

Chapter 1 Statistics and Data Science

1.4 Software for Statistical Analysis

Jung Jin Lee

Professor of Soongsil University, Korea

Visiting Professor of ADA University, Azerbaijan

1.4 Software for Statistical Analysis

- **Computer software is essential for Statistics & Data Science**
 - **Elementary: Excel**
 - **Advanced: **statistical packages** such as SAS, SPSS, R, Stata**
for advanced user
no educational module
expensive except R
not an web/mobile

1.4 Software for Statistical Analysis

- **eStat Development Project (2015 ~ 2020)**
 - by Jung Jin Lee and others in Korea
- **Features**
 - freeware
 - web based software : anytime and anywhere
 - easy user interface
 - dynamic graphs
 - various modules for statistics education
 - all statistical distributions are online
 - from middle school to university students

1.4 Software for Statistical Analysis - *eStat*

© Technology & Manpower for *eStat*

- HTML5
- CSS3
- JavaScript
- D3.js for dynamic graphs
- Handson table sheet
- Statistical distribution library
 - include nonparametric distributions
- Professors in statistics, statistical computing
Professors in mathematics education
Elementary, middle, high school teachers

1.4 Software for Statistical Analysis - *eStat*

© *eStat* modules

- Elementary School



- Middle School



- High School

Binomial, Normal, Sampling Distribution, Law of Large Number, Confidence Interval

1.4 Software for Statistical Analysis - *eStat*

University

eStatU - University Statistics Education SW

Uniform Random Number

Binomial Experiment

Binomial Distribution

Poisson Distribution

Geometric Distribution

HyperGeometric Distribution

Exponential Distribution

Normal Experiment

Normal Distribution

t Distribution

ChiSquare Distribution

F Distribution

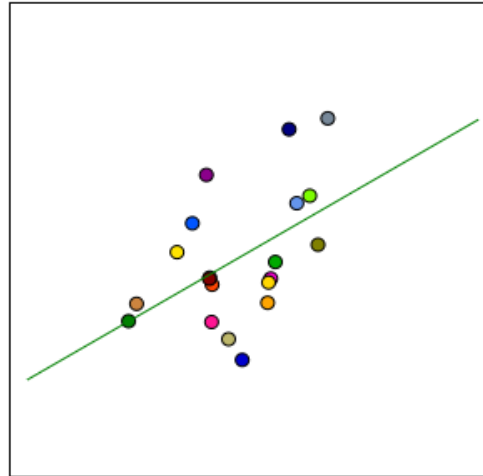
Wilcoxon Signed Rank Sum Dist.

Wilcoxon Rank Sum Distribution

Kruskal-Wallis H Distribution

Friedman S Distribution

HSD Studentized Range Dist.



