Bivariate Statistics with Categorical Variables

7

Abstract

In this part, we will discuss three types of bivariate statistics: first, an independent samples *t*-test measures if two groups of a continuous variable are different from one another; second, an *f*-test or ANOVA measures if several groups of one continuous variable are different from one another; third, a chi-square test gauges whether there are differences in a frequency table (i.e., two-by-two table or two-by-three table). Wherever possible we use money spent partying per week as the dependent variable. For the independent variables, we employ an appropriate explanatory variable from our sample survey.

7.1 Independent Sample *t*-Test

An independent samples *t*-test assesses whether the means of two groups are *statistically* different from each other. To properly conduct such a *t*-test, the following conditions should be met:

(1) The dependent variable should be continuous.
(2) The independent variable should consist of mutually exclusive groups (i.e., be categorical).
(3) All observations should be independent, which means that there should not be any linkage between observations (i.e., there should be no direct influence from one value within one group over other values in this same group).
(4) There should not be many significant outliers (this applies the more the smaller the sample is).
(5) The dependent variable should be more or less normally distributed.
(6) The variances between groups should be similar.

D. Stockemer, *Quantitative Methods for the Social Sciences*,
https://doi.org/10.1007/978-3-319-99118-4_7

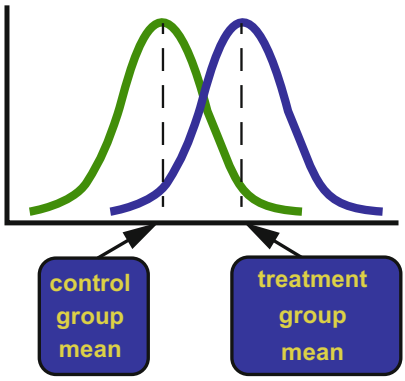

Fig. 7.1 The logic of a *t*-test

For example, for our data we might be interested whether guys spend more money than girls while partying, and therefore our dependent variable would be money spent partying (per week) and our independent variable gender. We have relative independence of observations as we cannot assume that the money one individual in the sample spends partying directly hinges upon the money another individual in the sample spends partying. From Figs. 6.23 and 6.25, we also know that the variable money spent partying per week is approximately normally distributed. As a preliminary test, we must check if the variance between the two distributions is equal, but a SPSS or Stata test can later help us detect that.

Having verified that our data fit the conditions for a *t*-test, we can now get into the mechanics of conducting such a test. Intuitively, we could first compare the means for the two groups. In other words, we should look at how far the two means are apart from each other. Second, we ought to look at the variability of the data. Pertaining to the variability, we can follow a simple rule; the less there is variability, the less there is overlap in the data, and the more the two groups are distinct. Therefore, to determine whether there is a difference between two groups, two conditions must be met: (1) the two group means must differ quite considerably, and (2) the spread of the two distributions must be relatively low. More precisely, we have to judge the difference between the two means relative to the spread or variability of their scores (see Fig. 7.1). The *t*-test does just this.

Figure 7.2 graphically illustrates that it is not enough that two group means are different from one another. Rather, it is also important how close the values of the two groups cluster around a mean. In the last of the three graphs, we can see that the two groups are distinct (i.e., there is basically no data overlap between the two groups). In the middle graph, we can be rather sure that these two groups are similar (i.e., more than 80% of the data points are indistinguishable; they could belong to either of the two groups). Looking at the first graph, we see that most of the observations clearly belong to one of the two groups but that there is also some overlap. In this case, we would not be sure that the two groups are different.

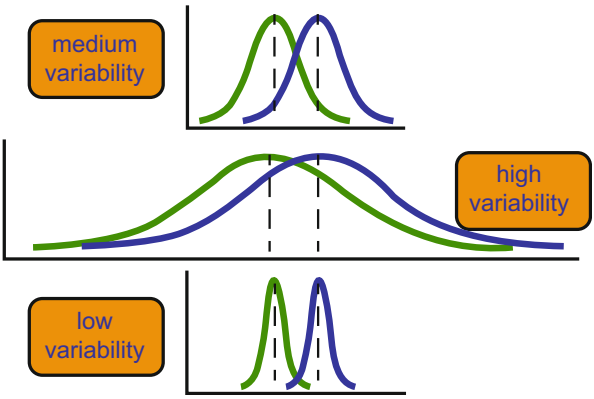

Fig. 7.2 The question of variability in a *t*-test

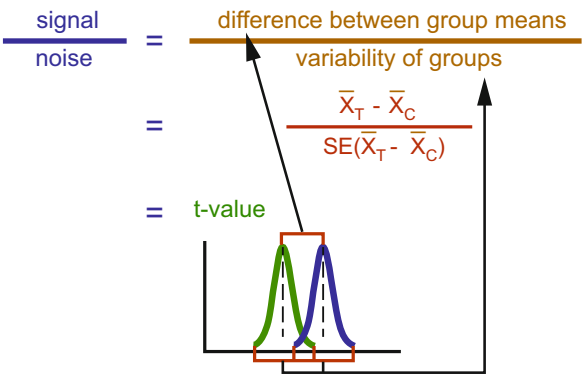

Fig. 7.3 Formula/logic of a *t*-test

Statistical Analysis of the *t*-Test

The difference between the means is the signal, and the bottom part of the formula is the noise, or a measure of variability; the smaller there are differences in the signal and the larger the variability, the harder it is to see the group differences. The logic of a *t*-test can be summarized as follows (see Fig. 7.3):

The top part of the formula is easy to compute—just find the difference between the means. The bottom is a bit more complex; it is called the **standard error of the difference**. To compute it, we have to take the variance for each group and divide it by the number of people in that group. We add these two values and then take their square root. The specific formula is as follows:

$$\mathrm{SE}\left(\bar{X}_T - \bar{X}_C\right) = \sqrt{\frac{\mathrm{Var}_T}{n_T} + \frac{\mathrm{Var}_C}{n_C}}$$

The final formula for the *t*-test is the following:

$$t = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{\text{Var}_T}{n_T} + \frac{\text{Var}_C}{n_C}}}$$

The *t*-value will be positive if the first mean is larger than the second one and negative if it is smaller. However, for our analysis this does not matter. What matters more is the size of the *t*-value. Intuitively, we can say that the larger the *t*-value the higher the chance that two groups are statistically different. A high *T*-value is triggered by a considerable difference between the two group means and low variability of the data around the two group means. To statistically determine whether the *t*-value is large enough to conclude that the two groups are statistically different, we need to use a test of significance. A test of significance sets the amount of error, called the alpha level, which we allow our statistical calculation to have. In most social research, the "rule of thumb" is to set the alpha level at 0.05. This means that we allow 5% error. In other words, we want to be 95% certain that a given relationship exists. This implies that, if we were to take 100 samples from the same population, we could get a significant *T*-value in 95 out of 100 cases.

As you can see from the formula, doing a *t*-test by hand can be rather complex. Therefore, we have SPSS or Stata to do the work for us.

7.1.1 Doing an Independent Samples *t*-Test in SPSS

Step 1: Pre-test—Create a histogram to detect whether the dependent variable—money spent partying—is normally distributed (see Sect. 6.8). Despite the outlier (200 $/month), the data is approximately normally distributed, and we can proceed with the independent samples *t*-test (see Fig. 7.4).

Step 2: Go to Analyze—Compare Means—Independent Samples *T*-Test (see Fig. 7.5).

Step 2: Put your continuous variable as test variable and your dichotomous variable as grouping variable. In the example that follows, we use our dependent variable—money spent partying from our sample dataset—as the test variable. As the grouping variable, we use the only dichotomous variable in our dataset—gender. After dragging over gender to the grouping field, click on Define Groups and label the grouping variable 1 and 0. Click okay (see Fig. 7.6).

