FACTOR ANALYSIS

```
proc import out=campaign
  datafile=".\cleaned.csv"
   dbms=csv replace;
run;

proc factor data=campaign method=ml nfactors=4 heywood;
  var Income Kidhome Teenhome Recency NumDealsPurchases NumWebPurchases
NumCatalogPurchases NumStorePurchases NumWebVisitsMonth AcceptedCmp3
AcceptedCmp4 AcceptedCmp5 AcceptedCmp1 AcceptedCmp Complain Response
CustomerAge TimeASCustomer;
run;
```

Significance Tests Based on 2212 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	153	11104.6437	<.0001
HA: At least one common factor			
H0: 4 Factors are sufficient	87	1175.7603	<.0001
HA: More factors are needed			

Low p-value so a significant lack of fit. Need more factors.

Recalling: $p(m+1) \leq p(p+1)/2$, so if we are using 18 variables we do not want to exceed 8 factors.

```
proc factor data=campaign method=ml nfactors=8 heywood;
  var Income Kidhome Teenhome Recency NumDealsPurchases NumWebPurchases
NumCatalogPurchases NumStorePurchases NumWebVisitsMonth AcceptedCmp3
AcceptedCmp4 AcceptedCmp5 AcceptedCmp1 AcceptedCmp Complain Response
CustomerAge TimeASCustomer;
run;
```

Significance Tests Based on 2212 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	153	11104.6437	<.0001
HA: At least one common factor			
H0: 8 Factors are sufficient	37	159.2050	<.0001
HA: More factors are needed			

Similar result, 8 factors is not sufficient.

Reducing the number of variables used, and using 2 factors

```
proc factor data=campaign method=ml nfactors=2 heywood;
  var Income Kidhome Teenhome Recency CustomerAge TimeASCustomer;
run;
```

Significance Tests Based on 2212 Observations			
Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	15	1164.4171	<.0001
HA: At least one common factor			
H0: 2 Factors are sufficient	4	12.2987	0.0153
HA: More factors are needed			

P-value has increased, but still not significant.

Another method to attempt is factor rotation:

```
proc factor data=campaign method=ml nfactors=4 heywood
      rotate=varimax;
  var Income Kidhome Teenhome Recency NumDealsPurchases NumWebPurchases
NumCatalogPurchases NumStorePurchases NumWebVisitsMonth AcceptedCmp3
AcceptedCmp4 AcceptedCmp5 AcceptedCmp1 AcceptedCmp Complain Response
CustomerAge TimeASCustomer;
run;
```

Significance Tests Based on 2212 Observations			
Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	153	11104.6437	<.0001
HA: At least one common factor			
H0: 4 Factors are sufficient	87	1175.7603	<.0001
HA: More factors are needed			

Still not significant. Even running it with 8 factors is not significant. Continuing to interpretation anyway.

Rotated Factor Pattern				
	Factor1	Factor2	Factor3	Factor4
Income	0.86092	0.18557	0.12693	-0.07795
Kidhome	-0.62285	-0.04545	-0.09019	0.00801
Teenhome	-0.07033	-0.10392	0.98311	0.13325
Recency	0.01985	-0.08378	0.00802	-0.00602
NumDealsPurchases	-0.09867	-0.06946	0.31344	0.48405
NumWebPurchases	0.52528	0.10202	0.14246	0.52329
NumCatalogPurchases	0.75450	0.18164	-0.04441	0.02348
NumStorePurchases	0.74643	-0.01475	0.07630	0.18886
NumWebVisitsMonth	-0.72195	0.08113	0.00981	0.59282
AcceptedCmp3	-0.03451	0.21529	-0.02793	0.03461
AcceptedCmp4	0.15696	0.41564	0.08498	0.06649
AcceptedCmp5	0.32431	0.58149	-0.08859	-0.14846
AcceptedCmp1	0.27233	0.51109	-0.06405	-0.07479
AcceptedCmp2	0.05263	0.36949	0.02786	-0.00689
Complain	-0.02608	-0.01475	0.00308	0.01014
Response	0.11305	0.51868	-0.10682	0.09357
CustomerAge	0.18742	-0.03077	0.37662	0.01244
TimeASCustomer	0.01625	0.08190	-0.03394	0.46213

Interpretation from factor rotation:

Factor 1 is primarily a measure of Income, Number of Kids, Number of Web Purchases, N. Catalog Purchases, N. Store Purchases, N. Web Visits

Factor 2 is primarily a measure of Accepting AD campaigns 5, 1, and the most recent ad campaign.

Factor 3 is primarily a measure of Number of Teenagers

Factors 4 overlaps with pieces of Factors 1: measure of Web Purchases and Web Visits