0.2830	0.2807, 0.2833	0.2805, 0.2835	0.2801, 0.2837	0.2799, 0.2839
1750	0.2811, 0.2830	0.2808, 0.2832	0.2806, 0.2834	0.2803, 0.2836	0.2801, 0.2838
2000	0.2812, 0.2829	0.2809, 0.2831	0.2807, 0.2833	0.2804, 0.2835	0.2802, 0.2837

For each significance level, α is given a pair of critical values. If the calculated D is $\leq$ the first member of the pair, or $\geq$ the second, then, the null hypothesis of population normality is rejected.

Kolmogorov-Smirnov
Shapiro-Wilks

Kolmogorov-Smirnov:

- Not sensitive to problems in the tails.
- For data sets > 50.

Shapiro-Wilks:

- Doesn't work well if several values in the data set are the same.
- Works best for data sets with < 50, but can be used with larger data sets.

Some characteristics of statistical test of significance

Mann Whitney U

- Alternate to the independent t-test
- Must have at least ordinal data.
- Counts the comparative ranks of scores in two samples (from highest to lowest)
- The null hypothesis is that the two samples are randomly distributed.
- *Use U sampling distribution tables for small sample sizes (1-8) and medium sample sizes (9-20) and the Z test for large samples*

Wilcoxon Matched Pairs (signed rank test)

- Alternate to the paired t-test.
- Used for repeated measures on the same individual.
- Requires a measurement between ordinal and interval scales – the scores must have some real meaning.
- *Use a T table.*
- *If the T is less than or equal to the T in the table, the null hypothesis is rejected.*

Parametric Measures of Association

- These answer the question, "within a given population, is there a relationship between one variable and another variable?" **A measure of association can exist only if data can be logically paired. It can be tested for significance.**
  - *Correlation – answers the question, "What is the degree of relationship between "x" and "y" – Use Pearson Product Moment Correlation (Pearson r)*

Measures of Association

The Pearson Correlation Coefficient (Pearson *r*)

- The *r* examines the relationship between two quantitative sets of scores.
- The *r* varies from – 1.00 to + 1.00
- The *r* is not a proportion and cannot be multiplied by 100 to get a percentage.

- *To think of the r as a percentage, it needs to be converted to the "Coefficient of Determination" or R^2. An r of .50 is 25% better than an r of 0.00*

Non-parametric tests for association

- **The Spearman Rank Order** Correlation (r)– "To what extent and how strongly are two variables related?"
- **Phi coefficient** – it can be used with nominal data, but should have ordinal data
- **Kendall's Q** – can be used with nominal data

Prediction Parametric

- Parametric Prediction – using a correlation, if you know score "x", you can predict score "y" for one person – Use regression analysis
  - Simple linear regression – allows the prediction from one variable to another – at least interval level data
  - Multiple linear regression – allows the prediction of one variable from several other variables. The dependent variable must be on the interval scale

Non-parametric Prediction

– Measures the extent to which you can reduce the error in predicting the dependent variable as a consequence of having some knowledge of the independent variable such as, predicting income by education.

- **Kendall's Tau** – with ordinal data and ranking - is better than the Gamma because it takes ties into account
- **Gamma** - with ordinal data to predict the rank of one variable by knowing rank on another variable
- **Lambda** – with nominal data

Parametric Multiple Comparisons

- The analysis of variance (ANOVA) most common test.
- Compares observed with expected values to discover whether the means of several populations are equal.
- Compares 2 estimates of the population variance.

- *One estimate is based on variance within each sample (within groups).*
- *Other is based on variation across samples (between groups).*
- The between group variance is the explained variance (due to the treatment) and the variation within each group is the unexplained variance (the error variance).

Parametric Multiple Comparisons: ANOVA

- The ratio of the explained scores to the unexplained scores gives the F statistic.
- If the variance between the groups is larger, giving an *F ratio greater than 1, it may be significant depending upon the degrees of freedom*.

- If the *F ratio is approximately 1, it means that the null hypothesis is supported and there was no significant difference between the groups.*

ANOVA: Post hoc

- If the null hypothesis is rejected, then one would be interested in determining which groups showed a significant difference.
- *The best way to check this is to conduct a post hoc test such as the Tukey, Bonferrioni, or Scheffe.*

Parametric Multiple Comparisons : Two-Way Analysis of Variance

- Classifies participants in two-ways
- Results answer three questions
  - Two main effects
  - An interaction effect

Non-parametric Multiple Comparison: Kruskal-Wallis Test

- An alternative to the one-way ANOVA.
- The scores are ranked and the analyses compare the mean rank in each group.
- It determines if there is a difference between groups.