Step 3: Verifying the equal variance assumption—before we conduct and interpret the *t*-test, we have to verify whether the assumption of equal variance is met. Columns 1 and 2 in Table 7.1 display the Levene test for equal variances, which measures whether the variances or spread of the data is similar between the two groups (in our case between guys and girls). If the *f*-value is not significant (i.e., the significance level in the second column is larger than 0.05), we do not violate the assumption of equal variances. In this case, it does not matter whether we interpret the upper or the lower row of the output table. However, in our case the significance value in the second column is below 0.05 ($p = 0.018$). This implies that the assumption of equal variances is violated. Yet, this is not dramatic for

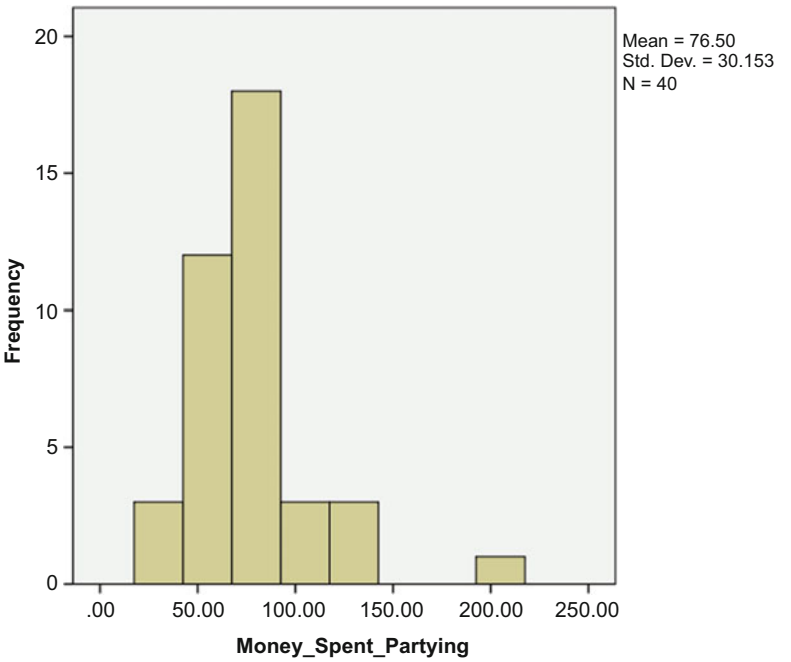

Fig. 7.4 Histogram of the variable money spent partying

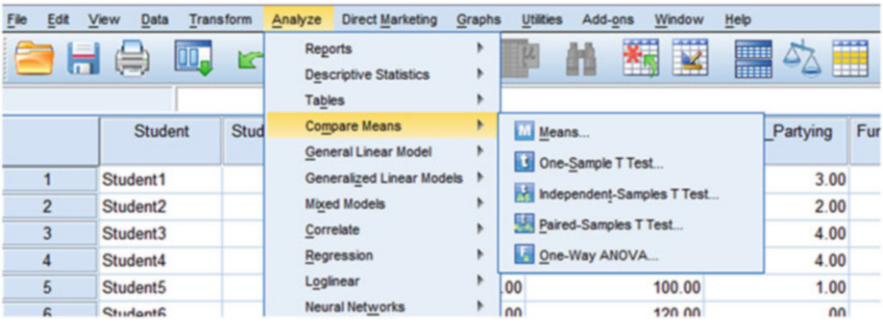

Fig. 7.5 Independent samples *t*-test in SPSS (second step)

interpreting the output, as SPSS offers us an adapted *t*-test, which relaxes the assumption of equal variances. This implies that, in order to interpret the *t*-test, we have to use the second row (i.e., the row labeled equal variances not assumed). In our case, it is the outlier that skews the variance, in particular in the girls' group.

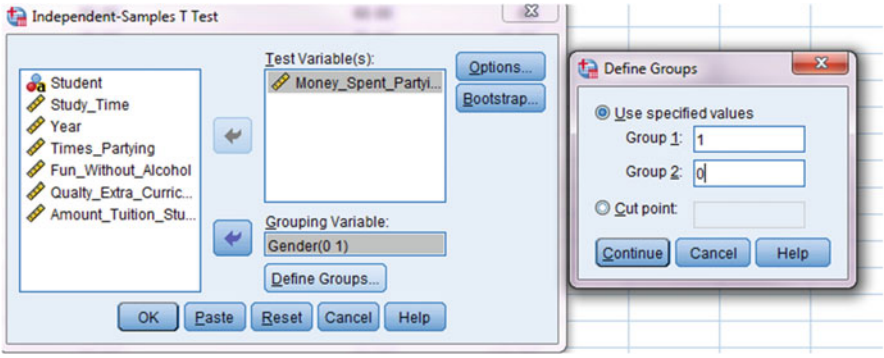

Fig. 7.6 Independent samples *t*-test in SPSS (third step)

Table 7.1 SPSS output of an independent samples *t*-test

T-Test

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Money_Spent_Partying	1.00	21	79.2857	39.18819	8.55156
	.00	19	73.4211	15.63939	3.58792

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
Money_Spent_Partying	Equal variances assumed	6.156	.018	.609	38	.546	5.86466	9.62520	-13.62055	25.34987
	Equal variances not assumed			.632	26.740	.533	5.86466	9.27375	-13.17215	24.90147

The significance level determines the alpha level. In our case, the alpha level is superior to 0.05. Hence, we would conclude that the two groups are not different enough to conclude with 95% certainty that there is a difference

7.1.2 Interpreting an Independent Samples *t*-Test SPSS Output

Having tested the data for normality and equal variances, we can now interpret the *t*-test. The *t*-test output provided by SPSS has two components (see Table 7.1): one summary table and one independent samples *t*-test table. The summary table gives us

the mean amount of money that girls and guys spent partying. We find that girls (who were coded 1) spend slightly more money when they go out and party compared to guys (which we coded 0). Yet, the difference is rather moderate. On average, girls merely spend 6 dollars more per week than guys. If we further look at the standard deviation, we see that it is rather large, especially for group 1 featuring girls. Yet, this large standard deviation is expected and at least partially triggered by the outlier. Based on these observations, we can take the educated guess that there is, in fact, no significant difference between the two groups. In order to confirm or disprove this conjecture, we have to look at the second output in Table 7.1, in particular the fifth column of the second table (which is the most important field to interpret a *t*-test). It displays the significance or alpha level of the independent samples *t*-test. Assuming that we take the 0.05 benchmark, we cannot reject the null hypothesis with 95% certainty. Hence, we can conclude that there is no statistically significant difference between the two groups.

7.1.3 Reading an SPSS Independent Samples *t*-Test Output Column by Column

Column 3 displays the actual *t*-value. (Large *t*-values normally trigger a difference in the two groups, whereas small *t*-values indicate that the two groups are similar.)

Column 4 displays what is called degrees of freedom (*df*) in statistical language. The degrees of freedom are important for the determination of the significance level in the statistical calculation. For interpretation purposes, they are less important. In short, the *df* are the number of observations that are free to vary. In our case, we have a sample of 40 and we have 2 groups, girls and guys. In order to conduct a *t*-test, we must have at least one girl and one guy in our sample, these two parameters are fixed. The remaining 38 people can then be either guys or girls. This means that we have 2 fixed parameters and 38 free-flying parameters or *df*.

Column 5 displays the significance or alpha level. The significance or alpha level is the most important sample statistic in our interpretation of the *t*-test; it gives us a level certainty about our relationship. We normally use the 95% certainty level in our interpretation of statistics. Hence, we allow 5% error (i.e., a significance level of 0.05). In our example, the significance level is 0.103, which is higher than 0.05. Therefore, we cannot reject the null hypothesis and hence cannot be sure that girls spend more than guys.

Column 6 displays the difference in means between the two groups (i.e., in our example this is the difference in the average amount spent partying between girls and guys, which is 5.86). The difference in means is also the numerator of the *t*-test formula.

Column 7 displays the denominator of the *t*-test, which is the standard error of the difference of the two groups. If we divide the value in column 6 by the value in column 7, we get the *t*-statistic (i.e., $5.86/9.27 = 0.632$).

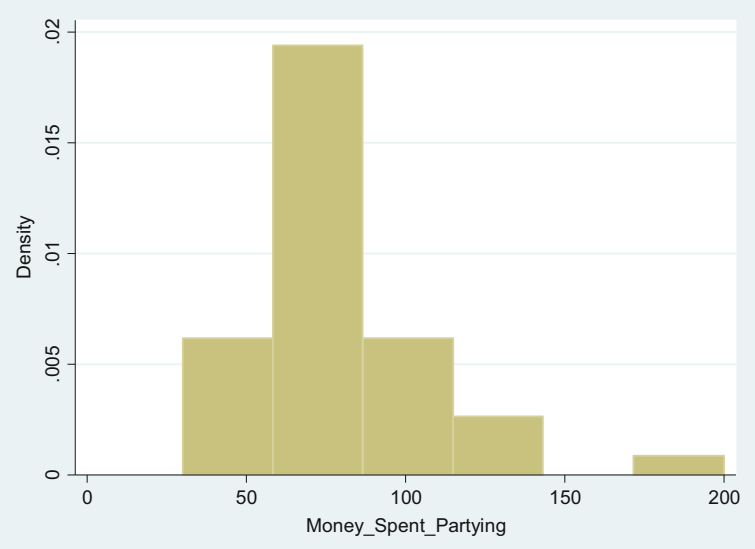

Fig. 7.7 Stata histogram of the variable money spent partying

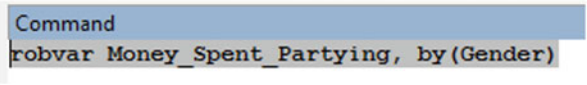

Fig. 7.8 Levene test of equal variances

Column 8 This final split column gives the confidence interval of the difference between the two groups. Assuming that this sample was randomly taken, we could be confident that the real difference between girls and guys lies between –0.321 and 3.554. Again, these two values confirm that we cannot reject the null hypothesis, because the value 0 is part of the confidence interval.

