

There are many “standard” tests like the t-tests and analyses of variance that are commonly used. They rest on assumptions like normality, which can be hard to assess: for example, if you have small samples, how do you assess that the population(s) they came from are normal, when all you have to go on is your small samples? Also, the various kinds of analysis of variance require equal spreads of the observations within each group, and it is hard to be sure whether unequal sample variances mean that the population variances could be equal or not.

Historically, there was also a desire for tests where the calculation was easy (in the days before computers), so various so-called “non-parametric” tests were devised, which carried fewer assumptions than the standard tests. If the data really were normal, these tests would be less powerful than the standard ones, but if the data were sufficiently non-normal to make the standard tests invalid, the non-parametric tests would still be good. Many of these tests, instead of using the original data, rank the data values first in some way, then use the ranks, which obviously reduces the influence of outliers, and usually makes the data distributions less non-normal.

In many situations, there is a non-parametric test that corresponds to the standard test, as described below:

Situation	Standard test	Non-parametric test
One sample, matched pairs	One-sample t-test	Sign test, signed rank test
Two independent samples	Two-sample t-test	Rank sum test
Two samples, comparing spread	F-test	Siegel-Tukey test
One-way ANOVA	F-test	Kruskal-Wallis test, median test
Randomized block design	F-test	Friedman test
Ordered alternative in ANOVA	-	Terpstra-Jonckheere test
Correlation	Pearson correlation	Spearman, Kendall correlation
Regression	Regression	Iman-Conover regression

Also, these are all tests, whereas we might be looking for confidence intervals. There is a standard procedure that can obtain a confidence interval from a test result, which can be applied to non-parametric tests as well.

Since we are using SAS, simplicity of computation is not an issue for us, but the easy-to-compute (by hand) tests often perform surprisingly well. Where possible, we will compare the standard test with the non-parametric one.

To start with single-sample tests for location: here are some data on time spent by orders in the administration department of a company (in working hours). Interest is in how the “typical” time compares to 8 hours.

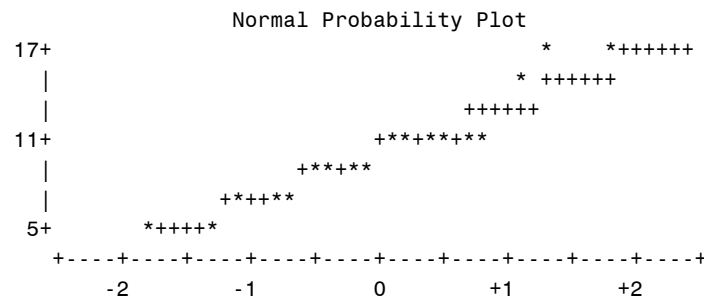
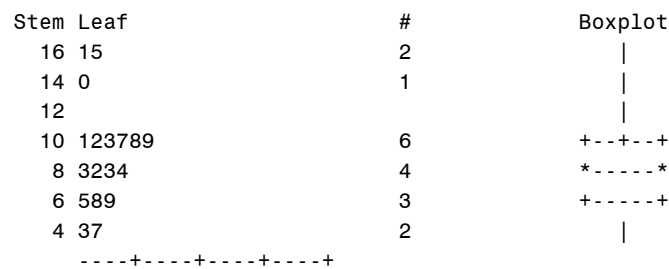
16.5 4.3 10.1 10.7 14.0 7.9 7.5 8.3 5.7 10.3 9.3 10.2 9.2 7.8 11.9 11.8 16.1 9.4

SAS's PROC UNIVARIATE does all of the one-sample tests, standard and non-parametric. We have to tell it that the null-hypothesis mean is 8, like this:

```
data admin;
  infile "admin.dat";
  input x @@;

proc univariate plot mu0=8;
```

The data values are all on one line, hence the @@. Looking at the plots first is probably helpful:



The data are mildly non-normal and seem skewed to the right, so maybe the median is a better measure of location than the mean. The sign test and signed-rank test are both testing the *median*, but the signed-rank test has the extra assumption that the distribution is symmetric, which is questionable here. The test results are:

Tests for Location: Mu0=8			
Test	-Statistic-	-----p Value-----	
Student's t	t 2.727938	Pr > t	0.0143
Sign	M 4	Pr >= M	0.0963
Signed Rank	S 55	Pr >= S	0.0142

The t-test and the signed-rank test both give significant results (very similar P-values), but we have doubts about both normality and symmetry, so maybe we should be looking at the non-significant sign

test.