Contact: jjlee@ssu.ac.kr
© eStat.org, Korea

Law of Large Number

Population vs Sample

Dist of Sample Means

Confidence Interval

Correlation Coefficient

Regression Experiment

Testing Hypothesis μ

Testing $\mu - C, \beta$

Testing $\mu - C, n$

Testing Hypothesis σ^2

Testing Hypothesis p

Testing Hypothesis μ_1, μ_2

Testing Hypothesis σ_1^2, σ_2^2

Testing Hypothesis p_1, p_2

Testing Hypothesis ANOVA

Sign Test

Signed Rank Sum Test

Rank Sum Test

Kruskal-Wallis Test

Friedman Test

Goodness of Fit Test

Testing Independence

1.4 Software for Statistical Analysis - eStat

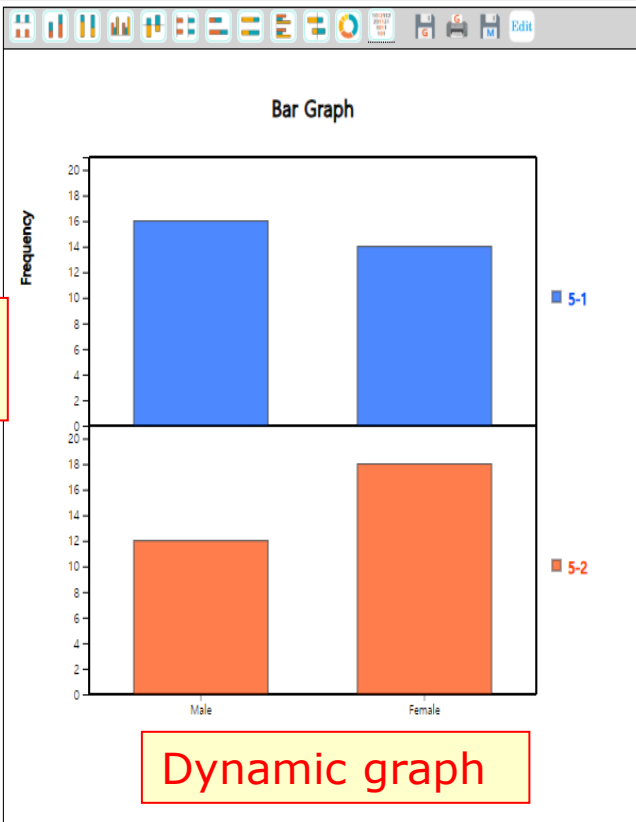
Easy UI - Icon only design

Language Selection

Examples

Level Selection

Easy UI - mouse clicking only



Summary Data Frequency Table	Group Variable	(V2 V3)	
Analysis Var (Sex)	5-1	5-2	Total
Male	16 57.1%	12 42.9%	28 100%
Female	14 43.8%	18 56.3%	32 100%
Total	30 50.0%	30 50.0%	60 100%
Missing Observations		0	

Log Window

Dynamic graph

1.4 Software for Statistical Analysis - *eStat*

© Data and Dynamic Graph

- Support csv and json format
- Support summary and raw data for data processing
- Dynamic graph

File: 000Summary_StudentBySex.csv

Analysis Var: 1: Sex (Selected data: Summary Data)

by Group: 3: 5-2 (Summary Data: Multip)

SelectedVar: V1 by V2,V3,

	Sex	5-1	5-2	V4	V5
1	Male	16	12		
2	Female	14	18		

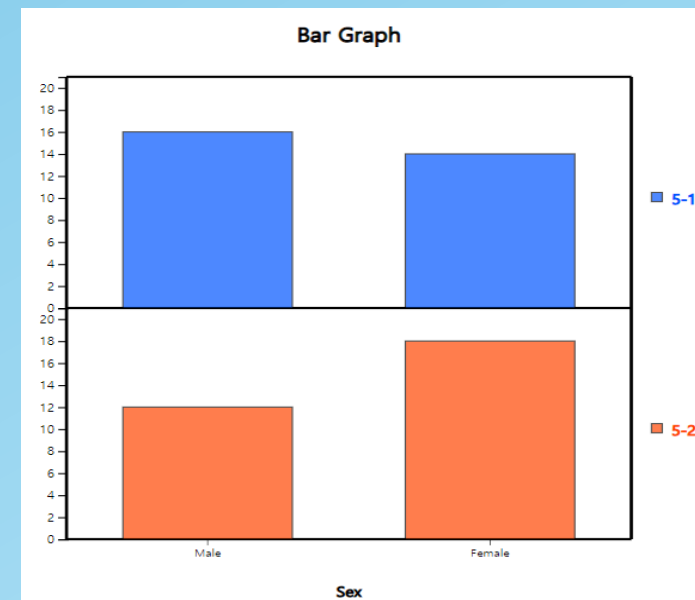
File: 021Discrete_MathPreference

Analysis Var: 2: MathPref (Selected data: Raw Data)

by Group: 1: Sex (Summary Data: Mu)

