

SAN JOSÉ STATE UNIVERSITY

College of Social Work
S. W. 242
Spring 2009
Edward Cohen

Week 9

March 20, 2009

- Statistics review
- Correlation—Pearson r
- Lab 2b Independent t -test
- Final Paper discussion—Choosing a statistical procedure
- Next week: Happy Spring Break!

Concepts you should know:

- Review univariate, bivariate, and multivariate analysis
- Correlation, or Pearson r , or Product-moment coefficient
- Assumptions of Pearson r
- Positive correlation
- Negative correlation

I. Review

A. Univariate, bivariate, or multi-variate analysis implied here?

1. The average age in your sample is 38, with a standard deviation of 12.82
2. Chi square test of association between gender and political party affiliation
3. Effects of group therapy (vs. individual therapy) on increasing score on a ratio level social support scale
4. Effects of group therapy (vs. individual therapy) on increasing score on ratio level social support scale, controlling for gender, ethnicity and SES

B. Name that statistic:

If ...	Independent variable is:	And dependent variable is:	Statistical test is:
	Categorical (nominal or ordinal)	Categorical (nominal or ordinal)	_____?
	Categorical (nominal—dichotomous, two independent groups)	Ratio or interval	_____?
	Categorical (two dependent groups)	Ratio or interval	_____?
	Categorical (nominal or ordinal) <i>more than two attributes</i>	Ratio or interval	_____?
New!	Continuous	Continuous	Correlation (Pearson r)

II. What is correlation?

A. Correlation tests the relationship between a continuous independent variable and a continuous dependent variable.

B. Examples of research scenarios:

1. The number of outpatient therapy sessions utilized is positively correlated with improvement in depression as measured by the total score of the Beck Depression Inventory
2. The number of hours students spend preparing for the 242 exam is positively correlated with the test score
3. The number of disruptive behaviors by children in residential care is negatively correlated with the number of strength-based supportive comments from staff

C. Compare to other statistical tests:

1. Chi-Square: “The variables are associated...”
2. t -test: “The group means differ...”
3. ANOVA: “The group means differ...”
4. Correlation: “There is a positive [or negative] correlation between the two variables.”
5. (Preview to next class) Multivariate regression: “The independent variables ____, ____, ____, and ____ are related or predictive of the dependent variable ____, controlling for ____.”

D. Calculation of the Pearson r

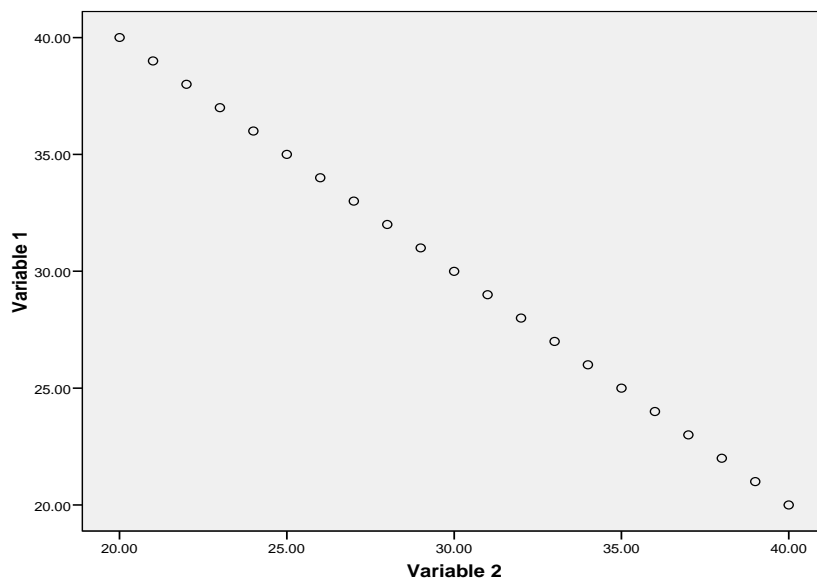
Subject #	Variable X	Variable Y	Calculation of r_{xy}
1	X_1	Y_1	$\left(\frac{x_1 - \text{mean}_x}{SD_x}\right)\left(\frac{y_1 - \text{mean}_y}{SD_y}\right)$
2	X_2	Y_2	$\left(\frac{x_2 - \text{mean}_x}{SD_x}\right)\left(\frac{y_2 - \text{mean}_y}{SD_y}\right)$
3	X_3	Y_3	$\left(\frac{x_3 - \text{mean}_x}{SD_x}\right)\left(\frac{y_3 - \text{mean}_y}{SD_y}\right)$
.	.	.	.
.	.	.	.
.	.	.	.
N	X_N	Y_N	$r_{xy} = \frac{1}{n-1} \sum \left(\frac{x_i - \text{mean}_x}{SD_x}\right)\left(\frac{y_i - \text{mean}_y}{SD_y}\right)$

