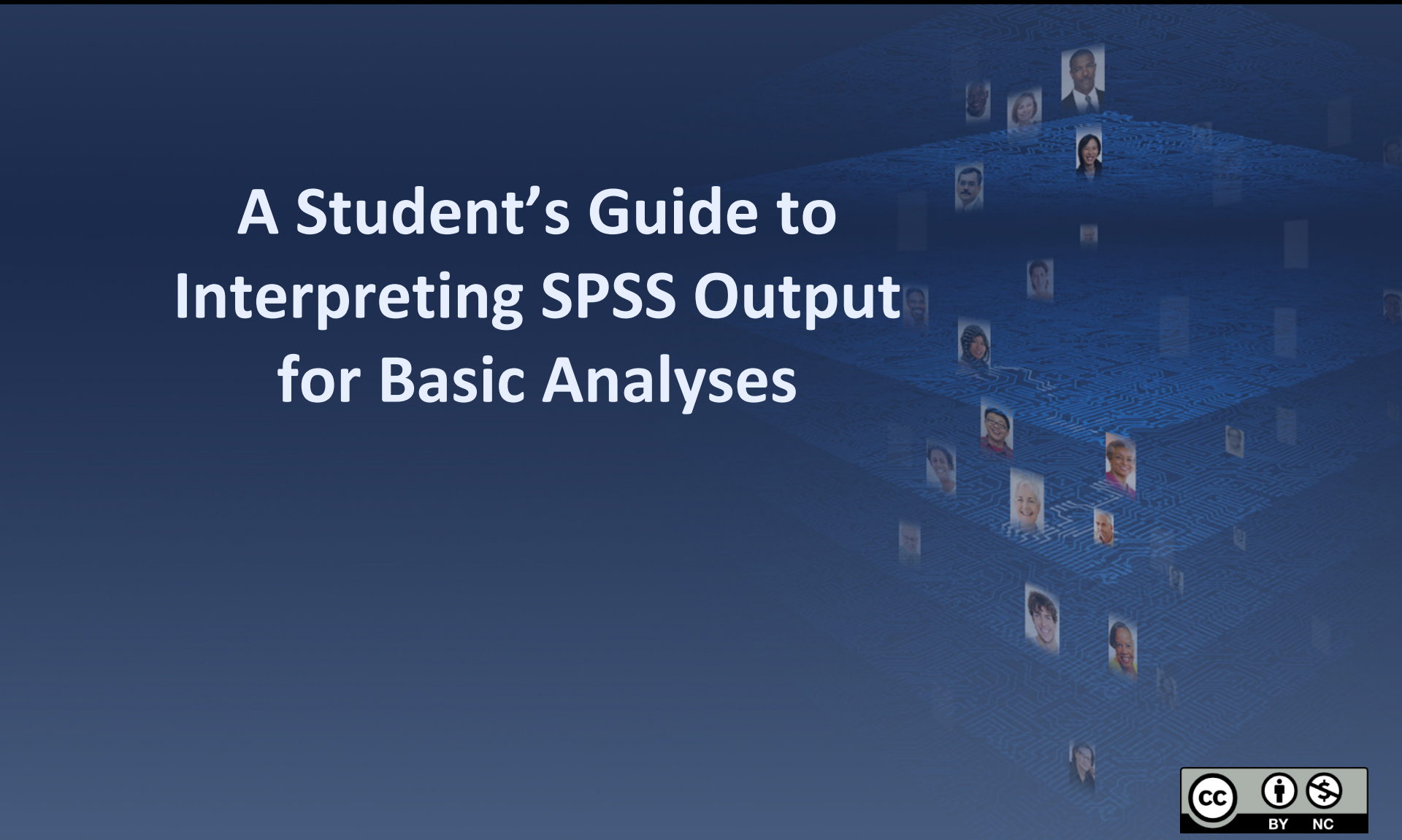
ICPSR

Inter-university Consortium for Political and Social Research

A Student's Guide to Interpreting SPSS Output for Basic Analyses



ICPSR

Inter-university Consortium for Political and Social Research

These slides give examples of SPSS output with notes about interpretation. All analyses were conducted using the *Family Exchanges Study, Wave 1* (target dataset)[1] from ICPSR. The slides were originally created for Intro to Statistics students (undergrads) and are meant for teaching purposes only[2]. For more information about the data or variables, please see: http://dx.doi.org/10.3886/ICPSR36360.v2[3]

[1] Fingerman, Karen. Family Exchanges Study Wave 1. ICPSR36360-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-04-14. http://doi.org/10.3886/ICPSR36360.v2

[2] The text used for the course was *The Essentials of Statistics: A Tool for Social Research* (Healey, 2013).