SelectedVar: V2 by V1,

	Sex	MathPref	V3	V4
1	1	3		
2	2	1		
3	1	3		
4	2	1		
5	1	3		
6	1	1		
7	1	2		
8	2	2		
9	2	3		
10	1	2		



1.4 Software for Statistical Analysis - *eStat*

© Graphical Result of Statistical Analysis - ANOVA

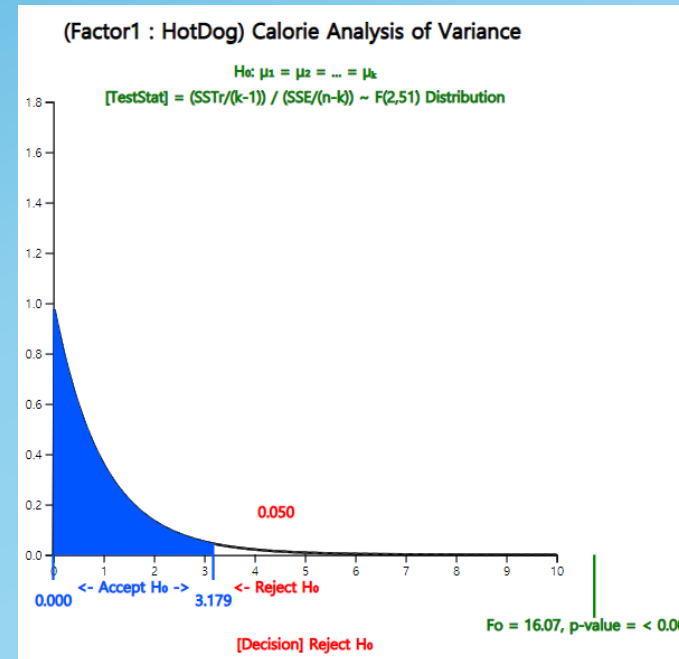
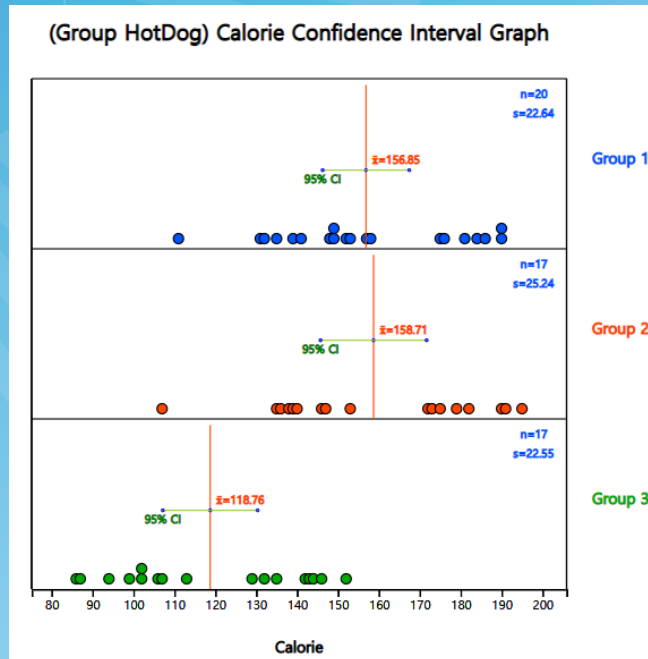
File: 033Cont_CalorieByHotDogT