7.1.4 Doing an Independent Samples *t*-Test in Stata

Step 1: Pre-test—Histogram to detect whether the dependent variable—money spent partying per week—is normally distributed. Despite the outlier (200 $/month), the data is approximately normally distributed, and we can proceed with the independent samples *t*-test (Fig. 7.7).

Step 2: Pre-test—Checking for equal variances—write into the Stata Command field: robvar Money_Spent_Partying, by(Gender) (see Fig. 7.8)—this command will conduct a Levene test of equal variances; if this test turns out to be significant, then the null hypothesis of equal variances must be rejected (to interpret the

Table 7.2 Stata Levene test of equal variances

Gender	Summary of Money_Spent_Partying Mean	Std. Dev.	Freq.
0	73.421053	15.639394	19
1	79.285714	39.188191	21
Total	76.5	30.153454	40

W0 = 6.1560718 df(1, 38) Pr > F = 0.01763701

W50 = 4.4595904 df(1, 38) Pr > F = 0.04133734

W10 = 5.2486258 df(1, 38) Pr > F = 0.02760046

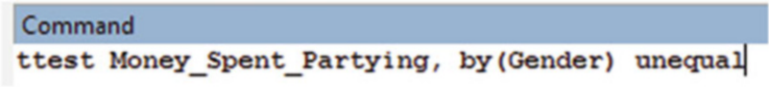

Fig. 7.9 Doing a *t*-test in Stata

Levene test, use the test labeled WO). This is the case in our example (see Table 7.2). The significance level (PR > $F = 0.018$) is below the bar of 0.05.

Step 3: Doing a *t*-test in in Stata—type into the Stata Command field: "ttest Money_Spent_Partying, by(Gender) unequal" (see Fig. 7.9). (Note: if the Levene test for equal variances does not come out significant, you do not need to add equal at the end of the command.)

7.1.5 Interpreting an Independent Samples *t*-Test Stata Output

Having tested the data for normality and equal variances, we can now interpret the *t*-test. The *t*-test output provided by Stata has six columns (see Table 7.3):

Column 1 labels the two groups (in our case group 0 = guys and group 1 = girls).

Column 2 gives the number of observations. In our case, we have 19 guys and 21 girls.

Column 3 displays the mean spending value for the two groups. We find that girls spend slightly more money when they go out and party compared to guys. Yet, the difference is rather moderate. On average, girls merely spend roughly 6 dollars more per week than guys.

Columns 4 and 5 show the standard error and standard deviation, respectively. If we look at both measures, we see that they are rather large, especially for group 1 featuring girls. Yet, this large standard deviation is expected and at least

Table 7.3 Stata independent samples *t*-test output

```
Two-sample t test with unequal variances

   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      19    73.42105    3.587923    15.63939    65.88311    80.959
       1 |      21    79.28571    8.551564    39.18819    61.44746    97.12396
---------+--------------------------------------------------------------------
combined |      40        76.5     4.76768    30.15345    66.85646    86.14354
---------+--------------------------------------------------------------------
    diff |           -5.864662     9.27375               -24.90147    13.17215
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -0.6324
Ho: diff = 0                     Satterthwaite's degrees of freedom =  26.7404

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.2663         Pr(|T| > |t|) = 0.5325          Pr(T > t) = 0.7337
```

The significance level determines the alpha level. In our case, the alpha level is superior to .05. Hence, we would conclude that the two groups are not different enough to conclude with 95 percent certainty that there is a difference

partially triggered by the outlier. (Based on these two observations—the two means are relatively close each other and the standard deviation/standard errors are comparatively large—we can take the educated guess that there is no significant difference in the spending patterns of guys and girls when they party.)

Column 6 presents the 95% confidence interval. It highlights that if these data were randomly drawn from a sample of college students, the real mean would fall between 65.88 dollars per week and 80.96 dollars per week for guys (allowing a certainty level of 95%). For girls, the corresponding confidence interval would be between 61.45 dollars per week and 97.12 dollars per week. Because there is some large overlap between the two confidence intervals, we can already conclude that the two groups are not statistically different from zero.

In order to statistically determine via the appropriate test statistic whether the two groups are different, we have to look at the significance level associated with the *t*-

test (see arrow below). The significance level is 0.53, which is above the 0.05 benchmark. Consequently, we cannot reject the null hypothesis with 95% certainty and can conclude that that there is no statistically significant difference between the two groups.

7.1.6 Reporting the Results of an Independent Samples *t*-Test

In Sect. 4.12 we hypothesized that guys like the bar scene more than girls do and are therefore going to spend more money when they go out and party. The independent samples *t*-test disconfirms this hypothesis. On average, it is actually girls who spend slightly more than guys do. However, the difference in average spending (73 dollars for guys and 79 dollars for girls) is not statistically different from zero ($p = 0.53$). Hence, we cannot reject the null hypothesis and can conclude that the spending pattern for partying is similar for the two genders.

7.2 *F*-Test or One-Way ANOVA

T-tests work great with dummy variables, but sometimes we have categorical variables with more than two categories. In cases where we have a continuous variable paired with an ordinal or nominal variable with more than two categories, we use what is called an *f*-test or one-way ANOVA. The logic behind an *f*-test is similar to the logic for a *t*-test. To highlight, if we compare the two graphs in Fig. 7.10, we would probably conclude that the three groups in the second graph are different, while the three groups in the first graph are rather similar (i.e., in the first graph, there is a lot of overlap, whereas in the second graph, there is no overlap, which entails that each value can only be attributed to one distribution).

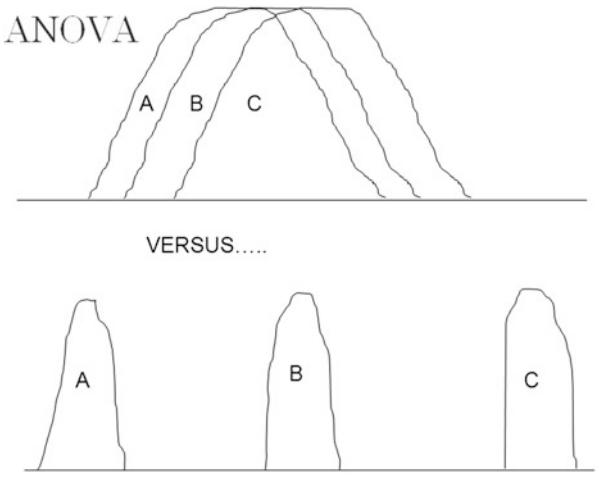

Fig. 7.10 What makes groups different?

Table 7.4 Within and between variation (as in other occassions, make sure to put the table heading and then the table)

	Sample A			Sample B		
•	High	Med	Low	High	Med	Low
	50	40	30	30	28	12
	51	41	31	40	32	18
	52	42	32	55	40	30
	53	43	33	65	50	45
	54	44	34	70	60	55
Mean	**52**	**42**	**32**	**52**	**42**	**32**

While the logic of an *f*-test reflects the logic of a *t*-test, the calculation of several group means and several measures of variability around the group means becomes more complex in an *f*-test. To reduce this complexity, an ANOVA test uses a simple method to determine whether there is a difference between several groups. It splits the total variance into two groups: between variance and within variance. The between variance measures the variation between groups, whereas the within variance measures the variation within groups. Whenever the between variation is considerably larger than the within variation, we can say that there are differences within groups. The following example highlights this logic (see Table 7.4).

Let us assume that Table 7.4 depicts two hypothetical samples, which measure the approval ratings of Chancellor Merkel based on social class. In the survey, an approval score of 0 means that respondents are not at all satisfied with her performance as chancellor. In contrast, 100 signifies that individuals are very satisfied with her performance as chancellor. The first sample consists of young people (i.e., 18–25) and the second sample of old people (65 and older). Both samples are split into three categories—high, medium, and low. High stands for higher or upper classes, med stands for medium or middle classes, and low stands for the lower or working classes. We can see that the mean satisfaction ratings for Chancellor Merkel per social strata do not differ between the two samples; that is, the higher classes, on average, rate her at 52, the middle classes at 42, and the lower classes at 32. However, what differs tremendously between the two samples is the variability of the data. In the first sample, the values are very closely clustered around the mean throughout each of the three categories. We can see that there is much more variability between groups than between observations within one group. Hence, we would conclude that the groups are different. In contrast, in sample 2, there is large within-group variation. That is, the values within each group differ much more than the corresponding values between groups. Therefore, we would predict for sample 2 that the three groups are probably not that different, despite the fact that their means are different. Following this logic, the formula for an ANOVA analysis or *f*-test is **between-group variance/within-group variance**. Since it is too difficult to calculate the between- and within-group variance by hand, we let statistical computer programs do it for us.

