5

Types of Analyses

As psychological research has been expanding it has become increasingly specialized, carving out ever more narrow topic areas with the result that findings are often decontextualized from each other and from lived experience. The particular research design, the experimental condition and instrumentation used, may generate interesting results but without much "ecological validity"—in other words, they often only apply within the conditions of the study and not beyond.

One of the unfortunate consequences of this trend is that specialization extends to research methodologies. At the most basic level, most research-ers would recognize the "qualitative" versus "quantitative" distinction and would readily place themselves within one or the other research paradigm. Many could go much further and name specific qualitative or quantitative methodologies they routinely employ in their studies. Those with this sharp distinction in mind will then develop a first impression that ESM is not just another quantitative method, it is a quantitative researcher's dream. After all, the sheer volume of numbers produced by an ESM study lends itself well to some complex statistical analyses. Aspects of personality manifested in daily life, such as self-esteem, can be measured by multiple items over mul-tiple time points across multiple contexts. Thus the utility of ESM for quan-titative research on human lives and conditions can hardly be overstated.

Yet ESM can also be a valuable qualitative research tool, letting us know, in the participants' own voices, what they are doing, thinking, and feeling, and how they are perceiving their social and physical environment. The questions asked on each ESM form can be fully tailored by the researcher and can be

made more or less open ended as the researcher sees fit. For example, the question, "How do different classroom activities affect the engagement of high school students in the curricular content?" could be addressed with a purely qualitative approach. This would involve analyzing responses students provided during classes to the question "As you were beeped, what were you thinking?" for narrative content to reveal connections (or disconnections) between the thoughts of students and the topic of the class. On the other hand, the same question could be addressed through purely quantitative means by comparing the average ratings students gave during different classroom activities on scales measuring "interest," "enjoyment," or "concentration."

This two-pronged example is not meant to imply that qualitative researchers should design their own qualitative ESM and quantitative researchers should focus only on numeric rating scales and percentages. In fact, ESM is best used when both the qualitative and quantitative information it provides is brought to bear on a broad question. By linking together quantitative and qualitative approaches in one instrument, ESM transcends the dichotomy between the two and encourages researchers to break the habit of addressing research problems with only one narrow methodology. ESM research can also be combined with lab studies to address both qualitative and quantitative questions, as Feldman Barrett (e.g., Feldman Barrett, Gross, Christensen, & Benvenuto, 2001) has done.

In this chapter, types of analyses that can be useful with ESM data will be described and examples in past research of specific applications of each analysis will be provided. The order of presentation will follow the continuum from qualitative to quantitative. The choice of which analyses to perform is obviously dependent on the kind of question one is interested in. However, for any given purpose, even if it involves a precisely worded question about persons or about situations (see Chapter 3 and Larson & Delespaul, 1992), there will be multiple analytical options, as evidenced by the previous example. Although performing more analyses and more complex analyses of different types is likely to produce the richest picture of the contents and contexts of human lives, there is also something powerful about a simple graph of one participant's mood changes over one week, denoted with her concurrent thoughts and activities.

Qualitative Approaches

One of the most widely used qualitative functions of ESM data is to illustrate a particular experiential pattern through the use of the detailed description of a single case. Three books, each describing a different ESM study, provide the best examples of the use of this approach (Csikszentmihalyi &

Larson, 1984; Csikszentmihalyi, Rathunde, & Whalen, 1993; Larson & Richards, 1994a). In each book a description of a single individual (or family) is presented within the first five pages. The descriptions focus in on a precise moment in time and provide details on what the respondents were doing, thinking, and feeling, as well as where they were and who they were with. From there, the discussion broadens to consider the participant's entire week, and how this pattern of experience is both illustrative of a wider pattern and unique to a particular individual. This qualitative technique serves not only as a means of capturing the reader's interest by giving life to the data in a way that a numerical table could not. It also reminds both the reader and the researcher that the actual lives behind the data are much more complex, messy, and nuanced than even the most sophisticated statistical analyses could show. Finally, the process of doing this qualitative work may lead the researcher to discover trends or patterns that had not previously been considered. Recognizing the power of single-case studies, Valsiner and Molenaar (2005) recently launched a journal dedicated to the analysis of single cases. Of course, reading through individual ESM forms, unlike split-second statistical computations, requires a tremendous amount of time. But the advantages of doing so usually make the effort worthwhile.

