

It can be shown that the Pearson correlation between a continuous variable (Y) and a dichotomous variable (X) is functionally related to the independent samples t-test, and thus, the squared Pearson correlation is related to the ANOVA F-ratio.

```
proc ttest data=demo; class x; var y; run;
```

The TTEST Procedure

Variable: Y

X	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	8	17.0000	6.0000	2.1213	6.0000	23.0000
1	7	23.0000	5.2915	2.0000	15.0000	28.0000
Diff (1-2)		-6.0000	5.6840	2.9417		

X	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
0		17.0000	11.9839 22.0161	6.0000	3.9670 12.2116
1		23.0000	18.1062 27.8938	5.2915	3.4098 11.6522
Diff (1-2)	Pooled	-6.0000	-12.3552 0.3552	5.6840	4.1206 9.1571
Diff (1-2)	Satterthwaite	-6.0000	-12.2987 0.2987		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	13	-2.04	0.0623
Satterthwaite	Unequal	12.996	-2.06	0.0602

```
data demo;
input Y X;
cards;
23 0
20 0
13 0
19 0
21 0
12 0
22 0
6 0
25 1
16 1
28 1
24 1
26 1
15 1
27 1
```

```
proc corr data=demo pearson; var Y ; with X ; run;
```

Pearson Correlation Coefficients, N = 15

Prob > |r| under H0: Rho=0

	Y
X	0.49237 0.0623

Thus, R² = 0.242424

The t-test for a Pearson correlation is

$$t_{(N-2)} = \frac{r_{YX}}{\sqrt{(1-r_{YX}^2)/(N-2)}}$$

$$= 0.49237 / [(1-.242424)/(15-2)]^{1/2}$$

$$= 0.49237 / 0.2414 = 2.04$$

Squaring the t -test for a Pearson correlation is $t_{(N-2)} = \frac{r_{YX}}{\sqrt{(1-r_{YX}^2)/(N-2)}}$ results in an F-ratio with 1 and $N - 2$ dfs .

$[t_{(N-2)}]^2 = F_{(1,N-2)} = \frac{r_{YX}^2}{(1-r_{YX}^2)/(N-2)} = \frac{r_{YX}^2(N-2)}{(1-r_{YX}^2)}$, which is the a reduced form of the general F-ratio:

$$F = \frac{[R_{FULL}^2 - R_{REDUC}^2][N - df_{FULL} - 1]}{[1 - R_{FULL}^2][df_{FULL} - df_{REDUC}]}$$

The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square
Model	1	134.4000000	134.4000000
Error	13	420.0000000	32.3076923
Corrected Total	14	554.4000000	

R-Square	Coeff Var	Root MSE	Y Mean
0.242424	28.70700	5.683986	19.80000

Level of X	N	Mean	Std Dev
0	8	17.0000000	6.0000000
1	7	23.0000000	5.29150262

The F -test for a Pearson correlation is

$$F_{(1,N-2)} = \frac{r_{YX}^2(N-2)}{(1-r_{YX}^2)}$$

$$= 0.242424 * (15-2) / (1-.242424)$$

$$= 3.151512 / 0.775765 = 4.16$$

Note: The SS_{Between} is calculated as:

$$SSB = 8(17 - 19.8)^2 = 62.72$$

$$7(23 - 19.8)^2 = 71.68$$

$$\mathbf{134.40}$$

One of the most popular nonparametric alternatives is to use Rank-based tests. Most of these tests involve using a form of the OLS procedure on the ranks.

```
proc rank data=demo out=demo;var Y X;ranks ranky rankx;
proc print data=demo;run;
```

Spearman correlation is simply Pearson correlation performed on ranks.

```
proc corr data=demo pearson spearman;
var Y ranky;with X rankx;run;
```

Obs	Y	X	ranky	rankx
1	23	0	10	4.5
2	20	0	7	4.5
3	13	0	3	4.5
4	19	0	6	4.5
5	21	0	8	4.5
6	12	0	2	4.5
7	22	0	9	4.5
8	6	0	1	4.5
9	25	1	12	12.0
10	16	1	5	12.0
11	28	1	15	12.0
12	24	1	11	12.0
13	26	1	13	12.0
14	15	1	4	12.0
15	27	1	14	12.0

Pearson Correlation Coefficients, N = 15

Prob > |r| under H0: Rho=0

	Y	ranky
X	0.49237	0.55673
	0.0623	0.0311
rankx	0.49237	0.55673
Rank for Variable X	0.0623	0.0311

Spearman Correlation Coefficients, N = 15

Prob > |r| under H0: Rho=0

	Y	ranky
X	0.55673	0.55673
	0.0311	0.0311
rankx	0.55673	0.55673
Rank for Variable X	0.0311	0.0311

Note that X only has two values so ranking really does not affect it.