E. Pearson r (or Pearson product-moment coefficient)

1. In the formula, the r is the average of the cross-product between standardized X and Y values (or their z scores)
2. The Pearson r is a statistic that specifies the strength and direction of a relationship between two continuous (interval or ratio) variables
3. An r value is between -1.0 and +1.0

a) An r value less than zero is a negative correlation—as the values of one variable increase, the values of the other variable decrease. The variables are *inversely* related.

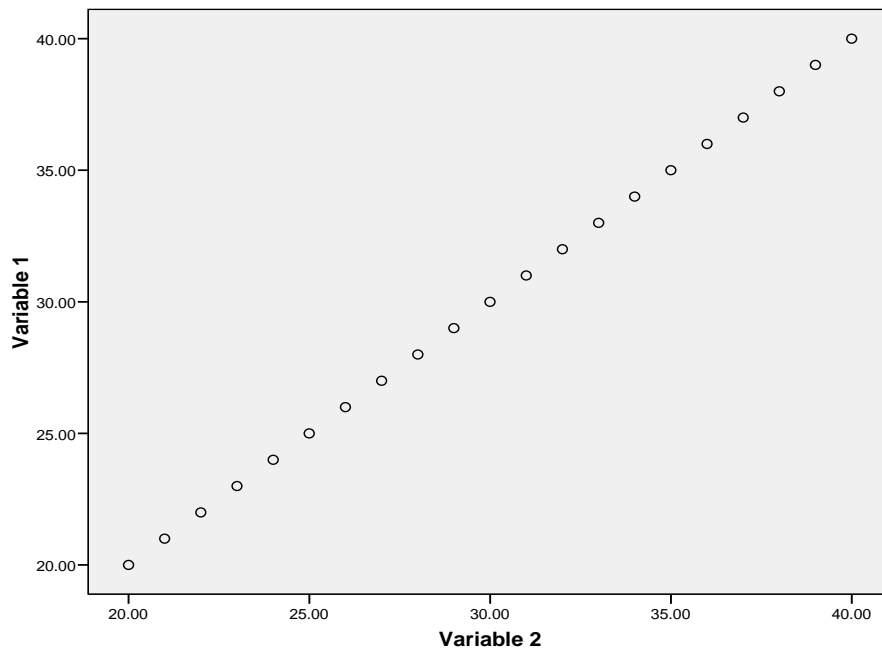
This is a graph is a perfect negative correlation, $r = -1.0$



***Give example of two continuous variables with a negative correlation.

- b) An r value greater than zero is a positive correlation--as the values of one variable increase, the values of the other variable increase

