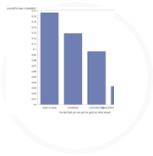


Data Analysis and Interpretation coursera blog

HW2 Chi Square Testing

Model Interpretation for Chi-Square Tests:

When examining the association between self-reported health status (categorical response) and self-reported disability status (categorical explanatory), a chi-square test of independence revealed that among youth surveyed for the AddHealth dataset (my sample), those who identify as being disabled (~10%) were not more likely to report poorer health compared to those without disabilities (~90%), X2 =88.60, 1 df, p=0001.

The df or degree of freedom we record is the number of levels of the explanatory variable -1. Here the df is 4 (health status has 5 levels, so df 5-1=4).

A Chi Square test of independence revealed that among youth surveyed (my sample), self-reported health status (out of a range of 5 options ranging from 1=excellent to 5=poor) and self-reported disability status (binary categorical variable, yes/no) were NOT significantly associated, X2 =2.36, 4 df, p=.67

Therefore, because the Chi Square Test of Independence revealed no significant relationship between the response and explanatory variables of interest, post hoc comparisons were not indicated.

SAS output can be viewed here:
file:///C:/Users/ajozkowski/Downloads/Chi%20square%20testing%20hw-results.html

SAS code is below:

```
PROC IMPORT DATAFILE ='/home/ajozkowski0/sasuser.v94/addhealth_pds.csv' OUT = imported
REPLACE;
RUN;

data imported;
set imported;
LABEL H1GH1='In general, how is your health?'
H1PL37="Do you consider yourself to have a disability?";
run;

Data imported; set imported;
IF H1GH1>=6 THEN H1GH1=.;
IF H1PL37>=6 THEN H1PL37=.;
if h1gh1 ne . then do;
if h1pl37 ne . then do;
end; end;
run;

proc freq; tables H1pl37*h1gh1/chisq;
run;
```

Feb 21st, 2016

HW1: Running an ANOVA

Model Interpretation for ANOVA:

When examining the association between Body Mass Index (BMI; quantitative response) and self-reported health status (categorical explanatory), an Analysis of Variance (ANOVA) revealed that among teens surveyed for the AddHealth dataset (my sample), those with lower (healthier) BMIs (means 21.58 & 22.10) reported better perceived health compared to those with higher (less healthy) BMIs (means 23.41, 25.36, & 26.55), F(4, 6289)=92.17, p<.0001.

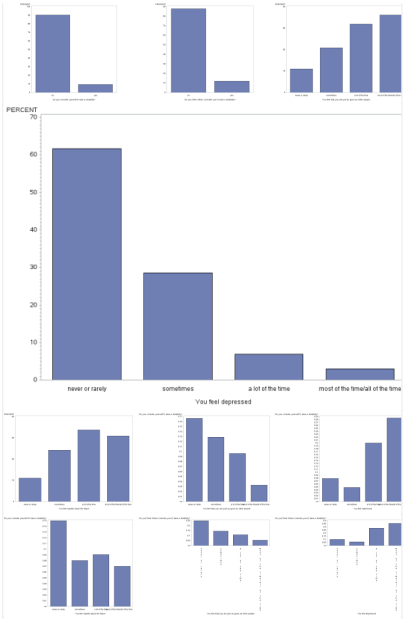
Model Interpretation for post hoc ANOVA results:

Analysis of Variance (ANOVA) revealed that among teens surveyed for the AddHealth dataset (my sample), BMI and perceived health status were significantly associated F(4, 6289)=92.17, p<.0001. Post hoc comparisons of BMI by self-reported health status categories revealed that those individuals who considered themselves to be significantly more healthy (stating that they were in “excellent” or “very good” overall health) had significantly lower (healthier) BMIs compared to all other groups. In addition, those in “good,” “fair,” and “poor” health each had significantly different BMIs from the other groups, with those in the poorest health having the highers (least healthy) BMIs and the other comparisons corresponding accordingly in the expected direction.