Non-parametric Multiple Comparison

- **McNemar Test**

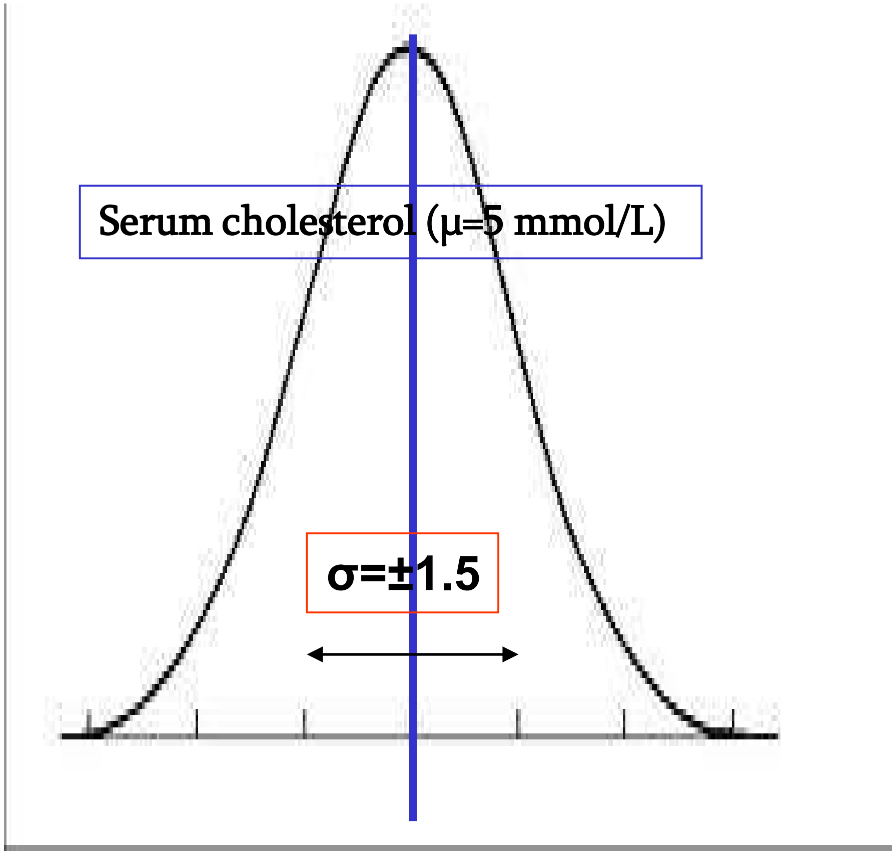
- an adaptation of the Chi-square that is used with repeated measures at the nominal level.

- **Friedman Test**

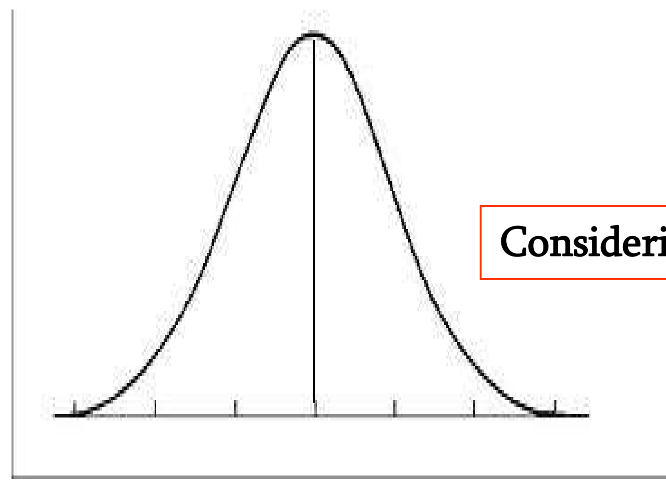
- an alternative to the repeated ANOVA.
- Two or more measurements are taken from the same subjects. It answers the questions as to whether the measurement changes over time.

Examples: conducting tests of significance

The *Z*-test for one sample



Diabetic patients, mean cholesterol > 5



Considering σ=±1.5?

Is there any difference between diabetes free population and the diabetic patients regarding serum cholesterol? Let's perform Z test.