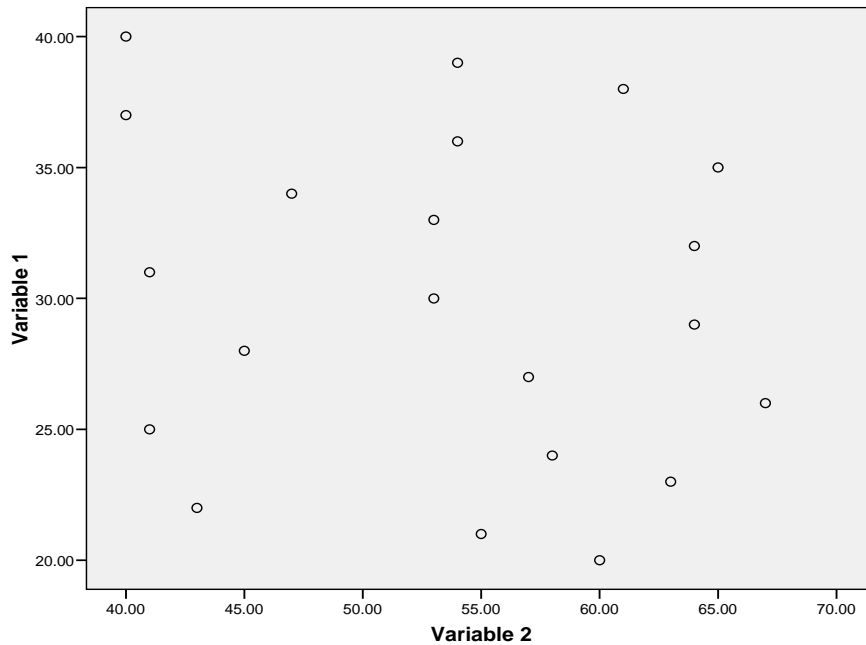
This is a graph of a perfect positive correlation, $r = +1.0$



- c) An r value of zero is no correlation

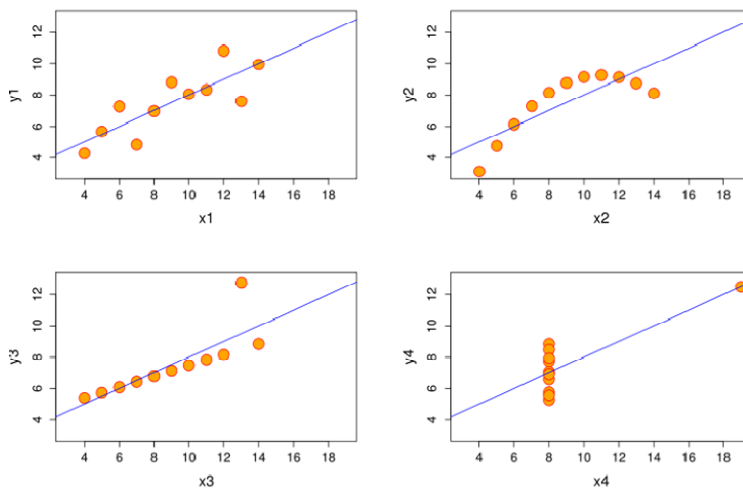
***Give an example of two continuous variables with a positive correlation.

This is a graph of no correlation (the observations are randomly distributed), $r \approx 0$



d) The higher the r , the more likely there is a linear relationship between X and Y . *However*— when the distribution of either variable deviates from normality, or when there are outliers, or when the variance of the correlation changes with larger or smaller values of X , then the r does not necessarily show a linear relationship:

$r = 0.81$,
but— these are very different distributions!



So—correlation is most accurate when the assumptions listed above are met. When they're not, we need other (non-parametric) statistics to understand the relationship between X and Y .

e) Rarely in social sciences research do we see perfect positive or negative correlations, or where $r = 0$.

F. Inference testing with the Pearson r : how do we know if the r we obtain is statistically significant (i.e. whether the two variables are really correlated in the population?)