[3] Some variables have been recoded so that higher numbers mean more of what is being measured. In those cases, an "r" is appended to the original variable name.

ICPSR

Inter-university Consortium for Political and Social Research

Frequency Distributions

Frequencies show how many people fall into each answer category on a given question (variable) and what percentage of the sample that represents.

Number of people who responded that "child1" was married

Percent of the total sample who answered that "child1" was married

Percent of those with non-missing data on this question who answered that "child1" was married

Number of people with valid (non-missing) answers to the question

Total number of people in the survey sample

Cumulative percent adds the percent of people answering in one category to the total of those in all categories with lower values. It is only meaningful for variables measured at the ordinal or interval/ratio level.

Statistics

S1C MARITAL STATUS-CHILD1

N	Valid	630
	Missing	3

S1C MARITAL STATUS-CHILD1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 MARRIED	125	19.7	19.8	19.8
	2 REMARRIED	6	.9	1.0	20.8
	3 COHABITING	54	8.5	8.6	29.4
	4 SINGLE,NEVER MARRIED	409	64.6	64.9	94.3
	5 DIVORCED	21	3.3	3.3	97.6
	7 SEPARATED	11	1.7	1.7	99.4
	76 OTHER	4	.6	.6	100.0
	Total	630	99.5	100.0	
Missing	-8 DK	3	.5		
Total		633	100.0		

Crosstabulation Tables

"Crosstabs" are frequency distributions for two variables together. The counts show how many people in category one of the first variable are also in category one of the second and so on.

d34r Amount of help adult children need recoded * A1A ARE ALL CHILDREN BIOLOGICAL? Crosstabulation

Count

		A1A ARE ALL CHILDREN BIOLOGICAL?		
		1 YES	5 NO	Total
d34r Amount of help adult children need recoded	1.00 less help	167	52	219
	2.00 about the same	266	56	322
	3.00 more help	51	23	74
Total		484	131	615

Number of people who answered that their children are biologically related to them *and* that their children need less help than others their age.

Marginal: Total number of people who answered that all of their children are biologically related to them.

Marginal: Total number of people who answered that their children need more help than others their age.

Marginal: Total number of people who had valid data on both D34r and A1A.

Crosstabulation Tables (Column %)

Crosstabs can be examined using either row or column percentages and the interpretation differs depending on which are used. The rule of thumb is to percentage on your independent variable.

d34r Amount of help adult children need recoded * A1A ARE ALL CHILDREN BIOLOGICAL? Crosstabulation

			A1A ARE ALL CHILDREN BIOLOGICAL?		
			1 YES	5 NO	Total
d34r Amount of help adult children need recoded	1.00 less help	Count	167	52	219
		% within A1A ARE ALL CHILDREN BIOLOGICAL?	34.5%	39.7%	35.6%
	2.00 about the same	Count	266	56	322
		% within A1A ARE ALL CHILDREN BIOLOGICAL?	55.0%	42.7%	52.4%
	3.00 more help	Count	51	23	74
		% within A1A ARE ALL CHILDREN BIOLOGICAL?	10.5%	17.6%	12.0%
Total		Count	484	131	615
		% within A1A ARE ALL CHILDREN BIOLOGICAL?	100.0%	100.0%	100.0%

Percent of the sample whose children are all biologically related to them that said their children need less help than others their age.

Marginal: Percent of sample who feel that their children need less help than others their age.

Marginal: 100% here tells you that you've percentaged on columns.

Can you interpret this number?
(17.6% of those whose children were not all biologically related to them felt their children needed more help than others their age.)

Crosstabulation Table (Row %)

d34r Amount of help adult children need recoded * A1A ARE ALL CHILDREN BIOLOGICAL? Crosstabulation

			A1A ARE ALL CHILDREN BIOLOGICAL?		
			1 YES	5 NO	Total
d34r Amount of help adult children need recoded	1.00 less help	Count	167	52	219
		% within d34r Amount of help adult children need recoded	76.3%	23.7%	100.0%
	2.00 about the same	Count	266	56	322
		% within d34r Amount of help adult children need recoded	82.6%	17.4%	100.0%
	3.00 more help	Count	51	23	74
		% within d34r Amount of help adult children need recoded	68.9%	31.1%	100.0%
Total		Count	484	131	615
		% within d34r Amount of help adult children need recoded	78.7%	21.3%	100.0%

Percent of those who said their children need more help than others their age whose children are all biologically related to them.