Analysis Var: 2: Calorie by Group: 1: HotDog

(Selected data: Raw Data) (Select up to two g

SelectedVar: V2 by V1,

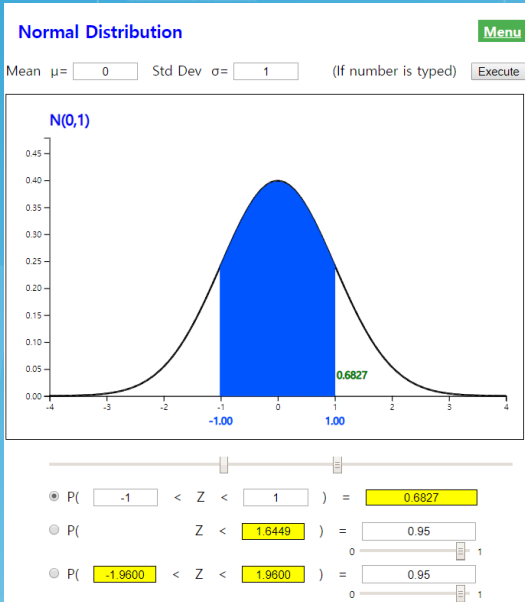
	HotDog	Calorie	V3	V4
1	1	186		
2	1	181		
3	1	176		
4	1	149		
5	1	184		
6	1	190		
7	1	158		
8	1	139		
9	1	175		
10	1	148		
11	1	152		
12	1	111		
13	1	141		
14	1	153		
15	1	190		



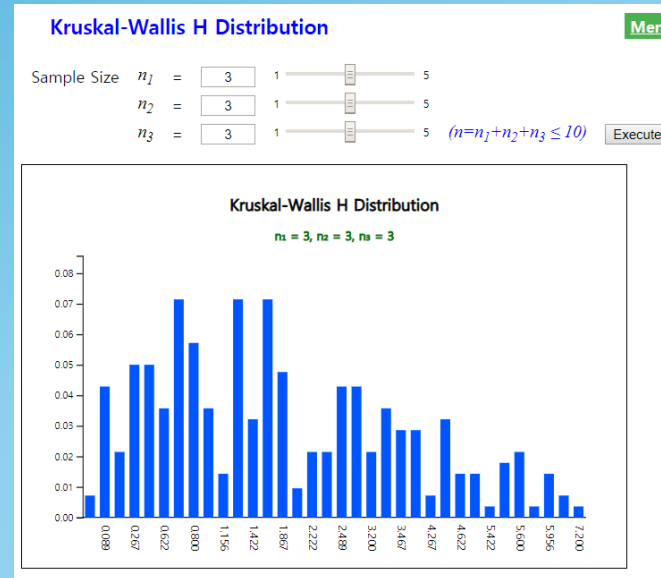
Analysis of Variance					
Factor	Sum of Squares	deg of freedom	Mean Squares	F value	p value
Treatment	17692.195	2	8846.098	16.074	< 0.0001
Error	28067.138	51	550.336		
Total	45759.333	53			

