4
A Consulting Project from A to Z

In this chapter our aim is to try to reproduce the entire consultation process for a particular project from initial contact with the client to the final written report and postcompletion followup. The actual consultation took place in a university environment which we refer to as the SCP (see Section 1.4.5). We have modified certain aspects of this project to maintain anonymity of the client, and to keep the length of this presentation within reasonable limits.

4.1 Prior Information

Arrangements for the time and place of the initial consultation session were made through telephone contact with the client. The project was briefly discussed and we instructed the client to bring relevant information, such as printouts of the data, to our forthcoming meeting. This contact provided us with the opportunity to obtain some prior information about the project:

- The project was postexperiment.
- The client wanted us to perform the analysis and provide graphs.
- The project was the client's dissertation study.
- It had something to do with teaching methods and learning styles.
- The experiment involved $n = 87$ teachers.

Remarks

An obvious question is, "Why didn't we gather more information?" One answer is simply that we will be asking the client to start from the beginning, which effectively means reiterating anything that was discussed prior to the initial consultation session. In our experience, trying to obtain too much information prior to the first session tends to introduce more chance for misunderstandings. This not only detracts from the objectivity needed by the statistical consultant; it can also be quite difficult to readjust preconceived notions about the data and direction of analysis.

Establishing contact with the client prior to the initial consultation session does have certain advantages, of course. It starts the communication process and provides us with the opportunity to become acquainted with the general nature of the project. In some cases, we may find that the project does **not** warrant setting up an initial consultation session. Obviously we want to extend our discussion with the client to make sure we have not misinterpreted the information that was provided by the client. We therefore need to obtain enough information to form a good opinion of the extent or level of (statistical) sophistication involved in the project. Some of the reasons why we might decide to defer or terminate setting up an initial consultation session are:

- In legal cases, a conflict of interest may arise with respect to the client's project simply because of current or previous work we performed for another client.
- The sample size is too small to justify a statistical analysis.
- It would better if the client performed certain tasks before setting up a meeting: collecting, entering, or formatting the data.
- We do not have enough resources, or expertise, with respect to the statistical analysis that would be required for the client's project. Could we design and implement a clinical trial? Do we know how to interpret a market model?

For this particular project, the prior information we obtained from the client warranted setting up a consultation session and enabled us to form the following opinions about the project.

1. The design and implementation of the experiment would need to be carefully diagnosed to ensure it satisfied the principles of experimental design: control, randomization, and replication.
2. The analysis would probably require standard ANOVA and t-test procedures. If there are problems with the data, nonparametric procedures may be necessary. The sample size may be of concern if too many factors are involved.

3. Quality graphics may be required since the project was the client's dissertation study. It was also likely that the project could be regarded as a pilot study — significant results will need to be interpreted with caution.

4. We will need to learn what "learning styles" really means.

4.2 Financial Issues

We can also approach the question, "Why didn't we gather more information?" from a purely economic point of view. An equally obvious answer is that the statistical consultant expects to be *paid* for their time and professional advice! While this may not necessarily be the answer the reader expected, it would certainly be prudent for a client to establish what costs are involved *before* committing to an initial consultation session. Thus, we will need to address the client's question:

> *"How much do you charge?"*

Providing the client with a standard hourly rate will usually suffice, but for short-term projects the client is often looking for an overall cost. They may also want to know how "soon" we can complete the analysis. To realistically deal with these issues would require that we know the full details of the project, details which we clearly want to defer until the actual consultation session. So how do we respond? The key is to get the client to agree to set up an initial consultation session without encumbering ourselves with unrealistic cost or time estimates. Some possible strategies are:

- Provide the first consultation hour free of charge, but under a no-obligation clause. That is, we reserve the option to decline the client's project during this period, but the client is not charged for that hour whether or not we exercise this option.

- Offer a contract option to client. After the initial consultation session we provide the client with a fixed total amount for the project: data processing, analysis, report writing, and graphics as required.

In a few isolated cases, the best option may be to simply adopt a *take-it-or-leave-it* approach. We may lose the client, but our instincts suggest that this might be for the best.

Adopting this approach may become necessary whenever the previous two approaches have failed. If the client is not willing to discuss the project face-to-face, then it's probably not worth getting involved.

Remarks

While it is possible that we may completely solve a client's problem within an hour, the *first-hour-free* option is well suited to the situation where we have prior information about the client's project. This approach is often a good way to get the client to quickly agree to setting up an initial consultation meeting, avoiding the need to have a protracted discussion on specific rates and charges. These can be discussed in detail at the first consultation session.

Estimating the total time (hence the total cost) to complete all aspects of a client's project is not easy. Data processing and report writing often take much longer than we expect. Employing graduate students certainly helps to defray the overall cost and provides the student with the opportunity to gain experience, but we are ultimately responsible for the analysis. Contracts work well for small-scale projects where all the requirements of the project can be specified explicitly. The terms and conditions of the contract will need to be documented and signed by both parties.

Although university consulting programs have the ability to involve graduate students in a wide variety of projects, clients sometimes assume graduate students can be exploited on the basis that they are providing the student with "experience." This is certainly true and some flexibility is usually required on our part; students are often prepared to work short-term for minimal rates as long as the experience is beneficial. Our job is clearly to filter out the more extreme cases.

4.3 Session I: The First Meeting

We are about to meet our client face-to-face for the first time and have made appropriate preparations for the meeting. Specifically, we should:

- Relax! The client is coming to us for advice and probably feels even more nervous than we do.
- Make sure the meeting will take place in an environment that is conducive to a focused verbal interaction with the client. The student center, scheduling the meeting during teaching-related office hours, or having no space on the desk to look at printouts, clearly do not provide good environments for consulting.
- Be attired in a manner that reflects a professional standard of service. The fact that we provide consulting services within a university environment does not alter the usual business protocol.
- Have the client's file in front of us. This should contain our notes from any prior contact, sign-in form for compiling the project summary

(see Figure 4.2 later), as well as paper, pens, and pencils[1] ready for taking notes during the consultation session.

- Be punctual. Our doctorate is only in statistics.
- Make any additional preparations in advance. For example, the presence of other personnel who might be involved in the consultation: graduate student, coconsultants, or experts from other disciplines.

Based on the prior information we obtained from the client, it seemed quite likely that a graduate student would be able to perform most of the analysis involved in the client's project. We therefore arranged to have the student present for this initial consultation session. This provided the student with the opportunity to participate in the "consulting process" and also allowed the client to see "who" the student was. We address the reason why this is important later.

In addition to the routine preparations listed above, there is one more important item that we need to prepare. Us! That is, we should always try to approach the consultation meeting with some type of agenda in mind. A good way to formulate an agenda is to ask ourselves the question:

What do we expect to achieve in this session?

With practice, it will become easier to anticipate the general pattern of our consultation sessions and our agenda may simply consist of mental footnotes. Otherwise, we should write down our proposed agenda and have it with us during the meeting.

So what do we expect to achieve in this example? Since the client's study was relatively small (only $n = 87$ observations), it was possible that we could be in a position to begin performing the statistical analysis for the project by the end of the session. To achieve this best-case scenario, the following items on our agenda would all need to be properly dealt with in sequential fashion:

- A clear definition of the problem and variables associated with the client's project.
- The objectives of the study can be supported by a statistical analysis.
- The specific contributions required for this project can be clearly stated.
- The time frame and terms of payment are mutually acceptable.

[1]Make sure to ask clients for permission *before* writing on any of their printouts, and always use pencil. We once made the mistake of not doing this and the client almost got up and left! Fortunately they didn't, but this did make the remainder of the consultation session somewhat tense.

Remark

In this particular example, we managed to cover all of the above items in one consultation session. That is, by the end of this session we had achieved our objective of being in a position to perform the statistical analysis for the project. This will not always be possible, and we should certainly not expect to be able to resolve every client's project within a single consultation session. Indeed, the discerning reader may already be somewhat suspicious about our claim in this example. Did we *really* get through everything that we present below in one session? We answer this question at the end of the presentation.

Initial Contact

The client arrived on time for our consultation session. What happens if they don't? If the client is early (as is common), introductions can be made and, if appropriate, we can ask the new client (A) to stay while we finish up with our current client (B). The purpose of this is to allow client A to hear and see the type of interaction that they will be encountering shortly. If the client is late (by more than 20 to 30 minutes), we should still try to accommodate them on that day. It is usually better to obtain at least some information about the client's project rather than simply arrange a new meeting time. Our schedule should therefore allow some flexibility: stacking up too many clients on a single day is more likely to cause rescheduling problems. In this example, our dialogue begins with the initial introductions:

client: Hi, I'm Another Client.

cons: Hi. I'm Zee Consultant and this is Affine Student who will be helping us with your project. Now, I understand you have brought some information for us to look at. But first, perhaps we could start by having you describe your project ... ?

client: Okay. ... We wanted to show that incorporating LSI preferences in technology training helped long-term retention. In the experiment we augmented traditional methods by auditory, visual, kinesthetic and tactile preferences of the participants. We used the same instruction formats in each session and our instruments were pretest, posttest, and SDS score. Most of the workshop participants were female elementary school teachers. ... We've collected the data and I have some printouts for you to see. ...

cons: (interrupts client) ... Good, but before we do that I'd just like to go back over some details about ...

client: ... my advisor also said something about "analysis of variance"? — which is why I'm here !

cons: Well, let's find out if ANOVA is needed.

Remarks

Choosing *when* to interrupt a client is not always easy. If we jump in too soon, the client may feel we are not giving them a fair opportunity to explain their project. It is worth remembering that we will expect the client to listen carefully to our explanation of certain statistical issues later. On the other hand, it is important to keep the discussion focused and we may need to interrupt the client to avoid backtracking too far. That is, we need to start processing "batches" of information. These verbal "cues" serve the dual purpose of not letting the client get too far ahead of our questions and understanding of the project, and it helps the client learn something about what we *really* need to know.

