

Using Parallel Analysis to Determine the Dimensionality of the Computer-Based Prostate Cancer Screen

Abbas S. Tavakoli, DrPH, MPH, ME¹, Otis L. Owens, PhD, MPH², Nikki R. Wooten, PhD, LISW-CP²

¹University of South Carolina, College of Nursing

²University of South Carolina, College of Social Work

Abstract

Prostate cancer (PrCA) incidence and mortality rates are higher among African-American men than any other racial group. Informed decision making about prostate cancer screening could result in early detection and potentially reduce cancer health disparities. Currently, there are some, but few, computer-based decision aids to facilitate PrCA decisions of African-American men, but no scale has been validated to assess the extent to which African-American men will accept and use a computer-based PrCA screening decision aid. Using parallel analysis, this study determined the dimensionality of the Computer-Based Prostate Cancer Screening Decision Aid and Acceptance Scale using data from a purposive sample of 352 African-American men aged 40 years and older who resided in South Carolina. Exploratory factor analysis was conducted using maximum likelihood, squared multiple correlations, and Promax rotation. Internal consistency reliability was assessed using Cronbach's alpha. Pearson's correlation assessed the association between factors and subscale items. Parallel analysis was used to determine the dimensionality of the scale using SAS Macro. Results showed the optimal factor structure of the Computer-Based Prostate Cancer Screening Decision Aid among African American men was a 24-item, 3-factor model. Factor loadings ranged from 0.32 to 0.94 with 11 items loading on Factor 1, 8 items on Factor 2, and 5 items on Factor 3. There was a negative weak correlation between Factors 1 and 3, a strong positive correlation between Factors 1 and 2, and no correlation between Factors 2 and 3. There were also positive correlations between factors, the total scale, and subscales ranging from .32 to .85. Internal consistency reliability of each subscale ranged from 0.87 to 0.94 and was 0.87 for the total scale. Parallel analysis is a valuable method for determining the dimensionality of the Computer-Based Prostate Cancer Screening Decision Aid.

Keywords: SAS, parallel analysis, exploratory factor analysis, psychometrics

Background

Prostate cancer (PrCA) incidence and mortality rates are higher among African-American men than any other racial group.¹ PrCA screening could result in early detection and reduce mortality burden among these men. According to the American Urological Association, it is important to make an informed decision about PrCA screening with a healthcare provider after talking about the risks and potential benefits of screening.² Prior studies indicate that men participating in informed decision making interventions (e.g., computer-based decision aids) for PrCA screening have increased knowledge and higher decision self-efficacy,³ which can enhance their ability to engage in shared decision making. While computer-based technologies are effective methods for disseminating PrCA information to diverse populations including African-American men,⁴ no current scales have assessed the conditions under which African-American men will accept and routinely use computerized PrCA decision aids.

Purpose

This study used parallel analysis to determine the dimensionality of Computer-Based Prostate Cancer Screening Decision Aid Acceptance Scale (PCBDAA)⁵ among African-American men.

Methodology

This study used baseline data from a purposive sample of 352 self-identified African-American men aged 40 years and older who were residents of South Carolina. Inclusion criteria for this study were the ability to speak and comprehend English, no personal history of PrCA, and no self-reported history of cognitive decline. The 24-item PCBDAA was developed using the Unified Theory of Acceptance and Use of Technology⁶ and measures acceptance and use of a specific technology for PrCA screening decision making using a 5-point Likert scale from 1 (*strongly agree*) to 5 (*strongly disagree*). A sample item asks respondents whether they believe a specific technology will help them accomplish a given task. Factor dimensionality was assessed through parallel analysis. Parallel analysis has been demonstrated to more accurately determine factor dimensionality than the traditional Kuder-Richardson (need reference). Parallel analysis produces correlation matrices from a randomly chosen simulated dataset that has a similar number of observations as the original dataset.⁷ The observations in the simulated dataset have the same sampling error as observations in the original dataset.⁷ Eigenvalues were computed for both the simulated and original data and compared to determine the point at which the eigenvalue in the simulated dataset was higher than in the original dataset. The number of factors before this transition point denoted the number of factors that were retained. A scree plot was also created to compare eigenvalues from simulated and original datasets to corroborate our determination of the number of factors to retain.