The sign test just looks at how many observations are above 8 (13) and below 8 (5), without considering how far above or below. The signed-rank test is essentially a t-test on the ranked data.

[The by-hand procedure: subtract the null-hypothesis mean from each observation, and rank the differences from smallest to largest *without considering whether they are plus or minus*. Then add up the ranks that correspond to the plusses or minuses, whichever is smaller.]

Matched-pair data (for example, before and after measurements on a number of subjects) are handled as above: in the DATA step, calculate the differences between the two measurements for each subject, and then test the null hypothesis that the mean (median) is zero for the differences.

Next, we have some data on the percentage of Japanese and British cars that break down in their first year (8 Japanese models and 12 British ones):

```
japan 3
japan 7
japan 15
japan 10
japan 4
japan 6
japan 4
japan 7
britain 19
britain 11
britain 36
britain 8
britain 25
britain 23
britain 38
britain 14
britain 17
britain 41
britain 25
britain 21
```

Are Japanese cars more reliable than British ones? The t-test is testing that the mean proportions are equal, but the test is ignoring that some of the British cars appear much less reliable than the rest (outliers). The rank sum test is based on this idea: if you pick a Japanese car and a British car at random, is it likely that the Japanese car's reliability will be better than the British one's (the answer appears to be "yes").

SAS has to do the two tests in two different PROCs. SAS's name for the rank sum test is "Wilcoxon":

```

data cars;
  infile "cars.dat";
  input type $ breakdowns;

proc ttest;
  class type;
  var breakdowns;

proc npar1way wilcoxon st;
  class type;
  var breakdowns;

```

(the ST thing I'll explain in a minute), with this output:

The TTEST Procedure

Variable: breakdowns

type	N	Mean	Std Dev	Std Err	Minimum	Maximum
britain	12	23.1667	10.5816	3.0546	8.0000	41.0000
japan	8	7.0000	3.9279	1.3887	3.0000	15.0000
Diff (1-2)		16.1667	8.6270	3.9377		

type	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
britain		23.1667	16.4435 29.8899	10.5816	7.4959 17.9662
japan		7.0000	3.7162 10.2838	3.9279	2.5970 7.9944
Diff (1-2)	Pooled	16.1667	7.8939 24.4394	8.6270	6.5187 12.7579
Diff (1-2)	Satterthwaite	16.1667	9.0150 23.3184		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	18	4.11	0.0007
Satterthwaite	Unequal	15.009	4.82	0.0002

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	11	7	7.26	0.0148

The equality of variances fails, so we should look at the Satterthwaite version of the test. The P-value is very small, so we'd reject the hypothesis that the means are equal.

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable breakdowns
Classified by Variable type

type	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
japan	8	40.0	84.0	12.946855	5.000000
britain	12	170.0	126.0	12.946855	14.166667

Average scores were used for ties.

```

Wilcoxon Two-Sample Test

Statistic                40.0000

Normal Approximation
Z                        -3.3599
One-Sided Pr < Z        0.0004
Two-Sided Pr > |Z|      0.0008

t Approximation
One-Sided Pr < Z        0.0016
Two-Sided Pr > |Z|      0.0033

```

Z includes a continuity correction of 0.5.

```

Kruskal-Wallis Test

Chi-Square              11.5499
DF                      1
Pr > Chi-Square         0.0007

```

Compare the “sum of scores” with “expected under H_0 ” to see that the Japanese figures are on average quite a bit lower than the British ones. All the approximations say that the P-value is very small, so the null is rejected again.

In PROC TTEST, SAS gives two-sided P-values. We were doing a one-sided test (are the Japanese cars more reliable?), so we check that the sample means were the correct way around (they were) and then divide the P-value given by SAS, 0.0002, by 2. In PROC NPAR1WAY, the one-sided P-values are given; they are both small.

[By hand: rank all the data, then sum up the ranks of the values that came from the first sample, or the second sample, whichever is smaller.]

The Kruskal-Wallis test is used for one-way ANOVA situations, but just as you can also do ANOVA when comparing only 2 means, so you can use Kruskal-Wallis in the same situation. It gives a similar P-value here.

