Data Screening for Between Groups Differences

Sparky Flame

School of Behavioral Sciences, Liberty University

Author Note

Sparky Flame

I have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to

Sparky Flame

Email: sparky.flame@liberty.edu

Data Screening of Categorical and Continuous Variables

Data screening was accomplished for the variables gender, a dichotomous (categorical) variable with two groups of male and female, and IPIP-Neuroticism, an interval level of measurement variable, from the EDCO 745 course dataset in preparation to conduct an independent-measures t-test. Frequency tables were created for gender and IPIP-Neuroticism (see Tables 1-2). Results of the frequency tables indicated slightly more male ($N = 704$) than female ($N = 596$) participants (Table 1). Tests of assumptions were conducted for an independent-measures t-test.

1. You have one dependent variable that is measured at the continuous level (i.e., the interval or ratio level).
   As described in the hypothesis and variable definition, the analysis is a test of whether there is a difference of Neuroticism, the dependent variable, between groups of gender. By research design, the dependent variable is Neuroticism, which is measured by the IPIP-50 with an interval level of measure, which is continuous. Therefore, Assumption 1 is satisfied.
2. You have one independent variable that consists of two categorical, independent groups (i.e., a dichotomous variable).
   As described in the hypothesis and variable definition, the test of group differences will be between groups of gender, which are male and female. By research design, the independent variable is gender, a dichotomous variable. Therefore, Assumption 2 is satisfied.
3. You have independence of observations.
   By research design, the two groups are unique and separate. As such, members from one group (e.g., males) cannot contribute data to the other group (e.g., females). By research design, Assumption 3 is met.
4. There are no significant outliers in the two groups of your independent variable in terms of the dependent variable.
   Outliers are tested by boxplots (Figure 1). An outlier (more than 1.5 IQR) is described in SPSS as a circle and an extreme outlier (more than 3 IQR) is described as an asterisk. To test this assumption, boxplots of DV (Neuroticism) between the IV groups (male and female) were developed. There were two outliers for males and four for females. Because the cases were not considered extreme outliers, the cases were kept for analysis.
5. Your dependent variable is approximately normally distributed for each group of the independent variable.

Normality was tested using histograms with separate figures for the distribution of Neuroticism for males (Figure 1) and females (Figure 2). Distributions for both genders were approximately normal.

6. You have homogeneity of variances.

Homogeneity of variances is a test to ensure the variance of the dependent variable for each group is approximately the same. The homogeneity of variances is tested with Levene's test, which is part of the independent-measures t-test output. A Levene's test has a null hypothesis of homogeneity of variances between groups, so that a result that is not statistically significant ($p \geq .05$) indicates homogeneity of variances, meeting Assumption 6. The Levene's test for Neuroticism was statistically significant, $F(1,1298) = 11.31$, $p = .001$, indicating the variances are not equal. Therefore, the "equal variances not assumed" analysis (Welch's t-test) in SPSS will be used.

Table 1

Frequency for Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	704	54.1	54.2	54.2
	Female	596	45.8	45.8	100.0
	Total	1300	99.9	100.0	
Missing		2	.1		
Total		1302	100.0		

Table 2

Frequency for IPIP-Neuroticism

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3.00	1	.1	.1	.1
	4.00	46	3.5	3.5	3.6
	5.00	34	2.6	2.6	6.2
	6.00	25	1.9	1.9	8.1
	7.00	53	4.1	4.1	12.2
	8.00	75	5.7	5.7	17.9
	9.00	110	8.4	8.4	26.3
	10.00	156	11.9	11.9	38.3
	11.00	186	14.2	14.2	52.5
	12.00	232	17.8	17.8	70.3
	13.00	166	12.7	12.7	83.0
	14.00	92	7.0	7.0	90.0
	15.00	48	3.7	3.7	93.7
	16.00	29	2.2	2.2	95.9
	17.00	22	1.7	1.7	97.6
	18.00	14	1.1	1.1	98.7
	19.00	9	.7	.7	99.4
	20.00	8	.6	.6	100.0
	Total	1306	100.0	100.0	