1.4 Software for Statistical Analysis - *eStat*

© All tables of statistical distributions are on smart-phone



Normal Distribution		$\mu = 0$	$\sigma = 1.000$												
x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)	x	P(X ≤ x)
-3.99	0.0000	-2.99	0.0014	-1.99	0.0233	-0.99	0.1611	0.01	0.5040	1.01	0.8438	2.01	0.9778	3.01	0.9987
-3.98	0.0000	-2.98	0.0014	-1.98	0.0239	-0.98	0.1635	0.02	0.5080	1.02	0.8461	2.02	0.9783	3.02	0.9987
-3.97	0.0000	-2.97	0.0015	-1.97	0.0244	-0.97	0.1660	0.03	0.5120	1.03	0.8485	2.03	0.9788	3.03	0.9988
-3.96	0.0000	-2.96	0.0015	-1.96	0.0250	-0.96	0.1685	0.04	0.5160	1.04	0.8508	2.04	0.9793	3.04	0.9988
-3.95	0.0000	-2.95	0.0016	-1.95	0.0256	-0.95	0.1711	0.05	0.5199	1.05	0.8531	2.05	0.9798	3.05	0.9989
-3.94	0.0000	-2.94	0.0016	-1.94	0.0262	-0.94	0.1736	0.06	0.5239	1.06	0.8554	2.06	0.9803	3.06	0.9989
-3.93	0.0000	-2.93	0.0017	-1.93	0.0268	-0.93	0.1762	0.07	0.5279	1.07	0.8577	2.07	0.9808	3.07	0.9989
-3.92	0.0000	-2.92	0.0018	-1.92	0.0274	-0.92	0.1788	0.08	0.5319	1.08	0.8599	2.08	0.9812	3.08	0.9990
-3.91	0.0000	-2.91	0.0018	-1.91	0.0281	-0.91	0.1814	0.09	0.5359	1.09	0.8621	2.09	0.9817	3.09	0.9990
-3.90	0.0000	-2.90	0.0019	-1.90	0.0287	-0.90	0.1841	0.10	0.5398	1.10	0.8643	2.10	0.9821	3.10	0.9990
-3.89	0.0001	-2.89	0.0019	-1.89	0.0294	-0.89	0.1867	0.11	0.5438	1.11	0.8665	2.11	0.9826	3.11	0.9991
-3.88	0.0001	-2.88	0.0020	-1.88	0.0301	-0.88	0.1894	0.12	0.5478	1.12	0.8686	2.12	0.9830	3.12	0.9991
-3.87	0.0001	-2.87	0.0021	-1.87	0.0307	-0.87	0.1922	0.13	0.5517	1.13	0.8708	2.13	0.9834	3.13	0.9991
-3.86	0.0001	-2.86	0.0021	-1.86	0.0314	-0.86	0.1949	0.14	0.5557	1.14	0.8729	2.14	0.9838	3.14	0.9992
-3.85	0.0001	-2.85	0.0022	-1.85	0.0322	-0.85	0.1977	0.15	0.5596	1.15	0.8749	2.15	0.9842	3.15	0.9992
-3.84	0.0001	-2.84	0.0023	-1.84	0.0329	-0.84	0.2005	0.16	0.5636	1.16	0.8770	2.16	0.9846	3.16	0.9992
-3.83	0.0001	-2.83	0.0023	-1.83	0.0336	-0.83	0.2033	0.17	0.5675	1.17	0.8790	2.17	0.9850	3.17	0.9992



Kruskal-Wallis H Distribution	k = 3		
	$n_1 = 3$	$n_2 = 3$	$n_3 = 3$
x	P(X = x)	P(X ≤ x)	P(X ≥ x)
0.000	0.0071	0.0071	1.0000
0.089	0.0429	0.0500	0.9929
0.089	0.0214	0.0714	0.9500
0.267	0.0500	0.1214	0.9286
0.356	0.0500	0.1714	0.8786
0.622	0.0357	0.2071	0.8286
0.622	0.0714	0.2786	0.7929
0.800	0.0571	0.3357	0.7214
1.067	0.0357	0.3714	0.6643
1.156	0.0143	0.3857	0.6286
1.156	0.0714	0.4571	0.6143
1.422	0.0321	0.4893	0.5429
1.689	0.0714	0.5607	0.5107
1.867	0.0476	0.6083	0.4393
1.867	0.0095	0.6179	0.3917
2.222	0.0214	0.6393	0.3821
2.400	0.0214	0.6607	0.3607

1.4 Software for Statistical Analysis - eStat

© Modules for Home Work Assignment - eStatU

Testing Hypothesis μ_1, μ_2

Menu

[Hypothesis] $H_0: \mu_1 - \mu_2 = D$

$H_1: \mu_1 - \mu_2 \neq D$ $H_1: \mu_1 - \mu_2 > D$ $H_1: \mu_1 - \mu_2 < D$

[Test Type] t test, Variance Assumption $\sigma_1^2 = \sigma_2^2$ $\sigma_1^2 \neq \sigma_2^2$

Significance Level $\alpha =$ 5% 1%

Sampling Type independent sample paired sample

[Sample Data] *Input either sample data using BSV or sample statistics*

Sample 1

Sample 2

[Sample Statistics]