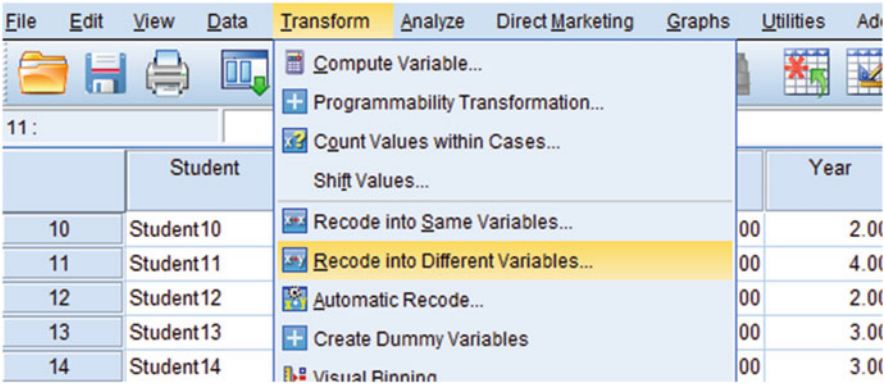

Fig. 7.11 Recoding the variable times partying 1 (first step)

A standard *f*-test or ANOVA analysis illustrates if there are differences between groups or group means, but it does not show which specific group means are different from one another. Yet, in most cases researchers want to know not only that there are some differences but also between which groups the differences lie. So-called multiple comparison tests—basically *t*-tests between the different groups—compare all means against one another and help us detect where the differences lie.

7.2.1 Doing an *f*-Test in SPSS

For our *f*-test we use the variable money spent partying per week as the dependent variable and the categorical variable times partying per week as the factor or grouping variable. Given that we only have 40 observations and given that there should be at least several observations per category to yield valid test results, we will reduce the six categories to three. In more detail, we cluster together no partying and partying once, partying twice and three times, and partying four times and five times and more together. We can create this new variable by hand, or we can also have SPSS do it for us. We will label this variable times partying 1.

Step 1: Creating the variable times partying 1—go to Transform—Recode into Different Variable (see Fig. 7.11).

Step 2: Drag the variable Times_Partying into the middle field—name the Output Variable Times_Partying_1—click on Change—click on Old and New Values (see Fig. 7.12).

Step 3: Include in the Range field the value range that will be clustered together—add the new value in the field labeled New Value—click Add—do this for the all three ranges—once your dialog field looks like the dialog field below click Continue. You will be redirected to the initial screen; click okay and then the new variable will be added to the SPSS dataset (see Fig. 7.13).

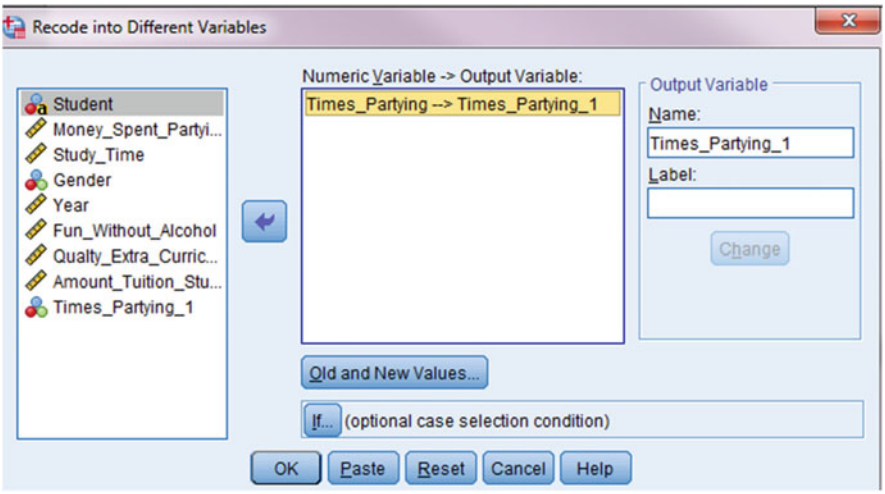

Fig. 7.12 Recoding the variable Times Partying 1 (second step)

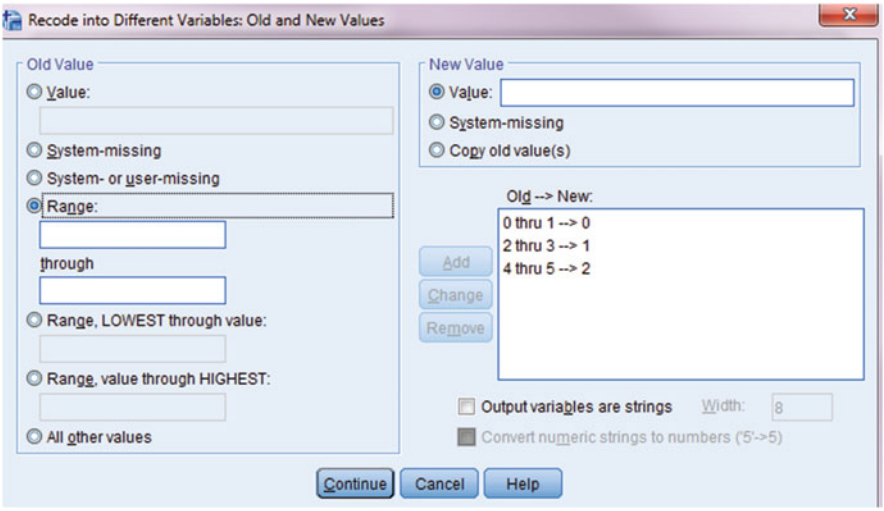

Fig. 7.13 Recoding the variable Times Partying 1 (third step)

Step 4: For doing the actual *f*-test—go to Analyze—Compare Means—One-Way ANOVA (see Fig. 7.14).

Step 5: Put your continuous variable (money spent partying) as Dependent Variable and your ordinal variable (times spent partying 1) as Factor. Click okay (see Fig. 7.15).

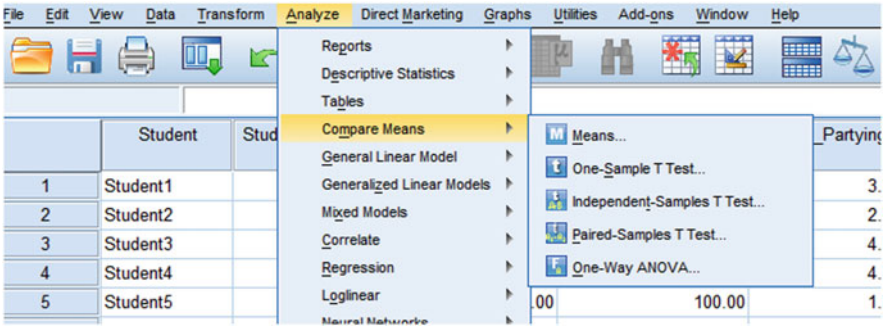

Fig. 7.14 Doing an *f*-test in SPSS (first step)

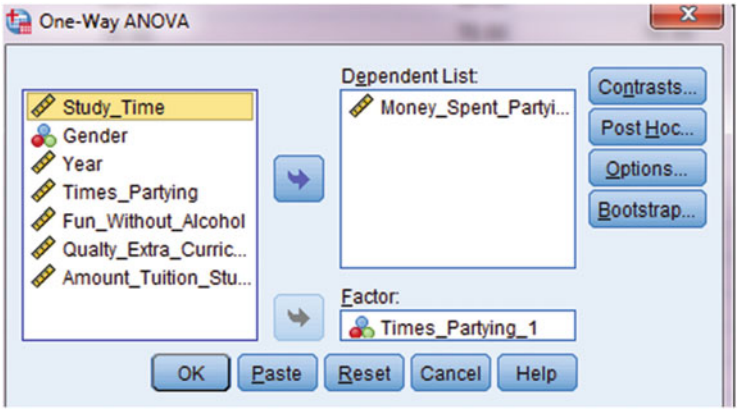

Fig. 7.15 Doing an *f*-test in SPSS (second step)

7.2.2 Interpreting an SPSS ANOVA Output

The SPSS ANOVA output contains five columns (see Table 7.5). Similar to a *t*-test, the most important column is the significance level (i.e., Sig). It tells us whether there is a difference between at least two out of the however many groups we have. In our example, the significance level is 0.000, which means that we can tell with nearly 100% certainty that at least two groups differ in the money they spent partying per week. Yet, the SPSS ANOVA output does not tell us which groups are different; the only thing it tells us is that at least two groups are different from one another.