Research hypothesis

The research hypothesis would be

The mean cholesterol of diabetics is > 5mmol/L

Null hypothesis

H0: μ=sample mean=5

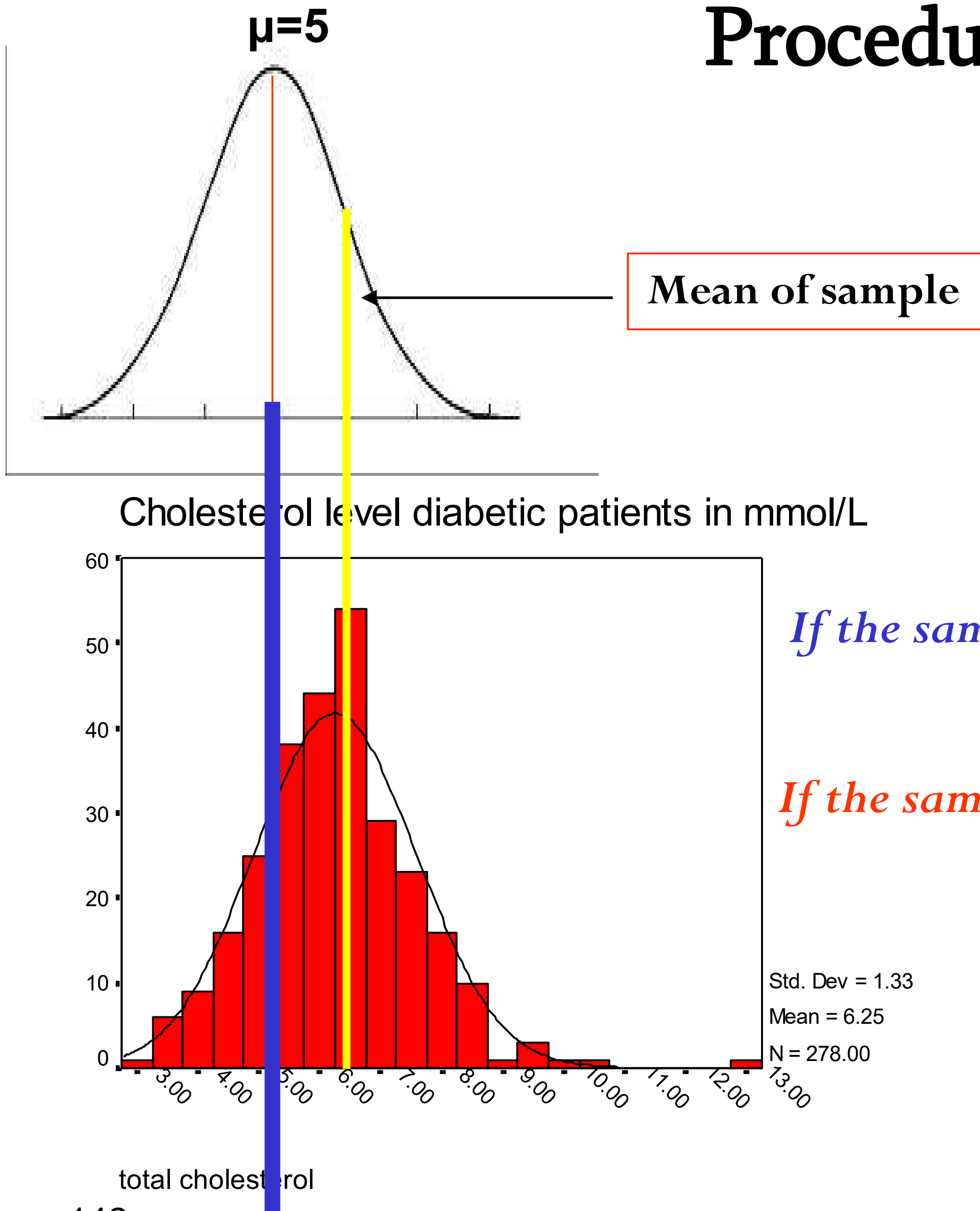
Alternative hypothesis

H1: μ >5 (one sided)

Or

H1: μ≠5 (two sided)

Procedure



If the sample mean close to the population mean
The null hypothesis is TRUE

If the sample mean differs from population mean
We REJECT the null

The ά level (P value)

The probability to obtain /achieve the null hypothesis

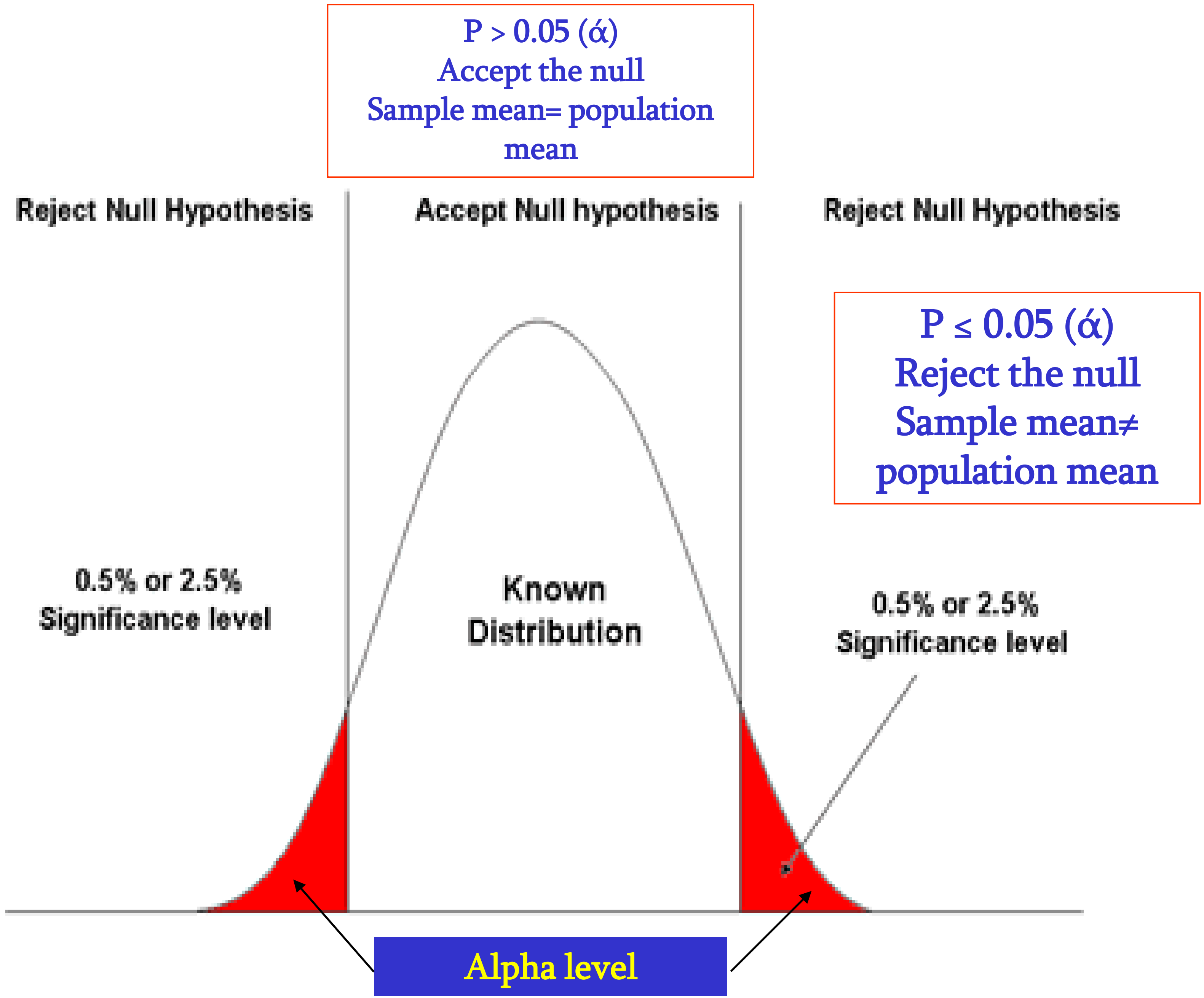
The probability that Population mean=sample mean

There no difference between the population and sample mean.