Data Analysis

Exploratory factor analysis was conducted using squared multiple correlations and prior communality estimates. Maximum likelihood (ML) estimation was used to extract factors followed by Promax (oblique) rotation. First, parallel analysis using a SAS macro, %parallel, was used to determine the dimensionality of the PCBDAA.^{8,9} Second, the Scree plot, eigenvalues, and proportion of eigenvalues were examined. An eigenvalue greater than one determined if a factor was retained in the factor structure. Third, a series of factor rotations were examined. Results of both varimax and promax rotation methods indicated an optimal factor structure of three factors. When interpreting the rotated factor pattern, an item was determined to load on a given factor if the factor loading was 0.40 or greater for that factor, and was less than 0.40 for other factors.⁸ internal consistency reliability was assessed using Cronbach's alpha for the total scale and each of the three subscales. Pearson's correlation assessed the associations between factors and subscale items. PROC MEAN and PROC FREQ were used to conduct descriptive statistics. PROC CORR and PROC FACTOR were used to conduct exploratory factor analysis, compute Cronbach's alpha, and estimate correlations. All data analyses were performed using SAS/STAT[®] statistical software, version 9.4.¹⁰

Results

Table 1 displays the frequency distribution of item 1 for the PCBDAA, which indicates that about 92% of participants agree or strongly agree that the computer-based decision aid was useful.

Table 1. Frequency distribution of item 1 of the PCBDAA

I find the CBDA useful.				
PCBDAA 1	Frequency	Percent	Cumulative Frequency	Cumulative Percent
		2	0.56	2
Strongly Disagree	3	0.85	5	1.41
Disagree	3	0.85	8	2.26
Neither	21	5.93	29	8.19
Agree	126	35.59	155	43.79
Strongly Agree	199	56.21	354	100.00

Table 2 presents the frequency and percentage of missing values for each item in the PCBDAA. Missing values ranged from 0.56 percent to 2.54 percent for PCBDAA items.

Table 2. Frequency and percentage of missing values for PCBDAA (N=354)

Items	N	%
PCBDAA1	2	0.56
PCBDAA2	5	1.41
PCBDAA3	3	0.85
PCBDAA4	3	0.85
PCBDAA5	3	0.85
PCBDAA6	3	0.85
PCBDAA7	5	1.41
PCBDAA8	5	1.41
PCBDAA9	3	0.85
PCBDAA10	4	1.13
PCBDAA11	3	0.85
PCBDAA12	4	1.13
PCBDAA13	4	1.13
PCBDAA14	3	0.85
PCBDAA15	5	1.41
PCBDAA16	3	0.85
PCBDAA17	2	0.56
PCBDAA18	3	0.85
PCBDAA19	2	0.56
PCBDAA20	2	0.56
PCBDAA21	3	0.85
PCBDAA22	2	0.56
PCBDAA23	2	0.56
PCBDAA24	9	2.54

Table 3 report frequency, means, and standard deviation of PCBDAA items. The mean ranged from 2.2 (PCDBAA 23) to 4.54 (PCDBAA 9).

Table 3. Frequency, means, and standard deviation for PCBDAA items.