Another qualitative approach is to focus on a single situation, rather than a single person. Csikszentmihalyi and Larson (1984), for example, focused on one world history class over four days in the lives of eight members of their high school sample. They described the activation level, thoughts, actions, and comments of each of the students during this class period. This approach is not as frequently used as person descriptions, perhaps because it is difficult to find well-defined situations shared by several members of a sample over several days. One way to gather more information about specific situations and issues is to supplement ESM with individual semistructured interviews. If the interviews are conducted at the end of the ESM signaling period, the interviewer can read through the participant's responses, asking for elaboration on particularly interesting moments in order to stimulate a conversation about specific issues. Csikszentmihalyi and Schneider (2000) used this technique to bring both ESM and interview data to shed light on the tension between what students enjoy in school and what they believe will be important for their future.

Often a week of ESM responses will include unexpected, unique situations that could not be studied any other way. For instance while Ann Wells (1985) was collecting data for her dissertation, a couple decided to get divorced in the middle of the week; this made it possible to compare in detail the changes in mood and emotion the two partners underwent before and after the event. In a study of high school math students by Rick Robinson (1986), one of the teenage boys who had been depressed for several days made a serious suicide

attempt that fortunately was foiled by his sister noticing an empty bottle of pills near his bed. Following this boy's activities, thoughts, and feelings for the several days preceding the event gives a powerful insight into the dynamics of suicidal behavior. Two Himalayan climbers, members of an expedition studied by colleagues at the University of Milan, were buried in their tents by an avalanche for two days—their ESM reports provide a unique window into extremely stressful situations (Delle Fave, Bassi, & Massimini, 2003). Such unexpected opportunities may not interest those who want only a "yes-or-no" answer to a specific question, but they can be illuminating to anyone who cares to understand the complexity of human experience in all its facets.

Graphic and Numeric Descriptive Information

To supplement their qualitative analysis of the world history class, Csikszentmihalyi and Larson (1984) provided a simple graphic depicting each person's self-rated level of activation in the class, as well as the average activation level for the class as a whole. Elsewhere in the book, they plotted one student's mood ratings over the course of 36 moments in 1 week and recorded her thoughts and activities corresponding to each point on the graph. These graphs are reproduced here as Figures 5.1 and 5.2. Graphs such as these bridge the gap between qualitative and quantitative approaches. Numbers from ESM are tabled or graphed to make the qualitative description of the person or situation more complete. These numbers are derived from the rating scales and from percentages of moments spent in different categories of thought, activity, location, or companionship.

Response-Level Data

In addition to describing a single case, numbers and graphs can also be used descriptively to summarize characteristics of a sample. Before discussing these approaches further, we need first to consider what the "sample" is. At one level (what some have called the "beep" or "response-level"), the sample is comprised of a collection of moments in time in the lives of several individuals. When data are initially entered, the structure of the database is usually designed as a response-level file, wherein one "case," or row of data, represents one self-report form completed by one person in response to one ESM signal. A summary of these moments can be created in tables or graphs of mean ratings of continuous scales and relative frequencies (percentages) of categorical variables. This strategy was used by Csikszentmihalyi and Larson (1984), who used pie charts to depict how adolescents allocated their time among different activities and locations (see Figure 5.3).
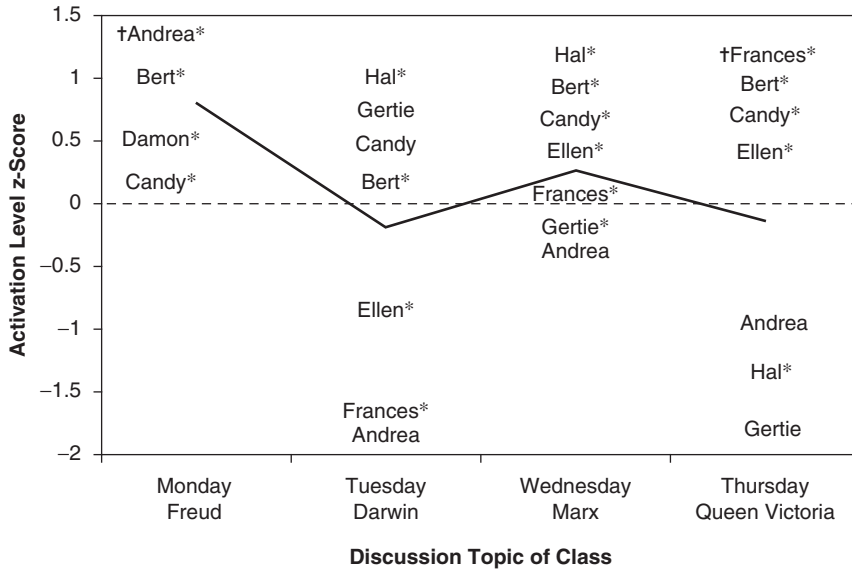
Figure 5.1   A Profile of Activation Levels in One History Class Over Four Days