In the above dialogue, we were able to interrupt the client at a convenient point: the client's description of the project had already raised several questions that needed to be resolved; they had moved on to the subject of data and printouts. In practice, this opportunity is quite often presented by clients who, understandably, try to shift quickly from old news (project description) to the current status of the project (data analysis).

Before presenting the questions that we need to resolve with the client, it is worth considering the client's reference to ANOVA. This provided some indication of their statistical knowledge: they knew enough to seek our help, but we need to be careful with our response. To embrace a client's methodological suggestion at this early stage is generally unwise: we may end up explaining why their "great" suggestion is **not** appropriate and we *both* come out feeling somewhat foolish. In this type of situation, we should try to sound encouraging, but avoid committing ourselves to the statistical methods suggested by the client. If their suggestion turns out to be correct, they feel good; if not, then that's why they came to us in the first place. Now back to our questions.

1. Several terms were mentioned by the client which need further explanation: LSI, traditional methods, kinesthetic and tactile preferences, SDS score.

2. How does long-term retention relate to this study and how was it measured? Details concerning the pretest and posttest instruments will be needed.

3. To what extent are the factors *gender* and *school level* of interest in this study? We will need to know details about the sample sizes involved in these categories.

4. Our main concern is that the usual "control" versus "treatment" design setup is not obvious from the client's description. Furthermore, the reference to *session* formats adds a potentially complicating design factor into the analysis. We will need the client to carefully describe the design format and implementation methods that were employed in this study.

Defining the Problem

After presenting these questions to the client we obtained the following information. During the course of this discussion we also examined the printouts brought by the client.

Design The sample consisted of high school and elementary school teachers who were randomly allocated into two groups: Control and Experiment. The traditional instruction format (textbook) was used for the Control group. For the Experiment group, traditional instruction was augmented by activities specifically suited to the preferred learning styles of the participants. Both groups were split into four sessions (subgroups) and each session received the same instruction format.

Implementation The Learning Style Inventory (LSI) instrument was first used to classify the preferences of the participants for both groups. For the purposes of this study, a single preference was assigned to a participant based on his or her highest LSI score obtained in the (A)uditory, (V)isual, (K)inesthetic, or (T)actile categories, provided the LSI score exceeded 50; otherwise the participant was considered to have (N)o preference.

In the Experiment group, specific sound (A), sight (V), role playing (K), and construction (T) activities were employed for **all** the participants, relative to their assigned preference. Thus, the N preferenced participants were also involved in activities specific to their highest LSI score in the A,V,K,T categories.

Variables For assessment purposes, three quantitative measures were employed in this study. They consisted of: a pretest (PRE) given at the start of the instruction period, an attitude scale score based on the semantic differential scale (SDS) instrument which was given at the end of the instruction period; and a posttest (POST) given one month after the instruction. Other factors recorded were GENDER, SESSION, and school level (SLEVEL). Details concerning these variables are summarized below.

Quantitative Measures

PRE	Pretest:	Maximum mark = 100

POST	Posttest:	Maximum mark = 100
SDS	Attitude Score:	Maximum mark = 60

Categorical Factors

GROUP	Control / Experiment
GENDER	Male / Female
SLEVEL	Elementary / High School
SESSION	S1, S2, S3, S4 (within each GROUP)
PREF	Auditory, Kinesthetic, None, Tactile, Visual

Sample Sizes ($N = 87$)

Control	43	Experiment	44
Female	70	Male	17
Elementary	59	High school	28
Preferences:	A = 12, K = 4, N = 32, T = 23, V = 7		
Session:	*Unavailable — to be provided by client*		

Overall Issues and Objectives

So far, the direction and purpose of the consultation session has been primarily for our benefit. We have identified the important components of the study and established properties of the variables involved. Now we need to return to the purpose of the client's study. A formal statement of the research hypothesis[2] developed by the client as the objective of this study is given below.

***H*$_1$** Teachers in technology training sessions that utilize a processing activity that matches their perceptual learning-style preferences will demonstrate significantly greater long-term retention of content than teachers in a traditional setting that has not utilized that processing activity.

This essentially translates to $H_o : \mu_{\text{Cont}} = \mu_{\text{Expt}}$, in our terminology. *t*-tests and ANOVA procedures may be used to investigate this hypothesis. Our first task is to explain these statistical methods to the client.

***t*-Tests** These are used to assess whether a significant difference exists between the means associated with two independent samples. For example, we are interested in whether the Experiment group had a *significantly* higher average POST score than the Control group.

ANOVA To test the effects of several factors simultaneously, an *analysis of variance* procedure can be used. For example, we can test whether

[2]For brevity, we confine our attention to the issue of long-term retention (PRE and POST). A similar research hypothesis was also developed for attitude (SDS) differences.

> GENDER, SLEVEL, and SESSION also have an effect on POST, after the GROUP effect has been accounted for. Perhaps high school teachers performed better within each group?

It is worth noting that our "explanation" is really just a statement of how we can use the method, illustrated by an example in the context of the client's project. Remember, we haven't actually performed the analysis yet. In our experience, introducing the abstract concepts of statistical inference in the absence of results tends to produce little more than a lengthy, but rather vacuous, discussion.

The rosy examples we employed for illustrating the statistical methods will need to be tempered with a realistic appraisal of the assumptions and potential issues that may affect the statistical validity of the analysis. In particular, we **must** emphasize that significance does **not** imply causality. This is not always quite as simple as it sounds. Dealing with the issue of causality often represents the boundary between what the client wants to conclude versus what the statistical analysis will actually provide. Some of the issues that we need to discuss in more detail with the client are:

1. Assuming a significant t-test result for POST by GROUP, what is our interpretation (conclusion)?

2. How should we proceed if PRE by GROUP is significant?

3. The imbalance in the GENDER and PREF levels (class sizes) may adversely affect the ANOVA procedure.

4. What happens if SESSION turns out to be a significant factor?

5. Was the assignment of a *single* preference realistic?

Remarks

As can be seen from the client's research hypothesis, *long-term retention* is asserted as the outcome associated with a significant GROUP effect. Now comes the hard part. Should we insist that the client remove this reference to long-term retention on the basis of causality? In this example we did not.

We can certainly argue that a statistically significant result only provides evidence of an *association* between the "treatment" effect (teaching methods utilizing learning style preferences) and posttest performance; it does not *prove* that long-term retention occurred. But long-term retention is the client's contextual interpretation of what posttest performance implies, which seems reasonable given the experiment's design and purpose. Whether we necessarily agree with this type of subjective interpretation, we must be careful not to impose our expertise on the client's field of study.

Our responsibility is to make sure the client clearly understands that statistical evidence is strictly that; it is not proof. The subjective nature of a contextual interpretation is ostensibly the client's responsibility.

Since the study is postexperiment, the main focus of our appraisal of the statistical issues associated with the client's project is on identifying potential sources of nonrandom error. That is, can we find any type of bias effect that may seriously compromise the statistical validity of the analysis? In this situation, where we are reliant on the client's description of the design and implementation of the experiment, it can be useful to employ the interrogation approach.

The client provides their interpretation of a hypothetical result

which we pose for them, such as a significant t-test for POST versus GROUP, and then we ask "what if ..." scenarios.

The purpose of this exercise is **not** to try to "trip up" the client — no experiment is going to be perfect — but to provide an independent and objective assessment of the client's study. In many cases, both the client and consultant gain a much better understanding of the issues surrounding the investigation: we may hit on an issue that resolves a contextual problem for the client, and the client may remember details about the study that turn out to be statistically important.

So what did we both learn in this example? The following items correspond to the issues that we discussed in detail with the client.

1. Some of the issues involved with interpreting a significant result for the POST by GROUP t-test are:

   - As is common in small-scale studies, the sample profile tends to limit the extent to which a significant result can be applied to a larger population. Our conclusions will therefore need to based on the suggestive value associated with a "pilot" study.
   - The implication of long-term retention based on a significant POST versus GROUP result is clearly subject to debate. This is a contextual interpretation of the GROUP effect; the statistical result does not *prove* this assertion is necessarily correct. A secondary issue is whether one month really constitutes evidence of "long-term" retention. Both issues were deferred to the client's judgement.
   - This t-test only evaluates the knowledge of the participant at the one month time point. Does the client really know what the participants did during the intervening month? For example, suppose the participants in the Experiment group were so turned off by those ridiculous preference activities, they all enrolled in a

fast-track technology course utilizing traditional teaching methods.[3]

- The POST versus GROUP result does not account for different knowledge levels (baselines) that may exist between the participants prior to the instruction. The pretest will be employed to account for possible baseline differences.

2. For the client, a significant PRE by GROUP result would have presented some difficulties. It suggested their design was somehow flawed and complicated the interpretation of posttest performance. Neither of these is necessarily true, of course, but it is not unusual for clients to employ pretests with "crossed fingers": they know pretests are important; they just hope they're not significant in their study! In this situation, we should spend some extra time with the client to help clarify the following statistical aspects.

   - Randomization is used to avoid selection bias. For example, we would never consider $\{1, 2, 3, 4, 5\}$ to be a "random" lotto drawing, but this is just as likely as any other set of five numbers! The client's design is not flawed simply because they obtained two groups with different pretest performances; it's just the *luck of the draw*.
   - Posttest performance can be assessed by employing a *paired* t-test procedure on DIFF = POST − PRE. That is, we consider the *difference* between a participant's PRE and POST scores. This actually makes more sense since we would generally expect a positive relationship between these scores for each participant (i.e., high PRE $\Rightarrow$ high POST). The paired t-test accounts for this baseline effect.
   - The DIFF by GROUP t-test also has another advantage. If there is a wide range of POST scores within each group, this may mask the GROUP effect. By accommodating the participant's baseline, a meaningful GROUP effect can be more easily detected.