Program code (note: SAS code used to calculate BMI from Age, Height, and Weight using the AddHealth data was sourced here: http://www.cpc.unc.edu/projects/addhealth/data/code/bmi, an open-source resource for transforming the data, as almost all the original variables in the provided dataset were categorical).

```
PROC IMPORT DATAFILE
='/home/ajozkowski0/sasuser.v94/addhealth_pds.csv' OUT = imported
REPLACE;
RUN;

data imported;
set imported;
/* Calculate Age at Wave 1    */
if h1gi1m = 96 then h1gi1m = .;
if h1gi1y = 96 then h1gi1y = .;
```



Week 4 homework: Visualizing Data

Above are the univariate and bivariate charts I created for my 5 variables of interest (2 predictor variables and 3 outcomes variables). I also ran a chart to demonstrate the relationship between the two predictor variables.

My SAS code is below:

```
PROC IMPORT DATAFILE
='/home/ajozkowski0/sasuser.v94/addhealth_pds.csv' OUT = imported
REPLACE;
RUN;

data imported;
set imported;
label H1PL37='Do you consider yourself to have a disability?'
H1PL38='Do you think others consider you to have a disability?'
H1FS4='You feel that you are just as good as other people'
H1FS6='You feel depressed'
H1FS8='You feel hopeful about the future';
run;

PROC FORMAT;
VALUE disabfmt
0 = 'no'
1 = 'yes'
6 = 'refused'
7 = 'legitimate skip'
8 = 'dont know';
RUN;

Data imported; set imported;
/*set missing data*/
IF H1PL37>=6 THEN H1PL37=.;
IF H1PL38>=6 THEN H1PL38=.;
IF H1FS4>=6 THEN H1FS4=.;
IF H1FS6>=6 THEN H1FS6=.;
IF H1FS8>=6 THEN H1FS8=.;
run;
```

The FREQ Procedure

Do you consider yourself to have a disability?				
H1PL37	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	247	90.48	247	90.48
yes	26	9.52	273	100.00
Frequency Missing = 6231				

Do you think others consider you to have a disability?				
H1PL38	Frequency	Percent	Cumulative Frequency	Cumulative Percent
no	241	87.96	241	87.96
yes	33	12.04	274	100.00
Frequency Missing = 6230				

The FREQ Procedure

You feel that you are just as good as other people				
H1FS4	Frequency	Percent	Cumulative Frequency	Cumulative Percent
never or rarely	715	11.03	715	11.03
sometimes	1353	20.87	2068	31.90
a lot of the time	2070	31.93	4138	63.83
most of the time/all of the time	2345	36.17	6483	100.00
Frequency Missing = 21				

You feel depressed				
H1FS6	Frequency	Percent	Cumulative Frequency	Cumulative Percent
never or rarely	3994	61.60	3994	61.60
sometimes	1853	28.58	5847	90.18
a lot of the time	444	6.85	6291	97.02
most of the time/all of the time	193	2.98	6484	100.00
Frequency Missing = 20				

You feel hopeful about the future				
H1FS8	Frequency	Percent	Cumulative Frequency	Cumulative Percent
never or rarely	720	11.12	720	11.12
sometimes	1567	24.20	2287	35.32
a lot of the time	2185	33.75	4472	69.07
most of the time/all of the time	2003	30.93	6475	100.00
Frequency Missing = 29				

Assignment 3: Making Data Management Decisions

This week, I worked on writing code for pulling out the missing data (including unanswered/refused, “don’t know,” and responses skipped due to skip logic). I then generated the frequency tables for my five variables again, excluding the missing data. Because all my variables are categorical and have a limited number of response options, no collapsing, grouping, or recoding is necessary. However, I made sure to label each of the response options and the variable names (with the actual text of the survey question) so that the tables are easy to read and interpret.

As you can see, there is a much smaller subset of subjects that answered the disability-related questions (the first two listed), because most of the subjects did not complete this portion of the survey due to skip logic. However, because I have nearly 300 subjects to work with, I still feel like I can get some legitimate analysis use Chi Square analysis.

We also see that might greater percentages of subjects endorse more “positive” responses to the three self-concept questions, meaning that well over half likely have good mental health and self-concept. For each of these three variables, fewer than 30

These are the fre
generated from th
“Assignment 2: R
Program” **(post b**

As you can see, t
sample (N=6504)
disability-related
who answered th
identified themsel
disability (247 en
and 33 stated the
identify them as h
endorsed no disa
others). I am seei
a very small sam
is a shame becas
research questio
I hope I have eno
some differences
more about the st
to see why so ma
subjects did not r
questions (the ma
skipped,” meanin
presented with th
respond to them).