Marginal: Percent of the sample whose children are biologically related to them.

Marginal: 100% here shows that you are using row percentages.

Can you interpret this number?
(17.4% of those who said their children need about the same amount of help as others their age had children who were not all biologically related to themselves.)

ICPSR

Inter-university Consortium for Political and Social Research

Chi Square (X^2)

Based on crosstabs, X^2 is used to test the relationship between two nominal or ordinal variables (or one of each). It compares the actual (observed) numbers in a cell to the number we would expect to see if the variables were unrelated in the population.

d34r Amount of help adult children need recoded * A1A ARE ALL CHILDREN BIOLOGICAL? Crosstabulation

			A1A ARE ALL CHILDREN BIOLOGICAL?		
			1 YES	5 NO	Total
d34r Amount of help adult children need recoded	1.00 less help	Count	167	52	219
		Expected Count	172.4	46.6	219.0
	2.00 about the same	Count	266	56	322
		Expected Count	253.4	68.6	322.0
	3.00 more help	Count	51	23	74
		Expected Count	58.2	15.8	74.0
Total		Count	484	131	615
		Expected Count	484.0	131.0	615.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.938[a]	2	.019
Likelihood Ratio	7.687	2	.021
Linear-by-Linear Association	.082	1	.775
N of Valid Cases	615		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 15.76.

Actual count (f_o).

Count expected (f_e) if variables were unrelated in pop.

$f_e = \frac{row\ marginal\ x\ column\ marginal}{N}$

X^2 value (obtained) = $\sum \frac{(fo - fe)2}{fe}$ This could be compared to a critical value (with the degrees of freedom) but the significance here tells you that there appears to be a relationship between the perceived amount of help needed and whether the children are related to the R.

Degrees of freedom = ((# rows – 1)(# col – 1))

The X^2 test is sensitive to small expected counts, it is less reliable if f_e is < 5 for multiple cells.

Independent-samples T-test

A test to compare the means of two groups on a quantitative (at least ordinal, ideally interval/ratio) dependent variable. Acomputed variable (dmean_m) exists in this dataset that is the average amount of support R offers mother across all domains (range 1-8); that will be the dependent variable.

Group Statistics

	ZSEX SCREENER SEX	N	Mean	Std. Deviation	Std. Error Mean
DMEAN_M Mother Support Provided MEAN	0 Female	285	4.7233	1.41115	.08359
	1 Male	258	4.3656	1.50724	.09384

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DMEAN_M Mother Support Provided MEAN	Equal variances assumed	1.498	.222	2.855	541	.004	.35764	.12526	.11159	.60369
	Equal variances not assumed			2.846	526.621	.005	.35764	.12567	.11077	.60451

The number of females and males who have non-missing data for dmean_m.

The actual mean and standard deviation of dmean_m for each of the groups.

T-value for the difference between 4.7233 and 4.3656. $t = \frac{\overline{x_1} - \overline{x_2}}{\sigma_{\bar{x}-\bar{x}}}$ where $\sigma_{\bar{x}-\bar{x}} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$.

(Note, as long as the sig. of F is ≥ .05, use the "equal variances assumed" row, for reasons beyond the scope of these slides).

Significance (p) level for t-statistic. If p ≤ .05, the two groups have statistically significantly different means. Here, females provide more support to their mothers, on average, than do males.

Confidence interval for the difference between the two means. If CI contains 0, the difference will not be statistically significant.

Paired-samples T-test

Like the independent-samples t-test, this compares two means to see if they are significantly different, but now it is comparing the average of same people's scores on two different variables. Often used to compare pre- and post-test scores, time 1 and time 2 scores, or, as in this case, the differences between the average amount of help Rs give to their mothers versus to their fathers.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	DMEAN_M Mother Support Provided MEAN	4.2672	235	1.31714	.08592
	DMEAN_F Father Support Provided MEAN	3.3291	235	1.41857	.09254

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	DMEAN_M Mother Support Provided MEAN - DMEAN_F Father Support Provided MEAN	.93816	1.49744	.09768	.74571	1.13061	9.604	234	.000

Mean amount of support R provided to mothers.