In more detail, the different columns are interpreted as follows:

Column 1 displays the sum of squares or the squared deviations for the different variance components (i.e., between-group variance, within-group variance, and total variance).

Table 7.5 SPSS ANOVA output

ANOVA

Money_Spent_Partying

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12485.000	2	6242.500	10.053	.000
Within Groups	22975.000	37	620.946		
Total	35460.000	39			

Column 2 displays the degrees of freedom, which allows us to calculate the within and between variance. The formula for these degrees of freedom is number of groups (k) – 1 for the between-group estimator and the number of observations (N) – k for within group.

Column 3 shows the between and within variance or sum of squares. According to the f-test formula, we need to divide the between variance by the within variance ($F = 6242.50/620.95 = 10.053$).

Column 4 displays the f-value. The larger the f-value, the more likely it is that at least two groups are statistically different from one another.

Column 5 gives us an alpha level or level of certainty indicating a probability level that there is a difference between at least two groups. In our case, the significance level is 0.000 indicating that we can be nearly be 100% certain that at least two groups differ in, how much money the spent partying per week.

7.2.3 Doing a Post hoc or Multiple Comparison Test in SPSS

The violation of the equal variance assumption (i.e., the distributions around the group means are different for the various groups in the sample) is rather unproblematic for interpreting a one-way ANOVA analysis (Park 2009). Yet, having an equal or unequal variability around the group means is important when doing multiple pairwise comparison tests between means. This assumption needs to be tested before we can do the test:

Step 1: Testing the equal variance assumption—go to the One-Way ANOVA Screen (see step 5 in Sect. 7.2.1)—click on Options—a new screen with options opens—click on Homogeneity of Variance Test (see Fig. 7.16).

Step 2: Press Continue—you will be redirected to the previous screen—press Okay (see Fig. 7.17).

The equality of means test provides a significant result (Sig = 0.024) indicating that the variation around the three group means in our sample is not equal. We have to take this inequality of variance in the three distributions into consideration when conducting the multiple comparison test (see Table 7.6).

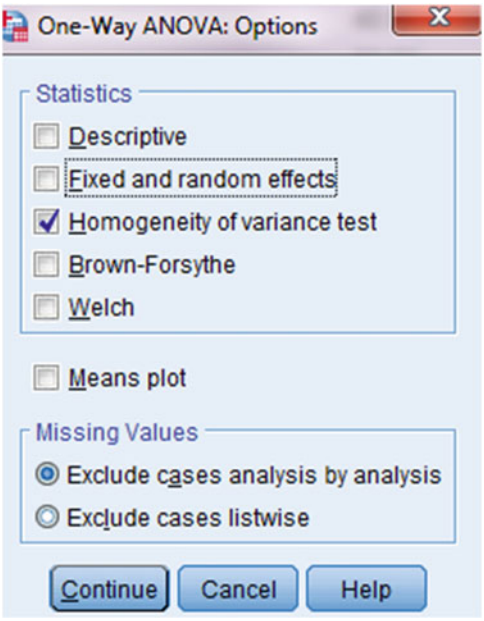

Fig. 7.16 Doing a post hoc multiple comparison test in SPSS (first step)

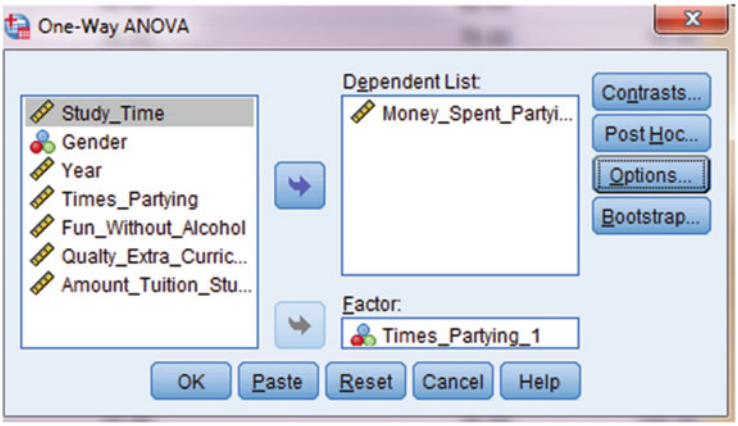

Fig. 7.17 Doing a post hoc multiple comparison test in SPSS (second step)

Step 3: Conducting the multiple comparison test—go to the One-Way ANOVA Command window (see Step 2)—click on the button Post Hoc and choose any of the four options under the label Equal Variances Not Assumed. Please note if your equality of variances test did not yield a statistically significant result (i.e. the sig value is not smaller than 0.05) you should choose any of the options under the label Equal Variances Assumed) (see Fig. 7.18).

Table 7.6 Robust test of equality of means

Test of Homogeneity of Variances

Money_Spent_Partying

Levene Statistic	df1	df2	Sig.
4.122	2	37	.024

ANOVA

Money_Spent_Partying

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12485.000	2	6242.500	10.053	.000
Within Groups	22975.000	37	620.946		
Total	35460.000	39			

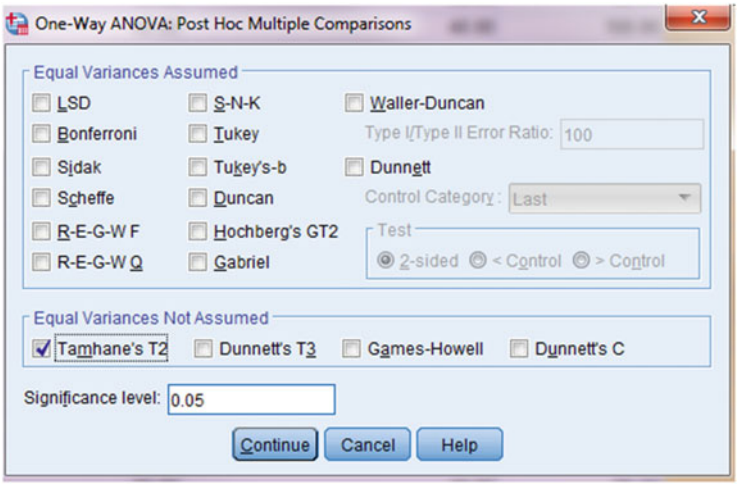

Fig. 7.18 Doing a post hoc multiple comparison test in SPSS (third step)

The post hoc multiple comparison test (see Table 7.7) allows us to decipher which groups are actually statistically different from one another. From the SPSS output, we see that individuals who either do not go out not at all or go out once per week (group 0) spend less than individuals who go out four times or more (group 2). The average mean difference is 43 dollars per week, this difference is statistically different from zero ($p = 0.032$). We also see from the SPSS output that individuals who go out and party two or three times (group 1) spend significantly less than individuals who party four or more times (group 2). In absolute terms, they are predicted to spend 39.5 dollars less per week. This difference is statistically different from zero ($p = 0.041$). In contrast, there is no statistical difference in the spending

Table 7.7 SPSS output of a post hoc multiple comparison test

Post Hoc Tests

Multiple Comparisons

Dependent Variable: Money_Spent_Partying

Tamhane

(I) Times_Partying_1	(J) Times_Partying_1	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
.00	1.00	-3.50000	7.43087	.955	-23.7758	16.7758
	2.00	-43.00000*	14.54877	.032	-82.6016	-3.3984
1.00	.00	3.50000	7.43087	.955	-16.7758	23.7758
	2.00	-39.50000*	13.30815	.041	-77.4683	-1.5317
2.00	.00	43.00000*	14.54877	.032	3.3984	82.6016
	1.00	39.50000*	13.30815	.041	1.5317	77.4683

*. The mean difference is significant at the 0.05 level.

patterns between groups 0 and 1. In absolute terms, there is a mere 3.5 dollars difference between the two groups. This difference is not statistically different from zero ($p = 0.955$).

7.2.4 Doing an *f*-Test in Stata

For our *f*-test we use money spent partying per week as the dependent variable, and as the independent variable, we use the categorical variable times partying per week. Given that we only have 40 observations and given that there should be at least several observations per category to yield valid test results, we reduce the six categories to three. In more detail, we cluster together no partying and partying once, partying twice and three times, and partying four times and five times or more. We can create this new variable by hand, or we can also have Stata do it for us. We will label this variable times partying 1.