3. The imbalance in the GENDER and PREF levels (class sizes) may adversely affect the ANOVA procedure. It is possible that we may obtain spurious results which are simply an artifact of the small sample size. Note that the number of K and V participants will be further split by GROUP in this analysis. A more likely outcome is that these factors will not be significant. This does not necessarily mean that these factors are unimportant; there just isn't enough information

[3]We were assured that this did not happen. The point here is that the consultant needs to be creative in generating "what if" scenarios.

in this study to make a meaningful determination. For the PREF factor, it may be worthwhile combining certain levels. The client recommended A + V and K + T.

4. Although the intention of this study was to provide identical instruction formats in the four sessions within each group, it is quite possible that the performance levels may vary significantly between SESSION. This was not an important concern for the client.

5. Two issues associated with the assignment of a *single* learning-style preference were:

   - Some participants could be multipreferenced. The decision to assign a specific preference to these participants therefore introduced some subjectivity into the study.
   - Nonpreferenced participants should be analyzed separately.[4]

Specific Contributions

Time to wrap things up. We have used up most of the time that was scheduled for this consultation and now need to set up the postsession agenda for the client. This is very important: the client should *always* be provided with a sense of what has been accomplished at the end of a consultation session. In this example, we are in a position to address the specific contributions of the client's project:

- What we will be doing for the client;
- What we may need from the client;
- Determining the time frame and costs.

Consultant From our discussion of the overall issues and objectives, the client knows how we intend to approach the statistical analysis of the project. Thus, the main purpose of this agenda is to provide the client with an outline of the steps involved in the analysis and to establish any specific requirements associated with the project. Based on this information, we will then determine a realistic time frame for completing the work.

1. Data processing: Database transfer and error-check analysis. Corrections to be provided by the client.
2. Exploratory data analysis: Summary statistics for all the variables in the database, overall and by GROUP.

[4]This was an important part of the project which we have left as an exercise.

3. Statistical analysis:
   t-tests: PRE, POST, and DIFF by GROUP
   ANOVA: PRE, POST, and DIFF with all explanatory factors: GROUP, SESSION, PREF, GENDER, and SLEVEL.
4. Written report: Will include an explanation of the statistical methods employed (requested by the client).
5. Presentation quality graphics: bar charts and histograms. Scatter plot (or equivalent display) for presenting the *t*-test results. Final selection to be made by client.

Client In this example, the client provided us with the database already in a suitable electronic format (Excel file on a disk). The printouts we had examined previously were self-explanatory, but did not include the SESSION codes for the participants. Since this was a relatively small sample, a simple approach was to have the client just email us the list of the SESSION codes (with the participant's ID code).

In general, clients are responsible for the data collection and data entry phase of the study. They will also need to be able to provide corrections for possible errors identified by the consultant during the error-checking analysis of the database. To assist with this initial phase of the analysis, the database should adhere to the following.

Software Excel spreadsheet (disk) or plain text document (email).

Format Rectangular. Each row is an observation, with an entry in each column (the variables).

Missing A special code (e.g., "−9") to be used for missing values. There should be **no** blank entries in the database.

Encoding: A key listing all the codes used in the database.

Time and Costs Involving graduate students in projects is one of the main aims (and advantages) of university consulting programs. This particular project was ideally suited for a graduate student and our primary role was to oversee the progress and provide assistance as necessary. Of course, our intention to employ a graduate student needs to be discussed with the client. Some of the main issues are:

1. We are ultimately responsible for the project; not the student.
2. The cost benefit to the client needs to be weighed against a more flexible time frame. The student will have other commitments

such as coursework, exams, assignments, and teaching duties which are equally important.

3. Students often find the written report to be the most difficult task and several drafts may be required. The client should **not** be charged for this learning experience!
4. Creating presentation quality graphics can be very time consuming and some students may not (yet) be proficient with high-level graphics in statistical software such as SAS and S-PLUS.
5. Confidentiality or other constraints associated with the project would extend to the student.

Session Summary

In this example, we were in a position to wrap things up and address the specific contributions for the project. That is, we had essentially resolved what needed to be done for the client's project and any further communication before our next meeting would be performed indirectly — by phone, fax, or email. Our next meeting would take place when the results and preliminary report were complete. These would be sent to the client prior to that meeting which would focus on the interpretation of the results.

Did We Really Do All This in One Session?

The answer to this question is yes ... and no. In reality, more than one consultation session would have been required to address the issues of this project in the detail presented here. So how can we claim only one session was needed? ... We took *extensive notes* during the session. That is, we relied on our notes to reconstruct certain details about the project *after* the consultation session was over. Our presentation of the consultation session is therefore the result of merging these two sources of information:

- Information we obtained during direct interaction with the client;
- Details that we reconstructed from our notes.

The main advantage of this approach is that it allows us to gain a good overview of the client's project. The obvious benefit for the client is that they only need to describe the details of their project once.

Additional Sessions

Even in small-scale projects, it may not be possible to resolve the issues associated with a client's project within a single consultation session. The discussion of the overall issues and objectives of the project may still be incomplete. Keeping track of time is therefore important since we should always try to reserve a sufficient amount of time at the end of a consultation

session to outline an agenda of future activity for the client. This should **not** just consist of scheduling another meeting with the client; they first deserve to know what was achieved from *this* meeting.

Briefly summarizing the main points of the client's project, or outlining the statistical approach that we are considering, are simple and effective ways for making the client feel the session was productive. In this situation, we should also try to set up a "task" for the client to perform before the next consultation session. Although this will depend on the particular nature of a project, some common examples are:

- Starting to create the database (or template for one)
- Getting copies of relevant references that the client cited
- Revising or creating a draft of a questionnaire for review
- Writing a protocol draft for an experiment
- Reporting pertinent issues back to an advisor/supervisor.

Students

Clients are sometimes apprehensive about having students involved in their project. If the client's demeanor suggests that this might be the case, act promptly to reassure the client in this situation. Their main concern — the quality of the analysis — may potentially lead to more negative feelings: they are being palmed off; that we do not have a vested interest in their project, and so on. Having the graduate student present during the consultation session is certainly helpful in this regard. The student obviously benefits by having the opportunity to participate in the "consulting process" and, the client knows "who" the student is. Whether the student is present or not, we will still need to emphasize our responsibility for the analysis to the client.

Graphics

Reasonable diagnostic graphics are usually provided as part of the output from a statistical procedure. Adding titles, labels, and other annotation may take a little more effort, but any serious modification for presentation purposes will be time consuming. This is really the key issue and our decision to produce presentation quality graphics for a client needs to be considered carefully.

In this example, the student had the opportunity to gain some experience with S-PLUS and so what we charged the client did not reflect the true costs (time) involved in creating the requested graphics. The ability to undercharge a client is not a luxury we can always afford, however, and some issues to consider with regard to presentation quality graphics are:

- What is the level of quality required?
- How complex are the graphics?
- Is color required?
- Aesthetics?
- How many graphs are needed?

Standard output from the statistical software with titles, labels, and other appropriate annotation, may actually be sufficient. If not, an example should be used to establish the level of quality required. In this situation, we should emphasize that simple displays can be quite effective and are easier to modify. Multivariate displays involving grouping variables, contours, or 3D representations require far more effort to "get it right." Note that grey-scale should be employed for displays that are likely to be photocopied; color requires additional resources and what we see on the screen is **not** necessarily reproduced on hardcopy.

What happens if the client doesn't like the design or aesthetics of our resulting display? We will need to be able to convey to the client what the graph will look like *before* creating it. Finally, creating a special (unique) graph takes time. The number of special graph "formats" should be kept to a minimum and each used for more than one display.

4.4 Documentation

The consultation session is over. The client has left with the outline of the work to be performed and we provided the client with a contract estimate for the cost of our services. As with any exchange of services on a contract basis, the agreement should be formalized in writing and signed by both parties. Do **not** perform contracted services solely on the basis of a verbal agreement. Otherwise, be prepared to experience that wonderful feeling of putting in a great effort, making the deadline — and not getting paid! A simple contract outline is presented in Figure 4.1.

We should emphasize that this type of contract is really just a "gentleman's agreement" and certain details will need expanding (e.g., the amount of the payment, when the payment should be remitted, etc.). It should **not** be used as a substitute for a proper legally binding contract. Such contracts are required where confidentiality, intellectual property, and liability issues need to be carefully addressed in legal terms. Small business consulting firms may also want to protect their clientele base when subcontracting projects to independent consultants.

The project summary outline shown in Figure 4.2 serves two purposes. It provides the client with a list of the main tasks we will perform, and it provides us with the type of information we can use for reference purposes. This

STATISTICAL CONSULTING PROGRAM
SCP Letterhead Information

CONTRACT

Date
Consultant's Name and Address

Client's Name and Address

Project Title
Project Description

Client agrees to the services and conditions to be provided by the SCP as detailed in the Project Summary attached. On completion of SCP services client agrees to remit payment ...

