

## Chapter 4 – Understanding Your Data and Checking Assumptions

Using the College Student data file, do the following problems. Print your outputs and circle the key parts of the output that you discuss.

- 4.1 For the variables with five or more ordered levels, compute the skewness. Describe the results. Which variables in the data set are approximately normally distributed/scale? Which ones are ordered but not normal?
- Select **Analyze => Descriptive Statistics => Descriptives.**
  - Move *student height, same sex parent's height, amount of tv watched per week, hours of study per week, student's current gpa, positive evaluation-institution, positive evaluation-major, positive evaluation-facilities, positive evaluation-social life, hours per week spent working* in the **Variables** box.
  - **Options** => Check **Skewness** (in addition to Mean, Std. Deviation, Minimum, and Maximum) => **Continue** => **OK.**

The **Valid N (listwise)** for the variables selected is 48. The **Means** all seem reasonable and within the expected range. The **Minimum** and **Maximum** values are all within the expected range, based on the codebook. The N for each variable makes sense and only two variables are missing values (positive evaluation-major and hours per week spent working).

The **Skewness Statistic** is utilized to determine which of these variables are approximately normally distributed. The guideline is that if the **Skewness Statistic** is between -1 and 1, the variable is at least approximately normal. In this case, all the variables with five or more ordered levels fall into that range and would be considered approximately normally distributed. For this dataset, the ordinal variables with five or more ordered levels (positive evaluation-institution, positive evaluation-major, positive evaluation-facilities, positive evaluation-social life) are all approximately normally distributed and we can assume they are more like scale variables and we can use inferential statistics that have the assumption of normality with them. None of the variables examined for this problem were not normal.

- 4.3 Which variables are nominal? Run frequencies for the nominal variables and other variables with fewer than five levels. Comment on the results.
- Select **Analyze => Descriptive Statistics => Frequencies.**
  - Move *gender of student, marital status, age group, does subject have children, television shows-sitcoms, television shows-movies, television shows-sports, television shows-news shows*

The table titled **Statistics** provides the number of participants for whom we have **Valid** data and the number of **Missing** data. No other statistics were requested because almost all of them are not appropriate to use with nominal and dichotomous data. Age group has three ordered levels so it is ordinal and the median would be appropriate.

The other tables are labeled **Frequency Table** and there is one for each of the variables selected. The left-hand column shows the **Valid** categories (or levels or values), **Missing** values, and **Total** number of participants. The **Frequency** column gives the number of participants who had each value. The **Percent** column is the percent who had each value, including missing values. For example, in the marital status table, 40.0% of ALL participants were single, 36.0% were married, 22.0% were divorced, and 2.0% were missing. The **Valid Percent** shows the percent of those with nonmissing data at each value; e.g. 40.8% of the 49 students with valid data were single. Finally, **Cumulative Percent** is the percent of the subjects in a category plus the categories listed above it.

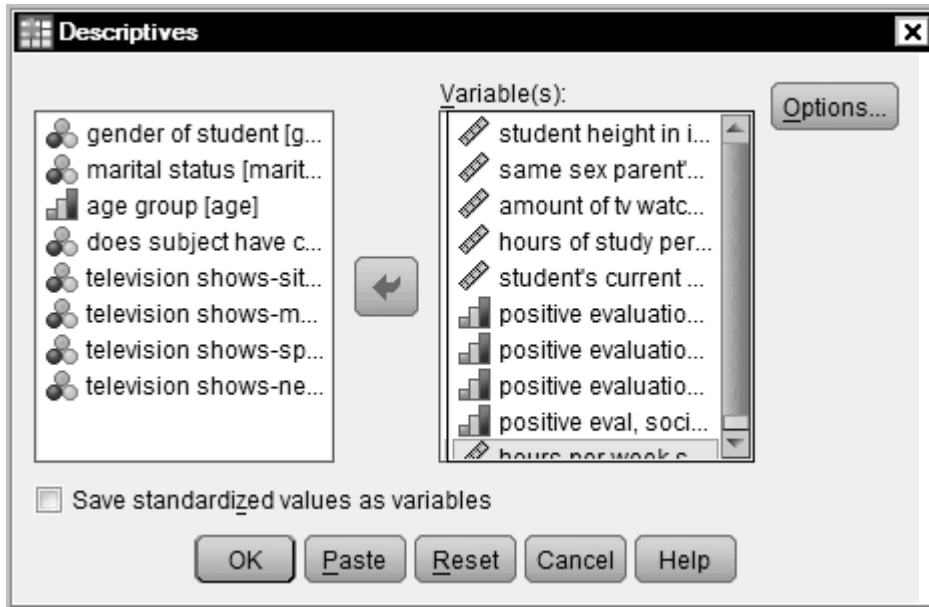


Fig. E.8

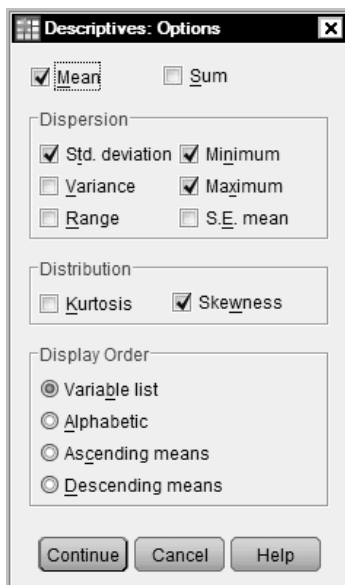


Fig. E.9

## Ch. 4 Output 4.1

```
DESCRIPTIVES VARIABLES=height pheight hrstv hrsstudy curr GPA evalinst
evalprog evalphys evalsoc1 hrswork
/STATISTICS=MEAN STDDEV MIN MAX SKEWNESS.
```