Research Scenario: Correlation

In the general child population, a number of contextual factors have been linked to emotional problems in children and youth, including:

- 1) Low income,
- 2) Negative parenting behavior (i.e. hostile or coercive parenting),
- 3) Family conflict (including family violence and verbal abuse), and
- 4) Low self-efficacy of the child's primary caregiver (i.e. the extent to which the primary caregiver feels mastery over her/his life)

Although the relationship between 1) income, 2) parenting, 3) family conflict, and 4) primary caregiver self-efficacy and children's emotional problems has been established in the general child population, much less is known about how these four factors might contribute to emotional problems among children in immigrant families.

You are interested in testing the relationship between income, parenting behavior, family conflict and primary caregiver self-efficacy among a randomly selected sample of 379 children of immigrant parents in Los Angeles County. Within each family one primary caregiver and one child is interviewed with a structured interview format. This sample includes children ages 9 to 15.

First, we will look at the **separate bivariate relationships between each independent variable and the dependent variable**. Then we will include all variables in a simultaneous multivariate model (reserve your seat now for the next exciting class!!)

Eight Steps to Hypothesis Testing

1. Identify the independent variables and their levels of measurement

There are four continuous independent variables in this research scenario:

Variable Name	As measured by...	Level of measurement
Poverty	Income-to-needs ratio (total family income divided by federal poverty threshold). A ratio of 1 means family	Continuous

Variable Name	As measured by...	Level of measurement
	income is exactly proportional to family's financial needs (by federal standards).	
Parenting behavior	Home Observation and Measurement of Environment (HOME) Inventory. The higher the score, the more positive the parenting behavior	Continuous
Family conflict	Family Conflict Scale—child response—questions about how family communicates and solves problems. The lower the score, the lower the perceived amount of conflict	Continuous
Primary caregiver (PGC) self efficacy	Pearlin Self-Efficacy Scale—parent response to questions about self efficacy (feeling that one has control over one's life). Higher scores indicate higher self-efficacy	Continuous

2. Identify the dependent variable and level of measurement

Variable Name	As measured by...	Level of measurement
Child's emotional problems	Behavior Problem Index, internalizing sub-scale, which includes a series of self-report questions that the parent answers in relation to the emotional well-being of their child (e.g. depression, anxiety, withdrawal etc). The higher the score, the more severe the emotional problems	Continuous

3. State the null hypotheses (there will be four, since there are four independent variables)

6. Results (SPSS output) ($\alpha \leq .05$)

Independent Variable: Poverty

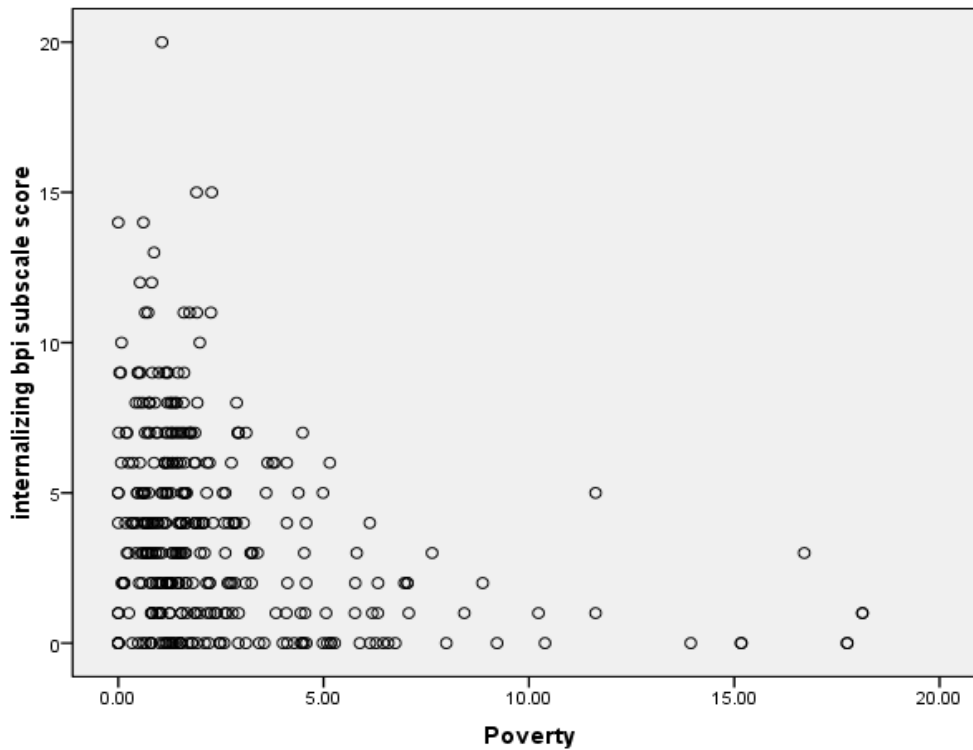
Dependent Variable: Children's emotional problems (BPI internalizing subscale score)