SOURCE: From *Being Adolescent: Conflict and Growth in the Teenage Years* by Mihaly Csikszentmihalyi. Copyright © 1984 by Basic Books, Inc. Reprinted by permission of Basic Books, a member of Perseus Books, L.L.C.

NOTES: The dark line shows the average for each class session. †Indicates student who is giving the presentation. *Students who report paying attention.

Using Z-Scores

As was described in Chapter 4, it is often desirable to apply a standardized metric to rating scale data. Standardized scores offer the advantage of controlling for individual differences in scale usage (favoring extremes versus middling responses, for example). Z-scored variables are not useful in examining the overall means of a sample because those means will be zero by definition. Where Z-scored variables can offer an advantage is in more fine-grained examinations of experience in different contexts. To see how this works, consider the ratings of happiness recorded by two hypothetical participants in an ESM study. In Figure 5.4, we plot their raw scores and corresponding Z-scores over ten ESM responses. Heather uses a smaller range of the response scale than Michael and favors more positive responses overall. If we are interested in the context of "TV watching," and both participants were watching TV during response 6, we would reach different conclusions depending on which measurement scale we used. When raw scores are used, it appears that TV is a neutral event for Heather, who rated

Mood (Raw Score)

| | | −24 | −16 | −8 | 0 | 8 | 16 | 24 | What she was doing and thinking about |

Monday 12:44 PM — Talking with friends at lunch about "the jerks who try out for Miss America"

6:45 — Talking with sister in TV room, thinking "what a liar my sister's friend is"

8:40 — Watching Miss America alone, thinking "what it would be like to be a model in Paris"

10:25 — Watching the news alone, thinking about "Trans Ams"

Tuesday 8:45 AM — Talking with friend in gym, thinking "how smart by brother is"

10:58 — In English Lit, thinking "Rime and rhyme are the same"

12:36 PM — Talking to friends at lunch about computer programming

2:00 — In Sociology class, discussing "how the number of marriages has grown"

5:15 — Talking to aunt on phone; "My aunt talks too much"
7:07 — Taking a shower, thinking, "What to do about my situation"

Wednesday 10:30 AM — In Spanish class: "What is that word?"
12:15 PM — Talking with friends about "whether you have to stop on the highway if a bus stops"

1:30 — In Psych class, thinking "how rotten a day this has been" Alone, buying a birthday present for
2:55 — sister: "When is the sales lady ever going to get done" With friend buying Coke; "I'm so mad right
4:20 — now, I'm going to get fat"

6:05 — Eating dinner alone; "When should I call Mrs. Querva" Watching "Charlie's Angels "with family;
8:55 — "When is my sister ever going to get off the phone"

Thursday 9:40 AM — Talking; watching friends fighting; "Why did Merri kiss Billy in the parking lot"

12:30 PM — Talking with brother in cafeteria, thinking "if I should kill Kirk"

2:00 — In Sociology class discussing "what to do about old people"

6:00 — Going upstairs to talk to brother
9:55 — Brushing teeth in bathroom; "What should I wear tomorrow?"

Friday 11:30 AM — In English class
12:15 PM — Talking to friends in the student center
1:20 — In Psychology class; "What is the name of that country?"