On the feelings-re
nearly the entire s
responded (<1%
of teens reported
least some of the
reported feeling h
the time. Feeling
similar distributio
“feeling just as go
over 88% of teen
way at least som
interesting to look

```
 intdt = mdy(imonth,iday,iyear);
 birthdt = mdy(h1gi1m,15,h1gi1y);
 age = int((intdt-birthdt)/365.25);

/* calculate bmi_w1r  Wave1 recalled
BMI  */
if h1gh59a lt 95 and h1gh59b lt 95 then
do;
 feetw1r = h1gh59a;
 inchw1r = h1gh59b;
 h_w1r = (12*feetw1r) + (inchw1r);
end;

attrib h_w1r   label = 'height in inches';

if h1gh60 lt 995 then do;
 w_w1r = h1gh60;
end;

attrib w_w1r   label = 'weight in pounds';

if h1gh59a lt 95 and h1gh59b lt 95 then
do;
 if h1gh60 lt 995 then do;
   /* weight in kilos wave 1 recall  */
   k_w1r = w_w1r * .454;
   /* height in meters wave 1 recall */
   m_w1r = h_w1r * .0254;
   /* meters squared wave 1 recall   */
   msq_w1r = m_w1r * m_w1r;
   /* BMI wave 1 recall              */
   bmi_w1r = k_w1r/msq_w1r;
   attrib bmi_w1r label = "Body Mass
Index Wave 1 Recall";

 end;
end;

Data imported; set imported;
LABEL bmi_w1r='BMI at Wave 1
timepoint'
H1GH1='In general, how is your
health?';
run;


Data imported; set imported;
/*set missing data*/
IF H1GH1>=6 THEN H1GH1=.;
run;

PROC ANOVA; CLASS H1GH1;
MODEL bmi_w1r=H1GH1;
MEANS H1GH1;
run;

PROC ANOVA; CLASS H1GH1;
MODEL bmi_w1r=H1GH1;
MEANS H1GH1/DUNCAN;
run;
```

Results and visual data: file:///C:/Users/ajozkowski/Dropbox/Coursera%20courses/Data%20Ananlysis%20Tools/SAS%20Program%20hypothesis%20testing%20ANOVA%20hw%20-results.html

```
PROC FREQ DATA=imported;
TABLES H1PL37 H1PL38;
FORMAT H1PL37 H1PL38 disabfmt.;
RUN;

PROC FORMAT;
VALUE scfmt
0 = 'never or rarely'
1 = 'sometimes'
2 = 'a lot of the time'
3 = 'most of the time/all of the time'
6 = 'refused'
8 = 'dont know';
RUN;

Proc freq data=imported;
tables H1FS4 H1FS6 H1FS8;
format H1FS4 H1FS6 H1FS8 scfmt.;
run;

DATA imported;
set imported;
label
H1PL37="Do you consider yourself to
have a disability?"
H1PL38="Do you think others consider
you to have a disability?"
H1FS4="You feel that you are just as
good as other people"
H1FS6="You feel depressed"
H1FS8="You feel hopeful about the
future";
run;

PROC GCHART data=imported; VBAR
H1PL37/Discrete type=PCT width=30;
FORMAT H1PL37 disabfmt.; IF
H1PL37<6 THEN output;
run;

PROC GCHART data=imported; VBAR
H1PL38/Discrete type=PCT width=30;
FORMAT H1PL38 disabfmt.; IF
H1PL37<6 THEN output;
run;

PROC GCHART data=imported; VBAR
H1FS4/Discrete type=PCT width=30;
format H1FS4 scfmt.; if H1FS4<6 then
output;
run;

PROC GCHART data=imported; VBAR
H1FS6/Discrete type=PCT width=30;
format H1FS6 scfmt.; if H1FS6<6 then
output;
run;

PROC GCHART data=imported; VBAR
H1FS8/Discrete type=PCT width=30;
format H1FS8 scfmt.; if H1FS8<6 then
output;
run;

PROC Gchart data=imported;
vbar H1PL37/discrete type=mean
width=30 sumvar=H1PL38;
format H1PL37 H1PL38 disabfmt.;
run;
```

respondents' values (out of over 6500 total) are missing, so these are nice complete samples.