Spearman correlation of Y and X is the same as the Pearson correlation of Rank(Y) and Rank(X)

Note: SAS incorrectly uses a t-test and t-distribution for the Spearman correlation.

Untied Ranks will always have a Mean of $M(R) = (N+1)/2$ a Sums of Squares of $SS(R) = [N(N-1)(N+1)]/12$

Thus the Spearman correlation reduces to Spearman rho = $1 - \frac{6 \sum [R(Y)_i - R(X)_i]^2}{N(N^2 - 1)}$

Noting the relationship between the squared Pearson correlation and the ANOVA F-ratio the Kruskal-Wallis test has a similar relationship to the squared Spearman correlation.

Note from above that the squared Spearman correlation is $(0.55673)^2 = 0.30995$

```
proc glm data=demo ; class x;model ranky = x;means x;run;
```

The GLM Procedure

Dependent Variable: ranky Rank for Variable Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	86.7857143	86.7857143	5.84	0.0311
Error	13	193.2142857	14.8626374		
Corrected Total	14	280.0000000			

R-Square	Coeff Var	Root MSE	ranky Mean
0.309949	48.19011	3.855209	8.000000

Level of X	N	Mean	Std Dev	Note: The SS_{Between} is calculated as:
0	8	5.7500000	3.37003603	$SSB = 8(5.75 - 8)^2 = 40.5000$
1	7	10.5714286	4.35343324	$7(10.57 - 8)^2 = 46.2857$
				86.7857

Untied Ranks will always have a Mean of $M(R) = (N+1)/2$ a Sums of Squares of $SS(R) = [N(N-1)(N+1)]/12$

The Reduced form of the Kruskal Wallis Test shown in many texts is calculated as:

$$KW = \frac{12}{N(N+1)} \sum_{j=1}^J n_j \bar{R}_j^2 - 3(N+1) = \frac{12}{N(N+1)} \sum_{j=1}^J n_j \left(\bar{R}_j - \frac{(N+1)}{2} \right)^2 = \frac{12[SSB(R)]}{N(N+1)}$$

Where \bar{R}_j = Mean Rank for j^{th} group.

Multiplying both numerator and denominator by $(N-1)/12$ results in:

$$KW = \frac{12[SSB(R)]}{N(N+1)} = \frac{[SSB(R)](N-1)}{N(N+1)(N-1)/12} = \frac{[SSB(R)](N-1)}{SST(R)} = R^2_{(R)} * (N-1)$$

```
proc npar1way wilcoxon data=demo;
class X; var y ;exact wilcoxon; run;
```

The NPAR1WAY Procedure
Wilcoxon Scores (Rank Sums) for Variable Y
Classified by Variable X

X	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
0	8	46.0	64.0	8.640988	5.750000
1	7	74.0	56.0	8.640988	10.571429

Kruskal-Wallis Test

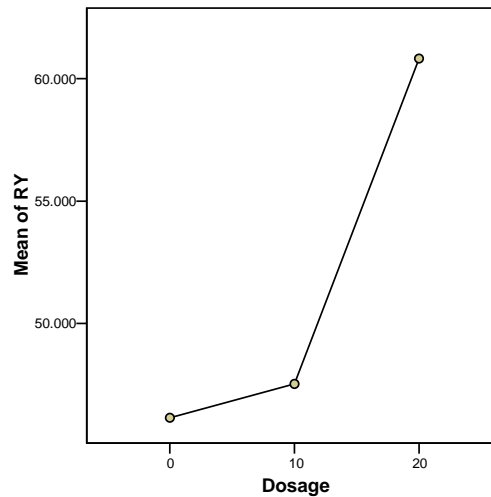
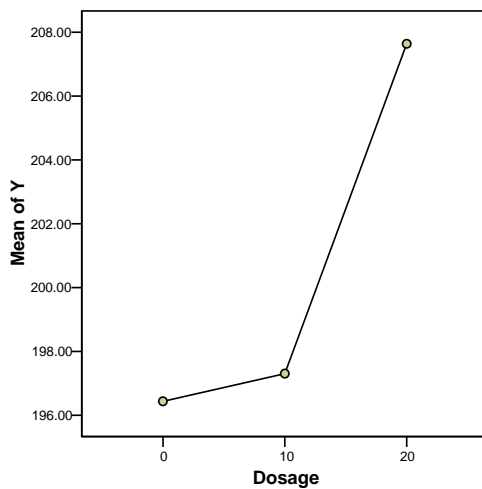
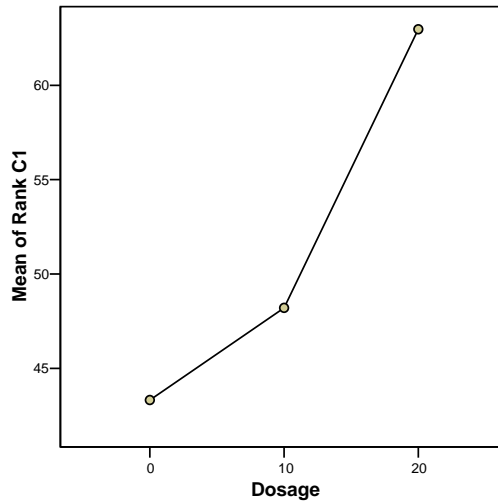
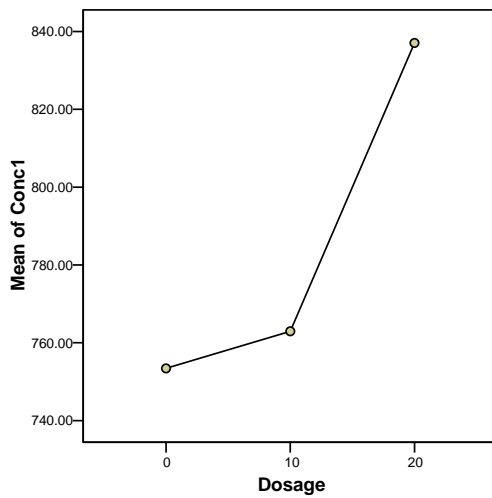
Chi-Square **4.3393** = Spearman R2*(N-1)