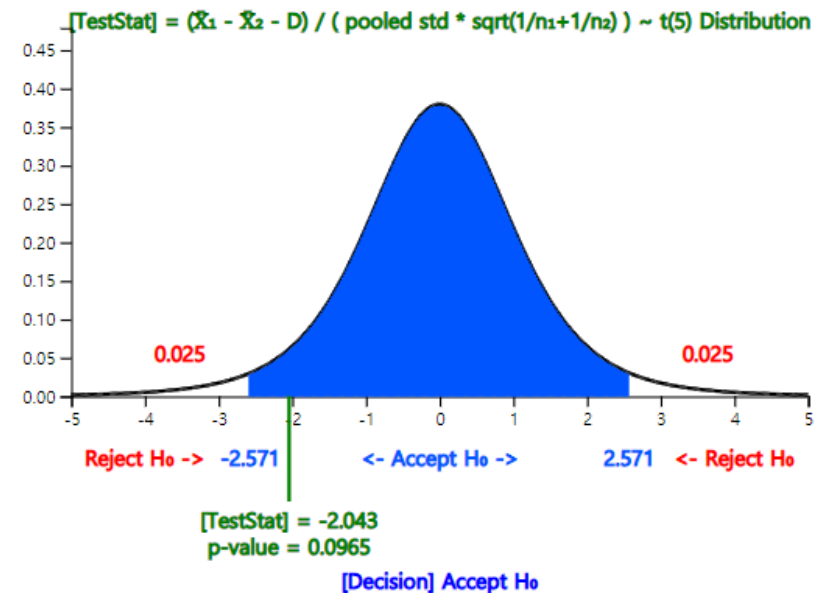
Sample Size $n_1 =$ $n_2 =$

Sample Mean $\bar{x}_1 =$ $\bar{x}_2 =$

Sample Variance $s_1^2 =$ $s_2^2 =$

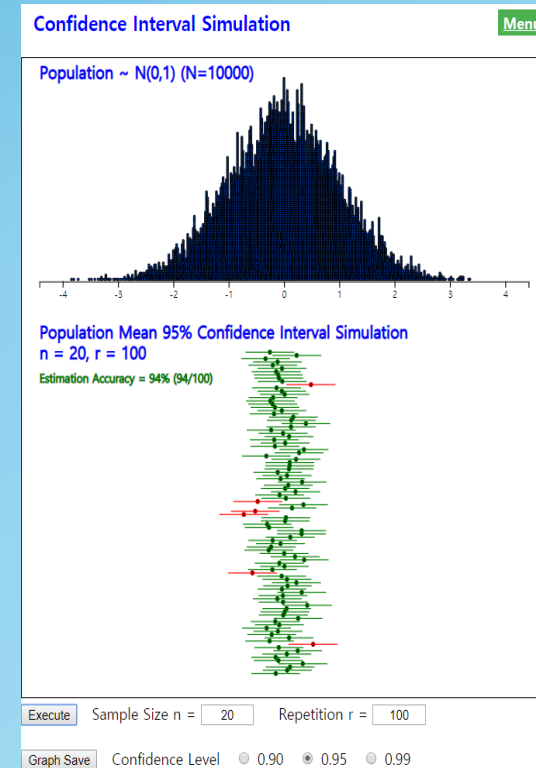
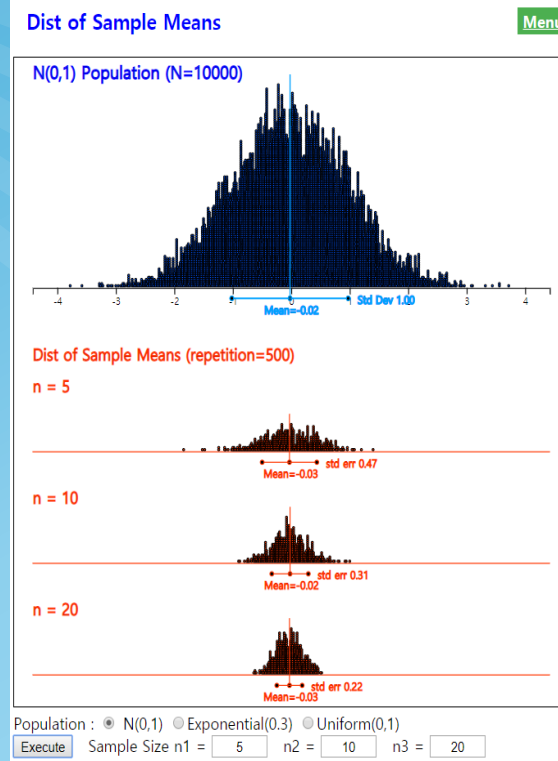
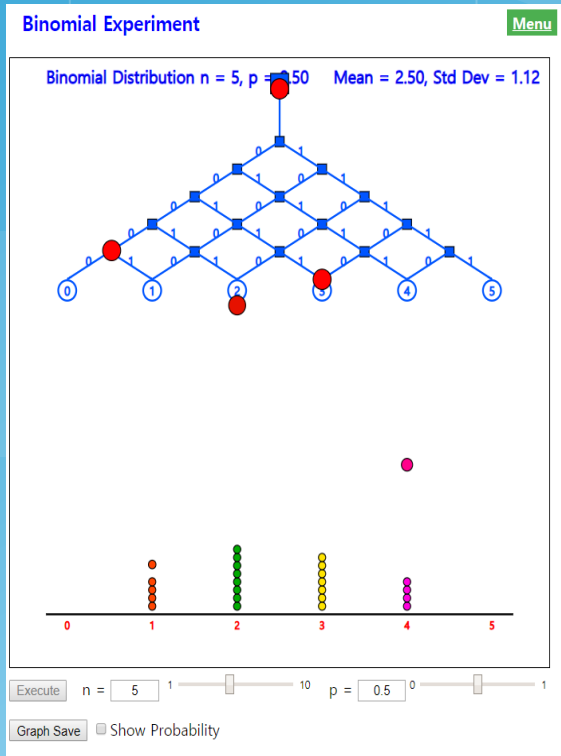
Execute

$H_0: \mu_1 - \mu_2 = 0.00$, $H_1: \mu_1 - \mu_2 \neq 0.00$



1.4 Software for Statistical Analysis - *eStat*

