EXPERIMENT: 7

(2K17/SE/79 PARV GUPTA)

AIM: Consider defect dataset and implement following statistical test using SPSS tool.

a. t-test
b. Chi-Square Test
c. Wilcoxon Signed Test
d. Friedman Test
e. Kruskal Wallis Test

THEORY:

a.) T-test: A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features. The t-test is one of many tests used for the purpose of hypothesis testing in statistics.

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

t = Student's t-test
m = mean
μ = theoretical value
s = standard deviation
n = variable set size

b.) Chi-Square Test: A chi-squared test, also written as χ^2 test, is a statistical hypothesis test that is valid to perform when the test statistic is chi-squared distributed under the null hypothesis, specifically Pearson's chi-squared test and variants thereof.

c.) Wilcoxin Signed: The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used to compare two

related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks differ.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

χ^2 = chi squared

O_i = observed value

E_i = expected value

$$W = \sum_{i=1}^{N_r} [\operatorname{sgn}(x_{2,i} - x_{1,i}) \cdot R_i]$$

W = test statistic

N_r = sample size, excluding pairs where x1 = x2

sgn = sign function

$x_{1,i}, x_{2,i}$ = corresponding ranked pairs from two distributions

R_i = rank i

d.) Friedman Test: The Friedman test is the non-parametric alternative to the oneway ANOVA with repeated measures. It is used to test for differences between groups when the dependent variable being measured is ordinal. It can also be used for continuous data that has violated the assumptions necessary to run the one-way ANOVA with repeated measures (e.g., data that has marked deviations from normality).

e.) Kruskal-Wallis: The Kruskal–Wallis test by ranks, Kruskal Wallis H test, or oneway ANOVA on ranks is a non-parametric method for testing whether samples originate from the same distribution. It is used for comparing two or more independent samples of equal or different sample sizes.

Dataset: Software Defect prediction dataset is used to evaluate these tests.

Features	Description
CBO	Coupling between objects. Counts the number of dependencies a class has.
WMC	Weight Method Class or McCabe's complexity. It counts the number of branch instructions in a class.
DIT	Depth Inheritance Tree. It counts the number of "fathers" a class has. All classes have DIT at least 1 (everyone inherits java.lang.Object).
rfc	Response for a Class. Counts the number of unique method invocations in a class.

lcom	Lack of Cohesion of Methods. Calculates LCOM metric.
totalMethods	Counts the number of methods.
totalFields	Counts the number of fields.
NOSI	Number of static invocations. Counts the number of invocations to static methods.

LOC	Lines of code. It counts the lines of count, ignoring empty lines.
returnQty	Quantity of returns. The number of return instructions.
loopQty	Quantity of loops. The number of loops (i.e., for, while, do while, enhanced for).
comparisonsQty	Quantity of comparisons. The number of comparisons (i.e., == and !=).
tryCatchQty	Quantity of try/catches. The number of try/catches.
parenthesizedExpsQty	Quantity of parenthesized expressions. The number of expressions inside parenthesis.
stringLiteralsQty	String literals. The number of string literals (e.g., "John Doe").
numbersQty	Quantity of Number. The number of numbers (i.e., int, long, double, float) literals.

assignmentsQty	Quantity of Variables. Number of declared variables.
mathOperationsQty	Quantity of Math Operations: The number of math operations (times, divide, remainder, plus, minus, left shit, right shift).
variablesQty	Quantity of Variables. Number of declared variables.
maxNestedBlocks	Max nested blocks. The highest number of blocks nested together.
uniqueWordsQty	Number of unique words. Number of unique words in the source code.

OUTPUT:

Paired T-Test performed between CBO & totalMethods

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	cbo	27.49	6052	33.215	.427
	totalMethods	33.50	6052	53.557	.688

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	cbo & totalMethods	6052	.545	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	cbo - totalMethods	-6.004	45.095	.580	-7.140	-4.868	-10.358	6051	<.001

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	Upper
Pair 1	cbo - totalMethods	Cohen's d	45.095	-.133	-.158	-.108
		Hedges' correction	45.098	-.133	-.158	-.108

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Independent Sample-T test performed with cbo, wmc, dit, rfc.