Items	Without Imputation		
	N	Mean	SD
PCBDAA1	352	4.46	0.72
PCBDAA2	349	4.38	0.78
PCBDAA3	351	4.43	0.71
PCBDAA4	351	4.47	0.68
PCBDAA5	351	4.45	0.69
PCBDAA6	351	4.14	0.87
PCBDAA7	349	4.35	0.82
PCBDAA8	349	4.35	0.85
PCBDAA9	351	4.54	0.67
PCBDAA10	350	4.49	0.69
PCBDAA11	351	4.27	0.82
PCBDAA12	350	4.41	0.75
PCBDAA13	350	4.41	0.73
PCBDAA14	351	4.34	0.81
PCBDAA15	349	3.31	1.30
PCBDAA16	351	4.16	1.03
PCBDAA17	352	4.29	0.87
PCBDAA18	351	4.20	0.91
PCBDAA19	352	4.30	0.84
PCBDAA20	352	2.38	1.33
PCBDAA21	351	2.39	1.34
PCBDAA22	352	2.34	1.33
PCBDAA23	352	2.20	1.33
PCBDAA24	345	4.28	0.82

Figure 1 shows the plot of parallel analysis which suggested three underlying dimensions.

Figure1. Plot of Eigenvalues

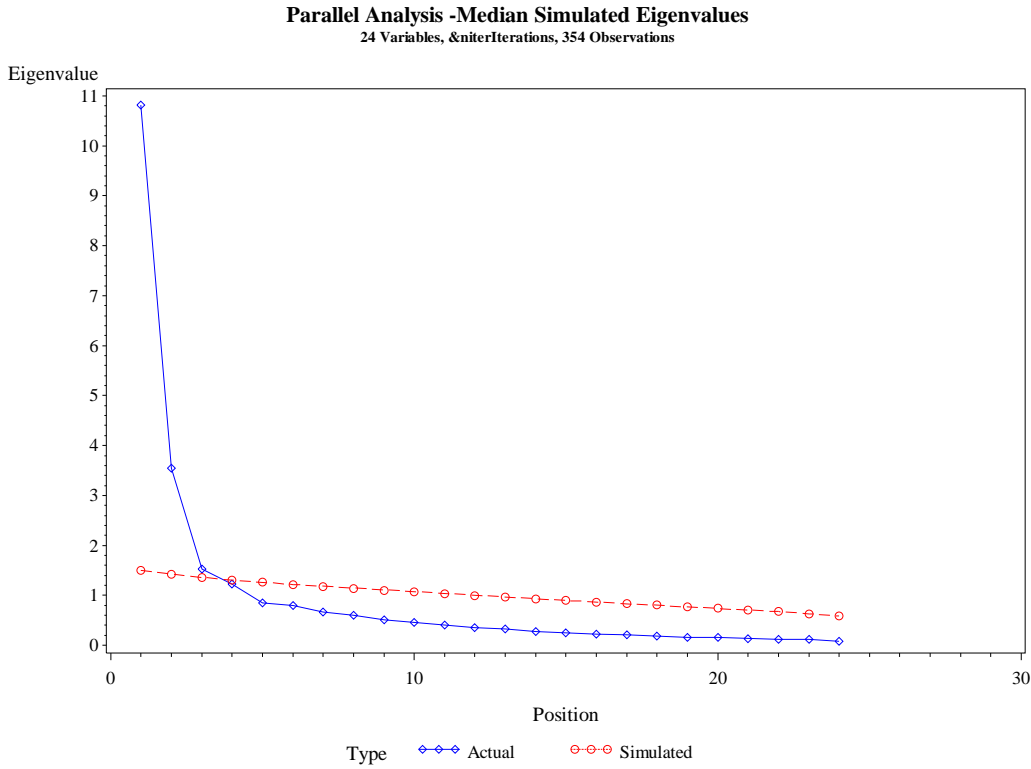


Table 4 reports eigenvalues for the PCBDAA scale. Three eigenvalues are above 1 and 93% of the variance was explained by these three factors.