Signed

Z. Consultant	A. Client

SCP/Client Additions:

- The SCP analysis is based on the information and database provided by the client. To the best of the SCP's knowledge the integrity of the database, and information provided by the client, is without prejudice.
- The statistical computing will be performed on a subcontracted basis by a graduate student in statistics under the supervision of the SCP consultant.
- The SCP is to provide the client with presentation quality graphical summaries.
- *Other items as necessary.*

FIGURE 4.1. Sample Contract Outline

STATISTICAL CONSULTING PROGRAM
SCP Letterhead Information

PROJECT SUMMARY

Date
Consultant
Student Assistant

Client
Client's Contact Information

Summary:

Project Title

Summary statistics as requested by client

Statistical analysis of project hypotheses

Written report to be provided to client

Graphical summaries as requested by client

Contract estimate: [ ]

Important:
The SCP assumes you accept the contract estimate with the understanding that the final SCP invoice may include additional charges. You will be informed of the need for any increase prior to the SCP performing ...

FIGURE 4.2. Outline of the Project Summary

documentation contains the pertinent information about the client and the project which can be easily retrieved from our "Client File" database. The project summary would also be included in the client's "Main File" which will eventually contain all the documentation associated with the project.

4.5 Project Analysis

The signed contract and missing SESSION codes were received from the client and a backup copy of the original database was made. Most of the project analysis was performed by the graduate student and consisted of the following:

- Data processing
- Exploratory data analysis (EDA)
- Statistical analysis
- Draft version of the written report
- Example of a presentation quality graph.

Data Processing

The first task was to download the client's data from Excel format into something more useful for statistical purposes. In this example, we employed the statistical software package JMP, which can import Excel files.[5] The SESSION codes were merged into the JMP database and a plain text (ASCII) version was output for later use in SAS and S-PLUS.

The numerical and graphical summary statistics provided by JMP did not indicate any obvious errors and agreed with the printout information that we had examined during the the consultation session. Thus, we were able to proceed directly to the exploratory phase of the analysis.

Remarks

Usually we will not be this fortunate. Our client just happened to be particularly thorough and avoided the type of common mistakes that can occur in practice (see Section 3.2, *Data Processing*). We had also seen a printout of the client's data which motivated our choice of JMP. Minitab and Statgraphics are other examples of sufficiently comprehensive, menu-driven packages that could be used in place of JMP. So what made us choose JMP?

[5] Version 4 of JMP can import an Excel file directly. JMP 3.x requires a "Save as text" version of the Excel file.

1. The sample size involved was relatively small.
2. The database was in rectangular format with no missing values.
3. The point-and-click, menu-driven interface of JMP would make it very easy to perform the exploratory and initial statistical analyses.
4. The standard diagnostic displays would provide the client with a useful basis for deciding on the presentation quality graphics.

Clearly, we were rather fortunate in this example. What about a more general situation? We consider the following aspects.

Plain Text In most cases, we would convert the client's database into plain text format and use SAS or S-PLUS to perform the data processing. One of the main advantages of plain text files is that they can be imported into (almost) any application irrespective of the computing platform being used. In particular, using plain text attachments in email ensures that both parties will be able to read the information.

An obvious disadvantage is that we lose the special formatting of the original database. Working efficiently with plain text files therefore requires a good editor such as Emacs — and, of course, knowledge of how to use it.

Client Database In the "unlikely" event[6] that the client provides us with a plain text file, what database format should we ask the client to provide us? Excel is widely available, can be used by a novice, and from our perspective, is easily converted into plain text format:

File → Save as ... → Text (Tab delimited)

If you, or your client, prefer to use something other than an Excel spreadsheet, make sure both parties have the desired application.[7] Translators do not always work well, especially across different platforms.

Coding Many of the data processing problems arise from improper coding. For example, survey questions that allow the respondent to select more than one option should be recoded with separate dummy variables for each option. Similarly, a dummy variable can be used to encode write-in comments or responses to "Other, explain" since these are usually of nonstatistical interest. We return to this issue in the case studies presented in Part II.

---

[6] Clients often use spreadsheets, such as Excel, but do not really understand that all that wonderful formatting usually needs to be discarded by the statistical consultant.

[7] One of our worst experiences was a case where the client had entered a large amount of data into an application that was completely outdated. Eventually, we were able to write a C program to extract the database directly from the ASCII octal code!

EDA

The numerical and graphical diagnostics from JMP indicated the presence of two potential outliers in the Control group for POST (low scores). However, both these participants also had low PRE scores and no outliers were present in the derived variable, DIFF = POST − PRE. For the categorical variables, PREF and GENDER showed the most disparity in their class sizes. CPREF was therefore created by combining the PREF classes, A + V and K + T, with CPREF = N making up the third class.

We used JMP to perform t-tests and to fit various ANOVA models (see *Statistical Analysis* below). The main purpose of this exercise was to gain some insight into the results that would be investigated in more detail using SAS, and to evaluate the assumptions underlying these statistical procedures. As can be seen from Figure 2.3 in Chapter 2, the distribution of the POST scores by Group exhibits nonnormality. However, the two-sample t-test is robust against departures from distributional assumptions.

Significant results with respect to the GROUP factor were only obtained for POST and DIFF. None of the other factors were significant in the ANOVA models (including CPREF). The JMP results and diagnostic graphics were printed and saved for review at our forthcoming meeting with the client.

Remarks

We could have actually completed the project analysis using JMP alone since the results above essentially formed the basis of the written report. However, a disadvantage of menu-driven systems such as JMP is that regenerating output usually requires repeating all the point-and-click interface interactions that we had performed previously. This is fine for exploratory purposes, but rapidly becomes rather tedious and inefficient when an entire analysis needs to be replicated.[8]

For report writing, a more serious disadvantage is that the output from JMP can not be saved as a plain text file. Hence we can't edit, delete, or modify parts of an output "file" to suit our purposes, nor can we email (readable) output from JMP to our client. Given all this, why did we even consider using JMP? Our answer is simply that JMP was the right tool for what we wanted to achieve in the exploratory phase of this analysis.

1. It is easy to use (and doesn't take long for a new student to learn).
2. The diagnostic summary statistics and graphics are good.
3. There is a comprehensive range of statistical methodology available.

[8]We once made the mistake of providing a client with our only copy of some JMP output. Unfortunately, we became a little enthusiastic in our explanation of the results and wrote numerous helpful comments on this version. At the end of the consultation session, the client requested a clean copy of the output!

There are other statistical software applications that also satisfy the three requirements above and could be used in place of JMP: Minitab and Statgraphics being two examples we have mentioned previously. The main point is that the statistical consultant really needs to be fluent in more than one application. Using the strengths and advantages of different software will help make the consulting process more productive.

Statistical Analysis

The *t*-tests and ANOVA models considered above were rerun using SAS. The results we obtained previously did not change, of course, but the SAS output could now be incorporated in the draft version of the written report. The following SAS code gives the basic steps of the analysis.

SAS code for analyzing client's data

```
data a ;
  infile 'client.dat' ;
  input id grp $ session $ pref $ gender $ slevel $
        pre post sds ;

proc freq ;
  tables grp session pref gender slevel
         (session pref gender slevel)*grp ;

proc means ;
  var sds pre post ;

proc ttest ;
  class grp ;
  var sds pre post ;

proc glm ;
  class grp session pref gender slevel ;
  model sds pre post = grp session(grp) pref gender slevel;
```

Although not shown in the SAS program above, the variables DIFF and CPREF were created and several variations of the ANOVA models were investigated. As in the JMP analysis, GROUP was found to be the only significant factor in the POST and DIFF models. Mean comparisons and residual diagnostics provided further support of this result. Satisfied that the statistical analysis was now complete, we obtained two versions of the output from the SAS program:

SCP Our version which contained the full range of summary statistics, diagnostic checks, and additional variables such as DIFF and CPREF in the *t*-tests and ANOVA models.

Client The client's version which only contained the pertinent results. This version was edited and incorporated in the preliminary report.

Remarks

As we indicated in Chapter 3 (see Section 3.8, *Statistical Software*), we are assuming that the reader is familiar with SAS programming statements as well as standard conventions (such as the need for `grp $` when the `grp` column contains character values). While we may get away with assuming what we like about the reader, this is not the case with our clients; they are certainly not expected to understand SAS code! However, they will need to be able to understand the output generated from this SAS code. For the benefit of the unassuming reader we briefly describe what this SAS program does.

SAS program summary:

`data a;` Reads 9 columns of data from the file `"client.dat"`

`proc freq;` Generates frequency tables (4 two-way tables all versus `grp`)

`proc means;` Computes means, variances, etc. for quantitative variables

`proc ttest;` Performs *t*-tests of SDS, PRE, and POST by GROUP

`proc glm;` Fits ANOVA models for SDS, PRE, and POST.[9]

In this example, the SESSION factor was "nested" within GROUP which is denoted by `session(grp)` in the model statement of the general linear models procedure, `proc glm`. This follows from the fact that session S1 of the Control group is unrelated to session S1 of the Experiment group. To compare different session classes across groups is clearly meaningless since the session "labels" were arbitrarily chosen.

Preliminary Report

Our intention is to email the client a draft report containing our conclusions based on results contained in their version of the SAS output. To avoid potential complications and confusion on the client's part, we performed additional editing of the output. Although the following items are specific

[9] As written, `proc glm` will perform a MANOVA analysis in addition to the (univariate) ANOVA models requested. Although we examined the MANOVA analysis, it was edited from the client's version of the output.

to SAS, they show that trying to make things easy for our client can take some effort. (See Section 4.7, *Final Report,* Tables 3 and 4.)

1. The F-test for homogeneous variances given below each t-test result was deleted.
2. The Type I sums of squares for the individual factor effects in the ANOVA models were removed.

Having taken care of the results presented in the output, we were left with the task of completing the report and addressing the issues below. These were dealt with as shown in the Final Report (Section 4.7).

- Significance. What results were significant and why? (P-values)
- `SESSION(GROUP)` appears in the ANOVA output. What was the reason for using a *nested* session effect?
- Type III sums of squares. Why are these used to test an individual factor?