Simulation Experiments

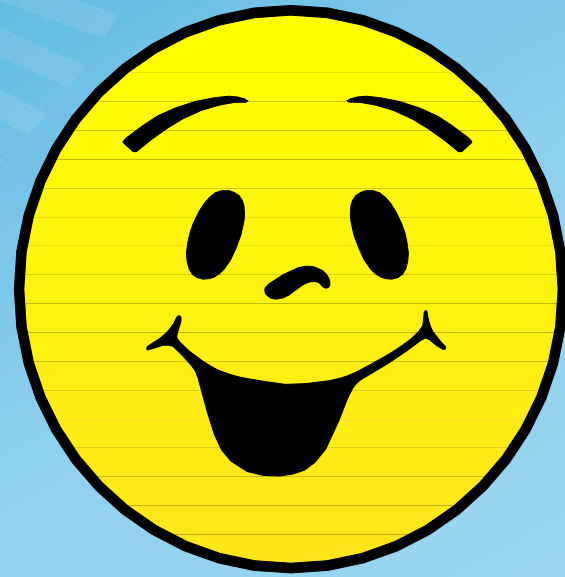


1.4 Software for Statistical Analysis - *eStat*

<http://www.estat.me>

eStat works 100% with Chrome

- 1) Enter system
- 2) Data input/save/open
- 3) Draw graph and data analysis
- 4) Save results / print results
- 5) Log out the system
- 6) Educational modules
- 7) Others



Thank you