Preliminary Procedure: Creating the variable times partying 1

Write in the Stata Command editor:

(1) generate Times_Partying_1 = 0
(this creates the variable and assigns the value of 0 to all observations) (see Fig. 7.19)

(2) replace Times_Partying_1 = 1 if(Times_Partying $\geq$ 2) and Times_Partying $\leq$ 3
(this assigns the value of 1 to all individuals in groups 2 and 3 for the variable times partying) (see Fig. 7.20).

(3) replace Times_Partying_1 = 2 if(Times_Partying $\geq$ 4) (see Fig. 7.21)
(this assigns the value of 2 to all individuals in groups 4 and 5).

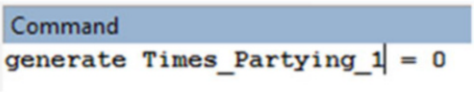

Fig. 7.19 Generating the variable Times_Partying 1 (first step)

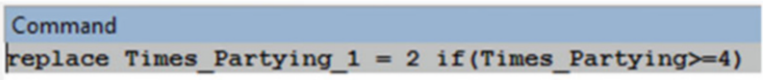

Fig. 7.20 Generating the variable Times_Partying 1 (second step)

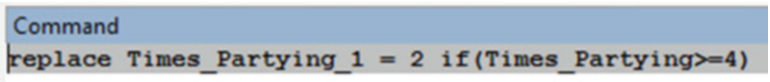

Fig. 7.21 Generating the variable Times_Partying 1 (third step)

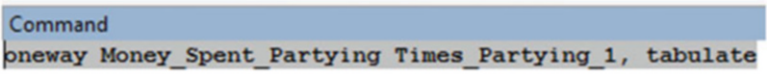

Fig. 7.22 Doing an ANOVA analysis in Stata

Main analysis: conducting the ANOVA

Write in the Stata Command editor: oneway Money_Spent_Partying Times_Partying_1, tabulate (see Fig. 7.22).

7.2.5 Interpreting an *f*-Test in Stata

In the first part of the table, Stata provides summary statistics of the three group means (Table 7.8). The output reports that individuals who go out partying once a week or less spend on average 64 dollars per week. Those who party two to three times per week merely spend 3.5 dollars more than the first group. In contrast, individuals in the third group of students, who party four or more times, spend approximately 107 dollars per week partying. We also see that the standard deviations are quite large, especially for the third group—students that party four or more times per week. These descriptive statistics also give us a hint that there is probably a statistically significant difference between the third group and the two other groups. The *f*-test (Prob $>$ $F = 0.0003$) further reveals that there are in fact differences between groups. However, the *f*-test does not allow us to detect where the differences lie. In order to gauge this, we have to conduct a multiple comparison test. However, before doing so we have to gauge whether the variances of the three

Table 7.8 Stata ANOVA output this table is misplaced here put below the text in 7.2.5

Times_Partying_1	Summary of Money_Spent_Partying Mean	Std. Dev.	Freq.
0	64	21.186998	10
1	67.5	14.372855	20
2	107	40.838435	10
Total	76.5	30.153454	40

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	12485	2	6242.5	10.05	0.0003
Within groups	22975	37	620.945946		
Total	35460	39	909.230769		

Bartlett's test for equal variances: chi2(2) = 14.3463 Prob>chi2 = 0.001

groups are equal or if they are dissimilar. The Bartlett's test for equal variance, which is listed below the ANOVA output, gives us this information. If the test statistic provides a statistically significant value, then we have to reject the null hypothesis of equal variances and accept the alternative hypothesis that the variances between groups are unequal. In our case, the Bartlett's test displays a statistically significant value (prob $>$ chi2 $=$ 0.001). Hence, we have to proceed with a post-stratification test with unequal variance.

7.2.6 Doing a Post hoc or Multiple Comparison Test with Unequal Variance in Stata

Since the Bartlett's test indicates some violation of the equal variance assumption (i.e., the distributions around the group means are different for the various groups in the sample), we have to conduct a multiple comparison test in which the unequal variance assumption is relaxed. A Stata module to compute pairwise multiple comparisons with unequal variances exists, but is not included by default and must be downloaded. To do the test we have follow a multistep procedure:

Step 1: Write in the Command field: search pwmc (see Fig. 7.23).
This brings us to a Stata page with several links—click on the first link and click on "(Click here to install)" to download the program (Once, downloaded, the program is installed and does not need to be downloaded again) (see Fig. 7.24).
Step 2: Write into the Command field: pwmc Money_Spent_Partying, over (Times_Partying_1) (Fig. 7.25).

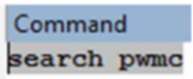

Fig. 7.23 Downloading a post hoc or multiple comparison test with unequal variance

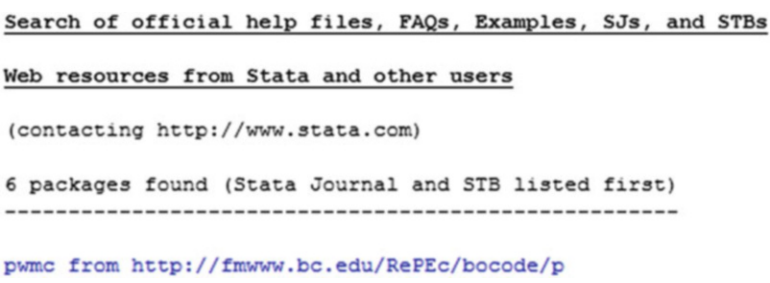

Fig. 7.24 Download page for the post hoc multiple comparison test

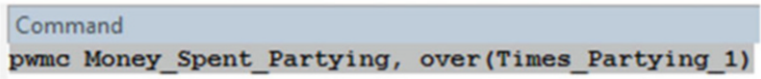

Fig. 7.25 Doing a post hoc or multiple comparison test with unequal variance in Stata

Stata actually provides test results of three different multiple comparison tests, all using some slightly different algebra (see Table 7.9). Regardless of the test, what we can see is that, substantively, the results of the three tests do not differ. These tests are also slightly more difficult to interpret because they do not display any significance level. Rather, we have to interpret the confidence interval to determine whether two means are statistically different from zero. We find that there is no statistically significant difference between groups 0 and 1. For example, if we look at the first of the three tests, we find that the confidence interval includes positive and negative values; it ranges from –16.90 to 23.90. Hence, we cannot accept the alternative hypothesis that groups 0 and 1 are different from one another. In contrast, we can conclude that groups 0 and 2, as well as 1 and 2, are statistically different from one another, respectively, because both confidence intervals include only positive values (i.e., the confidence interval for the difference between groups 0 and 2 ranges from 2.38 to 83.62 and the confidence interval for the difference between groups 1 and 2 ranges from 2.54 to 76.46).

Please also note that in case the Bartlett's test of equal variance (see Table 6.3) does not display any significance, a multiple comparison test assuming equal variance can be used. Stata has many options; the most prominent ones are probably the algorithms by Scheffe and Sidak. For presentation purposes, let us assume that the Bartlett test was not significant in Table 6.3.

In such a case, we would type into the Stata Command field:

Table 7.9 Stata output of post hoc multiple comparison test

Pairwise comparisons of means (unequal variances)

Money_Spent_Partying	Diff.	Std.Err	Dunnett's C [95% Conf.	Interval]
Times_Partying_1				
1 vs 0	3.5	7.43087	-16.89737	23.89737
2 vs 0	43	14.54877	2.379758	83.62024
2 vs 1	39.5	13.30815	2.538827	76.46117

Money_Spent_Partying	Diff.	Std.Err	Games and Howell [95% Conf.	Interval]
Times_Partying_1				
1 vs 0	3.5	7.43087	-16.06879	23.06879
2 vs 0	43	14.54877	4.766046	81.23395
2 vs 1	39.5	13.30815	3.095694	75.90431

Money_Spent_Partying	Diff.	Std.Err	Tamhane's T2 [95% Conf.	Interval]
Times_Partying_1				
1 vs 0	3.5	7.43087	-16.77576	23.77576
2 vs 0	43	14.54877	3.398387	82.60161
2 vs 1	39.5	13.30815	1.531734	77.46827

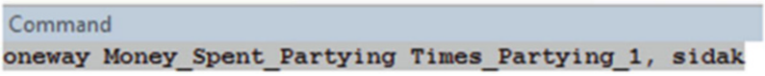

Fig. 7.26 Multiple comparison test according to Sidak

oneway Money_Spent_Partying Times_Partying_1, sidak (see Fig. 7.26).