The draft report was completed and emailed to the client along with a request for scheduling our next meeting. The client responded to this request and also added that they had several questions concerning P-values and significance that they wanted to ask at the forthcoming meeting. Prior to this meeting, our one remaining task was to produce an example of a presentation quality graph for the client.

Remarks

Why did we bother with editing the output? Two reasons:

1. If the client doesn't understand something in the output, we will need to explain it.
2. We needed to do this for the Final Report anyway.

Let's go back to the first reason for a moment, and assume that we didn't edit this information. What will our explanation be? "Don't worry about it. It's not important!" — [Client] "Then why do I need to have it?" ... Perhaps we could try our first response again? The point is that clients **do** worry. They have no way of knowing whether something is irrelevant. That's why they come to us.

Presentation Quality Graphics

We used S-PLUS to produce the presentation quality graphs for the client. Although we had hoped to be able to email a PostScript version of the graph to the client, they did not have the resources (PostScript printer) to be able to print the file. The following S-PLUS code provides a very rough outline of how Figure 2.3 in Chapter 2 was produced.

S-PLUS code for generating Figure 2.3

```
yy <- read.table("client.dat")   # read in data to S-PLUS
y  <- yy$post                    # extract the POST scores
ys <- yy$grp                     # extract the GROUP codes
g.hist2.fun(y,ys,...)       # call our customized graphics
                            # function (options not shown)
```

Outline of our customized graphics function

```
g.hist2.fun <- function(y,ys,signif=T,...)
{
  yb <- split(y,ys)              # split POST by GROUP
  xm <- c(mean(yb[[1]]),         # get POST means by GROUP
          mean(yb[[2]]))

  par(new=F,mar=c(6,4,3,4)+0.1)  # Figure Region Margins

  # ------------------------ This is the key step:
                                 # Create 2 sub-Figure
  fg <- list(c(0,1,0.4,1),       # Regions. One takes 60%
             c(0,1,0,0,6))       # of the top; the other
                                 # takes 60% of the bottom

  for(i in 1:2){                 # Loop through the two
       par(new=T,fig=fg[[i]])    # histgram plots, using
       hist(yb[[i]], ... )       # the sub-Figure Regions.
       legend( ... )             # Add Legend (indexed) and
       abline(v=xm[i])           # put vertical line at mean
  }
                                 # If significant, put this
  if(signif)                     # text on plot (position
      text( ... ,                # arguments not shown)
           "** Means Differ Significantly ** ")

  polygon( ... )                 # Draw the bowtie polygon

  box()                          # Box around everything
  title( ... )                   # Title and subtitles

  invisible()                    # Make plot invisible ??
}
```

This function was also used to create histogram displays for the non-significant PRE and SDS by GROUP t-test results. Setting `signif=F` in the options line: `g.hist2.fun(yy$pre,ys,signif=F,...)` would skip the

`text()` step when creating the PRE by GROUP display. The default option was to print the text as shown in Figure 4.3. Many other options and arguments were required to produce this display, of course, but the key steps are shown above. Note that the statement `invisible()` does not really make the plot invisible. S-PLUS complains if a function doesn't return some value; `invisible()` returns a special "nothing" value which stops it from complaining.

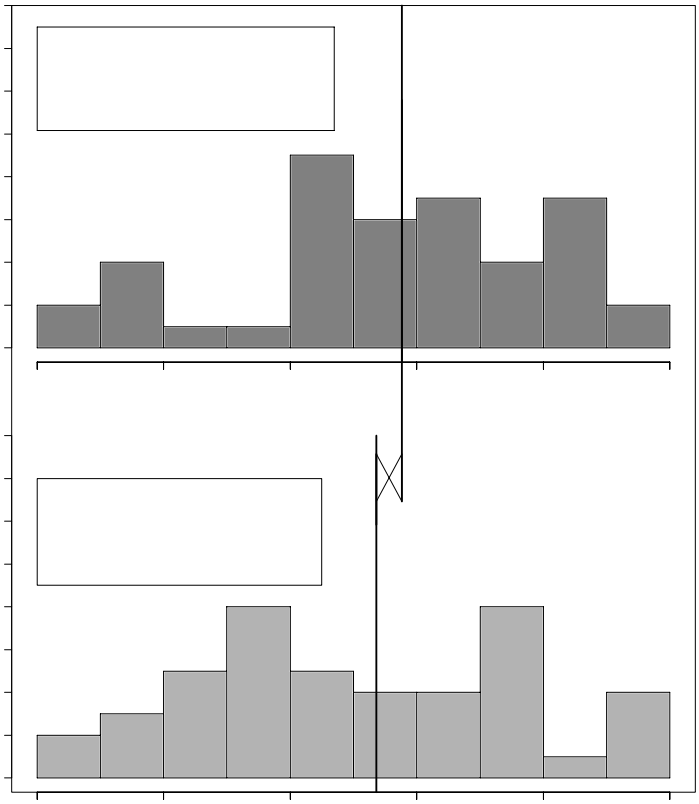

FIGURE 4.3. The Nonsignificant t-Test Display

Since the PRE and POST by GROUP displays would provide us with a very effective way of visually illustrating the t-test results to the client, we also produced Figure 4.3 for the forthcoming consultation meeting.

4.6 Session II: Presenting the Results

The client arrived on time for the consultation session, preliminary report in hand, and questions armed and ready. ... (Somewhere along the way we did actually exchange greetings!)

client: I've read through the report and I'm afraid I have a lot of questions to ask you ...

cons: (jumping in quickly and not pausing) Really? I must have sent you the wrong report! No, seriously. I would have been very surprised if you *didn't* have a lot of questions. So what I thought we could do is show you some of the output that Affine Student has prepared for us first, and see if that answers some of your questions as we go along. Then we can look at the report that you brought with you. How does that sound?

client: Okay ... great. But can I just ask one question before we start?

cons: Sure.

client: I'm really confused about the "Equal/Unequal" parts in Table 3 ... and the ANOVA tables too; which P-value am I supposed to be looking at? And ...

cons: (interrupts client) ... Wow, long question! I thought you might find those parts a bit confusing. But I think if we start with the easier parts first, the t-test and ANOVA results won't seem nearly as bad as they look when we get to them. Any other questions before we begin?

client: Well, um ... No.

Phew! Getting the client to relax and hold back their questions at the beginning of the meeting is often the hardest part of the (interpretation) consultation session. We need to be reasonably firm about setting the agenda for the meeting, but not overpowering. Hence, we must let the client ask the inevitable "just one ..." question. Our response to this question should again be a reiteration of the agenda for the meeting. Clients usually only need one round of this to realize that we seem rather determined about this agenda thing. Adding a little humor (if appropriate — don't force it) can help take the edge off our determination.

The Agenda

Our agenda for this meeting is fairly straightforward: we present the results and tell the client what they mean. Well ... yes and no. The agenda is certainly obvious, but presenting the results so the client actually understands them requires good preparation. Simply telling the client whether a certain result is significant does little to enhance the client's understanding of the analysis. They will more than likely contact us in a day or so and ask why such-and-such a result was significant. Result: a wasted consultation session for both us and the client.

So how should we prepare? As might be expected, visual explanations provide the best mechanism for explanation and ease of understanding. Where possible, use graphs instead of tables of statistical output. The advantage with graphs is that even complex displays (such as mosaic plots) can be interpreted much more easily than tables. Detecting signature patterns in a table takes experience; something we clearly do not have time to properly develop in a client new to the game.

Graphs are great, but only part of the story. ANOVA tables, for example, are rather difficult to visualize graphically. We will need to refer to tables at some point and so an important part of our preparation is to provide a logical and cohesive presentation. In this example, the main aim of our agenda was to progress slowly from graphical displays to the numerical output. The agenda therefore consisted of the following sequence.

1. Univariate summary statistics. Bar charts and histograms.
2. Bivariate statistics. The two-way frequency table.
3. t-tests. The presentation quality graphics we produced for the PRE and POST by GROUP results.
4. t-tests. The SAS output explained with reference to these displays.
5. The ANOVA tables. No graphics.

Since the student had worked on the project, this was a good opportunity for the student to gain some experience with presentations and learn to interact directly with the client. In this case, it also has the serendipitous effect of taking the client's focus away from us for a while, given that we have just railroaded them into accepting our agenda for the meeting.

Student's Presentation

The student started by showing the client the univariate summary statistics: bar charts, histograms, and frequency tables. As expected, the client had no difficulty following this part of the analysis, the main purpose being to initiate discussion about which displays the client thought should be converted to presentation quality.

The student needed to briefly explain about the cell row, column, and total percentages in the two-way tables, but the client decided this format was more than they needed. We suggested merging the GROUP breakdown into the one-way tables (see Table 1, page 185). Perfect! The client decided on bar charts split by GROUP for the PREF and GENDER variables. (The PREF by GROUP bar chart is shown in Figure 3.1, Chapter 3.)

The only problem with the means output was deciding how to present, graphically, the average scores for SDS, PRE, and POST by GROUP. We suggested three separate bar charts: one for each variable, with two bars on each plot, the heights of which would correspond to the mean scores from the two groups. The client agreed with this suggestion for SDS (since it was on a different scale), but was interested in the possibility of combining PRE and POST. This was certainly possible (and had the same "graph format" as the PREF, GENDER by GROUP bar charts above), but we were concerned that the POST bars would visually dominate such a plot. That is, the impact of the important *nonsignificant* difference between the PRE bars would be lost. The client understood our point but was not totally convinced. Both versions would therefore be produced.

t-Test Results

At this stage, the client was quite relaxed and seemed to have no problems with understanding the student's presentation. We now introduced the histogram displays of the t-test results for POST by GROUP (Figure 2.3, Chapter 2), and PRE by GROUP (Figure 4.3). The client was suitably impressed and no longer seemed concerned as we discussed the PRE and POST t-test results in Table 3, page 186, in relation to these two graphs.