Or

The maximum probability we accept to reject the null hypothesis falsely

ά = 0.05


P > 0.05 (ά)
Accept the null
Sample mean= population mean
Reject Null Hypothesis
Accept Null hypothesis
Reject Null Hypothesis
P ≤ 0.05 (ά)
Reject the null
Sample mean≠ population mean
0.5% or 2.5% Significance level
Known Distribution
0.5% or 2.5% Significance level
Alpha level

Calculation (σ=1.5)

SEM=$\mu/\sqrt{n}$=0.3

Z=(mean sample-μ)/σ

P (mean of the sample≥6)=P(Z ≥6-5)/0.3= 0.0005

Under the normal curve area of rejection >1.96 Z

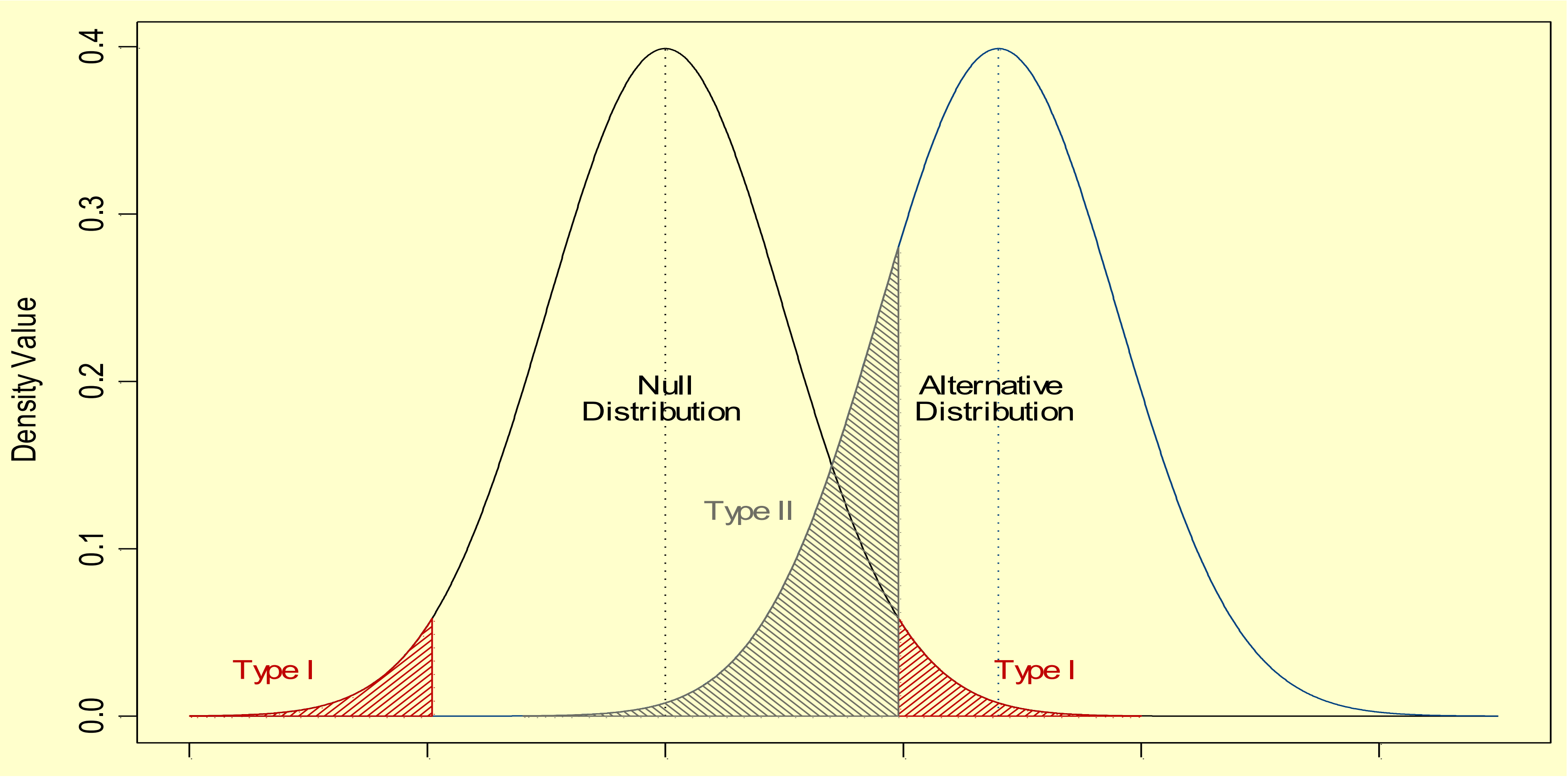
P=0.0005:
The cholesterol blood level of diabetic patients can coincide
with the population (disease free) 5 in 10,000 times
The two values could be the same in 5 times if we repeated this test 10,000 times
$P < 0.05$ so we reject the null
The diabetics have larger mean cholesterol level than the normal population

In reality

- It is unlikely that the σ (population SD) is known.
- In most of the cases, σ will be unknown and we will be able to apply neither the formula nor the table of normal distribution (areas under the curve=*Z* score).
- We resort to other statistical tests.