3:15 — Making a sandwich in kitchen alone, thinking "about how I was going to get to the party tonight"
6:40 — Waiting for the phone; "when is my friend's sister ever going to get off the phone"

10:22 — In Barry's car, looking for somewhere to go
11:07 — Drag racing, telling Barry to stop; wondering "where were we and why were we doing this"

Saturday 12:50 PM — Watching TV; father yelling at sister, thinking about "how much of a bastard he is to my sister"
2:15 — Playing tennis with friend; "Why was Lou mad"
5:00 — Getting a ride home; thinking "that this fat lady jogging down the street really needs it"
9:20 — Go-karting with friends and trying not to crash

11:20 — Drinking at friends house; "How am I going to hide this condition from my parents"

Sunday 1:20 AM — Watching a movie; talking with sister; thinking "how much my stomach hurts"
6:10 PM — Doing English alone in room; "watching some stupid movie"
8:35 — Watching "Poseidon Adventure" with family

Monday 10:40 AM — In Spanish Class; thinking "how tired I am"
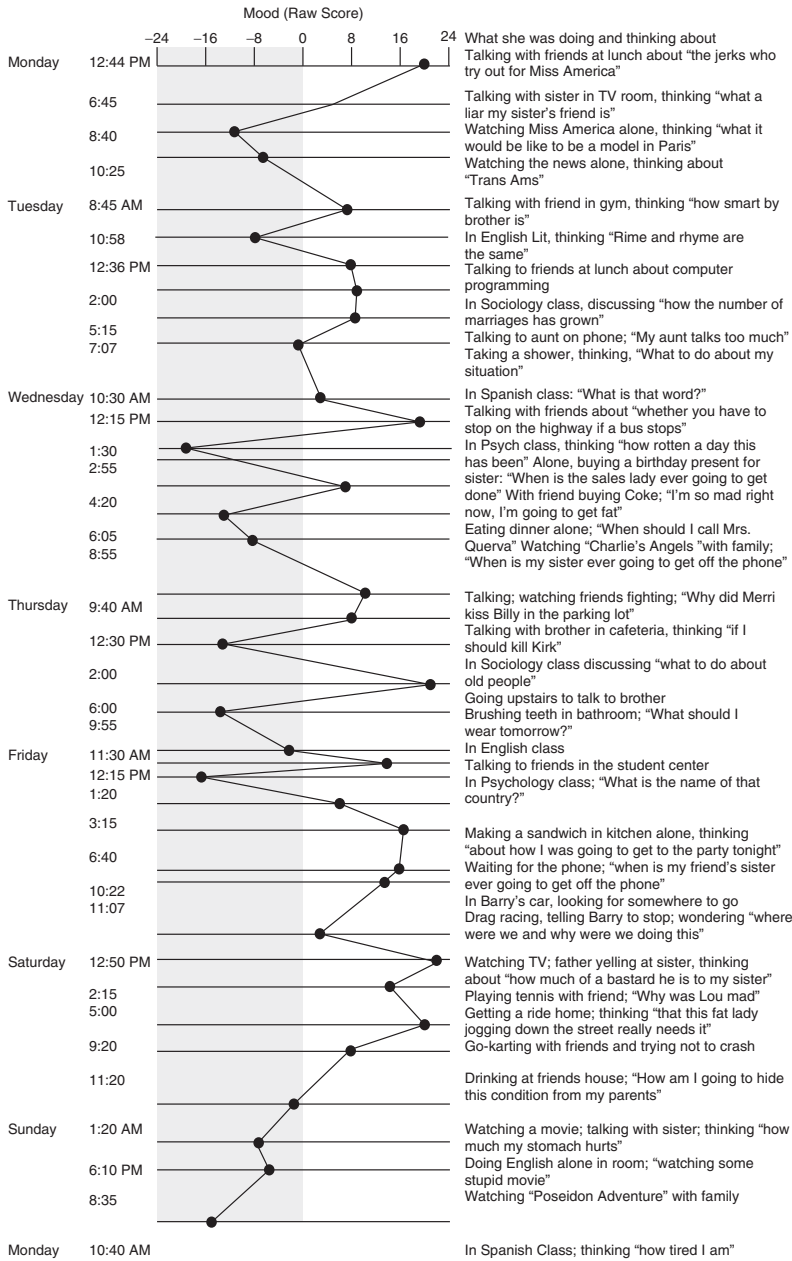

Figure 5.2   The Week of Lorraine Monawski

SOURCE: From *Being Adolescent: Conflict and Growth in the Teenage Years* by Mihaly Csikszentmihalyi. Copyright © 1984 by Basic Books, Inc. Reprinted by permission of Basic Books, a member of Perseus Books, L.L.C.