As we just saw, PROC TTEST does a 2-sample test for spread based on the sample variances. This could be questionable if variances are not the thing to use: if you have outliers, or if your population distributions are not normal. There is a non-parametric equivalent, which might be better for our car data (we were concerned about outliers), called the Siegel-Tukey test, which PROC NPAR1WAY will also do. (That's what the ST was.) The output looks like the rank sum test; in fact, the Siegel-Tukey test is just like the rank-sum test except that instead of assigning ranks to data values up from the smallest, it does so from the highest and lowest to the middle, so that if the two samples of the same size have the same spread, their Siegel-Tukey rank sums will be the same. (If they are not the same size, their rank *means* will be the same if the spreads are the same.)

The NPAR1WAY Procedure

Siegel-Tukey Scores for Variable breakdowns
Classified by Variable type

type	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
japan	8	74.0	84.0	12.868566	9.250000
britain	12	136.0	126.0	12.868566	11.333333

Average scores were used for ties.

Siegel-Tukey Two-Sample Test

Statistic	74.0000
Z	-0.7382
One-Sided Pr < Z	0.2302
Two-Sided Pr > Z	0.4604

Z includes a continuity correction of 0.5.

Siegel-Tukey One-Way Analysis

Chi-Square	0.6039
DF	1
Pr > Chi-Square	0.4371

This test says that there is no evidence of a difference in spreads, in contrast to the F-test. This surprises me, because the British values are “obviously” more spread out, however you measure spread – for example, the interquartile ranges are 15 and 4.5, hardly equal.

PROC NPAR1WAY will also do standard and non-parametric one-way ANOVA. Another study of the reliability of cars, this time from 4 different countries, gave this:

```

britain 21
britain 33
britain 12
britain 28
britain 41
britain 39
britain 24
britain 29
britain 30
britain 19
britain 27
britain 38
britain 23
japan 9
japan 13
japan 6
japan 3
japan 5
japan 10
japan 4
    
```

```

germany 18
germany 35
germany 8
germany 17
germany 22
germany 20
germany 37
germany 11
italy 41
italy 41
italy 48
italy 34

```

The differences in location seem to be pretty obvious, but there are also differences in spread that we need to worry about: the British and German values seem more spread out than the others. We'll do a standard ANOVA as well as the Kruskal-Wallis test (called, confusingly, "wilcoxon" by PROC NPAR1WAY) and the Mood median test (called "median" in PROC NPAR1WAY; the "mood" option there is something else! Thus, this code:

```

data cars;
  infile "cars2.dat";
  input country $ percent;

proc npar1way anova wilcoxon median;
  var percent;
  class country;

```

produces this output:

The NPAR1WAY Procedure

Analysis of Variance for Variable percent
Classified by Variable country

country	N	Mean
britain	13	28.000000
japan	7	7.142857
germany	8	21.000000
italy	4	41.000000

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Among	3	3410.017857	1136.672619	17.9523	<.0001
Within	28	1772.857143	63.316327		

Wilcoxon Scores (Rank Sums) for Variable percent
Classified by Variable country

country	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
britain	13	262.0	214.50	26.052870	20.153846
japan	7	33.0	115.50	21.929368	4.714286
germany	8	117.0	132.00	22.969826	14.625000
italy	4	116.0	66.00	17.543494	29.000000

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square 20.4582
DF 3
Pr > Chi-Square 0.0001

The NPAR1WAY Procedure

Median Scores (Number of Points Above Median) for Variable percent
Classified by Variable country

country	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
britain	13	10.0	6.50	1.411359	0.769231
japan	7	0.0	3.50	1.187977	0.000000
germany	8	2.0	4.00	1.244342	0.250000
italy	4	4.0	2.00	0.950382	1.000000

Average scores were used for ties.

Median One-Way Analysis

Chi-Square 16.2452
DF 3
Pr > Chi-Square 0.0010

The standard ANOVA produces a small P-value, but we should treat this with caution because of the unequal spreads within the groups. The Kruskal-Wallis test also produces a small P-value, which is more trustworthy because it is (essentially) a one-way ANOVA on the ranked data, which shouldn't be troubled by outliers.