T-Test

Group Statistics

	totalFields	N	Mean	Std. Deviation	Std. Error Mean
cbo	1	494	14.26	18.061	.813
	2	410	15.53	19.816	.979
wmc	1	494	38.88	58.346	2.625
	2	410	34.85	59.021	2.915
dit	1	494	2.16	4.064	.183
	2	410	2.25	3.958	.195
rfc	1	494	30.86	36.929	1.662
	2	410	32.11	42.536	2.101

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
cbo	Equal variances assumed	2.001	.158	-1.012	902	.312	-1.277	1.261	-3.752	1.198
	Equal variances not assumed			-1.004	837.242	.316	-1.277	1.272	-3.773	1.220
wmc	Equal variances assumed	4.695	.031	1.027	902	.305	4.023	3.918	-3.668	11.713
	Equal variances not assumed			1.026	867.875	.305	4.023	3.923	-3.676	11.722
dit	Equal variances assumed	.472	.492	-.342	902	.733	-.092	.268	-.618	.435
	Equal variances not assumed			-.343	879.321	.732	-.092	.268	-.617	.434
rfc	Equal variances assumed	.508	.476	-.473	902	.636	-1.251	2.644	-6.439	3.937
	Equal variances not assumed			-.467	815.896	.641	-1.251	2.678	-6.508	4.006

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
cbo	Cohen's d	18.877	-.068	-.199	.063
	Hedges' correction	18.893	-.068	-.198	.063
	Glass's delta	19.816	-.064	-.195	.067
wmc	Cohen's d	58.653	.069	-.062	.200
	Hedges' correction	58.702	.069	-.062	.199
	Glass's delta	59.021	.068	-.063	.199
dit	Cohen's d	4.016	-.023	-.154	.108
	Hedges' correction	4.019	-.023	-.154	.108
	Glass's delta	3.958	-.023	-.154	.108
rfc	Cohen's d	39.570	-.032	-.163	.099
	Hedges' correction	39.603	-.032	-.162	.099
	Glass's delta	42.536	-.029	-.160	.102

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

Chi-Square Test

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum	Percentiles 25th	50th (Median)	75th
cbo	6052	27.49	33.215	0	419	9.00	18.00	34.00
wmc	6052	86.18	136.077	0	1714	18.00	45.00	100.00
dit	6052	4.60	9.288	1	285	1.00	2.00	4.00
rfc	6052	68.95	85.772	0	1203	18.00	44.00	89.00

Chi-Square Test

Frequencies

cbo

	Observed N	Expected N	Residual
0	23	30.7	-7.7
1	95	30.7	64.3
2	130	30.7	99.3
3	161	30.7	130.3
4	181	30.7	150.3
5	219	30.7	188.3
6	196	30.7	165.3
7	220	30.7	189.3
8	185	30.7	154.3
9	260	30.7	229.3
10	191	30.7	160.3
11	166	30.7	135.3
12	145	30.7	114.3
13	177	30.7	146.3
14	143	30.7	112.3
15	134	30.7	103.3
16	160	30.7	129.3
17	163	30.7	132.3
18	149	30.7	118.3
19	142	30.7	111.3
20	107	30.7	76.3

wmc

	Observed N	Expected N	Residual
0	16	14.1	1.9
1	54	14.1	39.9
2	44	14.1	29.9
3	54	14.1	39.9
4	92	14.1	77.9
5	77	14.1	62.9
6	119	14.1	104.9
7	104	14.1	89.9
8	99	14.1	84.9
9	113	14.1	98.9
10	95	14.1	80.9
11	113	14.1	98.9
12	100	14.1	85.9
13	84	14.1	69.9
14	98	14.1	83.9
15	77	14.1	62.9
16	86	14.1	71.9
17	70	14.1	55.9
18	73	14.1	58.9
19	76	14.1	61.9
20	71	14.1	56.9

dit

	Observed N	Expected N	Residual
1	2245	89.0	2156.0
2	1386	89.0	1297.0
3	557	89.0	468.0
4	371	89.0	282.0
5	240	89.0	151.0
6	190	89.0	101.0
7	149	89.0	60.0
8	126	89.0	37.0
9	99	89.0	10.0
10	97	89.0	8.0
11	73	89.0	-16.0
12	76	89.0	-13.0
13	58	89.0	-31.0
14	40	89.0	-49.0
15	24	89.0	-65.0
16	20	89.0	-69.0
17	30	89.0	-59.0
18	17	89.0	-72.0
19	23	89.0	-66.0
20	17	89.0	-72.0

rfc

	Observed N	Expected N	Residual
0	147	15.8	131.2
1	35	15.8	19.2
2	96	15.8	80.2
3	55	15.8	39.2
4	87	15.8	71.2
5	63	15.8	47.2
6	66	15.8	50.2
7	84	15.8	68.2
8	74	15.8	58.2
9	90	15.8	74.2
10	96	15.8	80.2
11	107	15.8	91.2
12	93	15.8	77.2
13	72	15.8	56.2
14	83	15.8	67.2
15	73	15.8	57.2
16	97	15.8	81.2
17	73	15.8	57.2
18	77	15.8	61.2
19	63	15.8	47.2
20	66	15.8	50.2

Test Statistics

	cbo	wmc	dit	rfc
Chi-Square	18016.244[a]	14453.945[b]	79129.146[c]	12848.642[d]
df	196	428	67	383
Asymp. Sig.	.000	.000	.000	.000

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 30.7.

b. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 14.1.

c. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 89.0.

d. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 15.8.