Mean amount of support R provided to fathers.

Number of cases with valid data for both variables. SPSS also gives the correlation between the two dependent variables, that was left off here for space.

The difference between the average amount of support provided to mothers and fathers and accompanying standard deviation.

T-statistic for the difference between the two means and the significance. In this sample, respondents provide significantly more support to their mothers than to their fathers.

ICPSR

Inter-university Consortium for Political and Social Research

Oneway ANOVA

Another test for comparing means, the oneway ANOVA is used when the independent variable has three or more categories. You would typically report the F-ratio (and sig.) and use the means to describe the groups.

Number of cases in each group of the independent variable.

Average amount of support provided (and standard deviation) by those with incomes < $10k.

Average amount of support provided by all 521 people with valid data on dmean_m.

Confidence Interval: the range within which you can be 95% certain that the group's mean falls for the population.

Variability of means between the groups: Mean Square Between = $\frac{SSB}{dfb}$ where SSB= $\sum N_k(\bar{x}_k - \bar{x})^2$ and dfb = k-1. (k is number of groups; N_k is # people in a given group; $\bar{x}_k$ is mean for that group)

Total Sum of Squares (SST) = SSB + SSW

Variability within each group: Within Groups Mean Square = $\frac{SSW}{dfw}$ where SSW = SST-SSB and dfw = N-k.

Descriptives

DMEAN_M Mother Support Provided MEAN

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 < $10,000	30	4.9778	1.65671	.30247	4.3592	5.5964	1.33	7.83
2 $10,001 TO $25,000	38	5.6447	1.37828	.22359	5.1917	6.0978	2.83	8.00
3 $25,001 TO $40,000	39	4.6923	1.70848	.27358	4.1385	5.2461	1.17	8.00
4 $40,001 TO $75,000	157	4.6592	1.41300	.11277	4.4365	4.8820	1.00	7.67
5 $75,001 TO $100,000	101	4.3119	1.47359	.14663	4.0210	4.6028	1.00	8.00
6 > $100,000	156	4.2532	1.28396	.10280	4.0501	4.4563	1.00	8.00
Total	521	4.5630	1.46630	.06424	4.4368	4.6892	1.00	8.00

ANOVA

DMEAN_M Mother Support Provided MEAN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	73.074	5	14.615	7.203	.000
Within Groups	1044.940	515	2.029		
Total	1118.014	520			

F-statistic (and associated p-value) test the null hypothesis that all groups have the same mean in the population. A significant F means that at least one group is different than the others. Small within groups variance and large between groups produces a higher F –value: F= $\frac{Mean\ square\ between}{Mean\ square\ within}$. Here we see that at least one group's mean amount of support is significantly different than the others. Additional (post-hoc) tests can be run to determine which groups are significantly different than each other.

Correlation

Pearson's r: measures the strength and direction of association between two quantitative variables. Matrices are symmetric on the diagonal.

Correlation coefficient (r) tells how strong the relationship is and in what direction. Range is -1 to 1, with absolute values closer to 1 indicating stronger relationships. Here, the frequency of visits is moderately related to the amount of emotional support given to the mother; more visits correlate with more frequent emotional support.

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{[\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

Significance (p) tells how certain you can be that the relationship displayed is not due to chance. Typically look for this to be ≤ .05.

N=sample size: this number may be different in each cell if missing cases are excluded pairwise rather than listwise.

Can you interpret this number? (How often mother forgets to ask about R's life is negatively, albeit weakly, related to the amount of emotional support R gives mother – If mother forgets to ask a great deal of the time, she gets less emotional support from R.)