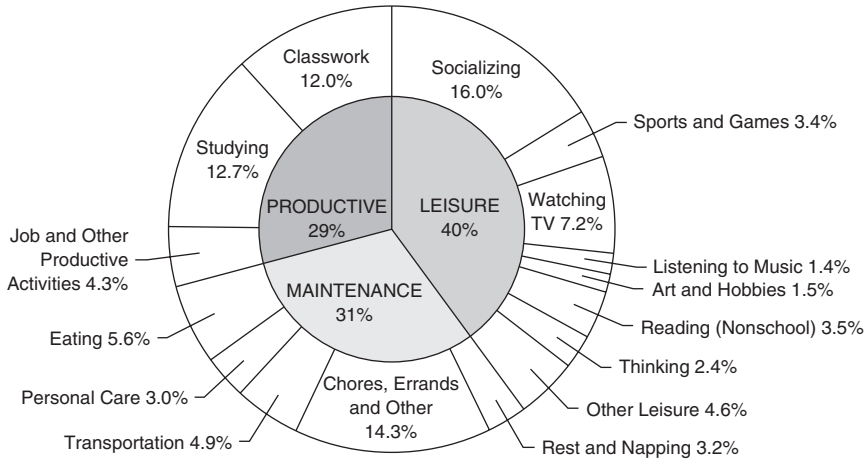NOTE: Chart shows Lorraine's mood score for each of her self-reports.

Figure 5.3   The External Landscape: What Adolescents Spend Their Time Doing

SOURCE: From *Being Adolescent: Conflict and Growth in the Teenage Years* by Mihaly Csikszentmihalyi. Copyright © 1984 by Basic Books, Inc. Reprinted by permission of Basic Books, a member of Perseus Books, L.L.C.

it a 4, the midpoint of the 7-point scale. Michael appears to be slightly unhappier during TV watching, as indicated by the 3 he recorded. We could stop with this interpretation and conclude that Heather is objectively happier than Michael. On the other hand, relative to Heather's other experiences, TV watching appears to be a pretty negative experience (Z-score of –1.2), whereas it is much less negative (–0.5) within the realm of Michael's experiences. Thus if we are interested in levels of happiness during different contexts as it is perceived within each person's own frame of reference, we would do well to use Z-scores. In a real-life example of this approach, Larson and Richards (1994a) plotted the Z-scored emotion ratings of working mothers and fathers at seven different times of the day, documenting the "six o'clock crash" of negative emotions experienced by mothers (but not fathers) when they return home from work.

Person-Level Data

In considering the research question underlying such a graph, one may have difficulty determining whether the relevant sample is the response-level sample of moments or the smaller sample of individuals. For analyses employing inferential statistics, it is usually best to focus on the sample of individuals, as we will explain shortly. For descriptive purposes, either sampling level can be used, depending on the research questions. To construct a
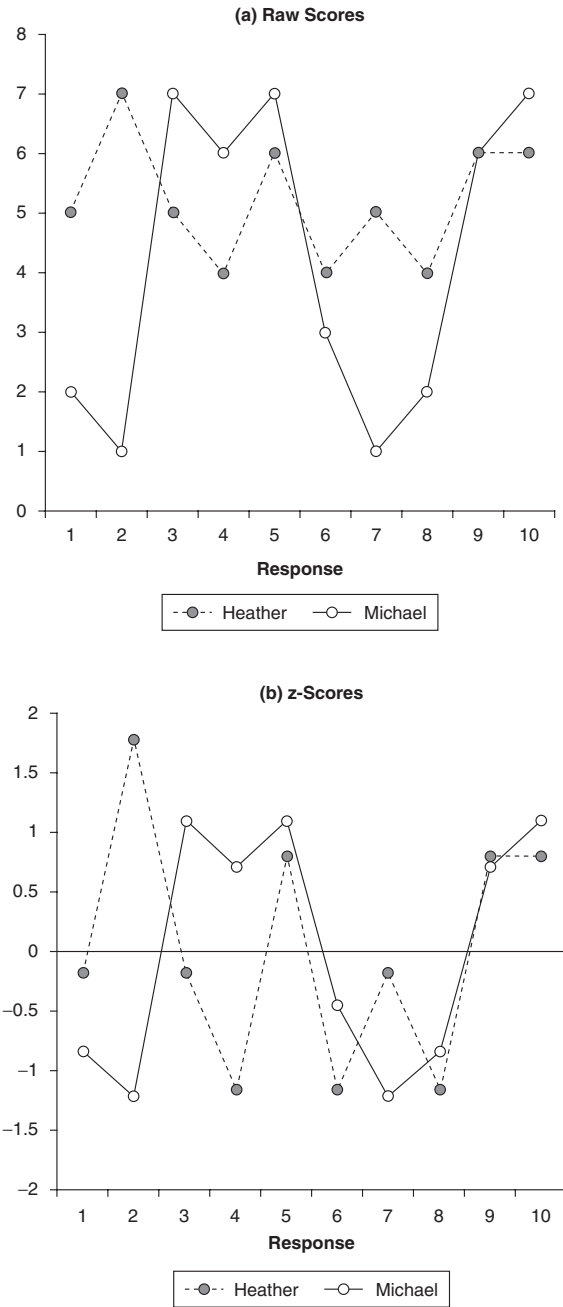
Figure 5.4   Comparison of Raw and *Z*-Scores Over Ten Responses of Two
                  Hypothetical Participants