My SAS code is below:

```
PROC IMPORT DATAFILE
='/home/ajozkowski0/sasuser.v94/addh
ealth_pds.csv' OUT = imported
REPLACE;
RUN;

data imported;
set imported;
label H1PL37="Do you consider
yourself to have a disability?"
H1PL38="Do you think others consider
you to have a disability?"
H1FS4="You feel that you are just as
good as other people"
H1FS6="You feel depressed"
H1FS8="You feel hopeful about the
future";
run;

PROC FORMAT;
VALUE disabfmt
0 = 'no'
1 = 'yes'
6 = 'refused'
7 = 'legitimate skip'
8 = 'dont know';
RUN;

Data addhealth_nomiss; set imported;
/*set missing data*/
IF H1PL37>=6 THEN H1PL37=.;
IF H1PL38>=6 THEN H1PL38=.;
IF H1FS4>=6 THEN H1FS4=.;
IF H1FS6>=6 THEN H1FS6=.;
IF H1FS8>=6 THEN H1FS8=.;

PROC FREQ
DATA=addhealth_nomiss;
TABLES H1PL37 H1PL38;
FORMAT H1PL37 H1PL38 disabfmt.;
RUN;

PROC FORMAT;
VALUE scfmt
0 = 'never or rarely'
1 = 'sometimes'
2 = 'a lot of the time'
3 = 'most of the time/all of the time'
6 = 'refused'
8 = 'dont know';
RUN;

Proc freq data=addhealth_nomiss;
tables H1FS4 H1FS6 H1FS8;
format H1FS4 H1FS6 H1FS8 scfmt.;
run;
```

relationships amc
feelings variables
disability-related i

New C
course

I just started this
be pretty straight
intro data analysi
work through the
so this is a good

1 note

```
PROC Gchart data=imported;
vbar H1FS4/discrete type=mean
sumvar=H1PL37;
format H1PL37 disabfmt. format H1FS4
scfmt.;
run;

PROC Gchart data=imported;
vbar H1FS6/discrete type=mean
sumvar=H1PL37;
format H1PL37 disabfmt. format H1FS6
scfmt.;
run;

PROC Gchart data=imported;
vbar H1FS8/discrete type=mean
sumvar=H1PL37;
format H1PL37 disabfmt. format H1FS8
scfmt.;
run;

PROC Gchart data=imported;
vbar H1FS4/discrete type=mean
sumvar=H1PL38;
format H1PL38 disabfmt. format H1FS4
scfmt.;
run;

PROC Gchart data=imported;
vbar H1FS6/discrete type=mean
sumvar=H1PL38;
format H1PL38 disabfmt. format H1FS6
scfmt.;
run;

PROC Gchart data=imported;
vbar H1FS8/discrete type=mean
sumvar=H1PL38;
format H1PL38 disabfmt. format H1FS8
scfmt.;
run;
```

As you can see, there is a strong relationship between self-identity as a person with a disability and reflective self appraisal of disability (endorsing that others view him/herself as having a disability). However, there are still some individuals who may identify as having a disability but who do not think others label them this way, or vice-versa.

Additionally, there does appear to be a relatively strong relationship between both disability identity variables (self-report and reflective self-appraisal) and the mental-health related outcomes, including “feeling just as good as other people,” “feeling depressed,” and “feeling hopeful about the future.” We know this because for each of the disability identities (yes or no), level of agreement with the mental health statements varies. For example, those who consider themselves to be disabled are more likely to endorse “never or rarely” feeling “just as good as others” than those who do not identify as disabled. These relationships match those that were predicted, with disabled

individuals reporting poorer mental health on all three survey items. The most dramatic result is for individuals who think that others label them as disabled. For this group, the greatest percentage (~40%) endorsed feeling depressed “most of the time or all of the time.” This is interesting, given that even though self-reported depression was high (~30% endorsing “some of the time or all of the time) among those who consider themselves to be disabled, it seems that there is something about others’ perceptions that may be related to increased likelihood of feeling depressed. Therefore, the perceptions of others may play a significant role in the mental health of individuals with disabilities.

Show more