Table 4. Part of Eigenvalues for PCBDAA

Preliminary Eigenvalues:				
Total = 66.7070721 Average = 2.77946134				
	Eigenvalue	Difference	Proportion	Cumulative
1	41.2324089	25.0127050	0.6181	0.6181
2	16.2197038	11.7620409	0.2431	0.8613
3	4.4576629	1.1695876	0.0668	0.9281
4	3.2880753	1.0535851	0.0493	0.9774
5	2.2344903	1.2329772	0.0335	1.0109
6	1.0015130	0.3789205	0.0150	1.0259
7	0.6225925	0.2495025	0.0093	1.0352
8	0.3730900	0.0229063	0.0056	1.0408

Figure 2 displays the scree plot of eigenvalues for the PCDBAA scale, which revealed that 3 to 5 factors are meaningful.

Figure 2: Scree Plot of Eigenvalues for PCDBAA Scale

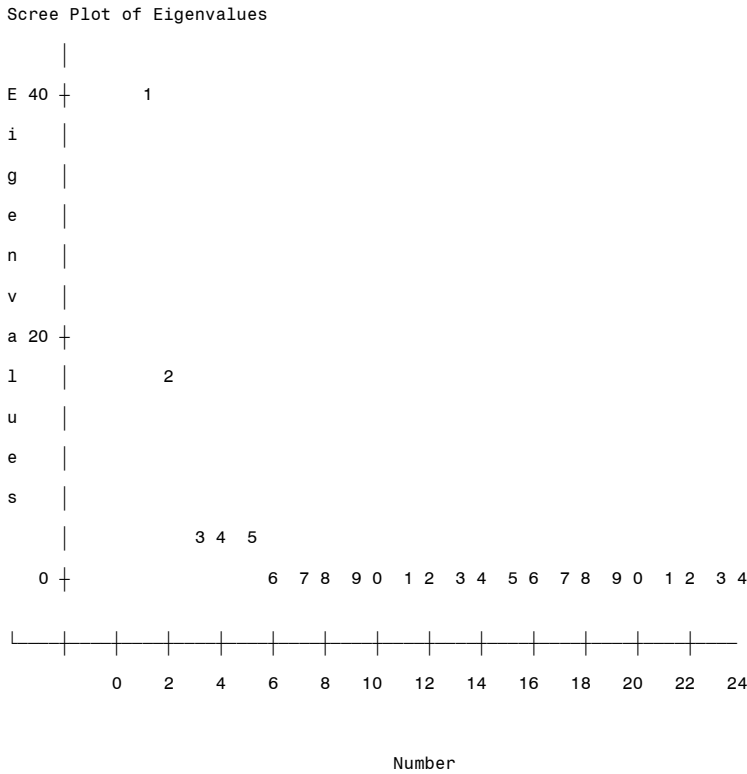


Table 5 reports the Rotated Factor Pattern (Standardized Regression Coefficients) for the PCBDAA. Parallel analysis showed there are three meaningful underlying factors. Eigenvalues and the proportion of variance explained by each factor also suggested three meaningful factors for the 24-item scale. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is 0.93, which is acceptable. The residuals are all small and the overall RMSR is 0.049, indicating that the factor structure explains most of the correlations.¹¹ All 24 items loaded positively on three different factors at 0.32 or above with factor loadings ranging from 0.32 to 0.94. Eleven items loaded on Factor 1, eight on Factor 2, and five on Factor 3.

Table 5. Rotated Factor Pattern (Standardized Regression Coefficients) for the PCBDAA

Items	Factor 1	Factor 2	Factor 3
PCBDAA4	86	3	5
PCBDAA1	84	-1	-2
PCBDAA3	80	8	5
PCBDAA2	79	1	5
PCBDAA7	79	9	-7
PCBDAA8	78	8	-10
PCBDAA5	70	20	-1
PCBDAA6	63	13	2
PCBDAA14	46	37	-8
PCBDAA16	41	22	-1
PCBDAA19	32	27	-2
PCBDAA11	-5	89	0
PCBDAA10	-2	84	-6
PCBDAA9	16	65	-10
PCBDAA12	31	65	3
PCBDAA13	28	55	5
PCBDAA18	16	41	12
PCBDAA24	26	40	13
PCBDAA17	24	38	8
PCBDAA22	-3	3	94
PCBDAA21	3	-3	93
PCBDAA23	-1	-2	92
PCBDAA20	-7	3	86
PCBDAA15	19	18	44

Note: Root Mean Square Residual (RMSR) = 0.049. Kaiser's Measure of Sampling Adequacy = 0.93.