One Sample

The distribution of $\bar{X}$ under the null and alternative hypotheses.



t-distribution

In real life situations we will estimate the unknown population SD using Sample SD.

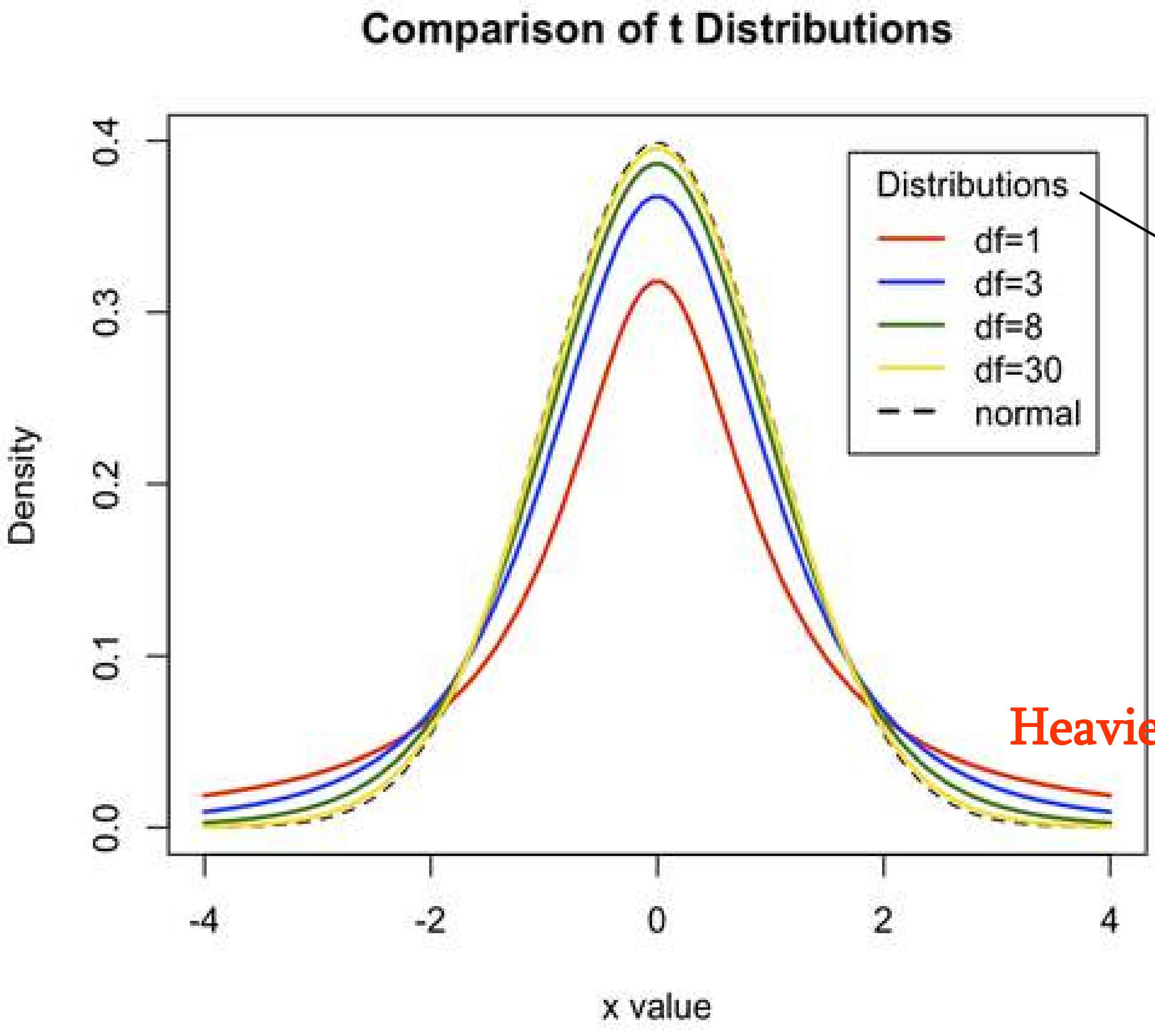
Results are standardized to the t-distribution:

$$t = \frac{\overline{\chi - \mu}}{s / \sqrt{n}}$$

$$Z = \frac{\overline{\chi - \mu}}{\sigma \sqrt{n}}$$

Z test for normal distribution
The population SD is known

t-distribution



df=No. of observations
(sample size)-1

Heavier tails than the Z distribution

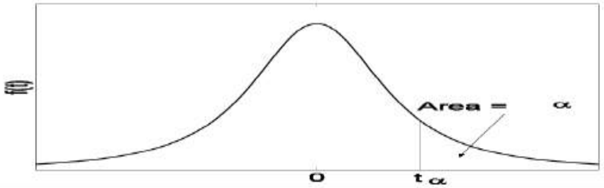
Degree of freedom (df)

For all sample statistics: variance, SD, we used n-1

All the observations in any given sample are free except one= Complementary effect.