The notion of P-value was initially discussed in terms of the likelihood that, given there was really no advantage in incorporating learning style preferences, the client's experiment produced a bowtie (difference between the means) of width observed in the histograms. The POST result clearly did not appear to support this hypothesis, whereas the PRE result did. The conventional value of 5% (P-value < 0.05) was introduced as the standard criterion for assessing significance. While the existence of this "magic number" for assessing significance was certainly not new to the client, interpretating the PRE and POST results in statistical terms was clearly the hard part. Thus, we slowly and carefully walked our way through the "statistical" conclusion that applied to each of these results. To check whether we had been successful, we asked the client to interpret the SDS result. The client was surprisingly good with the contextual interpretation and only needed a little prodding to check whether the standard deviations were comparable.

ANOVA Results

The slow but steady transition from graphical to numerical output seemed to have done the trick. The client was now ready for the ANOVA results (Table 4, page 187). We first discussed the issue of SESSION being a nested factor in the model:

```
POST = GROUP + SESSION(GROUP) + PREF + GENDER + SLEVEL
```

The client agreed that session S1 of the Control group bore no relation to S1 of the Experiment group, hence that session differences were only relevant within each group. This explained the presence of `SESSION(GROUP)` in the ANOVA output. The next step was to introduce the idea of ANOVA as a two-step process.

1. Did any of the factors have an effect? (The overall model F-test)
2. If so, which ones? (The Type III F-tests)

Step 1 was reinforced by the PRE and SDS results; we needn't go any further. That left the POST result. The model was significant (P-value = 0.0270) so it made sense to examine the Type III F-tests. We simply told the client that these so-called "Type III" F-tests were required for two reasons: none of the factors had an equal number of participants in their respective classes (the design was unbalanced); and we wanted to see whether a factor was still important even after accounting for the other factors (partial sums of squares). We suspected that the client didn't really understand the second reason, but they had no difficulty identifying GROUP as being the only significant factor in the POST model.

Conclusions

The client was somewhat disappointed that the PREF factor didn't show up to be significant (which added to our suspicion above). We quickly pointed out that the absence of a significant PREF (or CPREF) effect did *not* necessarily mean that this factor was unimportant. Rather, that there was insufficient evidence from the client's experiment to support this conclusion with regard to *individual* preferences. Since GROUP was significant, it followed that we could reasonably conclude that incorporating learning styles *did* have an impact on POST scores. Which particular preferences were more important was not indicated by this study. This seemed to help the client and eased their disappointment about PREF.

We tried (gently) to remind the client about the issues involved in associating the POST result with long-term retention, but this was a very short-lived discussion. The client's interpretation of posttest performance would be in terms of long-term retention. Instead, we switched the discussion towards the limitations of the inference and the extent to which our conclusions could be applied to a larger population profile. The client

agreed that their sample should not be considered as representative of the general population of elementary and high school teachers, but this concurred with our suggestion that the client's experiment be regarded as a pilot study.

Closure

The client was very pleased with what had been achieved in this meeting and made one further request: "Would we be able to review their methods and results chapter draft?" (for their dissertation). This still needed to be written by the client. We agreed to perform a review of the statistical aspects[10] of that chapter. We closed out the meeting by briefly discussing and summarizing what remained to be done for this project.

- Reiterating the specific graphs that were to be produced in presentation quality format
- Revising the preliminary report as per the client's requests
- Sending the Final Report, invoice, and graphics to the client
- The terms of payment
- Followup review of the client's methods and results chapter.

Remarks

And so another project draws to a close. There are still some loose ends to tie up, but we will have no further direct contact with this client. In our experience, this particular stage of the consulting process is where the "let-down effect" is most likely to occur. (The exact medical term probably has the word syndrome somewhere in it, but for our purposes the above will suffice.) What are we referring to here? Shouldn't we feel pleased with, perhaps even proud of, our efforts? Of course. But it's over, and sometimes it may be hard to just "detach" ourselves from this reality. In this situation, a good strategy is to try to divorce ourselves from that particular project, for example, by working on something completely different, or simply taking a break. On other hand, the "let-down effect" (if any) can also be a very positive experience: at last, we're almost finished.

The main point we are trying to make is that interacting with clients can take a lot of emotional energy. How that plays out after the client has gone obviously depends on many things and we may feel no effect at all. Just don't be surprised if there is.

---

[10]The emphasis on "statistical" makes it clear that we will be checking this aspect only. We will **not** rewrite the client's chapter.

Finishing up the Project

The report was revised and all the presentation quality graphs were produced. Apart from our promise to review the client's methods and results chapter when they had written it, the following documentation would finish off the project. Examples are shown in the figures indicated.

- Cover letter (Figure 4.4)
- Invoice (Figure 4.5)
- Title page (Figure 4.6)

STATISTICAL CONSULTING PROGRAM
SCP Letterhead Information

Date of Letter

Client's Name and Address

RE: SCP Report, invoice, and graphs

PROJECT: **Analysis of Dissertation Project:** Effects of Accommodating Perceptual Learning-Style Preferences on Long-Term Retention and Attitudes Toward Technology of Elementary and Secondary Teachers in Professional Development Training.

Dear A Client,

The SCP Report, invoice, and graphical summaries for the above project are enclosed. Please make your check payable to ...

If you have any questions, please do not hesitate to contact me.
Sincerely,

Z. Consultant
Title information

FIGURE 4.4. Example of a Cover Letter

Remarks

Adding our "title information" (academic degree, position) to the cover letter makes it clear that we are qualified to perform this type of statistical analysis. The client may not be the only person who reads our report.

The invoice example is provided for illustration purposes only. It does **not** reflect a complete accounting of the charges (hours) associated with the actual project, nor do we advocate this as a "standard" format for invoicing. Notwithstanding the "as-is" label, we note the following:

- The hourly rates (not shown) would include any overhead cost, unless this *must* be listed separately.
- NC (no charge) is included for the client's benefit. They can see that we honored the first-hour-free appointment and did not charge for the five minutes it took to convert and process their Excel file.
- We have charged for services in hour units which will not always be practical. So to avoid (vulgar?) fractions or double decimals, some consultants employ tenths-of-an-hour (six-minute) units — and a good watch, presumably.

4.7 The Final Report

STATISTICAL CONSULTING PROGRAM

CONSULTANTS:	Z.Consultant and A.Student
DATE:	January 1, 2000
CLIENT:	A. Client
PROJECT:	Dissertation Project

1. Introduction:

The aim of this study was to investigate whether instruction based on a person's learning-style preferences would improve retention of the material taught.

1.1 Study Design:

The sample consisted of high school and elementary school teachers who were randomly allocated into two groups: Control and Experiment. The traditional instruction format was used for the Control group. For the Experiment group, traditional instruction was augmented by activities specifically suited to the preferred learning styles of the participants. Both groups were split into four sessions and each session received the same instruction formats.

STATISTICAL CONSULTING PROGRAM
SCP Letterhead Information

INVOICE

Date *Client*
Consultant *Project*

Services	Hours	Amount
Subcontracted:		
Data Preparation	1	NC
Statistical Computing	2	
Documentation of Results	3	
Subcontract Total @ [ ] per hour	5	[ ]
SCP Consultant:		
Appointment: *Date*	1	NC
Statistical Analysis	1	
Report Preparation	1	
Appointment: *Date*	1	
Presentation Quality Graphics	2	
SCP Consultant Total @ [ ] per hour	5	[ ]
SCP Contract Total	10	[ ]

FIGURE 4.5. Example of an Invoice

Statistical Analysis of Dissertation Project

Effects of Accommodating Perceptual Learning-Style Preferences on Long-Term Retention and Attitudes Toward Technology of Elementary and Secondary Teachers in Professional Development Training

Report prepared for A. Client

by

Z. Consultant and A. Student
Statistical Consulting Program (SCP)

January 1, 2000

Executive Summary

The results of the SCP analysis suggest that augmentation of the teaching methods based on learning style preferences improves long-term retention of material taught. Further study is necessary to properly evaluate the longitudinal effect of the retention.

FIGURE 4.6. Title Page for the Final Report

1.2 Variables:

The Learning Style Inventory (LSI) instrument was first used to classify the preferences (PREF) of the participants for both groups (GROUP). For assessment purposes, three quantitative measures were employed in this study. They consisted of: a pretest (PRE), an attitude scale score (SDS), and a posttest (POST) given one month after the instruction. Other factors recorded were GENDER, SESSION, and school level (SLEVEL). Details concerning these variables are summarized below:

Quantitative Measures:

PRE	Pretest:	Maximum mark = 100
POST	Posttest:	Maximum mark = 100
SDS	Attitude Score:	Maximum mark = 60

Categorical Factors:

GROUP	Control / Experiment
GENDER	Male / Female
SLEVEL	Elementary / High School level
SESSION	S1,S2,S3,S4 Four sessions within each GROUP
PREF	Auditory, No preference, Tactile, Kinesthetic, Visual [coded as: A,N,T,K,V]

Note: For people with $\mathrm{PREF} = \mathrm{N}$ in $\mathrm{GROUP} = \mathrm{Experiment}$, the teaching method preference was assigned based on the highest LSI score with respect to the categories A, T, K, and V.