DF 1 = **0.309949 * (15-1)**

Pr > Chi-Square 0.0372 = **4.3393**

Wilcoxon Two-Sample Test	
Statistic (S)	74.0000
Normal Approximation	
Z	2.0252
One-Sided Pr > Z	0.0214
Two-Sided Pr > Z	0.0428
t Approximation	
One-Sided Pr > Z	0.0312
Two-Sided Pr > Z	0.0623
Exact Test	
One-Sided Pr >= S	0.0200
Two-Sided Pr >= S - Mean	0.0401
Z includes a continuity correction of 0.5.	

Dosage		Concen1	RANK of C1	Y	RANK of Y
0	Mean	753.4282	43.32	196.4393	46.14706
	N	34	34	34	34
	Std. Deviation	195.72879	27.257	19.11497	27.747926
	Kurtosis	2.315	-.103	-.497	-.969
	Skewness	1.643	1.048	.205	.270
10	Mean	762.9471	48.21	197.3066	47.52941
	N	34	34	34	34
	Std. Deviation	208.22749	30.180	22.30850	31.592323
	Kurtosis	-.562	-.928	-.819	-1.316
	Skewness	.036	.163	.025	.129
20	Mean	837.0703	62.97	207.6366	60.82353
	N	34	34	34	34
	Std. Deviation	206.98932	28.463	20.41367	27.860552
	Kurtosis	2.804	.270	-.023	-.904
	Skewness	-1.876	-1.188	.066	-.360
Total	Mean	784.4819	51.50	200.4608	51.50000
	N	102	102	102	102
	Std. Deviation	205.16868	29.589	21.07771	29.588849
	Kurtosis	.191	-1.200	-.487	-1.200
	Skewness	-.126	.000	.081	.000



ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Concen1 * Dosage	Between Groups	(Combined)	142583.097	2	71291.548	1.718	.185
		Linearity	118931.982	1	118931.982	2.866	.094
		Deviation from Linearity	23651.115	1	23651.115	.570	.452
	Within Groups		4108929.759	99	41504.341		
	Total		4251512.856	101			
RANK of C1 * Dosage	Between Groups	(Combined)	7115.529	2	3557.765	4.332	.016
		Linearity	6562.118	1	6562.118	7.990	.006
		Deviation from Linearity	553.412	1	553.412	.674	.414
	Within Groups		81309.971	99	821.313		
	Total		88425.500	101			
Y * Dosage	Between Groups	(Combined)	2638.888	2	1319.444	3.093	.050
		Linearity	2131.472	1	2131.472	4.997	.028
		Deviation from Linearity	507.416	1	507.416	1.189	.278
	Within Groups		42232.382	99	426.590		
	Total		44871.270	101			
RANK of Y * Dosage	Between Groups	(Combined)	4465.824	2	2232.912	2.633	.077
		Linearity	3661.779	1	3661.779	4.318	.040
		Deviation from Linearity	804.044	1	804.044	.948	.333
	Within Groups		83959.676	99	848.078		
	Total		88425.500	101			