To determine whether there is a significant difference between groups, we would interpret the two-by-two table (see the second part of Table 7.10). The table highlights that the mean difference between group 0 and group 1 is 3.5 dollars, a value that is not statistically different from 0 (Sig = 0.978). In contrast the difference in spending between group 0 and group 2, which is 43 dollars, is statistically significant (sig = 0.001). The same applies to the difference in spending (39.5 dollars) between groups 1 and 2 (sig = 0.001).

Table 7.10 Multiple comparison test with equal variance in Stata

```
                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups           12485       2        6242.5     10.05     0.0003
 Within groups           22975      37    620.945946
------------------------------------------------------------------------
    Total                35460      39    909.230769

Bartlett's test for equal variances:  chi2(2) =  14.3463  Prob>chi2 = 0.001
```

Comparison of Money_Spen~g by Times_Part~1
(Sidak)

Row Mean-Col Mean	0	1
1	3.5	
	0.978	
2	43	39.5
	0.001	0.001

7.2.7 Reporting the Results of an *f*-Test

In Sect. 4.12, we hypothesized that individuals who party more frequently will spend more money for their weekly partying habits than individuals that party less frequently. Creating three groups of party goers—(1) students, who party once or less, on average, (2) students who party between two and three times, and (3) students who party more than four times—we find some support for our hypothesis; that is, a general *f*-test confirms (Sig = 0.0003) that there are differences between groups. Yet, the *f*-test cannot tell us between which groups the differences lie. In order to find this out, we must compute a post hoc multiple comparison test. We do so assuming unequal variances between the three distributions, because a Bartlett's test of equal variance (sig = 0.001) reveals that the null hypothesis (get rid of the plural, I cannot delete the word hypotheses) hypotheses of equal variances must be rejected. Our results indicate that the mean spending average statistically significantly differs between groups 0 and 2 [i.e., the average difference in party spending is 43 dollars (sig = 0.032)]. The same applies to the difference between groups 1 and 2 [i.e., the average difference in party spending is 39.5 dollars (Sig = 0.041)]. In contrast, there is no difference in spending between those who party once or less and those who party two or three times (sig 0.955).

7.3 Cross-tabulation Table and Chi-Square Test

7.3.1 Cross-tabulation Table

So far, we have discussed bivariate tests that work with a categorical variable as the independent variable and a continuous variable as the dependent variable (i.e., an independent samples *t*-test if the independent variable is binary and the dependent variable continuous and an ANOVA or *f*-test if the independent variable has more than two categories). What happens if both the independent and dependent variables are binary? In this case, we can present the data in a crosstab.

Table 7.11 provides an example of a two-by-two table. The table presents the results of a lab experiment with mice. A researcher has 105 mice with a severe illness; she treats 50 mice with a new drug and does not treat 55 mice at all. She wants to know whether this new drug can cure animals. To do so she creates four categories: (1) treated and dead, (2) treated and alive, (3) not treated and dead, and (4) not treated and alive.

Based on this two-by-two table, we can ask the question: How many of the dead are either treated or not treated? To answer this question, we have to use the column as unit of analysis and calculate the percentage of dead, which are treated and the percentage of dead, which are not treated. To do so, we have to convert the column raw numbers into percentages. To calculate these percentages for the first field, we take the number in the field—treated/dead—and divide it by the column total (36/66 = 55.55%). We do analogously for the other fields (see Table 7.12). Interpreting Table 6.7, we can find, for example, that roughly 56% of the dead mice have been treated, but only 35.9% of those alive have undergone treatment.

Instead of asking the question how many of the dead mice have been treated, we might change the question and ask how many of the dead have been treated or alive? To get at this information, we first calculate the percentage of dead mice with treatment and of alive mice with treatment. We do analogously for the dead that are not treated and the alive that are not treated. Doing so, we find that of all treated, 72% are dead and only 28% are alive. In contrast, the not treated mice have a higher chance of survival. Only 55.55% have died and 44.45% have survived (see Table 7.13).

Table 7.11 Two-by-two table measuring the relationship between drug treatment and the survival of animals

	Dead	Alive
Treated	36	14
Not treated	30	25
Total	66	39

Table 7.12 Two-by-two table focusing on the interpretation of the columns

	Dead	Alive
Treated	36	14
Percent	55.55	35.90
Not treated	30	25
Percent	44.45	64.10
Total	66	39
Percent	100	100

Table 7.13 Two-by-two table focusing on the interpretation of the rows

	Dead	Alive	Total
Treated	36	14	50
Percent	72.00	28.00	100
Not treated	30	25	55
Total	55.55	44.45	100

Table 7.14 Calculating the expected values

	Dead	Alive	Total
Treated	36 (31.4)	14 (18.6)	50
Not treated	30 (34.6)	25 (20.4)	55
Total	66	39	105

7.3.2 Chi-Square Test

Having interpreted the crosstabs a researcher might ask whether there is a relationship between drug treatment and the survival of mice. To answer this research question, she might postulate the following hypothesis:

H_0: The survival of the animals is independent of drug treatment.
H_a: The survival of the animals is higher with drug treatment.

In order to determine if there is a relationship between drug treatment and the survival of mice, we use what is called a chi-square test. To apply this test, we compare the actual value in each field with a random distribution of values between the four fields. In other words, we compare the real value in each field with an expected value, calculated so that the chance that a mouse falls in each of the four fields is the same.

To calculate this expected value, we use the following formula:

$$\frac{\text{Row Total} \times \text{Column Total}}{\text{Table Total}}$$

Table 7.14 displays the observed and the expected values for the four possible categories: (1) treated and dead, (2) treated and alive, (3) not-treated and dead, and

(4) not-treated and alive. The logic behind a chi-square test is that the larger the gap between one or several observed and expected values, the higher the chance that there actually is a pattern or relationship in the data (i.e., that treatment has an influence on survival).

The formula for a chi-square test is:

$$X^2 = \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

In more detail the four steps to calculate a chi-square value are the following:

(1) For each observed value in the table, subtract the corresponding expected value (O–E).
(2) Square the difference $(O–E)^2$.
(3) Divide the squares obtained for each cell in the table by the expected number for that cell
    $(O–E)^2/E$.
(4) Sum all the values for $(O–E)^2/E$. This is the chi-square statistic.

Our treated animal example gives us a chi-square value of 3.418. This value is not high enough to reject the null hypothesis of a no effect between treatment and survival. Please note that in earlier times, researchers compared their chi-square test value with a critical value, a value that marked the 5% alpha level (i.e., a 95% degree of certainty). If their test value fell below this critical value, then the researcher could conclude that there is no difference between the groups, and if it was above, then the researcher could conclude that there is a pattern in the data. In modern times, statistical outputs display the significance level associated with a chi-square value right away, thus allowing the researcher to determine right away if there is a relationship between the two categorical variables.

A chi-square test works with the following limitations:

- No expected cell count can be less than 5.
- Larger samples are more likely to trigger statistically significant results.
- The test only identifies *that* a difference exists, not necessarily *where* it exists. (If we want to decipher where the differences are, we have look at the data and detect in what cells are the largest differences between observed and expected values. These differences are the drivers of a high chi-square value.)

7.3.3 Doing a Chi-Square Test in SPSS

The dependent variable of our study, which we used for the previous tests (i.e., *t*-test and *f*-test)—money spent partying—is continuous, and we cannot use it for our chi-square test. We have to use two categorical variables and make sure that there are

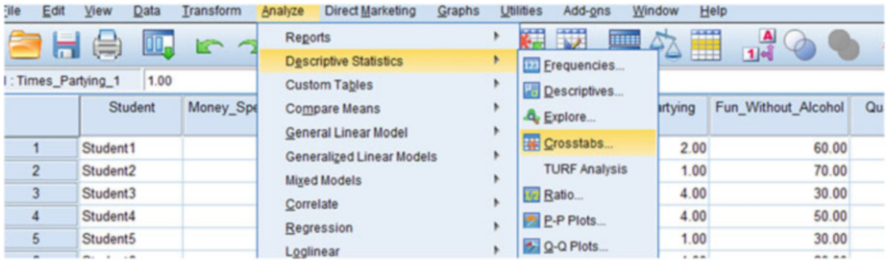

Fig. 7.27 Doing crosstabs in SPSS (first step)

least five observations in each cell. The two categorical variables we choose are gender and times partying per week. Since, we have only 40 observations, we further contract the variable times partying and create 2 categories: (1) partying a lot (three times or more a week) and partying moderately (less than 3 times a week). We name this new variable times partying 2. To create this new variable, see Sect. 7.1.2.