Correlations^a

		internalizing bpi subscale score	Poverty
internalizing bpi subscale score	Pearson Correlation	1.000	-.306**
	Sig. (2-tailed)		.000
Poverty	Pearson Correlation	-.306**	1.000
	Sig. (2-tailed)	.000	

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=352



Independent Variable: Parenting behavior

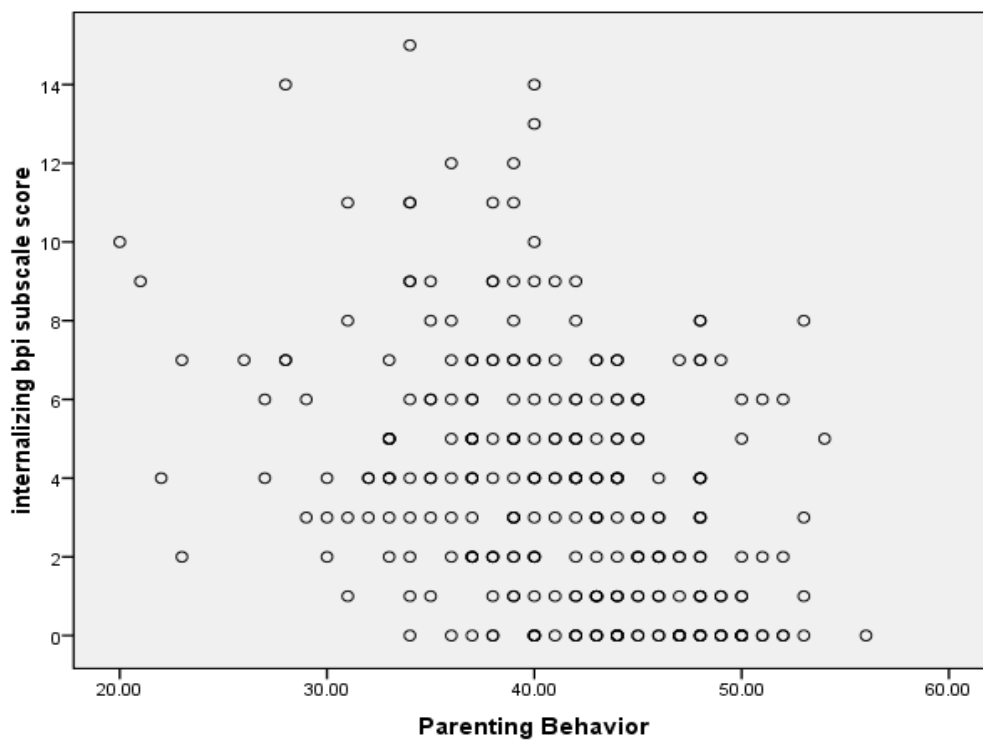
Dependent Variable: Children's emotional problems (BPI internalizing subscale score)

Correlations^a

		internalizing bpi subscale score	Parenting Behavior
internalizing bpi subscale score	Pearson Correlation	1.000	-.382**
	Sig. (2-tailed)		.000
Parenting Behavior	Pearson Correlation	-.382**	1.000
	Sig. (2-tailed)	.000	

