PHY340 Data Analysis Feedback:

Group A01 doing Problem A1

# Data Analysis

The data analysis appears to be valid, although it is very poorly explained in places; however, the calculations of uncertainties and χ2 are not only poorly explained (if at all) but certainly are *not* valid. In parti­cu­lar, the errors quoted in tables 1, 2 and 3 are not explained anywhere, and in some cases clearly do not make sense: for example, the bulge masses for NGC 4051, 4151 and 5548 cannot possibly have errors that are larger than the central value (what would a negative bulge mass mean?). Although sources are given for these values, the sources themselves do not in fact quote any formal error on the bulge masses. Wandel (2002) says that he assumes that the error in the bulge luminosity is a factor of 2, i.e. (−50%, +100%), and that he calculates the mass using . This, however, makes no sense (since *M*/*M*⊙ is of order 1010, subtracting 1.11 would be a pointless exercise), and in fact his cited source, Magor­rian (1998), actually quotes . Since the numbers in Wandel’s Table 1 match this relation, we can assume that the version without the logs is a typo. Assuming a factor 2 uncertainty in *L* then gives uncertainties of ±0.30 in log(*L*/*L*⊙) (since log 2 = 0.301), and hence ±0.35 in log(*M*/*M*⊙). Bian and Zhao (2003) do not quote any uncertainty for Akn 564, but they are using the same relation as Wandel (2002) so it would seem sensible to assume the same uncertainty.

Furthermore, the uncertainties quoted in different places in the report are not self-consistent. For example, the error bars in figure 4 certainly do not match any of the errors quoted in table 3, and the various values quoted for the black hole mass in NGC 5548 in the text (page 2), table 1 and table 2 do not seem to bear any relation to each other; the values quoted in tables 3 and 4 have the same central value, but wildly different uncertainties. This does not give the reader any confidence in your analysis.

As explained in the lectures, the χ2 of a fit is given by   
where *yi* is the *i*th measured point, *y*fit is the corresponding value calculated from the fit, and *σi* is the error on *yi*. It is *not* possible to quote a “χ2” for an individual point, so I have no idea what the values in the last column of table 3 are supposed to be; also, the quoted expression for χ2,   
is *only* valid when the *y* values are Poisson variables (so that ). As this is not the case here, this expression gives meaningless nonsense, which you should have realised (since χ2 per degree of freedom should be of order 1, not of order 0.001).

The appropriate way to weight data with different error bars for fitting, assuming that they are approximately Gaussian, is by , the inverse square of the *absolute* error. Weighting by the inverse square of the *fractional* error makes no sense at all (for example, consider what would happen if the central value were 0.0 or very close to it). This was also explained in the lectures. However, the best way to deal with weighted data is sim­ply to use a fitting program which can take the errors of individual points into account, such as Python’s curve\_fit. An example of the use of curve\_fit was presented in the lectures.

Finally, *one can fit a straight line to* ***any*** *dataset*: this does *not* imply that the data are well des­cribed by a straight line. It is clear to any reasonable person that the straight line fit in your figure 4 does *not* do a good job of describing the data: in fact, the data do not show any dis­cernible trend at all. This is not entirely your fault, although some of your black hole masses do not seem very accurate: the data from Wandel (2002) do not show much of a trend either (see figure 1 below).

*Figure 1: BH−bulge relation using data from Wandel (2002), except Akn 564 (red point) from Bian and Zhao (2003). The uncertainties are a factor of 3 on the black hole mass and a factor of 2 on the bulge luminosity, as suggested by Wandel (2002). The straight line has a gradient of 0.29±0.22 and an intercept of 8.2±1.6; this is essentially consistent with no scaling at all.*

Note that the gradient of the straight line is consistent with zero: given the uncertainty of ±0.22, the probability of obtaining a result at least as far from zero as 0.29 is 19%, which is not signi­ficant by any measure. Therefore, it is certainly *not* possible to claim that this result “demon­strates the correlation between black hole mass and bulge mass”, since it is entirely consistent with there being no correlation whatsoever. (The fit result is not consistent with the result of quoted by Wandel (2002), but he is using a much larger dataset.) You do not quote the gradient of the fit you show in figure 4, much less its uncertainty, but it looks very similar, and I am sure that the conclusion is the same: these data do not show a statistically significant correlation.

Average grade for this section: 24.75/50.

# Data Presentation

The figures and tables are properly numbered, and have captions—though in many cases the captions could usefully present more information (if you look at captions in published papers, you will see that yours are much too short). Giving the references to values taken from the li­tera­ture in the figure captions is good practice, although in some cases the references are a little misleading: you may have taken the central values of your bulge masses in table 3 from Wandel, but certainly not the nonsensical error values! However, there are some significant issues.

The key points in presenting data are that the presentation should be *clear* and *accurate*, and that it should be as easy as possible for the reader to understand. Generally, graphical presen­tations are easier to grasp than long tables: it is therefore common for plots to be in the main body of the text and tables in the appendix (or, these days, in the “additional online material”). It is therefore unclear why figure 4 is in an appendix while table 3 is in the main body. It is also unclear why figure 1 is repeated as figure 2: there is absolutely no point to this. The axis labels on the figures are too small, and in figure 1/2 no units are given. The parameters of the straight line fit in figure 4 should be given, either in the caption or in the text, and compared with values from the literature (they will not agree, but that is not a good reason not to quote them).

None of the error estimates is explained, some are clearly nonsense, and all are inconsistent be­tween tables 2 and 3. In table 1 and the text on page 2, the error bars and the central value are not quoted to appropriate numbers of significant figures: “1.03±0.0285” makes no sense (what are the last two figures in the error being added to or subtracted from?). Where explanations *are* present, they are often unclear (what do you mean by “a two-dimensional data set”, and why should this result in incorrect values for the wavelength axis?) or garbled (“it was required that only values for the flux were transferred into the software” can­not *possibly* be right—how can you fit a line shape with only fluxes and not wavelengths?).

Average grade for this section: 15.8/30.

# Style

There is a clear attempt to present the report in the appropriate style: the section headings are reasonable and you have tried to write in the correct formal English. However, there is evidence of lack of proof-reading (the report should *not* have been submitted with the last paragraph reading “…by the equation (equation linking velocity dispersion to Mbh)”!), and some reorganisation was called for: if you are going to quote the (wrong!) expression for χ2, you should quote it where it is first used, not append it randomly at the end; figure 1 should not be repeated as figure 2; there should be a conclusion; the statement “χ2 = 0.124” should not appear at random in the middle of the text. The early part of the report is under-referenced: the introduction should have at least five additional references, for example. There are some clumsy sentences and a number of grammatical errors, along with some odd word choices: you might have benefited from a session with the Writing Advisory Service. As noted above, the choice of what to include in the main text and what to relegate to an appendix is a bit strange.

Average grade for this section: 10.9/20.

Overall average grade: 51.45%.