Figure 1

Box plots of IPIP-Neuroticism for Gender Groups
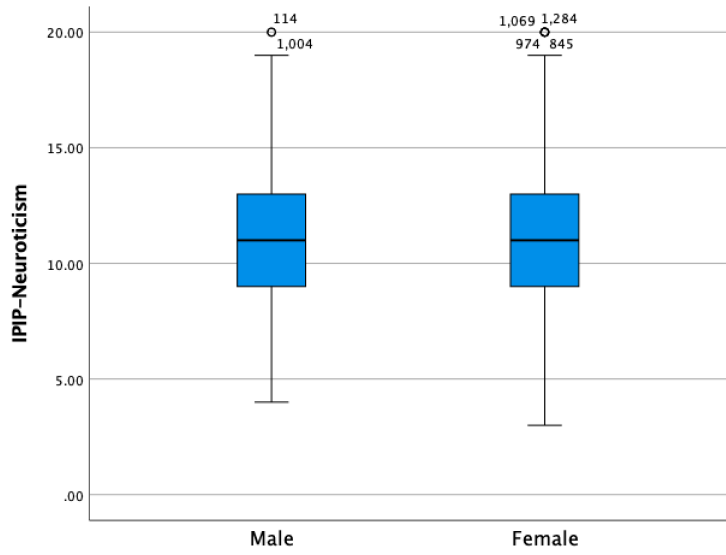

Figure 2

Histogram for IPIP-Neuroticism Males
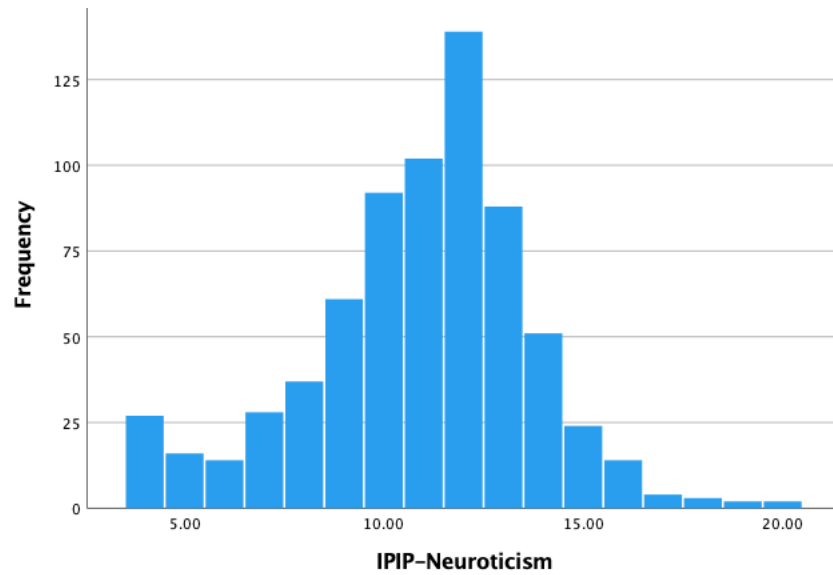
Figure 3

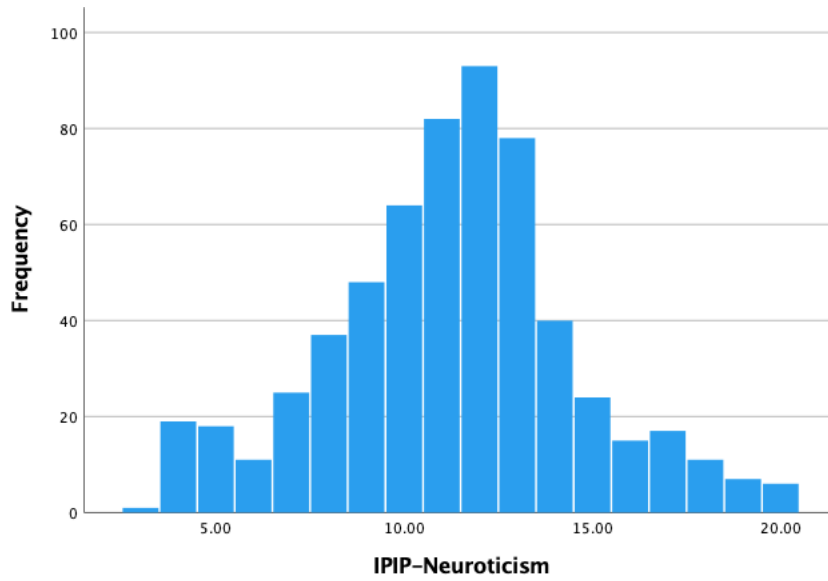*Histogram for IPIP-Neuroticism for Females*

Table 3

Differences of Neuroticism Between Genders

	Levene's Test		*t*-test for Equality of Means						
	F	Sig.	*t*	df	Sig.	Mean difference	Std Error Difference	Lower	Upper
Equal variances assumed	11.315	.001	-2.181	1128	.029	-.37061	.16994	-.70399	-.03722
Equal variances not assumed			-2.153	1179.256	.031	-.37061	.17211	-.70828	-.03293

Note: Independent variable is gender, dependent variable is Neuroticism