Table B

Critical Values of the Student st Distribution



d.f.	α						
	0.100	**0.050**	**0.025**	**0.010**	**0.005**	**0.001**	**0.0005**
1	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088	636.6193
2	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271	31.5991
3	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.9240
4	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.0247	4.4370
12	1.3562	1.7823	2.1788	2.6810	3.0545	3.9296	4.3178
13	1.3502	1.7709	2.1604	2.6503	3.0123	3.8520	4.2208
14	1.3450	1.7613	2.1448	2.6245	2.9768	3.7874	4.1405
15	1.3406	1.7531	2.1314	2.6025	2.9467	3.7328	4.0728
16	1.3368	1.7459	2.1199	2.5835	2.9208	3.6862	4.0150
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.6458	3.9651
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.6105	3.9216
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.5794	3.8834
20	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518	3.8495
21	1.3232	1.7207	2.0796	2.5176	2.8314	3.5272	3.8193
22	1.3212	1.7171	2.0739	2.5083	2.8188	3.5050	3.7921
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.4850	3.7676
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.4668	3.7454
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.4502	3.7251
26	1.3150	1.7056	2.0555	2.4786	2.7787	3.4350	3.7066
27	1.3137	1.7033	2.0518	2.4727	2.7707	3.4210	3.6896
28	1.3125	1.7011	2.0484	2.4671	2.7633	3.4082	3.6739
29	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
40	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510
60	1.2958	1.6706	2.0003	2.3901	2.6603	3.2317	3.4602
120	1.2886	1.6577	1.9799	2.3578	2.6174	3.1595	3.3735
∞	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905

t-distribution

t-test-steps to determine the statistical difference

When? descriptive statistics: mean ± standard deviation

Number of samples

One sample vs. population mean

$$t = \bar{\chi} - \mu / SD \Big/ \sqrt{n}$$

Two independent samples

$$\chi_1 - \chi_2 / \sqrt{\frac{SD_1^2}{n_1}} + \sqrt{\frac{SD_2^2}{n2}}$$

Two dependent (t-paired):
Repeated measures
Matched pairs

$$t_{dependent} = \frac{d^-}{SE(d^-)}$$

<u>Steps:</u>

1- State the hypothesis to be tested: Null (non-directional-two tailed)
mean= mean
Alternative (unidirectional-one tail)
mean ≠ mean

2- Find the calculated t value: using the formulae.

3- Find the degree of freedom: all = n-1 (two sample independent df=n1-1+n2-1 (n1+n2-2).

4- Find the P value using the tables of t-distribution.

5- Conclude: if < 0.05 = rejection. If > 0.05 the null is accepted.

t-test (student's t-test) one sample

$$t = \bar{\chi} - \mu / SD \Big/ \sqrt{n}$$

Using diabetes data: Is the mean age of diabetics > 65 years?

H_0:μ=65
H_1:μ≠65

$t_{\text{one sample}}$ =67.24-65/SD/√n=3.18

Statistics

age (years)

N	Valid	278
	Missing	0
Mean		67.24
Std. Error of Mean		.704
Std. Deviation		11.743
Variance		137.902

t distribution P=0.002
Reject the null
Diabetics are significantly older than 65 years

P value (two sided)

One-Sample Test

	Test Value = 65					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference: Lower	Upper
age (years)	3.182	277	.002	2.24	.85	3.63

Degree of *freedom*

Assuming that the distribution of age is normal
Population SD is unknown (σ)

t-test for comparison of means of two independent samples

H_0: Smoking has no effect on systolic blood pressure
Mean S= Mean NS or Mean S-mean NS=0
H_1: smoking has an effect
Mean S≠ Mean NS or Mean S-Mean NS≠0

Assumptions:
- Independent observations (2 samples)
- Normally distributed
- Equal variances (for the pooled t-test)

Three formulae

Expected difference if H_0 is true

Standardized

$$t = \frac{\overline{\chi_1} - \overline{\chi_2} - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

SD of the difference

If SDs are equal

$$t = \frac{\overline{\chi_1 - \chi_2}}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Pooled SD

If SDs are not equal

$$t = \frac{\overline{\chi_1 - \chi_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Decision based on Levene's test

Variances are apparently equal

Group Statistics

	SMOKING	N	Mean	Std. Deviation	Std. Error Mean
syst. blood pressure at start	no	214	153.11	21.995	1.504
	smokers	64	144.82	20.934	2.617

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
syst. blood pressure at start	Equal variances assumed	.006	.936	2.674	276	.008	8.29	3.100	2.188	14.392
	Equal variances not assumed			2.747	107.982	.007	8.29	3.018	2.308	14.272

Two separate t-test

Not significant it means equal variances

P value <0.05, reject H_0

Paired t-test

- If we have paired data (two repeated measurements on the same subjects) or before and after

- If the difference of the paired observations are Normally distributed.

Paired samples (dependent)

(Paired / dependent 2-sample *t*-test)

- To compare observations collected form the same group of individuals on 2 separate occasions (dependent observations or paired samples).
- <u>The paired t statistics is calculated by:</u>

- Calculate the difference between the 2 measurements taken on each individual.
- Calculate the mean of the differences. **m d**
- Calculate the SE of the observed differences. **SE d**
- Under the null hypothesis of no difference or difference = 0, the paired t statistic takes the form.
- t= Mean difference / SE of the difference.

$$t = \frac{m_d - 0}{SE_d}$$

- It has a normal distribution with degrees of freedom = (n-1)

Example

Four students had the following scores in 2 subsequent tests.

