**Ziyun Ding**

**data** work.interaction;

length filename $ **40** type $ **20**; /\*7a\*/

infile 'W:\interactions.dat' dlm='09'x dsd missover; /\*7b,7d, 7e\*/

input PROT1 $ PROT2 $ RANK\_MAJOR RANK\_MINOR TYPE $

PDB\_ID $ BIO\_UNIT CHAIN1 $ MODEL1 SEQ\_IDENT1

COVERAGE1 SEQ\_BEGIN1 $ SEQ\_END1 $ DOMAIN1 $ CHAIN2 $

MODEL2 SEQ\_IDENT2 COVERAGE2 SEQ\_BEGIN2 SEQ\_END2

DOMAIN2 $ FILENAME $; /\*7c\*/

coverage1=round(coverage1); /\*20a\*/

coverage2=int(coverage2); /\*20b\*/

totoal\_coverage=sum(coverage1,coverage2); /\*20c\*/

maximum\_coverage=max(coverage1,coverage2); /\*20d\*/

miss\_domain1=cmiss(domain1); /\*20e\*/

**run**;

**proc** **print** data=work.interaction; /\*7f\*/

**run**;

===========================================================

libname orion 'w:\'; /\*1a\*/

**proc** **contents** data=orion.\_ALL\_ NODS; /\*1b\*/

**run**;

**proc** **contents** data=work.interaction; /\*1c\*/

**run**;

**proc** **print** data=work.interaction label;

label prot1='protein1' prot2='protein2'; /\*1d\*/

**run**;

===========================================================

title 'Protein Protein Interaction Data in Arabidopsis'; /\*3c\*/

footnote 'Downloaded from Interactome3D Website'; /\*3d\*/

**proc** **print** data=work.interaction split='\*'; /\*3a, 3b\*/

label seq\_ident1='sequence\*identity 1' seq\_ident2='sequence\*identity 2'; /\*3b\*/

**run**;

title; footnote;

===========================================================

**data** model domain structure; /\*14a\*/

keep prot1 prot2 type domain1 rank\_major /\*14c\*/

bio\_unit seq\_ident1 seq\_begin1 seq\_end1;

set work.interaction (firstobs=**10** obs=**1000**); /\*14d\*/

select(Type); /\*14b\*/

when('Model') output model;

when('Dom\_dom\_model') output domain;

when('Structure') output structure;

otherwise;

end;

**run**;

===========================================================

libname orion 'W:\';

ods pdf file="W:\interaction\_freq.pdf"; /\*13d\*/

**proc** **freq** data=model nlevels; /\*13a\*/

tables \_all\_ / noprint;

**run**;

ods pdf close;

libname orion 'W:\';

ods rrf file="W:\interaction\_mean.rtf"; /\*13d\*/

**proc** **means** data=interaction nmiss min max sum; /\*13b\*/

var seq\_ident1;

class type;

**run**;

ods rtf close;

libname orion 'W:\';

ods html file="W:\interaction\_univar.html"; /\*13d\*/

**proc** **univariate** data=interaction nextrobs=**5**; /\*13c\*/

var seq\_ident1;

**run**;

ods html close;

===========================================================

**data** extract(keep=interaction type template filename file\_length categary1);

set interaction;

filename=propcase(filename,'.'); /\*18d\*/

template=scan(filename,**4**,'-'); /\*18c\*/

categary1=substr(prot1,**1**,**1**); /\*18a\*/

file\_length=length(filename); /\*18b\*/

type=upcase(type); /\*18e\*/

interaction=catx('-',prot1,prot2); /\*18f\*/

**run**;

===========================================================

**data** model\_converted;

set model;