2. Methodology:

A statistical analysis of this experiment was performed by the SCP using the statistical software package SAS [1]. Three statistical procedures were used in this analysis: exploratory data analysis (EDA) was used to summarize the data, t-tests were used to detect differences between the average test scores between the Control and Experiment groups, and analysis of variance (ANOVA) was used to detect significant factor effects. Details concerning the methodology and interpretation of these statistical procedures are briefly discussed below. Further information is given in [2].

2.1 Exploratory Data Analysis:

EDA techniques are used to summarize the data. Frequency tables, bar charts, and histograms effectively display the distribution of the variables under consideration. Bar charts were employed for the categorical variables: GROUP, SESSION, SLEVEL, PREF, and GENDER; histograms for the quantitative variables: PRE, POST, and SDS.

2.2 t-Tests:

t-tests are used to assess whether a significant difference exists between the means associated with two independent samples (e.g., the average POST scores of the Experiment vs. Control Groups). Significance is based on the P-value associated with the t-test. By convention, a P-value < 0.05 (5%) is considered to provide

sufficient evidence of a significant difference. A P-value less than 1% would suggest strong evidence of a statistically significant result.

2.3 ANOVA:

To test the effects of several factors simultaneously, an analysis of variance model can be used. For example, the appropriate model to test whether any of the pertinent factors had an effect upon POST would be:

```
POST = GROUP + SESSION(GROUP) + PREF + GENDER + SLEVEL ,
```

where POST is the response variable, and GROUP, SESSION(GROUP), PREF, GENDER, and SLEVEL are all factors that could potentially affect the response. In this study, SESSION is said to be a "nested" factor since its levels were allocated $within$ each GROUP. The statistical effect of SESSION is therefore assessed as the nested factor, SESSION(GROUP).

The overall significance of the model is first used to determine whether any of the factors had a significant effect on the response. To determine the effect of an individual factor, the Type III sums of squares (SS) is used. These show the individual effect of a factor when the contributions from all the other factors have already been taken into account.

3. Results:

Table 1 consists of several frequency tables giving the breakdown of the participants with respect to the categorical factors. Tables 2A and 2B present summary statistics for the PRE, POST, and SDS variables overall, and with respect to GROUP.

The t-tests are presented in Table 3. It was found that there is no significant difference between the Experiment and Control groups when SDS and PRE are considered (P-value > 0.05). This suggests that there was no meaningful difference between the Control group and the Experiment group as far as prior knowledge and attitude in concerned. However, POST is strongly significant which suggests that the means of the two groups are significantly different as far as knowledge retained is concerned.

The analysis of variance results for the variables PRE, SDS, and POST are presented in Table 4. The same factors were employed in each model as follows.

```
(1)  PRE  = GROUP SESSION(GROUP) PREF GENDER SLEVEL
(2)  SDS  = GROUP SESSION(GROUP) PREF GENDER SLEVEL
(3)  POST = GROUP SESSION(GROUP) PREF GENDER SLEVEL
```

Of these models, only (3) provided a significant result with a P-value of 0.0270. From the P-values ($Pr > F$) associated with the Type III SS, only GROUP was found to be significant with a P-value of 0.0006.

4. Conclusions

The SCP found that there were no significant results for the variables PRE and SDS which suggests that the participants as a whole did not differ significantly with respect to these variables. There was a significant GROUP effect for the variable POST, suggesting that augmentation of teaching methods through the use of learning-style

preferences would improve retention of the material for a similar seminar. It is the conclusion of the SCP that the investigation performed by the client has produced some significant results that would be worth pursuing in a larger-scale study.

References:

[1] SAS Institute Inc. (1990) *SAS/STAT User's Guide.* Version 6. Cary, NC.
[2] Moore, D.S. and McCabe, G.P. (1993) *Introduction to the Practice of Statistics.* 2nd ed., Freeman Press, NY.

Appendix 1: Tables

List of Tables

Table 1:

Frequency Tables for Categorical Variables
Total Number = 87

GROUP	Number	Percent
Control	43	49.4
Experiment	44	50.6

SESSION	Control	Expt	Number	Percent
S1	11	11	22	25.3
S2	11	12	23	26.4
S3	11	11	22	25.3
S4	10	10	20	23.0

PREF	Control	Expt	Number	Percent
A	12	9	21	24.1
K	2	2	4	4.6
N	17	15	32	36.8
T	11	12	23	26.4
V	1	6	7	8.0

GENDER	Control	Expt	Number	Percent
F	33	37	70	80.5
M	10	7	17	19.5

SLEVEL	Control	Expt	Number	Percent
E	31	28	59	67.8
H	12	16	28	32.2

Table 2A:

Summary Statistics of Test Scores
Total Number = 87

	PRE	POST	SDS
Min.	5.00	36.00	21.00
1st Qu.	40.00	76.00	42.50
Median	55.00	80.00	50.00
Mean	55.63	80.87	48.05
3rd Qu.	75.00	90.00	55.00
Max.	100.00	100.00	60.00
Std. Dev.	25.31	12.11	8.56

Table 2B:

Summary Statistics of Test Scores by GROUP
Total Number: Control = 43 , Experiment = 44

	Control			Experiment		
	PRE	POST	SDS	PRE	POST	SDS
Min.	5.00	36.00	21.00	5.00	64.00	21.00
1st Qu.	45.00	72.00	42.00	35.00	80.00	44.25
Median	60.00	80.00	49.00	50.00	88.00	51.50
Mean	57.67	76.00	47.12	53.64	85.64	48.95
3rd Qu.	80.00	84.00	55.00	75.00	96.00	44.25
Max.	95.00	96.00	59.00	100.00	100.00	60.00
Std. Dev.	24.67	11.58	8.48	26.04	10.75	8.64

Table 3:

t-Test Results
Difference in Average Test Scores across GROUP
Total Number 87: Control = 43, Experiment = 44

Variable: SDS

GROUP	N	Mean	Std Dev	Std Error	Min	Max
Control	43	47.116	8.477	1.293	21	59
Experiment	44	48.954	8.637	1.302	21	60

Variances	T	DF	Prob> $\|T\|$
Unequal	-1.0018	85.0	0.3193
Equal	-1.0016	85.0	0.3194

Variable: PRE

GROUP	N	Mean	Std Dev	Std Error	Min	Max
Control	43	57.674	24.673	3.763	5	95
Experiment	44	53.636	26.045	3.926	5	100

Variances	T	DF	Prob> $\|T\|$
Unequal	0.7425	84.9	0.4598
Equal	0.7421	85.0	0.4601

Variable: POST (*** Significant)

GROUP	N	Mean	Std Dev	Std Error	Min	Max
Control	43	76.000	11.580	1.766	36	96
Experiment	44	85.636	10.751	1.621	64	100

Variances	T	DF	Prob> $\|T\|$	
Unequal	-4.0202	84.2	0.0001	***
Equal	-4.0237	85.0	0.0001	***

Table 4:
ANOVA Results
General Linear Models Procedure
Number of Observations = 87

$$\left.\begin{array}{ll} \text{Model 1:} & \text{SDS} \\ \text{Model 2:} & \text{PRE} \\ \text{Model 3:} & \text{POST} \end{array}\right\} = \begin{array}{l} \text{GROUP} + \text{SESSION(GROUP)} + \\ \quad \text{PREF} + \text{GENDER} + \text{SLEVEL} \end{array}$$

Model 1: SDS (No Significant Factors)

Source	DF	Sum of Squares	Mean Square	F-Value	$Pr > F$
Model	13	876.6456	67.4343	0.91	0.5492
Error	73	5423.1705	74.2900		
Corrected Total	86	6299.8161			

Source	DF	Type III SS	Mean Square	F-Value	$Pr > F$
GROUP	1	65.0331	65.0331	0.88	0.3526
SESSION(GROUP)	6	707.4197	117.9033	1.59	0.1631
PREF	4	80.4978	20.1245	0.27	0.8958
GENDER	1	5.4309	5.4309	0.07	0.7876
SLEVEL	1	10.6014	10.6014	0.14	0.7067

Model 2: PRE (No Significant Factors)

Source	DF	Sum of Squares	Mean Square	F-Value	$Pr > F$
Model	13	12281.0661	944.6974	1.61	0.1017
Error	73	42809.1638	586.4269		
Corrected Total	86	55090.2299			

Source	DF	Type III SS	Mean Square	F-Value	$Pr > F$
GROUP	1	23.3249	23.3249	0.04	0.8425
SESSION(GROUP)	6	4339.8085	723.3014	1.23	0.2993
PREF	4	4465.2543	1116.3136	1.90	0.1189
GENDER	1	6.4008	6.4008	0.01	0.9171
SLEVEL	1	689.2741	689.2741	1.18	0.2819

Model 3: POST (** GROUP ** Factor Fignificant)

Source	DF	Sum of Squares	Mean Square	F-Value	$Pr > F$	
Model	13	3389.7547	260.7504	2.06	0.0270	**
Error	73	9231.8545	126.4638			
Corrected Total	86	12621.6092				

Source	DF	Type III SS	Mean Square	F-Value	$Pr > F$	
GROUP	1	1644.6794	1644.6794	13.01	0.0006	**
SESSION(GROUP)	6	880.1432	146.6905	1.16	0.3372	
PREF	4	295.3252	73.8313	0.58	0.6753	
GENDER	1	50.6251	50.6251	0.40	0.5289	
SLEVEL	1	12.4934	12.4934	0.10	0.7542	