Is there a significant difference in their performance?

Number	Name	Test 1	Test 2	Dif
1	Mike	35%	67%	-32
2	Melanie	50%	46%	4
3	Melissa	90%	86%	4
4	Mitchell	78%	91%	-13

Mean $_{Dif}$ = -9.25, S D $_{Dif}$ = 17.152, SE $_{Dif}$ = 8.58

Calculated Paired t = -9.25/8.58 = -1.078,

df=n-1 = 3

$$t = \frac{m_d - 0}{SE_d}$$

P value

df	Level of significance for one-tail test				
	[illegible]	0.05	0.02	0.01	0.005
	Level of significance for two-tail test				
	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	[illegible]	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.340	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
35	1.306	1.690	2.030	2.438	2.724
50	1.299	1.676	2.009	2.403	2.678
∞	1.282	1.645	1.960	2.326	2.576

The P value = 0.20, the null is accepted!

Conclusion

The observed difference can be encountered in 36 (actual P value =0.362 out of 100 cases.

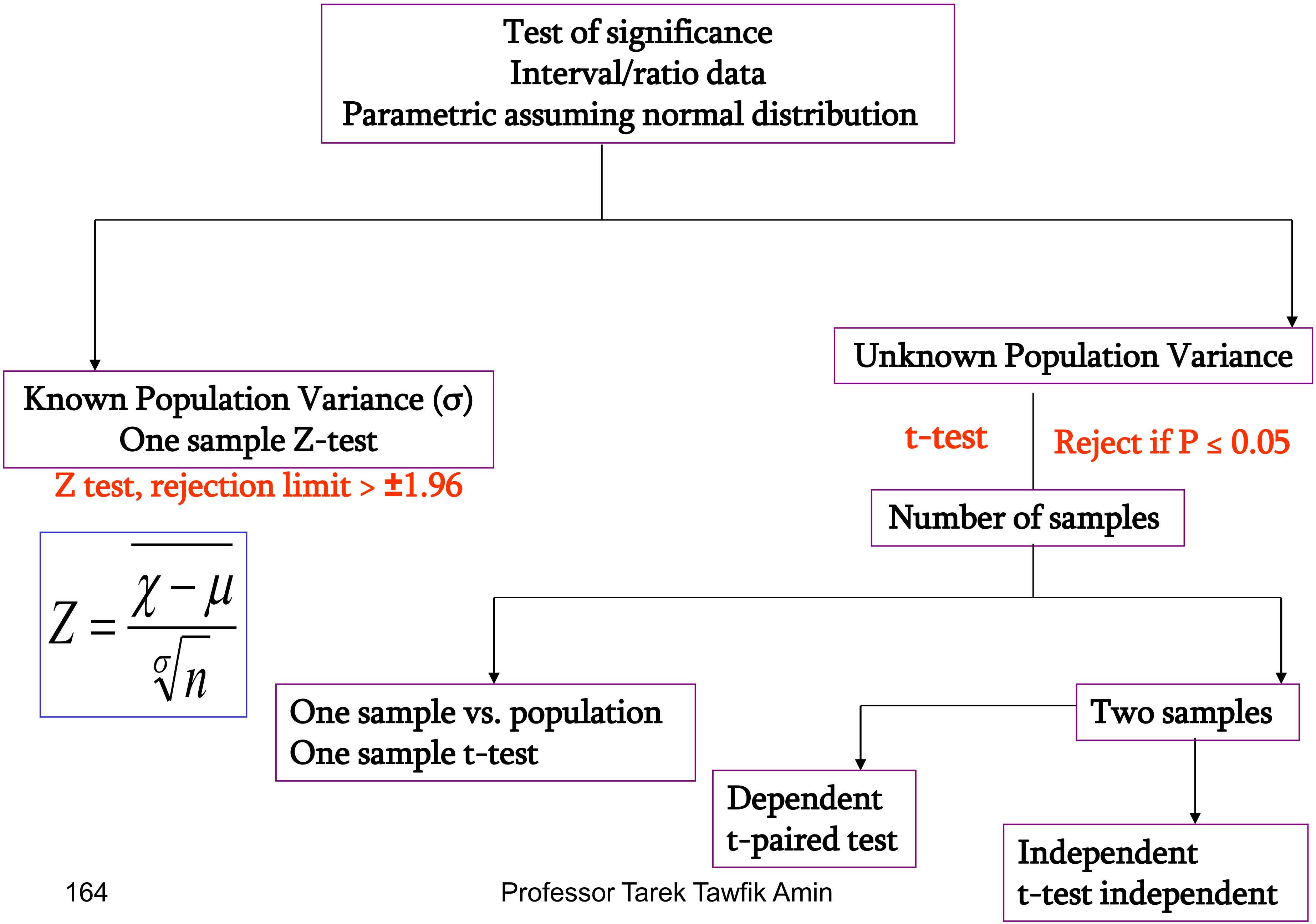
i.e. we accept the null hypothesis of no difference between first and 2nd test.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	syst. blood pressure at start	151.20	278	21.997	1.319
	syst. blood pressure after 2 years	153.83	278	29.076	1.744

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	syst. blood pressure at start - syst. blood pressure after 2 years	-2.63	17.920	1.075	-4.74	-.51	-2.443	277	.015


Test of significance
Interval/ratio data
Parametric assuming normal distribution
Known Population Variance (σ)
One sample Z-test
Z test, rejection limit > ±1.96
$Z = \frac{\overline{\chi - \mu}}{\sqrt[\sigma]{n}}$
Unknown Population Variance
t-test
Reject if P ≤ 0.05
Number of samples
One sample vs. population
One sample t-test
Two samples
Dependent
t-paired test
Independent
t-test independent

The Chi-Square test χ^2

Used for hypothesis testing for categorical variables

Many types depends on design, distribution of variables and objectives of testing

χ^2

Example:

Vaccination against Influenza deceases the risk to get the disease.