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=280



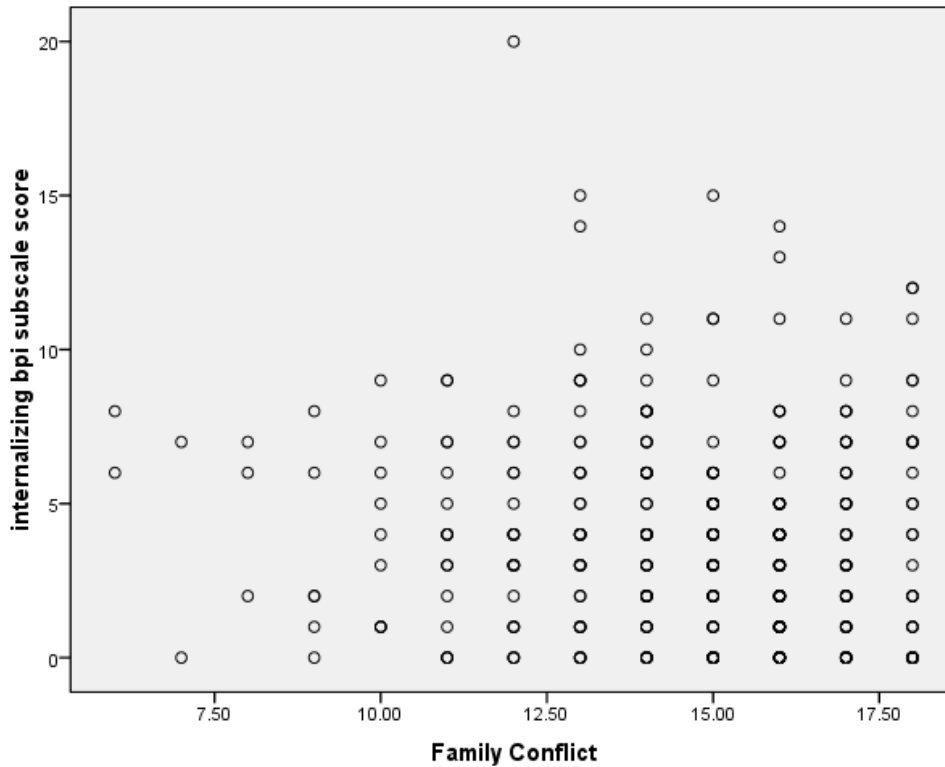
Independent Variable: Family Conflict

Dependent Variable: Children's emotional problems (BPI internalizing subscale score)

Correlations^a

		internalizing bpi subscale score	Family Conflict
internalizing bpi subscale score	Pearson Correlation	1.000	-.091
	Sig. (2-tailed)		.090
Family Conflict	Pearson Correlation	-.091	1.000
	Sig. (2-tailed)	.090	