Ndomain=input(seq\_begin1, **5.**); /\*21a\*/

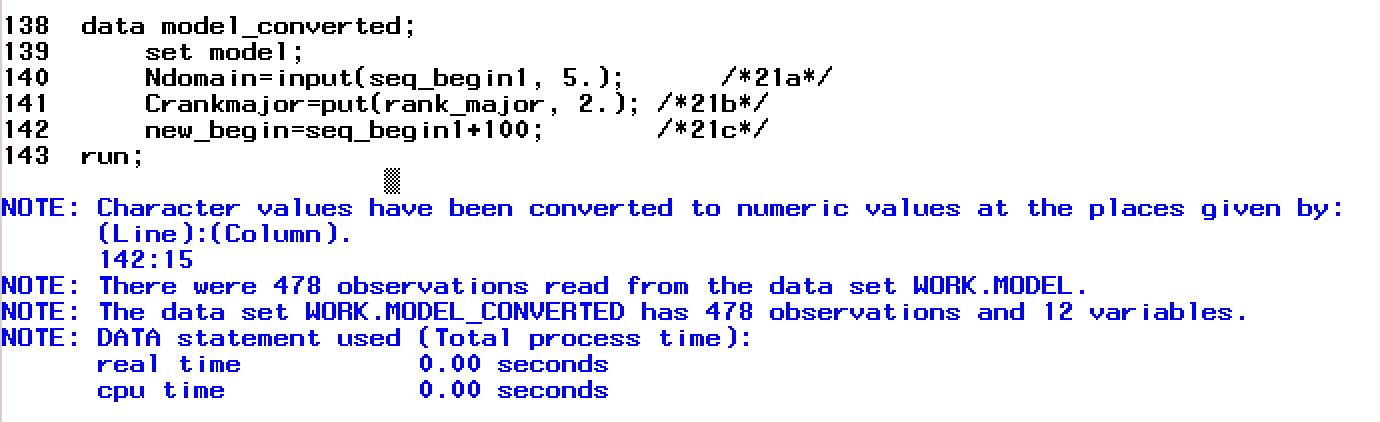
Crankmajor=put(rank\_major, **2.**); /\*21b\*/

new\_begin=seq\_begin1+**100**; /\*21c\*/

**run**;

**proc** **contents** data=model\_converted; /\*21d\*/

**run**;



===========================================================

**data** interaction\_2;

infile 'W:\interactions.dat' dlm='09'x; /\*8a\*/

input protein1 :$6. protein2 :$6. RANK\_MAJOR

RANK\_MINOR TYPE :$20. PDB\_ID :$4. BIO\_UNIT

CHAIN1 :$1. MODEL1 SEQ\_IDENT1 COVERAGE1 :**2.** SEQ\_BEGIN1 SEQ\_END1 DOMAIN1 CHAIN2 :$1. MODEL2 SEQ\_IDENT2 COVERAGE2 :**2.** SEQ\_BEGIN2 SEQ\_END2

DOMAIN2 FILENAME :$40.; /\*8b\*/

format coverage1 **4.1** converage2 **4.1**; /\*8c\*/

**run**;

**proc** **print** data=interaction\_2; /\*8d\*/

**run**;

===========================================================

**data** compare;

set structure;

array template{**4**} seq\_ident1 seq\_ident2 coverage1 coverage2;

array diff{**4**};

array aver{**4**} \_temporary\_ (**80**,**80**,**75**,**75**); /\*23a\*/

do i=**1** to dim(diff); /\*23b\*/

diff{i}=template{i}-aver{i};

end;

**run**;

**proc** **print** data=compare;

**run**;

===========================================================

%put Starting with topics from mac1; /\*25b\*/

%let filename=structure; /\*25c\*/

%let opt=noobs;

%let name=protein;

options symbolgen; /\*25a\*/

**proc** **print** data=&filename &opt label;

footnote "Created &systime &sysday, &sysdate9"; /\*25d\*/

label prot1="&name.1" prot2="&name.2"; /\*26a\*/

**run**;

%symdel opt name; /\*25e\*/

===========================================================

options mcompilenote=ALL; /\*27d\*/

**%macro** ***classify***; /\*27a\*/

proc means data=structure;

run;

**%mend** classify; /\*27b\*/

options mprint; /\*27c\*/

%***classify***