Correlations

		M_7 MOTHER FORGETS TO ASK ABOUT R'S LIFE	B3AR Seen Mother in Person REVERSE CODED	D1R R provides mother with emotional support REVERSE CODED	D22R R provides mother with financial support REVERSE CODED
M_7 MOTHER FORGETS TO ASK ABOUT R'S LIFE	Pearson Correlation	1	-.235	-.131	-.021
	Sig. (2-tailed)		.000	.002	.622
	N	540	540	540	539
B3AR Seen Mother in Person REVERSE CODED	Pearson Correlation	-.235	1	.476	.284
	Sig. (2-tailed)	.000		.000	.000
	N	540	543	542	542
D1R R provides mother with emotional support REVERSE CODED	Pearson Correlation	-.131	.476	1	.345
	Sig. (2-tailed)	.002	.000		.000
	N	540	542	542	541
D22R R provides mother with financial support REVERSE CODED	Pearson Correlation	-.021	.284	.345	1
	Sig. (2-tailed)	.622	.000	.000	
	N	539	542	541	542

Bivariate Regression (model statistics)

Examines the relationship between a single independent ("cause") variable and a dependent (outcome) variable. While it's good to look at all numbers, the ones you typically interpret/report are those boxes marked with an * (true for all following slides).

Regression line: $\hat{y} = a + b(x)$.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.306[a]	.094	.092	.91898

a. Predictors: (Constant), B3AR R has seen mother in person in year - reversed

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	47.249	1	47.249	55.947	.000[b]
	Residual	456.045	540	.845		
	Total	503.293	541			

a. Dependent Variable: N_10r Satis with amount of help given to mother - reversed

b. Predictors: (Constant), B3AR R has seen mother in person in year - reversed

Independent (predictor) and Dependent variables.

Multiple correlation (R): in bivariate regression, same as standardized coefficient

Coefficient of determination (R^2): the amount of variance in satisfaction with help given to mother that is explained by how often the R saw mother. R^2 = (TSS – SSE)/ TSS. *

Total Sum of Squares (TSS) = $\sum(y - \bar{y})^2$.

Residual sum of squared errors (or Sum of Squared Errors, SSE) = $\sum(y - \hat{y})^2$.

F-value (and associated p-value) tells whether model is statistically significant. Here we can say that the relationship between frequency of visits with ones mother and satisfaction with help given is significant; it is unlikely we would get an F this large by chance. *

Bivariate Regression (coefficients)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.770	.119		14.920	.000	1.537	2.003
	B3AR R has seen mother in person in year - reversed	.157	.021	.306	7.480	.000	.115	.198

a. Dependent Variable: N_10r Satis with amount of help given to mother - reversed

Y-intercept (a): value of y when x is 0.

Slope (b): how much y chances for each unit increase in x. Here, for every additional "bump up" in frequency of visits, satisfaction with the amount of help given to mother increases by .157. *

Standard error of the estimate: divide the slope by this to get the t-value. *

Standardized coefficient (β): influence of x on y in "standard units."

T-statistic (and associated p-value) tells whether the individual variable has a significant effect on dependent variable. *

Confidence Interval – the slope +/- (critical t-value * std. error) – shows that you can be 95% confident that the slope in the population falls within this range. If range contains 0, variable does not have an effect on y.

The regression equation (for this model would be $\hat{y} = 1.770 + .157(x)$.

ICPSR

Inter-university Consortium for Political and Social Research

Multiple Regression (OLS: model statistics)

Used to find effects of multiple independent variables (predictors) on a dependent variable. Provides information about the independent variables as a group as well as individually.

Regression line: $Y' = a + b_1x_1 + b_2x_2 + b_3x_3 \ldots$

R= multiple correlation. The association between the group of independent variables and the dependent variable. Ranges from 0-1.

R^2 and Adjusted R^2– how much of the variance in satisfaction with amount of help R provided mother is explained by the combination of independent variables in the model. Also called "coefficient of determination." R^2 = (TSS – SSE)/ TSS. Adjusted R^2 compensates for the effect that adding any variable to a model will raise the R^2 to some degree. About 14% of the variance in satisfaction is accounted for by financial and emotional support, seeing mother in person, and whether mother makes demands on R. *

Residual sum of squared errors (or Sum of Squared Errors, SSE) = $\sum(y - \hat{y})^2$.

Total Sum of Squares (TSS) = $\sum(y - \bar{y})^2$.