Wilcoxin Signed Test

Ranks

		N	Mean Rank	Sum of Ranks
totalMethods - cbo	Negative Ranks	3026[a]	2661.35	8053241.00
	Positive Ranks	2796[b]	3182.23	8897512.00
	Ties	230[c]		
	Total	6052		
dit - wmc	Negative Ranks	5954[d]	3015.31	17953128.5
	Positive Ranks	40[e]	347.16	13886.50
	Ties	58[f]		
	Total	6052		
lcom - rfc	Negative Ranks	2864[g]	2036.71	5833127.50
	Positive Ranks	3104[h]	3859.01	11978368.5
	Ties	84[i]		
	Total	6052		

a. totalMethods < cbo

b. totalMethods > cbo

c. totalMethods = cbo

d. dit < wmc

e. dit > wmc

f. dit = wmc

g. lcom < rfc

h. lcom > rfc

i. lcom = rfc

Test Statistics[a]

	totalMethods - cbo	dit - wmc	lcom - rfc
Z	-3.292[b]	-66.949[c]	-23.084[b]
Asymp. Sig. (2-tailed)	<.001	.000	<.001

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

c. Based on positive ranks.

Friedman Test

NPar Tests

Friedman Test

Ranks

	Mean Rank
cbo	13.90
wmc	18.14
dit	6.98
rfc	17.45
lcom	15.52
totalMethods	13.91
totalFields	9.80
nosi	4.73
loc	21.62
returnQty	11.69
loopQty	5.80
comparisonsQty	9.76
tryCatchQty	4.50
parenthesizedExpsQty	5.00
stringLiteralsQty	11.82
numbersQty	10.40
assignmentsQty	16.64
mathOperationsQty	8.00
variablesQty	15.37
maxNestedBlocks	7.48
uniqueWordsQty	20.79
defect	3.68

Test Statistics[a]

N	6052
Chi-Square	91337.006
df	21
Asymp. Sig.	.000

a. Friedman Test

Kruskal Wallis Test

Ranks

	totalFields	N	Mean Rank
cbo	1	494	1655.32
	2	410	1734.75
	3	411	1697.18
	4	417	2221.85
	5	356	2060.77
	6	268	2309.46
	7	258	2345.04
	8	212	2564.77
	9	170	2661.93
	10	182	2846.79
	11	143	2889.40
	12	145	2907.52
	13	147	3161.90
	14	120	3088.93
	15	105	3280.45
	16	94	3533.00
	17	97	3692.43
	18	117	3536.93
	19	69	3605.61
	20	88	3439.35
wmc	1	494	1588.96
	2	410	1503.97
	3	411	1535.29
	4	417	1995.41
	5	356	2081.31
	6	268	2128.48
	7	258	2324.39
	8	212	2548.22
	9	170	2569.37
	10	182	2715.01
	11	143	2915.40
	12	145	3057.42
	13	147	3310.17
	14	120	3169.70
	15	105	3209.66
	16	94	3537.80
	17	97	3655.73
	18	117	3737.84
	19	69	3670.29
	20	88	3858.55
dit	1	494	1849.87
	2	410	1862.83
	3	411	1821.44
	4	417	2097.34
	5	356	2097.12
	6	268	2414.12
	7	258	2397.76
	8	212	2535.19
	9	170	2544.13
	10	182	3219.52
	11	143	2659.31
	12	145	2943.32
	13	147	2984.28
	14	120	3226.70
	15	105	3224.72
	16	94	3441.59
	17	97	3583.77
	18	117	3371.88
	19	69	3335.65
	20	88	3070.97
rfc	1	494	1552.31
	2	410	1571.10
	3	411	1636.04
	4	417	2189.11
	5	356	2091.46
	6	268	2135.43
	7	258	2427.61
	8	212	2507.00
	9	170	2576.94
	10	182	2752.69
	11	143	2900.47
	12	145	2968.47
	13	147	3315.47
	14	120	3234.03
	15	105	3173.82
	16	94	3557.42
	17	97	3640.56
	18	117	3719.88
	19	69	3677.68
	20	88	3658.43

Test Statistics[a,b]

	cbo	wmc	dit	rfc
Kruskal-Wallis H	1932.136	2483.392	1772.069	2253.392
df	110	110	110	110
Asymp. Sig.	.000	.000	<.001	.000

a. Kruskal Wallis Test

b. Grouping Variable: totalFields

CONCLUSION:

Through this experiment I learned how to perform hypotheses tests such as parametric test and non parametric test which includes kruskal Wallis, paired t-test, chi-square test, wilcoxin signed rank test. I also explore the SPSS tool (by IBM) to evaluate these tests on Software defect prediction dataset.