a. Listwise N=346



Independent Variable: Primary Caregiver Self-Efficacy

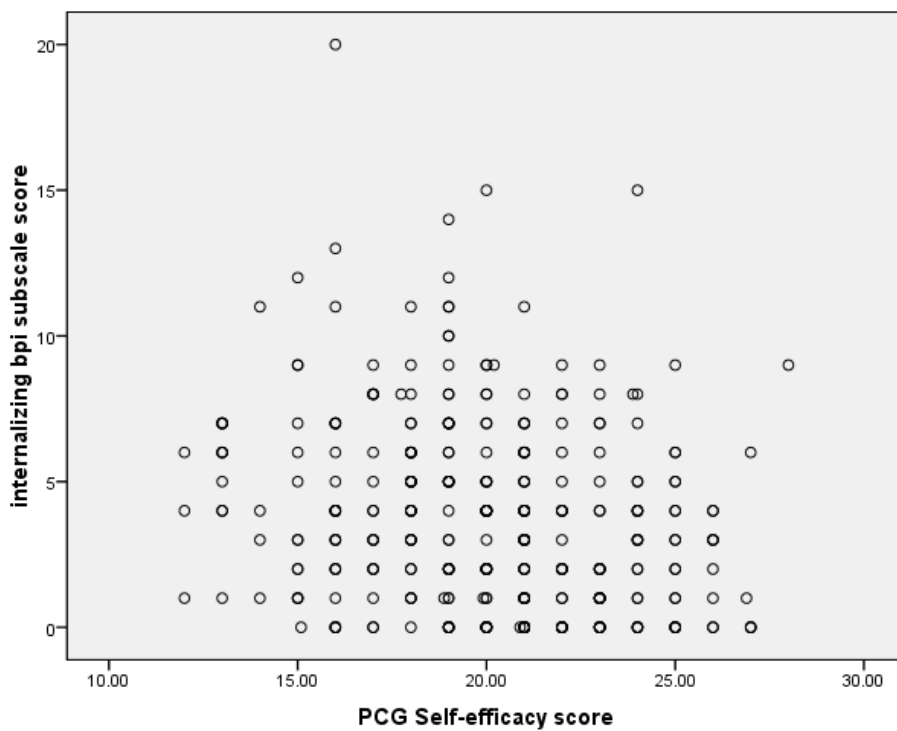
Dependent Variable: Children's emotional problems (BPI internalizing subscale score)

Correlations^a

		internalizing bpi subscale score	PCG Self- efficacy score
internalizing bpi subscale score	Pearson Correlation	1.000	-.222**
	Sig. (2-tailed)		.000
PCG Self-efficacy score	Pearson Correlation	-.222**	1.000
	Sig. (2-tailed)	.000	

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=350



7) Describe results and decision to accept or reject the null hypotheses (use APA). Text in the boxes is suitable for reporting in the Results section. Note the format for reporting the r , sample size for each analysis, and the p value.

For poverty and BPI:

Results support the hypothesis that poverty is negatively related to the child's emotional problems (BPI scores), $r(352) = -.306, p < .001$. The p value of is less than the alpha of .05, and we can reject the null hypothesis and conclude that as the income/needs ratio decreases, children's emotional problems increase.

For parenting behavior and BPI:

Results also support the hypothesis that parenting behavior is negatively related to the child's emotional problems, $r(280) = -.382, p < .001$. Since the p value is less than the alpha of .05, we can reject the null hypothesis and conclude as positive parenting behavior decreases, children's emotional problems increase (this is a negative correlation).

For family conflict and BPI:

The hypothesis that family conflict will be positively related to the child's emotional problems, however, was not supported, $r(346) = -.091, p = .090$. Since the p value is greater than the alpha of .05, there is no evidence to reject the null hypothesis in favor of the alternative hypothesis, and conclude that there is not a relationship between family conflict and children's emotional problems.

Caregiver self efficacy and BPI:

The hypothesis that caregiver self-efficacy is negatively related child's emotional problems is supported by the findings, $r(350) = -.222, p < .001$. Since the p vale is less than the alpha of .05, we can reject the null hypothesis and conclude that as primary caregiver self-efficacy decreases, children's emotional problems increase.

8) Provide a discussion of these results

Results of the bivariate analyses using Pearson correlation indicated a statistically significant relationship between children's emotional problems and income, parenting behavior and primary caregiver self-efficacy. Specifically the lower a family's income to needs ratio, the higher the level of emotional problems for the child; the lower a parent's level of parenting behavior (i.e. low levels of emotional support and cognitive stimulation), the higher the level of emotional problems for the child; and the lower the level of primary caregiver self-efficacy, the higher the level of emotional problems for the child. There was no significant correlation between family conflict and children's emotional problems.

What else can we say about:

Meaning and implications? Limitations/areas for future research? Transcultural contexts?