Table 6 reports means, standard deviations, minimums, and maximums of total scale and subscales of the PCBDAA. The results showed the average of the total scale and each subscale were 94.90, 47.64, 34.68, and 12.58, respectively.

Table 6: Frequency, Mean, and Standard Deviation for total scale and subscales of the PCDBA

Variable	N	Mean	Std Dev	Min	Max
Factor 1	352	47.64	7.28	11.00	55.00
Factor 2	352	34.68	5.09	6.00	40.00
Factor 3	352	12.58	5.57	4.00	25.00
Total Scale	352	94.90	12.24	23.00	120.00

Table 7 reports Pearson correlations for the total scale and three subscales of the PCBDAA. A negative weak correlation was found between Factors 1 and 3 ($r=-0.17$), a strong positive correlation was found between Factors 1 and 2 ($r=0.78$), and no correlation

between Factors 2 and 3 ($r=-.08$). A positive correlation was found between factors and the total subscale ranging from .32 to .85. Scale reliability in this study was examined by estimating the Cronbach's alpha. The reliability coefficient for total scale was 0.87. The Cronbach's alpha for the subscales ranged from 0.88 to 0.94.

Table 7: Pearson correlation and Cronbach's alpha for total scale and subscales PCDBAA

Pearson Correlation Coefficients ($N = 352$) Prob > r under $H_0: \rho=0$				
	mcbdaf1	mcbdaf2	mcbdaf3	ptcbda
Factor 1	0.94	0.78 <.0001	-0.17 0.0014	0.84 <.0001
Factor 2		0.88	-0.08 0.1344	0.85 <.0001
Factor 3			0.89	0.32 <.0001
Total Scale				0.87

Note: Cronbach's alphas are in the diagonal.

Conclusion

We used the SAS® macro, %parallel, to determine the dimensionality of the PCBDAA among African-American men. Parallel analysis indicated three factors were the optimal solution for PCBDAA. All loading from rotated factor pattern (standardized regression coefficient) were positive for 24 items with factor loadings ranging from 0.32 to 0.94. Eleven items loaded for Factor 1 (Technology Use and Expectancy and Intention), 8 on Factor 2 (Technology Use Anxiety), and 5 on Factor 3 (Technology Use Self-Efficacy). The Pearson correlation resulted in a negative weak correlation between Factors 1 and 3, a strong positive correlation between Factors 1 and 2, and no correlation between Factors 2 and 3. Results also showed a positive correlation between factors, the total scale, and subscales ranging from .32 to .85. The Cronbach's alpha of each subscale ranged from 0.87 to 0.94 with a total scale reliability of 0.87. Parallel analysis is a valuable method for determining the dimensionality of the Computer-Based PrCA Screening Decision Aid.