(a) shows the raw scores, and (b) shows how the same scores would look when transformed
into *Z*-scores.

person-level data file, that is, a dataset in which the individual person is the fundamental unit, data are aggregated within each person by taking the mean of each rating scale across all observations and by calculating the percentage of the person's observations in each of several relevant categories of location, activity, and thought content (see Chapter 4 for further discussion). The person-level data file can also contain mean (raw and/or Z-scored) ratings of scale variables while the person is in specific contexts (e.g., with friends, at work, at 6 p.m., or while watching television). For more details on creating these files, see also Larson and Delespaul's (1992) excellent guidebook.

Planning for Statistical Analyses

Given the complexity of ESM data, researchers may develop the impression that only complex statistical analyses are appropriate in ESM research. Although complex multivariate strategies show much promise for particular applications, we would advise against relying on any one technique to make sense of ESM data. Instead, researchers would do well to pursue several qualitative and quantitative approaches to address a set of related questions.

As a way of organizing the multiple statistical procedures that should be considered in any analysis plan, we have arranged our discussion from the most basic and traditional strategies to the newer and more complex analyses. A useful way of categorizing these strategies would divide them into those based on the ordinary least squares (OLS) linear model, such as analysis of variance (ANOVA) and OLS regression, and those based on multilevel modeling or hierarchical linear modeling. One of the major decisions necessary when using OLS strategies, which is whether to analyze the data using the response-level or person-level approach, becomes irrelevant in a multilevel approach that considers responses and persons simultaneously. For this reason, ESM researchers are increasingly turning to multilevel strategies. Nevertheless, a basic grounding in issues and techniques surrounding the use of OLS strategies with ESM data is essential.

Choosing the Response or Person-Level Approach in OLS Strategies

Whether the research question is one about situations or about persons, analyses appropriate to the question can be conducted at the response level or person level. In most cases, the person-level approach is to be preferred if p-values are going to be examined. One of the problems of response-level analysis is that it violates the assumption of independence of observations, an assumption required in order to validly infer from a sample to a population. Observations within each person are not independent, and observations that

are adjacent in time are likely to be interrelated even more. It should be noted, however, that although response-level data are not independent because many responses come from the same person, they are independent in another sense: they come from different situations and different points in time. Traditional testing involves placing many subjects (who are independent from each other) in one large hall or laboratory at one time (negating the possible effects of different times and places). Thus both ways of gathering data have their advantages and disadvantages.

One feature of response-level analysis, the large N, might be seen as an advantage to researchers bent on achieving statistical significance, but can actually be a problem if trivial effects are given too much credence. *P*-values are partly dependent on sample size, so that with, say, 3,000 total responses, the tiniest of differences will turn out to be statistically significant. Therefore if results from response-level analyses are reported, they should be accompanied by effect-size information so the reader can judge the "meaningful" or "clinical" significance of the result alongside the statistical significance. A further problem of response-level analysis is that some individuals contribute more data than others to the dataset, and the individuals who complete the most self-reports may not be representative of the whole sample.