Study:

Compare the effectiveness of 5 vaccines with respect to the probability to get influenza.

Comparison will be in respect to a nominal variable (getting influenza: yes or no)

Effectiveness of Five Vaccines

Data cross tabulated 2X5: response variable: Influenza

Frequency

Vaccines	Influenza No	Influenza Yes	Total
1	237	43	280
2	198	52	250
3	245	25	270
4	212	48	260
5	233	57	290
Total	1125	225	1350

% within Vaccines

Vaccines	Influenza No	Influenza Yes	Total
1	84.6	15.4	100
2	79.2	20.8	100
3	90.7	9.3	100
4	81.5	18.5	100
5	80.3	19.7	100
Total	83.3	16.7	100

The probability to get influenza

The null hypothesis states that the probability to get influenza is independent of the vaccines
The alternative states that a dependency exists

Effectiveness of Five Vaccines

If H_0 is true:

The probability to influenza in every group should be the same=

the probability in the total population,

Equal to: 225/1350=0.167 (16.7%)

Vaccine 1 used in 280, if H_0 is true,

we expect that 16.7% (≈47) to get influenza.

However this is not true

Expected frequencies

*For any cell: Expected Frequency= Row total*column total / grand total*

Vaccines	Influenza No	Influenza Yes	Total
1-Observed	237	43	280
Expected	233.3	46.7	
2-Observed	198	52	250
Expected	208.3	41.7	
3-Observed	245	25	270
Expected	225.0	45.0	
4-Observed	212	48	260
Expected	216.7	43.3	
5-Observed	233	57	290
Expected	241.7	48.3	
Total	1125	225	1350

Row total

280X225/1350

260*1125/1350

Grand total

Column total

Pearson Chi-square test

Calculate the expected frequencies (assuming H_0 is true) for all the ten cells.

Calculate Chi square: O_f= observed frequency
E_f= Expected frequency

$$\chi^2 = \sum \frac{(O_f - E_f)^2}{E_f}$$

Reject H0 if $\chi 2$ is large
Use the Chi-square distribution
After determining the degree of freedom (df)
df= (r-1)*(c-1)

Critical values for Chi-square

df	Level of Significance									
	0.99	0.90	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.00016	0.0158	0.148	0.455	1.074	1.642	2.706	3.841	6.635	10.827
2	0.0201	0.211	0.713	1.386	2.408	3.219	4.605	5.991	9.210	13.815
3	0.115	0.584	1.424	2.366	3.665	4.642	6.251	7.815	11.341	16.268
4	0.297	1.064	2.195	3.357	4.878	5.989	7.779	9.488	13.277	18.465
5	0.554	1.610	3.000	4.351	6.064	7.289	9.236	11.070	15.086	20.517
.										
.										
30	14.953	20.599	25.508	29.336	33.530	36.250	40.256	43.773	50.892	59.703

$\chi^2_{critical} = 9.488$

Calculated=16.555

$df=(2-1)(5-1)=4$

P=0.002

There is a relation (dependence) between type of vaccine and influenza prevention

SMOKING * SEX Crosstabulation

			SEX		Total
			male	female	
SMOKING	no	Count	90	124	214
		% within SMOKING	42.1%	57.9%	100.0%
	smokers	Count	55	9	64
		% within SMOKING	85.9%	14.1%	100.0%
Total		Count	145	133	278
		% within SMOKING	52.2%	47.8%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	38.017[b]	1	.000		
Continuity Correction[a]	36.279	1	.000		
Likelihood Ratio	41.649	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	37.880	1	.000		
N of Valid Cases	278				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 30.62.

At least 80% of cells must have $E_f > 5$

We can't use Pearson Chi-square if the expected frequency is <5

In this case we use Fisher's Exact test

status * SEX Crosstabulation

Count

		SEX		Total
		male	female	
status	alive	24	15	39
	died from CVD	4	1	5
	other cause of death	2	2	4
Total		30	18	48

E f=5*18/48=1.875 (<5)

Expected f=4*30/48=2.5 (<5)

Fisher Exact test provides correction

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	.935[a]	2	.626
Likelihood Ratio	.991	2	.609
Linear-by-Linear Association	.004	1	.951
N of Valid Cases	48		

a. 4 cells (66.7%) have expected count less than 5. The minimum expected count is 1.50.

Chi-square is not valid

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	38.017[b]	1	.000		
Continuity Correction[a]	36.279	1	.000		
Likelihood Ratio	41.649	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	37.880	1	.000		
N of Valid Cases	278				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 30.62.

McNemar test

Paired data in a cross tabulation

54 eczematous persons on both arms use ointment A or B (randomized)

	Ointment B +	Ointment B No	Total
Ointment A			
+	16	10	26
No	23	5	28
Total	39	15	54

$X^2=(23-10)^2/23+10$
df=1

McNemar test only take the discordant pairs into account

Questions