Df1 = # of independent variables

Df2 = # of cases – (# of independent variables +1)

F-value and associated p-value tells whether model is statistically significant (chance that at least one slope is not zero in the population). This combination of independent variables significantly predicts satisfaction with amount of help R gives mother. *

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.387[a]	.150	.143	.89280

a. Predictors: (Constant), D22R R provides mother with financial support REVERSE CODED, M_6 MOTHER MAKES DEMANDS ON R, B3AR Seen Mother in Person REVERSE CODED, D1R R provides mother with emotional support REVERSE CODED

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	75.119	4	18.780	23.560	.000[b]
	Residual	427.244	536	.797		
	Total	502.363	540			

a. Dependent Variable: N_10r Satis with amount of help given to mother - reversed

b. Predictors: (Constant), D22R R provides mother with financial support REVERSE CODED, M_6 MOTHER MAKES DEMANDS ON R, B3AR Seen Mother in Person REVERSE CODED, D1R R provides mother with emotional support REVERSE CODED

ICPSR

Inter-university Consortium for Political and Social Research

Multiple Regression (coefficients)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	1.477	.138		10.695	.000			
	B3AR Seen Mother in Person REVERSE CODED	.130	.023	.256	5.593	.000	.306	.235	.223
	M_6 MOTHER MAKES DEMANDS ON R	.195	.036	.225	5.413	.000	.281	.228	.216
	D1R R provides mother with emotional support REVERSE CODED	-.009	.023	-.018	-.373	.709	.181	-.016	-.015
	D22R R provides mother with financial support REVERSE CODED	.033	.023	.061	1.414	.158	.177	.061	.056

a. Dependent Variable: N_10r Satis with amount of help given to mother - reversed

Y-intercept (a): value of y when all Xs are 0.

Slope (b): how much satisfaction changes for each increase in frequency of visiting mother. *

Standard error of the estimate: divide the slope by this to get the t-value. *

Standardized coefficient (β): influence of x on y in "standard units." Can use this coefficient to compare the strength of the relationships of each of the independent variables to the dependent. Largest β = strongest relationship, so here frequency of visits has the strongest relationship to satisfaction, all else constant.

T-statistic (and associated p-value) tells whether the individual variable has a significant effect on dependent variable, controlling for the other independent variables. *

Zero-order correlation = Pearson's r. Bivariate relationship between frequency of visits and R's satisfaction w/ help R's given to mother.

Partial correlation: bivariate relationship between frequency of visits and R's satisfaction, *controlling for demands mother makes and emotional/financial support given*.

Regression Equation: $\hat{y} = 1.477 + .130(\text{b3ar}) + .195(\text{m\_6}) - .009(\text{d1r}) + .033(\text{d22r})$

OLS with Dummy Variables

Using a categorical variable broken into dichotomies (e.g., race recoded into white, black, other with each coded 1 if R fits that category and 0 if not) as predictors. In this case, the amount of help R perceives his/her adult child to need was recoded into 1 = "more than others" and 0 = "less or about the same as others." If the concept is represented by multiple dummy variables, leave one out as the comparison group (otherwise there will be perfect multicollinearity in the model).

Dummy variable. Children who need less or the same amount of help as their peers is the reference category (0).

Slope is interpreted as the amount of difference between the "0" group and the "1" group. Here, those who perceive their children needing more help than their peers are .321 more satisfied with the amount of help they give their mother than those whose children require less help, controlling for frequency of visiting mother, demands made, and emotional and financial support provided (and the difference is statistically significant).

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.402[a]	.161	.153	.88757

a. Predictors: (Constant), d34rr Children need more help than others their age, M_6 MOTHER MAKES DEMANDS ON R, B3AR Seen Mother in Person REVERSE CODED, D22R R provides mother with financial support REVERSE CODED, D1R R provides mother with emotional support REVERSE CODED

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	79.191	5	15.838	20.105	.000[b]
	Residual	412.009	523	.788		
	Total	491.199	528			

a. Dependent Variable: N_10r Satis with amount of help given to mother - reversed

b. Predictors: (Constant), d34rr Children need more help than others their age, M_6 MOTHER MAKES DEMANDS ON R, B3AR Seen Mother in Person REVERSE CODED, D22R R provides mother with financial support REVERSE CODED, D1R R provides mother with emotional support REVERSE CODED

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.447	.139		10.393	.000
	B3AR Seen Mother in Person REVERSE CODED	.128	.023	.252	5.459	.000
	M_6 MOTHER MAKES DEMANDS ON R	.198	.036	.228	5.456	.000
	D1R R provides mother with emotional support REVERSE CODED	-.008	.023	-.016	-.331	.741
	D22R R provides mother with financial support REVERSE CODED	.030	.024	.056	1.277	.202
	d34rr Children need more help than others their age	.321	.119	.108	2.700	.007

a. Dependent Variable: N_10r Satis with amount of help given to mother - reversed