Data analysis tools

Week 4 assignment

Work with moderation

I chose **addhealth** as my data set.

I want to find out if gender matters in the health situation. Three different analysis were conducted with **gender as moderator**:

1. ANOVA: general health and BMI.

2. how close with mother and how close with father.

3. frequency of feeling happy and BMI.

1. ANOVA

The ANOVA Procedure
Dependent Variable: BMI

gender=1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2355.15072	588.78768	32.27	<.0001
Error	3017	55054.14298	18.24798		
Corrected Total	3021	57409.29370			

R-Square	Coeff Var	Root MSE	BMI Mean
0.041024	18.85970	4.271765	22.65023

Source	DF	Anova SS	Mean Square	F Value	Pr > F
H1GH1	4	2355.150724	588.787681	32.27	<.0001

The ANOVA Procedure
Dependent Variable: BMI

gender=2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	4517.96533	1129.49133	61.62	<.0001
Error	3122	57227.08475	18.33026		
Corrected Total	3126	61745.05009			

R-Square	Coeff Var	Root MSE	BMI Mean
0.073171	19.19223	4.281386	22.30791

Source	DF	Anova SS	Mean Square	F Value	Pr > F
H1GH1	4	4517.965332	1129.491333	61.62	<.0001

The results show that no matter the respondents are male or female, the means of BMI at different levels of general health are not all equal to each other. While the F-value for male is 32.27 and for female is 61.62, which means **gender does moderate the relationship.**

Because there are more than 2 levels of general health, **post hoc test** was conducted.

The ANOVA Procedure
Duncan's Multiple Range Test for BMI

gender=1

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	3017
Error Mean Square	18.24798
Harmonic Mean of Cell Sizes	45.36658

Note: Cell sizes are not equal.

Number of Means	2	3	4	5
Critical Range	1.759	1.852	1.914	1.960

Means with the same letter are not significantly different.

Duncan Grouping		Mean	N	H1GH1
	A	26.9031	10	5
	B	24.9657	146	4
	B			
C	B	23.6934	651	3
C				
C		22.4245	1229	2
C				
C		21.8569	986	1

The ANOVA Procedure
Duncan's Multiple Range Test for BMI

gender=2

Note: This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	3122
Error Mean Square	18.33026
Harmonic Mean of Cell Sizes	59.11551

Note: Cell sizes are not equal.

Number of Means	2	3	4	5
Critical Range	1.544	1.626	1.680	1.721

Means with the same letter are not significantly different.

Duncan Grouping		Mean	N	H1GH1
	A	25.8565	13	5
	A			
	A	25.6244	227	4
	B	23.1309	849	3
	B			
C	B	21.8089	1255	2
C				
C		21.1950	783	1

We can see for male, level 1 and level 2 are similar to each other, in addition to this, all other pairs are different from each other. While for female, level 5 is similar to level 4, other pairs are different from each other. This further illustrate that **gender moderates their relationship**.

2. Chi square.

Statistics for Table of mother by father

Statistic	DF	Value	Prob
Chi-Square	1	65.2504	<.0001
Likelihood Ratio Chi-Square	1	61.6716	<.0001
Continuity Adj. Chi-Square	1	63.6273	<.0001
Mantel-Haenszel Chi-Square	1	65.1427	<.0001
Phi Coefficient		0.3281	
Contingency Coefficient		0.3118	
Cramer's V		0.3281	

Statistics for Table of mother by father

Statistic	DF	Value	Prob
Chi-Square	1	70.7614	<.0001
Likelihood Ratio Chi-Square	1	71.5040	<.0001
Continuity Adj. Chi-Square	1	69.5008	<.0001
Mantel-Haenszel Chi-Square	1	70.6673	<.0001
Phi Coefficient		0.3068	
Contingency Coefficient		0.2933	
Cramer's V		0.3068	

The results show that the p-values for both genders are less than 0.05. At 95% confidence level, we can say there is significant relationship between these two variables. The values for both gender does not differ from each other a lot, which mean the **gender may have no effect on their relationship.**

3. Correlation.

The CORR Procedure

gender=1

2 Variables:	H1FS11 BMI

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
H1FS11	3071	2.11527	0.78736	6496	0	3.00000	feeling happy
BMI	3023	22.64858	4.35951	68467	11.21973	56.43406	

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	H1FS11	BMI
H1FS11 feeling happy	1.00000 3071	-0.01968 0.2795 3023
BMI	-0.01968 0.2795 3023	1.00000 3023

The CORR Procedure

gender=2

2 Variables:	H1FS11 BMI

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
H1FS11	3258	2.14426	0.81666	6986	0	3.00000	feeling happy
BMI	3127	22.30791	4.44433	69757	12.02615	46.36417	

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	H1FS11	BMI
H1FS11 feeling happy	1.00000 3258	-0.04713 0.0084 3124
BMI	-0.04713 0.0084 3124	1.00000 3127

The results show that for male (gender = 1), the p-value is larger than 0.05, while for female(gender = 2), it is less than 0.05, which means for male, there feeling happy and BMI are not correlated, while for female they are correlated. The different correlation shows **gender does moderate their relationship.**

My code:

```
/*load data*/
LIBNAME mydata "/courses/d1406ae5ba27fe300" access=readonly;
data new; set mydata.addhealth_pds;

/*select respondents from grade 7 to grade 12*/
if H1GI20=97 then delete; if H1GI20=99 then delete; if H1GI20=96 then delete;
if H1GI20=98 then delete;

/*set aside the missing values*/
if H1GH1=6 then H1GH1=.; if H1GH1=8 then H1GH1=.;
if H1GH59A=96 then H1GH59A=.; if H1GH59A=98 then H1GH59A=.; if H1GH59A=99 then H1GH
if H1GH59B=96 then H1GH59B=.; if H1GH59B=98 then H1GH59B=.; if H1GH59B=99 then H1GH
if H1GH60=996 then H1GH60=.; if H1GH60=998 then H1GH60=.; if H1GH60=999 then H1GH6(
if H1FS11=6 then H1FS11=.; if H1FS11=8 then H1FS11=.;
if H1WP9=6 then H1WP9=.; if H1WP9=7 then H1WP9=.;
if H1WP9=8 then H1WP9=.; if H1WP9=9 then H1WP9=.;
if H1WP13=6 then H1WP13=.; if H1WP13=7 then H1WP13=.;
if H1WP13=8 then H1WP13=.; if H1WP13=9 then H1WP13=.;

/*calculate the height*/
H1GH59=H1GH59A * 12 + H1GH59B;

```

```
/*calculate the body mass index*/
BMI=H1GH60 * 0.454/(H1GH59 * 0.0254)**2;

/*add labels*/
label AID="respondent ID"
      BIO_SEX="gender"
      H1GH1="general health"
      H1GH59A="height in feet"
      H1GH59B="height in inch"
      H1GH60="weigt (pound)"
      H1FS11="feeling happy"
      H1WP9="how close with mother"
      H1WP13="how close with father"
      H1GH59="height (inch)";

/*ANOVA*/
data anovadata; set new;
proc sort; by BIO_SEX;
proc anova; class H1GH1;
model BMI=H1GH1;
means H1GH1; by BIO_SEX;
run;

/*chi square*/
data chidata; set new;
proc sort; by BIO_SEX;
proc freq; tables H1WP9*H1WP13/chisq; by BIO_SEX;
run;

/*correlation*/
data corrdata; set new;
proc sort;by BIO_SEX;
proc corr; var H1FS11 BMI; by BIO_SEX;
run;
```