Person-level analysis avoids these problems because each person is "counted" in the data just once. Inferential statistics are based on the number of persons and are consequently more conservative, making a Type I error less likely. However, person-level analysis is often more complex. The tricky part is in setting up the data to be analyzed, specifically in determining how to create a single aggregated variable that measures what it is intended to measure. For example, to measure (at the person-level) alertness while watching television, one would compute an aggregate alertness value for each person by averaging across only those self-reports on which the person reported watching television. This process does not entirely remove the problem of unequal weighting because the aggregated alertness ratings of some people will be based on many TV-watching moments, while others may have only one or none at all. Indeed, to conduct analyses of experience during rare contexts (such as during illegal drug use), researchers may be forced to use response-level data, as most respondents would have no data to aggregate.

OLS Statistical Techniques

Comparing Groups of People

One of the most basic types of analyses is to compare groups of people, such as females to males, adolescents to parents, and clinical to normative

Table 5.1   Mean Level of Moods Over the Course of a Week

Mood	Talented (N = 208)	Average (N = 41)
Affect		
Happy	4.87	5.15**
Cheerful	4.51	4.89***
Sociable	4.59	5.06***
Potency		
Alert	4.68	5.04**
Active	4.25	4.46
Strong	4.41	4.69**
Excited	4.10	4.43**
Concentration		
(How much were you concentrating)	4.39	4.82**
Motivation		
(Wish to be doing activity)	4.29	5.26***

SOURCE: Csikszentmihalyi, Rathunde, and Whalen. Copyright © 1993 Cambridge University Press. Reprinted with the permission of Cambridge University Press.

NOTES: The figures above are group-means based on individual means. The actual number of observations that contributed to them was over 7,000 for the talented group and over 1,500 for the average group.

Significance of *t*-tests between the two groups: *$p < .05$. **$p < .01$. ***$p < .001$.

samples. Groups can be compared on how much time they spend in various locations, with various types of companions, or doing different kinds of activities. The aggregate variables in these types of analyses are usually expressed as a percentage of the individual's total self-reports. Thus if a given individual submitted 30 self-reports and reported daydreaming on 3 of them, a researcher creating a "percent time daydreaming" variable would compute 10 percent for this person. Rating scales are another dimension on which groups can be compared. Csikszentmihalyi and colleagues (1993) compared talented to average adolescents on nine dimensions of experience by simply computing a person-level average rating across all self-reports for each individual on each dimension (see Table 5.1). They used t-tests to test for statistical significance, but if more than two groups are being compared, ANOVAs could be used instead. Analysis of covariance (ANCOVA) could also be used if the researcher wanted to control for a covariate.

Groups of individuals can also be compared on their ratings of experience within specific contexts, as in the previous example of alertness during TV viewing. In these cases, either raw scores or Z-scores can be used to create within-context aggregated variables. Which type of scores to use depends on the aims of the research and the assumptions underlying response scale usage. If all participants are assumed to map their internal states onto the raw scales in the same way, and the goal is to test for group differences using a common "objective" scale, then raw scores should be used. If the goal is to treat every individual's experience in a context relative to his or her own internal frame of reference, then Z-scores would be preferred. Whichever method is chosen, the process for computing within-context aggregated variables is the same. From a response-level file, compute individual averages of the desired variables using only those self-reports matching the desired context. Continuing with the comparison of their talented and average adolescent samples, Csikszentmihalyi and colleagues (1993) compared mean Z-scores of nine dimensions of experience during several different contexts (alone, with friends, while working on productive activities, etc.). Their analyses revealed interesting and theoretically meaningful differences, not only between the two samples, but also between overall (raw-scored) experience and relative quality of experience in the different contexts.

If Z-scores are used in the creation of within-context aggregated variables, one additional caveat must be noted. Because of the properties of Z-scores, the closer the context comes to encompassing all of an individual's responses, the closer the average Z-scored response within that context will come to zero. In a concrete example, if 90 percent of a student's responses were during school, and we aggregate across the school responses, the resulting average of Z-scores will necessarily be close to the average of Z-scores across all responses, which is by definition zero. When comparing groups of people within a context, this issue becomes a problem only if one group has a large majority of responses within the context while the other group does not. When comparing situations, as we discuss next, this issue must be considered if a very frequent context is compared to an infrequent one. In either case, a simple solution is to use sufficiently specific contextual categories so that participants do not have large proportions of responses within any one category.

Comparing Situations

Although it may sound confusing, person-level analyses also can (and should) be used to assess differences between situations. Before proceeding with an example, it will be helpful first to consider a response-level analysis. To compare mood at home versus at work one could use a response-level file