Step 1: Go to Analyze—Descriptive Statistics—Crosstabs (see Fig. 7.27).
Step 2: Put Gender in the Row(s) field and Times Partying 2 in the Columns field. Click on statistics (see Fig. 7.28).
Step 3: In the Crosstabs: Statistics field, check the box Chi-square—click continue—you will be redirected to the previous box—click okay (see Fig. 7.29).

7.3.4 Interpreting an SPSS Chi-Square Test

An SPSS chi-square test output consists of two parts: a cross-tabulation table and the actual chi-square test (see Table 7.15). The crosstab indicates that the sample consists of 21 guys and 19 girls. Within the 2 genders, partying habits are quite equally distributed; 9 guys party 2 times or less per week, and 12 guys party 3 times or more. For girls, the numbers are rather similar—ten girls party twice or less per week, on average, and nine girls three times or more. We already see from the chi-square table that there is hardly any difference between guys and girls. The Pearson chi-square test (i.e., first row of the chi-square test table) confirms this observation. The chi-square value is 0.382 and the corresponding significance level is 0.536, which is far above the benchmark of 0.05. Based on these test-statistics, we cannot reject the null hypothesis; this implies that we can conclude that the partying habits of guys and girls (in terms of the times per week both genders party) do not differ.

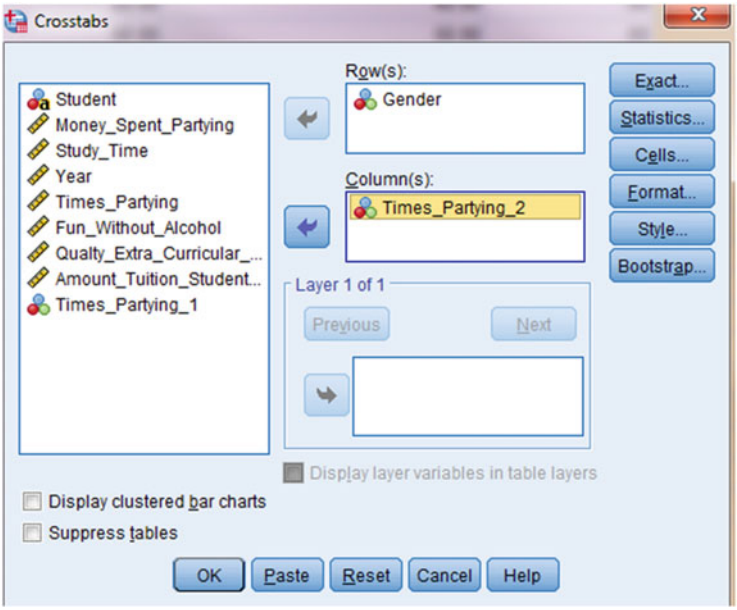

Fig. 7.28 Doing crosstabs in SPSS (second step)

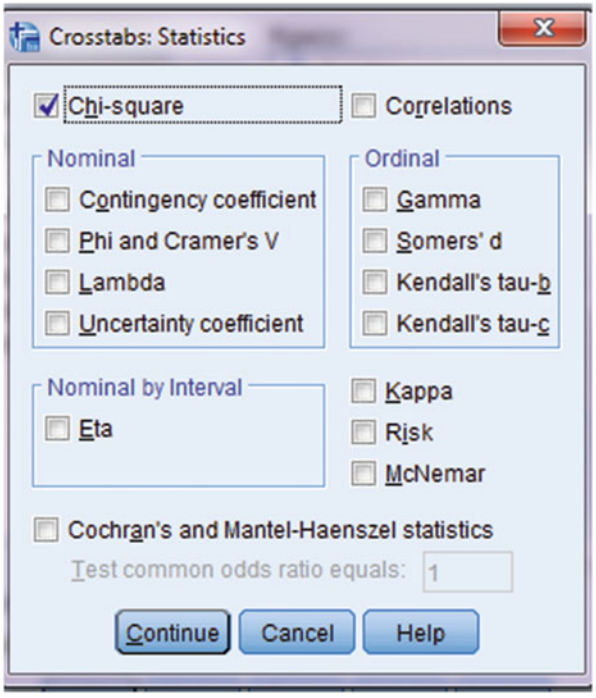

Fig. 7.29 Doing crosstabs in SPSS (third step)

Table 7.15 SPSS chi-square test output

Gender * Times_Partying_2 Crosstabulation

Count

		Times_Partying_2		Total
		.00	1.00	
Gender	.00	9	10	19
	1.00	12	9	21
Total		21	19	40

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.382[a]	1	.536		
Continuity Correction[b]	.091	1	.763		
Likelihood Ratio	.383	1	.536		
Fisher's Exact Test				.752	.382
Linear-by-Linear Association	.373	1	.542		
N of Valid Cases	40				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 9.03.

b. Computed only for a 2x2 table

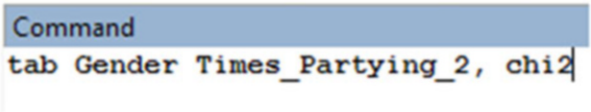

Fig. 7.30 Doing a chi-square test in Stata

7.3.5 Doing a Chi-Square Test in Stata

To do a chi-square test, we cannot use the dependent variable of our study—money spent partying—because it is continuous. For a chi-square test to function, we have to use two categorical variables and make sure that there are at least five observations in each cell. The two categorical variables we choose are gender and times partying per week. Since we have only 40 observations, we further collapse the variable times partying and create 2 categories: (1) partying a lot (3 times or more a week) and (2) partying moderately (less than 3 times a week). We name this new variable times partying 2. To create this new variable, see Sect. 8.1.3.

Step 1: Write into the Stata Command field: tab Gender Times_Partying_2 (see Fig. 7.30)

(the order in which we write our two categorical variables does not matter).

Table 7.16 Stata chi-square output

```
tab Gender Times_Partying_2, chi2
```

Gender	Times_Partying_2 0	1	Total
0	9	10	19
1	12	9	21
Total	21	19	40

Pearson chi2(1) = 0.3822 Pr = 0.536

The Stata test output in Table 7.16 consists of a cross-tabulation table and the actual chi-square test below. The crosstab indicates that the sample consists of 21 guys and 19 girls. Within the 2 genders, partying habits are quite equally distributed; 9 guys party 2 times or less per week, and 12 guys party 3 times or more. For girls, the numbers are rather similar: ten girls party twice or less per week, on average, and nine girls three times or more. We already see from the chi-square table that there is hardly any difference between guys and girls. The Pearson chi-square test result below the cross-table confirms this observation. The chi-square value is 0.382, and the corresponding significance level is 0.536, which is far above the benchmark of 0.05. Based on these test statistics, we can conclude that the partying habits of guys and girls (in terms of the times per week both genders party) do not differ.

7.3.6 Reporting a Chi-Square Test Result

Using a chi-square test, we have tried to detect if there is a relationship between gender and the times per week students party. We find that in absolute terms roughly about half the girls and guys, respectively, either party two times or less or three times or more. The chi-square test confirms that there is no statistically significant difference in the number of times either of two genders goes out to party (i.e., the chi-square value is 0.38 and the corresponding significance level is 0.534).

Reference

Park, H. (2009). *Comparing group means: T-tests and one-way ANOVA using Stata, SAS, R, and SPSS*. Working paper, The University Information Technology Services (UITS), Center for Statistical and Mathematical Computing, Indiana University.

Further Reading

Statistics Textbooks

Basically every introductory to statistics book covers bivariate statistics between categorical and continuous variables. The books I list here are just a short selection of possible textbooks. I have chosen these books because they are accessible and approachable and they do not use math excessively.

Brians, C. L. (2016). *Empirical political analysis: Pearson new international edition coursesmart etextbook*. London: Routledge (chapter 11). Provides a concise introduction into different types of means testing.

Macfie, B. P., & Nufrio, P. M. (2017). *Applied statistics for public policy*. New York: Routledge. This practical text provides students with the statistical tools needed to analyze data. It also shows through several examples how statistics can be used as a tool in making informed, intelligent policy decisions (part 2).

Walsh, A., & Ollenburger, J. C. (2001). *Essential statistics for the social and behavioral sciences: A conceptual approach*. Upper Saddle River: Prentice Hall (chapters 7–11). These chapters explain in rather simple forms the logic behind different types of statistical tests between categorical variables and provide real life examples.

Presenting Results in Publications

Morgan, S., Reichert, T., & Harrison, T. R. (2016). *From numbers to words: Reporting statistical results for the social sciences*. London: Routledge. This book complements introductory to statistics books. It shows scholars how they can present their test results in either visual or text form in an article or scholarly book