Measures of Association

	Linear R	Linear R Squared	Eta	Eta Squared
Concen1 * Dosage	.167254	.027974	.183131	.033537
RANK C1 * Dosage	.272416	.074211	.283671	.080469
Y * Dosage	.217949	.047502	.242508	.058810
RANK of Y * Dosage	.203497	.041411	.224730	.050504

Test Statistics(a,b)

	Concen1	Y
Chi-Square	8.127	5.101
df	2	2
Asymp. Sig.	.017	.078

Eta-Square for Rank C1 = .080469 = 8.127/100 = KW-Chi-Square/(N-1)

Eta-Square for Rank Y = .050504 = 5.101/100 = KW-Chi-Square/(N-1)

a Kruskal Wallis Test

b Grouping Variable: Dosage

KW-Chi-Square for Concen1 = 8.127 = .080469*101 = Eta-squared*(N-1) = (N-1)*SSB_(R)/SST_(R)

KW-Chi-Square for Y = 5.101 = .050504*101 = Eta-squared*(N-1) = (N-1)*SSB_(R)/SST_(R)

Multiple Comparisons

Dependent Variable		(I) Dosage	(J) Dosage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Concen1	Tukey HSD	0	10	-9.51885	49.41085	.980	-127.0907	108.0530
			20	-83.64209	49.41085	.213	-201.2139	33.9297
		10	0	9.51885	49.41085	.980	-108.0530	127.0907
			20	-74.12324	49.41085	.295	-191.6951	43.4486
		20	0	83.64209	49.41085	.213	-33.9297	201.2139
			10	74.12324	49.41085	.295	-43.4486	191.6951
	Bonferroni	0	10	-9.51885	49.41085	1.000	-129.8510	110.8133
			20	-83.64209	49.41085	.281	-203.9743	36.6901
		10	0	9.51885	49.41085	1.000	-110.8133	129.8510
			20	-74.12324	49.41085	.410	-194.4554	46.2089
20		0	83.64209	49.41085	.281	-36.6901	203.9743	
		10	74.12324	49.41085	.410	-46.2089	194.4554	
RANK C1	Tukey HSD	0	10	-4.882	6.951	.763	-21.42	11.66
			20	-19.647(*)	6.951	.016	-36.19	-3.11
		10	0	4.882	6.951	.763	-11.66	21.42
			20	-14.765	6.951	.090	-31.30	1.77
		20	0	19.647(*)	6.951	.016	3.11	36.19
			10	14.765	6.951	.090	-1.77	31.30
	Bonferroni	0	10	-4.882	6.951	1.000	-21.81	12.05
			20	-19.647(*)	6.951	.017	-36.57	-2.72
		10	0	4.882	6.951	1.000	-12.05	21.81
			20	-14.765	6.951	.108	-31.69	2.16
20		0	19.647(*)	6.951	.017	2.72	36.57	
		10	14.765	6.951	.108	-2.16	31.69	
Y	Tukey HSD	0	10	-.86729	5.00934	.984	-12.7869	11.0523
			20	-11.19735	5.00934	.070	-23.1170	.7222
		10	0	.86729	5.00934	.984	-11.0523	12.7869
			20	-10.33006	5.00934	.103	-22.2497	1.5895
		20	0	11.19735	5.00934	.070	-.7222	23.1170
			10	10.33006	5.00934	.103	-1.5895	22.2497
	Bonferroni	0	10	-.86729	5.00934	1.000	-13.0667	11.3322
			20	-11.19735	5.00934	.083	-23.3968	1.0021
		10	0	.86729	5.00934	1.000	-11.3322	13.0667
			20	-10.33006	5.00934	.125	-22.5295	1.8694
20		0	11.19735	5.00934	.083	-1.0021	23.3968	
		10	10.33006	5.00934	.125	-1.8694	22.5295	
RANK Y	Tukey HSD	0	10	-1.382353	7.063067	.979	-18.18874	15.42403
			20	-14.676471	7.063067	.100	-31.48285	2.12991
		10	0	1.382353	7.063067	.979	-15.42403	18.18874
			20	-13.294118	7.063067	.149	-30.10050	3.51227
		20	0	14.676471	7.063067	.100	-2.12991	31.48285
			10	13.294118	7.063067	.149	-3.51227	30.10050
	Bonferroni	0	10	-1.382353	7.063067	1.000	-18.58332	15.81861
			20	-14.676471	7.063067	.121	-31.87744	2.52449
		10	0	1.382353	7.063067	1.000	-15.81861	18.58332
			20	-13.294118	7.063067	.188	-30.49508	3.90685
20		0	14.676471	7.063067	.121	-2.52449	31.87744	
		10	13.294118	7.063067	.188	-3.90685	30.49508	

* The mean difference is significant at the .05 level.