Appendix 2: The Data

GROUP = Control							GROUP = Experiment						
Order: SESSION, PREF, GENDER, SLEVEL, PRE, POST, SDS													
S1	V	F	H	25	64	51	S1	A	F	H	40	96	52
S1	K	F	E	65	64	21	S1	T	F	E	5	64	29
S1	A	F	E	5	60	41	S1	N	F	E	20	64	54
S1	A	M	E	60	76	56	S1	N	M	E	95	92	57
S1	T	M	E	65	76	51	S1	T	F	H	75	92	53
S1	T	M	E	95	88	55	S1	A	F	E	70	100	46
S1	T	F	H	85	84	50	S1	T	F	E	40	84	21
S1	N	F	E	85	72	42	S1	N	M	H	80	96	37
S1	A	M	E	65	76	34	S1	N	F	E	25	76	52
S1	N	F	H	50	68	42	S1	V	F	E	20	92	38
S1	N	M	E	5	36	36	S1	V	F	E	5	96	44
S2	T	F	H	65	80	56	S2	A	F	E	15	76	54
S2	T	F	E	15	76	47	S2	N	F	H	50	80	48
S2	A	M	E	20	80	44	S2	V	M	H	50	88	57
S2	A	F	E	85	76	51	S2	N	F	H	25	88	41
S2	A	F	E	20	80	56	S2	T	F	H	50	96	48
S2	T	F	E	45	80	58	S2	N	F	E	80	92	49
S2	N	F	E	55	76	54	S2	N	M	E	60	68	43
S2	A	F	H	50	80	42	S2	T	F	E	40	92	52
S2	A	M	H	15	44	39	S2	N	F	E	35	96	32
S2	N	M	E	90	88	52	S2	N	F	E	90	92	60
S2	A	F	E	55	72	44	S2	V	F	H	25	88	52
							S2	A	F	H	50	80	42
S3	N	F	H	50	64	41	S3	A	F	H	70	72	56
S3	N	F	H	45	80	50	S3	T	F	H	35	68	49
S3	N	F	E	90	88	40	S3	N	F	H	75	96	48
S3	N	F	E	70	80	49	S3	A	M	H	25	100	45
S3	T	F	E	85	84	59	S3	N	F	H	40	84	56
S3	T	F	H	70	76	46	S3	K	F	E	75	96	60
S3	N	F	E	80	68	59	S3	N	F	E	80	96	51
S3	T	F	H	45	88	52	S3	A	F	E	45	92	45
S3	N	F	H	80	80	42	S3	T	F	E	95	80	57
S3	N	F	E	75	84	56	S3	T	F	E	60	88	37
S3	N	F	H	70	84	32	S3	T	F	E	100	96	59
S4	N	F	E	50	80	50	S4	A	F	E	80	92	56
S4	N	M	E	45	64	43	S4	T	M	H	65	92	58
S4	N	F	E	85	96	58	S4	A	F	H	30	80	52
S4	A	F	E	95	80	56	S4	N	F	E	60	64	49
S4	A	F	E	75	84	44	S4	K	F	E	70	64	60
S4	T	M	E	45	60	35	S4	N	M	E	40	80	54
S4	T	F	E	55	88	58	S4	V	F	E	35	80	52
S4	K	F	E	60	76	45	S4	T	F	E	100	96	51
S4	N	F	E	55	88	49	S4	V	F	E	60	84	43
S4	A	F	E	35	80	40	S4	T	F	E	75	80	55

4.8 Postscript

We received the client's methods and results chapter that we had agreed to review, along with a check for payment. (Clients can pull the right strings too!) In these types of projects, making ourselves available for certain post-completion services is good "PR" (public relations) and usually does not take much time or effort. (If it does, we can always charge for it.) Below is a modified transcript of the informal, first-name basis, email review we provided the client. The reference to Table 1 pertains to the client's document.

A,
Some notes/suggestions for your methods chapter ...
[page/line numbers as per the document you sent me]
Hope this is helpful.
Regards,
Z

---

page 1: *The dependent variables were the subjects' mean ...*

The dependent variables are really the "actual" scores since you only gave one POST/SDS-test; i.e., you didn't average the scores from "several" POST tests.

page 3: *... and (c) Analysis of Variance (ANOVA) ...*

What you have written seems fine. Remember that ANOVA is employed when:

- a factor has more than two levels
- the simultaneous effect of several factors is to be considered.

Note that the actual procedure we used was: general linear models (GLM) — the difference between GLM and ANOVA is simply that GLM takes account of the "unequal cell sizes" (i.e., GLM accounted for the different number of Male/Female, Elem/High School, etc.). ... In effect, GLM was nothing more than a "correct" ANOVA analysis.

page 3: [line 4 ↑] ... *experimental groups (P-value > 0.05).*

I would suggest expanding the concept of a "P-value" ... perhaps end the first sentence with:

" ... experiment group." Then add something like:

To assess the statistical significance (or lack of) associated with a particular test, such as the two-sample t-test employed in Table 1, we may use the "P-value" criterion. Standard statistical practice has adopted 0.05 as the cutoff criterion for assessing significance (P-value < 0.05). As can be seen from Table 1, both (*) values listed under "Prob$> |T|$" exceed 0.05. This implies that the Control and Experiment groups did not differ (significantly) in terms of ...

(*) [A possible footnote to explain why there are two t-tests.]

The two-sample t-test can be conducted under the assumption that the variability of pretest scores within the Control and Experiment groups are the same (denoted by "Equal" in Table 1). Although this assumption can be considered reasonable in this study (and is supported by comparable sample standard deviations from the two groups), the two-sample t-test can also be conducted without making the assumption that the variability is the same ("Unequal"). In either case, the P-value criterion can be applied to assess statistical significance.

page 4: [line 10] ... *different* ($\mathrm{p} < 0.0004$).

The P-value is equal to 0.0004 (not less than 0.0004). Rephrase as:

" ... different (P-value of 0.0004 (*))."

(*) [Possible footnote.]

This result is considered "strongly" significant. ... (since it is smaller than 0.01 which is often used to qualify significance beyond the standard 0.05 criterion).

page 5: [line 12] typo: (P-value of 0.0031) $\longrightarrow$ (P-value of 0.0013).

page 6: Seems fine.

In the next chapter you could postulate "further research directions" to investigate the SDS variable. Clearly, you did not have enough data in this study to perform a meaningful analysis of SDS by PREF (too few people in the K, V groups and not really enough in the A group once you split SDS by GROUP with $\mathrm{PREF} = \mathrm{N}$ excluded). ... Your next thesis maybe?

Questions

1. One of the main issues we excluded in our presentation of this case study example concerned the PREF = N paticipants.

   IMPORTANT: You may want to read the next question before starting any part of this question.

   (a) Discuss the statistical issues associated with the PREF = N class with the client. Recall that these participants were "assigned" a preference in the Experiment group.

   (b) An obvious approach is to analyze the response variables with respect to the separate subsets: PREF = N and PREF $\neq$ N. What problems arise? Is there a better way to approach the analysis?

   (c) Conduct the analysis suggested above. Are your conclusions different for the two subsets? Should they be?

2. Suppose you were asked to create a detailed invoice for the client. Your analysis of the PREF = N issue above will be very helpful here: document every task you perform, and note the time it takes to complete each task. Alternatively, just use our presentation of the case study entirely as your basis for creating the invoice. Assume both consultation sessions took a full hour.

   (a) The first step is to itemize every cost-related component of the consultation process. Where did we actively engage in working on the project? What task did we do?

   NOTE: Do not include capital costs. The client isn't going to pay for your new printer **and** your consultation fee!

   (b) Attach a time unit (use tenths of an hour) to each component. Don't forget little things like the 12 minutes we spent arranging the initial consultation session and getting the prior information. You will need to estimate times for many of these components.

   [Hence the value of documenting each step (and time taken) in your analysis of the PREF = N issue.]

   (c) Allocate your hourly rate, assuming you did all the work. Your rate would normally include an overhead which absorbs incidental costs and can be put towards capital improvements. Calculate the total cost to the client.

   Is this amount realistic for $n = 87$ observations?

   (d) Adjust this total cost by providing the first consultation hour free of charge and allocating appropriate tasks to an "assistant" (with a lower hourly rate). Divide the adjusted total cost by

$n = 87$. When might this per observation cost be useful? Not useful?

3. In view of our Remarks on page 171, it may seem surprising that we retained the "Unequal Variance" entries in the Preliminary and Final Reports. Note that the `Std Dev` values in Table 3 for the Experiment and Control groups are very similar within each response variable.

   (a) Is there any reason why we should **not** have deleted the Unequal Variance entries?

   (b) No test for normality appears to have been performed with regard to the t-test or ANOVA analyses. How would you explain the results of this test to the client?

      (i) What are the consequences of a significant result? (That is, the normality assumption is rejected.)

      (ii) How would you incorporate the issue of "Robustness" in your explanation?

4. For S-PLUS users.

   (a) Analyze these data using S-PLUS. Note that Tables 2A and 2B were actually produced using the S-PLUS `summary()` function. What are the differences in the output of the t-test and ANOVA results from S-PLUS as compared to SAS?

   (b) Complete the function `g.hist2.fun()` outlined on page 172. The objective is to be able to produce either the PRE by GROUP (Figure 4.3), or the POST by GROUP (Figure 2.3) histogram display using only the arguments provided in the options line of your `g.hist2.fun()` function.

   (c) The client requested presentation quality histograms for several variables. On each plot, the client wanted the value of the mean and standard deviation of the variable printed on the graph (within the plot region) with the labels: "Mean = `<value>`" and "Standard Deviation = `<value>`" Consider the following options.

      (i) Create a generic function that automatically determines a "good" place to put this text within the plot region of any histogram display.

      (ii) Position the text inside the plot region of each histogram by trial and error.

      Which approach should we take?

http://www.springer.com/978-0-387-98863-4

Statistical Consulting
Cabrera, J.; McDougall, A.
2002, XII, 390 p., Hardcover
ISBN: 978-0-387-98863-4