Mood's median test is the easiest one to do by hand: work out the overall median (of all the data), and count how many values are above and below the median in each group. If the groups are all about the same, the observations should be about 50-50 above and below the overall median in each group. Here the overall median is 22.5; all the Japan figures are below this, and all the Italy figures are above. The counts of below-median and above-median for each country form a two-way table, which is then analyzed using a chi-squared test. Mood's median test is a lot like the sign test generalized to two or more groups.

In the one-way ANOVA situation, you might know the order in which the groups should lie, if the null hypothesis is not true. A test that assesses the evidence against the null, while allowing for the alternative order, should be more powerful (should give you a greater chance of rejecting the null in favour of that specific alternative) than, say, the Kruskal-Wallis test. Such a test is the Jonckheere-Terpstra test; it is really a rank sum test done on each pair of groups, so that if the ordering is as claimed, there will be a bunch of small rank sum test statistics to add together.

Some data: these are lifetimes of light bulbs of three different brands. Brand A is bought in a discount store, brand B in a grocery store, and brand C in a hardware store. We might expect brand C lightbulbs to last longer on average than brand B, which in turn last longer than brand A. We can test an alternative that the lifetimes go A, B, C in that order.

```
a 619
a 35
a 126
a 2031
a 215
b 343
b 2437
b 409
b 267
b 1953
b 1804
c 3670
c 2860
c 502
c 2008
c 5004
c 4782
```

These lifetimes are obviously not normal (lifetimes of things are usually skewed to the right) so that even if we had a “standard” test to use against the ordered alternative, it wouldn't apply here, and we would want to use something non-parametric, which the Jonckheere-Terpstra test is.

SAS doesn't make it that easy to do this test: it is hiding as an option of the TABLES command in PROC FREQ (which is usually for making contingency tables, doing chi-squared tests and the like). Also, it is a good idea to arrange your labels (here a, b, c) in the order that they appear in the alternative hypothesis (SAS can understand which order you want, but I wasn't able to get it working reliably).

Here is my code:

```
data light;
  infile "lightbulbs.dat";
  input brand $ lifetime;

proc freq;
  tables brand*lifetime / jt;
```

Never mind that “lifetime” isn't in any sense a categorical variable, so that making a contingency table of it with “brand” seems kind of pointless. Here's what you get (which is SAS's best effort at making a contingency table anyway):

Statistics for Table of brand by lifetime

Jonckheere-Terpstra Test	
Statistic	81.0000
Z	2.9055
One-sided Pr > Z	0.0018
Two-sided Pr > Z	0.0037

Sample Size = 17

The table suggests that most of the short lifetimes went with Brand A and most of the long ones with Brand C, which is confirmed by the test: the P-value is very small, so we reject a null hypothesis that the “average” lifetimes are equal in favour of the alternative A, B, C.

You could do an ordinary ANOVA here, like this, with some trepidation:

```
proc glm;
  class brand;
  model lifetime=brand;
```

with the following output:

The GLM Procedure

Dependent Variable: lifetime

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	19879674.56	9939837.28	6.27	0.0114
Error	14	22203092.97	1585935.21		
Corrected Total	16	42082767.53			

R-Square	Coeff Var	Root MSE	lifetime Mean
0.472395	73.65824	1259.339	1709.706

Source	DF	Type I SS	Mean Square	F Value	Pr > F
brand	2	19879674.56	9939837.28	6.27	0.0114

Source	DF	Type III SS	Mean Square	F Value	Pr > F
brand	2	19879674.56	9939837.28	6.27	0.0114

This P-value of 0.0114 is bigger than we had from Jonckheere-Terpstra, for a couple of reasons: one, we shouldn't really have done an ordinary ANOVA here (the data are far from normal), and two, the ANOVA alternative is just “the means are not all the same”, which is not really the alternative that we want: we'd have to go on and do multiple comparisons to verify that the means came out in the right order.

The next step beyond one-way ANOVA is a randomized blocks design. An example: we want to see whether people's pulse measurements vary during the day. We have three patients available, and we measure each one's pulse at 6 and 10 am and pm. We could just think of the three patients as replicates and do a one-way ANOVA (or Kruskal-Wallis), but some people have “typically” different pulse rates than others, which means that an observed pulse rate could depend both on time of day and patient. This can be handled by treating patients as “blocks”, which means we are taking it for granted that patients will be different, and allowing for patient differences in assessing the time-of-day effect (like including a second variable in a multiple regression not because we care about it, but because we want to allow for its effect).