REFERENCES

1. Siegel, R. L., Miller, K. D., & Jemal, A. (2017). Cancer statistics, 2017. *CA: A Cancer Journal for Clinicians*, 67(1), 7-30. doi:10.3322/caac.21387
2. Carter HB, Albersten PC, Barry MJ, et al. (2013). Early Detection of Prostate Cancer: AUA Guideline. *The Journal of Urology* 2013;190:419-426
3. O'Brien BK, Whelan TJ, Villais-Keever M, Et al. (2209). Are Cancer- Related Decision Aids Effective? A Systemic Review and Meta-Analysis. *Journal of Clinical Oncology*. 2009;27:974-85
4. Lisette C, Yasavur U, De Leon C, Amini R, Rische N, Visser U. (2012). Building an on-International FLAIRS Conference, Marco Island, USA.
5. Owens O. L., Friedman D.B., Brandt H.M., Bernhardt J.M., Hébert J.R (2015). An iterative Process for Developing and Evaluating a Computer-Based Prostate Cancer Decision Aid for African American Men. 16(5):642-55, Epub 2015 May 16. doi: 10.1177/1524839915585737.
6. Venkatesh V, Morris MG, Dvis GB, Davis FD. (2003). User Acceptance of Information Technology: Toward a Unified View. *Mis Quart* 2003;27:425-78.
7. Çokluk, Ö., & Koçak, D. (2016). Using Horn's Parallel Analysis Method in Exploratory Factor Analysis for Determining the Number of Factors. *Educational Sciences: Theory and Practice*, 16(2), 537-551.
8. Kabacoff, R. (2003). Determining the Dimensionality of Data: A SAS Macro for Parallel Analysis, *SUGI 28*, March 30 – April 2.
9. Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185.
10. SAS Institute Incorporated. (2013). *SAS for Windows 9.4*. Cary, NC: SAS Institute Inc.
11. Browne, M. W. & Cudeck, R. (1993), "Alternative Ways of Assessing Model Fit," in K. A. Bollen and S. Long, eds., *Testing Structural Equation Models*, Newbury Park, CA: Sage Publications.

TRADEMARK NOTICE

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

Contact Information

Abbas S. Tavakoli, DrPH, MPH, ME
Clinical Associate Professor
College of Nursing
University of South Carolina
1601 Greene Street
Columbia, SC 29208-4001
Fax: (803) 777-5561
E-mail: abbas.tavakoli@sc.edu

SAS Syntax

Part of Data steps:

```
data one;
set psa.psadat17;

if maritr=1 then martgb=1;
else if maritr=2 then martgb=2;
  else if maritr=3 or maritr=4 or maritr=5 then martgb=2;

  label
    martgb ="MArital status"
run;

data two;
set one;
ptcbda = sum (of pcbdaa1 - pcbdaa24);
mcbdaf1 = sum (of PCBDAA1 PCBDAA2 PCBDAA3 PCBDAA4 PCBDAA5 PCBDAA6 PCBDAA7 PCBDAA8 PCBDAA14
PCBDAA16 PCBDAA19);
mcbdaf2 = sum (of PCBDAA9 PCBDAA10 PCBDAA11 PCBDAA12 PCBDAA13 PCBDAA17 PCBDAA18 pcbda24 );
mcbdaf3 = sum (of PCBDAA15 PCBDAA20 PCBDAA21 PCBDAA22 PCBDAA23 );
label
ptcbda = " total Computer based decision aid (CBDAA) scale/post"
cbdaf1 = " total Computer based decision aid (CBDAA) scale/post/factor 1"
cbdaf2 = " total Computer based decision aid (CBDAA) scale/post/ factor 2"
cbdaf3 = " total Computer based decision aid (CBDAA) scale/post/ factor 3"
;
run;
```

Procedures:

Parallel analysis Macro⁹:

```
ods rtf; ods listing close;
%macro parallel(data=_LAST_, var=_NUMERIC_,
niter=1000, statistic=Median);
data _temp; set &data; keep &var; run;
/* obtain number of observations and variables in dataset */
ods output Attributes=Params;
ods listing close;
proc contents data=_temp ; run;
ods listing;

data _NULL_; set Params;
if Label2 eq 'Observations' then call
symput('Nobs',Trim(Left(nValue2)));
```



```

else if Label2 eq 'Variables' then call
symput('NVar',Trim(Left(nValue2))); run;

/* obtain eigenvalues for actual data */
proc factor data=_temp nfact=&nvar noprint
outstat=E1(where=( _TYPE_ = 'EIGENVAL'));
var &var; run; data E1; set E1; array A1{&nvar} &var;
array A2{&nvar} X1-X&nvar;
do J = 1 to &nvar; A2{J} = A1{J}; end;
keep X1-X&nvar; run;

/* generate simulated datasets and obtain eigenvalues */
%DO K = 1 %TO &niter;
data raw;
array X {&nvar} X1-X&nvar; keep X1-X&nvar;
do N = 1 to &nobs; do I = 1 to &nvar; X{I} = rannor(-1); end; output; end; run;
proc factor data=raw nfact=&nvar noprint
outstat=E(where=( _TYPE_ = 'EIGENVAL')); var X1-X&nvar;
proc append base=Eigen
data=E(keep=X1-X&nvar); run; %END;

/* summarize eigenvalues for simulated datasets */
proc means data=Eigen noprint;
var X1-X&nvar; output out=Simulated(keep=X1-X&nvar) &statistic=;
proc datasets nolist; delete Eigen;
proc transpose data=E1 out=E1; run;
proc transpose data=Simulated out=Simulated; run;

/* plot actual vs. simulated eigenvalues */
data plotdata; length Type $ 9; Position+1;
if Position eq (&nvar + 1) then Position = 1;
set E1(IN=A) Simulated(IN=B);
if A then Type = 'Actual'; if B then Type = 'Simulated';
rename Col1 = Eigenvalue; run;
title height=1.5 "Parallel Analysis - &statistic Simulated Eigenvalues";
title2 height=1 "&nvar Variables, &niter Iterations, &nobs Observations";
proc print data = plotdata ; run;
symbol1 interpol = join value=diamond height=1 line=1 color=blue;
symbol2 interpol = join value=circle height=1 line=3 color=red ;
proc gplot data = plotdata;
plot Eigenvalue * Position = Type; run;quit;
%mend parallel; run;
%parallel(data=two, var=PCBDAA1 - PCBDAA24, niter=1000, statistic=Median);
%parallel(data=two, var=self1-self11, niter=1000, statistic=Median); run;

ods rtf close; ods listing; quit; run;

```

Other Procedures:

```

ods rtf; ods listing close;
proc freq data =one;
tables pcbdaa1 - pcbdaa24
title ' Frequency tables / ' ; title2 ' PSA Study '; run;

%macro avg (q,t);
proc means data=two maxdec=2;
var &q
TITLE 'means / ' &t; title2 ' PSA Study ';run;
%mend avg;
%avg (pcbdaa1-pcbdaa24 , post);
%avg (mcbdaf1 mcbdaf2 mcbdaf3 ptcbda);
run;
ods rtf close; ods listing; quit; run;

```

```
ods rtf; ods listing close;
proc corr data=two;
  var mcbdaf1 mcbdaf2 mcbdaf3 ptcbda;
  title 'correlation coefficient'; title2 'PSA Study '; run;
ods rtf close; ods listing; quit; run;
```

```
ods rtf; ods listing close;
%macro corr (q);
proc corr nocorr alpha nomiss data=two; var &q;;
  title 'Reliability coefficient'; title2 'PSA Study ';
  %mend corr;
%corr (PCBDAA1 - PCBDAA24);
%Corr (PCBDAA1 PCBDAA2 PCBDAA3 PCBDAA4 PCBDAA5 PCBDAA6 PCBDAA7 PCBDAA8 PCBDAA14 PCBDAA16
PCBDAA19);
%corr (PCBDAA9 PCBDAA10 PCBDAA11 PCBDAA12 PCBDAA13 PCBDAA17 PCBDAA18 PCBDAA24 );
%corr (PCBDAA15 PCBDAA20 PCBDAA21 PCBDAA22 PCBDAA23 );run;
ods rtf close; ods listing; quit; run;
```

```
ods rtf; ods listing close;
%macro fact (q,n,t);
proc factor data=two method=ml priors=smc scree rotate=promax reorder residual
  flag=.35 nfact=&n msa ;
  var &q ;
  title 'factor analysis ' &t; title2 'PSA Study ';
  %mend fact;
%fact (pcbdaa1 - pcbdaa24,3, Computer based decision aid );
run;
ods rtf close; ods listing; quit; run;
```