## Descriptives

Descriptive Statistics							
	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error
student height in inches	50	60.00	75.00	67.3000	3.93959	.163	.337
same sex parent's height	50	58.00	76.00	66.7800	5.10418	.333	.337
amount of tv watched per week	50	4	25	11.98	6.096	.645	.337
hours of study per week	50	2	38	15.62	8.310	.950	.337
student's current gpa	50	2.4	4.0	3.172	.3907	.147	.337
positive evaluation, institution	50	2	5	3.38	.945	.059	.337
positive evaluation, major	49	1	5	3.27	.953	-.115	.340
positive evaluation, facilities	50	1	5	2.76	1.061	-.136	.337
positive eval, social life	50	1	5	3.10	1.182	.031	.337
hours per week spent working	49	0	50	26.12	14.857	-.516	.340
Valid N (listwise)	48						

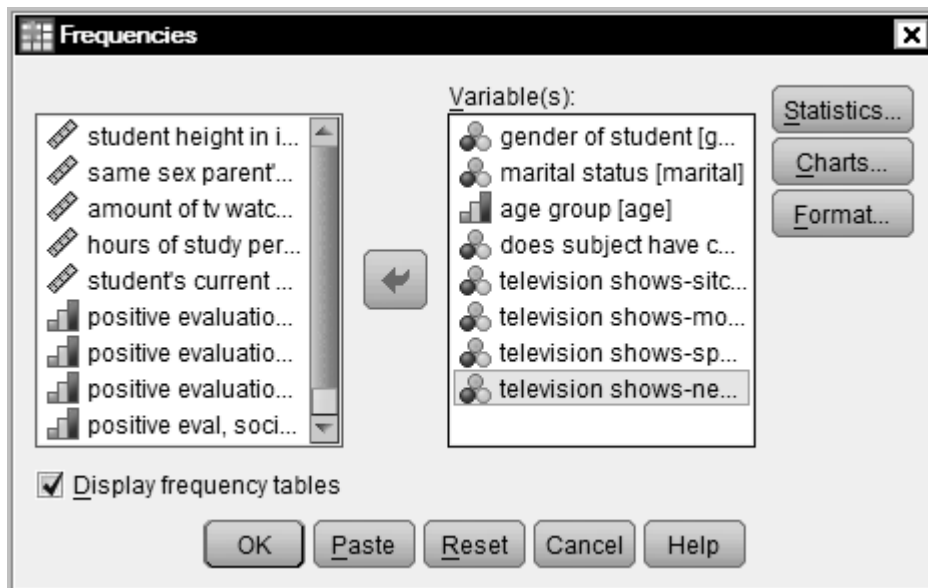


Fig. E.10

## Ch. 4 Output 4.3

```
FREQUENCIES VARIABLES=gender marital age children tvsitcom tvmovies
tvsports tvnews
/ORDER=ANALYSIS.
```

### Frequencies

Statistics									
		gender of student	marital status	age group	does subject have children	television shows-sitcoms	television shows-movies	television shows-sports	television shows-news shows
N	Valid	50	49	50	50	50	50	50	50
	Missing	0	1	0	0	0	0	0	0

### Frequency Table

gender of student					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	males	26	52.0	52.0	52.0
	females	24	48.0	48.0	100.0
	Total	50	100.0	100.0	

marital status					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	single	20	40.0	40.8	40.8
	married	18	36.0	36.7	77.6
	divorced	11	22.0	22.4	100.0
	Total	49	98.0	100.0	
Missing	System	1	2.0		
Total		50	100.0		

**age group**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less than 22	17	34.0	34.0	34.0
	22-29	18	36.0	36.0	70.0
	30 or more	15	30.0	30.0	100.0
	Total	50	100.0	100.0	

**does subject have children**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	24	48.0	48.0	48.0
	yes	26	52.0	52.0	100.0
	Total	50	100.0	100.0	

**television shows-sitcoms**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	18	36.0	36.0	36.0
	yes	32	64.0	64.0	100.0
	Total	50	100.0	100.0	

**television shows-movies**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	32	64.0	64.0	64.0
	yes	18	36.0	36.0	100.0
	Total	50	100.0	100.0	

**television shows-sports**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	24	48.0	48.0	48.0
	yes	26	52.0	52.0	100.0
	Total	50	100.0	100.0	

**television shows-news shows**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	no	27	54.0	54.0	54.0
	yes	23	46.0	46.0	100.0
	Total	50	100.0	100.0	