In the non-parametric world, we can use Friedman's test (named after Milton Friedman the economist, who was apparently doing some moonlighting). The idea is to use ranks *within* blocks. Here's some data:

	Patient		
Time	K	F	R
0600	60	62	55
1000	65	66	65
1800	70	80	75
2200	65	72	68

Patient F had the highest pulse at all the times, but we don't really care about that: what appears to be happening is that pulse rates rise to a maximum at 6 pm and then drop. Ranking within columns makes this clear:

	Patient		
Time	K	F	R
0600	1	1	1
1000	2.5	2	2
1800	4	4	4
2200	2.5	3	3

Even though the actual pulse rates are not the same, the rankings over time of day are almost identical, and certainly way more similar than you'd expect by chance.

SAS claims to be able to do the Friedman test, but it is a very obscure process, and I prefer to construct a “home-brew” test by obtaining the ranks and doing a randomized-blocks ANOVA (a two-way ANOVA without interaction) on the ranks.

Here are the data for SAS purposes:

```
0600 K 60
0600 F 62
0600 R 55
1000 K 65
1000 F 66
1000 R 65
1800 K 70
1800 F 80
1800 R 75
2200 K 65
2200 F 72
2200 R 68
```

We first need to obtain the ranks of the pulse rates within each patient, and to do *that* we need the data sorted by patient first. Once we've done that, we run the ANOVA on the ranks using PROC GLM. Here's all the code:

```
data pulse;
  infile "pulse.dat";
  input time $ patient $ pulse;

proc sort;
  by patient;

proc rank;
  by patient;
  var pulse;
  ranks rpulse;

proc print;

proc glm;
  class patient time;
  model rpulse=time patient;
```

PROC RANK produces a new data set that contains the ranked pulse values within each patient, with the stored ranks in the variable "rpulse". I then do a PROC PRINT so that you can see what the output of PROC RANK looks like, and then I run it through PROC GLM using the ranked-within-patient pulse rates as the response.

Here is the output from PROC RANK:

Obs	time	patient	pulse	rpulse
1	0600	F	62	1.0
2	1000	F	66	2.0
3	1800	F	80	4.0
4	2200	F	72	3.0
5	0600	K	60	1.0
6	1000	K	65	2.5
7	1800	K	70	4.0
8	2200	K	65	2.5
9	0600	R	55	1.0
10	1000	R	65	2.0
11	1800	R	75	4.0
12	2200	R	68	3.0

where you can compare the ranked pulse values for each patient with the table above. (Patient K had two values 65, which share the 2nd and 3rd ranks.) The ranks are very consistent over time, which should show up in the ANOVA of ranks:

The GLM Procedure

Dependent Variable: rpulse Rank for Variable pulse

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	14.16666667	2.833333333	51.00	<.0001
Error	6	0.333333333	0.055555556		
Corrected Total	11	14.50000000			

R-Square	Coeff Var	Root MSE	rpulse Mean
0.977011	9.428090	0.235702	2.500000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
time	3	14.16666667	4.722222222	85.00	<.0001
patient	2	0.00000000	0.00000000	0.00	1.0000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
time	3	14.16666667	4.722222222	85.00	<.0001
patient	2	0.00000000	0.00000000	0.00	1.0000

as indeed it does. The P-value for a difference due to times is as small as it could be. (We organized the ranks so that ranks 1, 2, 3, 4 appear once for each patient, apart from the tied values, so that even if there is a patient effect with the original data, there isn't with the ranks.)

If you hadn't allowed for differences between patients, you would have done Kruskal-Wallis and gotten this:

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable pulse
Classified by Variable time

time	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
0600	3	6.0	19.50	5.370373	2.000000
1000	3	17.0	19.50	5.370373	5.666667
1800	3	32.0	19.50	5.370373	10.666667
2200	3	23.0	19.50	5.370373	7.666667

Average scores were used for ties.

Kruskal-Wallis Test

Chi-Square	9.2837
DF	3
Pr > Chi-Square	0.0257

for which the P-value is also small. The P-value from Friedman's test is smaller, because it allows for differences due to patients as well. (In less fortunate circumstances, the Kruskal-Wallis test might not have been significant at all because the difference between patients might have overwhelmed the difference due to time of day.)

[There aren't specific non-parametric tests for two-way ANOVA and above, but you can always replace the data by their ranks and see whether it makes any difference. You rank *all* the data, and then look to see whether the mean ranks are significantly different.]

Our final procedure is correlation. There is the usual "Pearson" correlation, but there are also two kinds of non-parametric correlation. SAS PROC CORR handles them all. As an example: a company is trying to see what its sales staff can do to increase sales. One kind of visit a sales rep can do is a "demo", where the sales rep goes to a potential customer and demonstrates how the product works (and answers questions about it). A record is kept of the number of demos and number of sales for each sales rep. The data look like this, with demos in the first column and sales in the second:

```

96 17
104 25
113 33
115 34
126 36
135 41
137 42
142 42
143 43
145 43
149 44
151 44
156 45
161 45
175 46

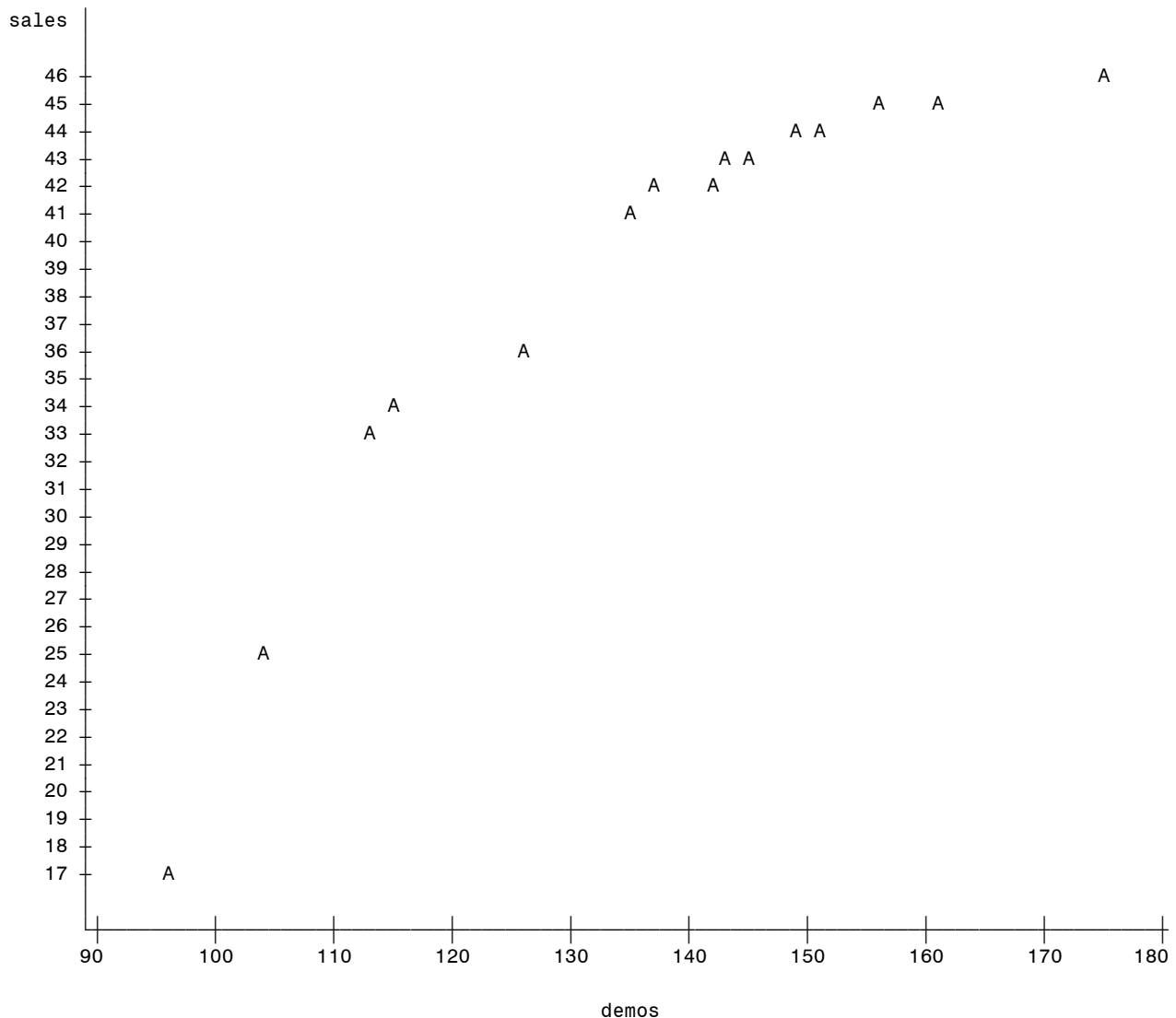
```

The first step with this kind of data is to draw a scatterplot to get a sense of the relationship. Then we have SAS calculate the correlations, the usual one called “Pearson”, and the two non-parametric ones called “Spearman” and “Kendall”. The code looks like this:

```
data demos;  
  infile "demos.dat";  
  input demos sales;  
  
proc plot;  
  plot sales*demos;  
  
proc corr pearson spearman kendall noprob;  
  var sales demos;
```

SAS by default does a test that each correlation is zero. A look at the plot reveals that there is obviously a relationship, so we skip the tests (with “noprob”):

Plot of sales*demos. Legend: A = 1 obs, B = 2 obs, etc.



The trend is clearly upward, but it appears to level off about about 140 demos, so there is a “diminishing returns” thing happening here. Recall that the usual Pearson correlation is only good for straight-line relationships, so we shouldn't trust it here. However, the Spearman and Kendall correlations are good for any “monotonic” relationship: one that keeps going up (like this one) or keeps going down, even if the trend is a curve.

The Spearman correlation uses the ranked demos and sales; it is the ordinary correlation but of the ranks. So it will be high if a low rank on one variable goes with a low rank on the other, and high with high: here, the smallest number of demos goes with the smallest number of sales, and the largest with the largest, so the Spearman correlation will be high even though the trend is curved.

The Kendall correlation is even “more non-parametric” than the Spearman one: take a pair of points, say the two at the bottom left. For the one where the demos is higher, is the number of sales higher? Here, the answer is “yes”; this pair is called “concordant”. This is then repeated for all pairs of points, and here almost all of them are concordant, so the Kendall correlation should also be close to 1:

The CORR Procedure

2 Variables: sales demos

Simple Statistics

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
sales	15	38.66667	8.33809	42.00000	17.00000	46.00000
demos	15	136.53333	22.03525	142.00000	96.00000	175.00000

Pearson Correlation Coefficients, N = 15

	sales	demos
sales	1.00000	0.92124
demos	0.92124	1.00000

Spearman Correlation Coefficients, N = 15

	sales	demos
sales	1.00000	0.99642
demos	0.99642	1.00000

Kendall Tau b Correlation Coefficients, N = 15

	sales	demos
sales	1.00000	0.98077
demos	0.98077	1.00000

The correlations are: Pearson 0.921, Spearman 0.996, Kendall 0.981. The two non-parametric ones are higher because the trend isn't a straight line. (Spearman usually comes out closer to 1 than Kendall does).

SAS does tests that all 3 correlations are zero; for Pearson and Spearman, you can also test a specific value in the null, say 0.9, like this:

```
proc corr pearson spearman fisher(rho0=0.9);
  var sales demos;
```

with this output:

Pearson Correlation Statistics (Fisher's z Transformation)

Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate
sales	demos	15	0.92124	1.59718	0.03290	0.91611

Pearson Correlation Statistics (Fisher's z Transformation)

Variable	With Variable	95% Confidence Limits		-----HO:Rho=Rho0----- Rho0	p Value
sales	demos	0.760959	0.972153	0.90000	0.7478

The CORR Procedure

Spearman Correlation Statistics (Fisher's z Transformation)

Variable	With Variable	N	Sample Correlation	Fisher's z	Bias Adjustment	Correlation Estimate
sales	demos	15	0.99642	3.16218	0.03559	0.99616

Spearman Correlation Statistics (Fisher's z Transformation)

Variable	With Variable	95% Confidence Limits		-----HO:Rho=Rho0----- Rho0	p Value
sales	demos	0.988138	0.998760	0.90000	<.0001

Look at the second of each pair of lines. The Pearson correlation is not significantly different from 0.9, and the 95% confidence interval includes 0.9, but the Spearman correlation *is* significantly different from 0.9, and the confidence interval shows that it is significantly higher.