Biost 517: Applied Biostatistics I
Emerson, Fall 2012

Homework #3
October 12, 2012

Written problems: To be handed in at the beginning of class on Friday, October 19, 2012.

In Homework #2, I gave you a dataset in which multiple measurements were made on each subject, but I (stupidly) had you generate statistics on all *measurements*, rather than on a *patient* basis. In this homework, we will better address such a situation.

This homework assignment deals with the audiology data set posted on the web pages (audio.doc and audio.csf). Unlike most data files I give you, this data file is supplied in "comma separated value" format. Instructions to read the data into Stata and R are given in the documentation file. Note that you will need to convert dates to Julian dates:

```
g intStartDate = date(StartDate, "MDY")
g intStopDate = date(StopDate, "MDY")
g intVisitDate = date(VisitDate, "MDY")
```

Then you will likely want to create a variable indicating whether measurements were made post randomization and whether a patient was currently taking the assigned drug at the time of a measurement being made. The following code could be used to create variables *postRand* and *onDrug* that indicate this:

```
g postRand = 0
replace postRand = 1 if intVisitDate > intStartDate
replace postRand=. if intVisitDate==. | intStartDate==.
g onDrug = 0
replace onDrug = 1 if postRand==1 & intVisitDate <= intStopDate
replace onDrug=. if intVisitDate==. | intStartDate==. | intStopDate==.
```

For no particular scientific reason, we are going to focus only on the measurements made for hearing levels at 4000 Hz in both the left and right ears. (I figure one frequency is enough to illustrate the issues with "longitudinal studies".) When giving instructions in Stata, I will only give examples for the right ear, figuring that you can also do the commands for the left ear on your own. Any time I refer to "hearing levels", I will mean at "hearing levels at 4000 Hz". Unless I specifically state otherwise, you should consider the right and left ears separately.

Questions for Biost 514 and Biost 517:

General Comments:

The webpages contain a file containing annotated Stata code that was used to answer this question. Note that not all analyses that I did made it into the key as a table. It is perfectly okay to do an analysis and summarize its results in a sentence. (I also include on the webpages a separate file containing a tutorial on how I used R to answer the questions of this homework.)

Also included in the annotated Stata code is an analysis that was directed toward assessing whether the earliest measurement on each subject was prior to randomization. This is of course a very important scientific issue, though I did not ask you to examine such things. Some of you did this spontaneously, which is a very good habit to get into. I comment more

on these issues below. Aspects that I did not necessarily expect you to explore I put in italics.

Also provided on the course webpages is an Excel spreadsheet that I used to format the data. Of special note is the formatting of tables I wanted in the key versus the formatting I got in Stata with the analyses that I ran.

- **I am willing to bet that most of you ran separate analyses for each set of variables (mean post randomization, mean difference from baseline post randomization, mean on drug, mean difference from baseline on drug, etc.).**
  - **If you did this, formatting of tables in Excel is straightforward if you just copy the tabular output in successive rows in Excel.**
  - **However, I used other approaches just to illustrate how you can re-order the row using Excel and the INDIRECT() function in Excel.**
  - **You may well never want to use the approach I did, so ignore this if you want.**
- **Note for ease of Stata programming, I often generated descriptive statistics for multiple variables by dose. This then groups all variables under dose.**
- **However, the greatest interest in a randomized clinical trial (RCT) is to compare variables across dose groups.**
- **I regard the measurements on the right and left to be more-or-less replications for the purposes of systemic drug toxicity, so I do not mind keeping the measurements on the R and L together under each dose.**
- **But I want to keep all the mean measurements post-randomization together, all the mean difference measurements post-randomization together, etc.**
- **By creating a column indicating the row of Stata output to be used in my table row, I could write one line of formatting instructions and then copy it to other rows.**

1. In Homework #2, you generated descriptive statistics using all measurements in the dataset. However, multiple measurements were made on each subject. This problem guides you through the process of using Stata to determine how many repeat measurements are made on each individual.

   The data file contains repeated measurements on each individual. When our interest is on how patients fare, we often combine such repeated measurements into a single summary. For instance, we might consider taking the average of the measurements, the maximum or minimum of the measurements, or only the last measurement. Stata provides a command "egen" that will allow us to easily abstract such summaries by patient.

   For instance, suppose we want the mean hearing level at 4000 Hz for each patient's right ear. We can obtain a variable *mnR4000* that will contain that by:
   - `egen mnR4000 = mean(R4000), by(Subject)`

   Each row will now have a value for variable *mnR4000* that is equal to the mean of all the hearing levels in the right ear at 4000 Hz for that patient. If you wanted to have instead the mean of hearing measurements made after randomization (so after the start date) you could use:

   ```
   egen mnR4000 = mean(R4000) if intVisitDate > intStartDate, by(subject)
   ```

   After this command, you would have a variable that had missing values for any rows corresponding to visits at or before the start date, and for all other rows, the value for

variable *mnR4000* would be equal to the mean of all hearing values made after the drug start date for that patient.

In this and the following problems you will need to use "egen" repeatedly in order to be able to perform analyses on a per patient rather than per measurement basis.

a. Use "egen" to generate a variable *nR4000* counting the number of non-missing hearing levels at 4000 Hz made for each individual's right ear, and provide suitable descriptive statistics for this variable using all cases in the datafile. The following Stata code will generate the variable:

```
egen nR4000= count(R4000) if R4000!=., by(Subject)
```

How many measurements in the datafile correspond to a patient having a maximum of 6 nonmissing hearing measurements? How many patients does this represent? You might consider either or both of the following Stata commands:

```
table nR4000
list Subject R4000 nR4000 if nR4000==6
```

Answer: There were 192 rows in the datafile corresponding to patients who had 6 visits. But as each such patient was represented 6 times, that means there were 192 / 6 = 32 patients who had 6 visits.

b. As can be seen in part a, doing descriptive statistics on the summarized variable is still complicated due to the number of repeated measurements on each individual. If we want to find out the distribution of *nR4000* across patients (rather than rows in the file), we will need to restrict our analysis to one row for each patient. Generate a variable *mindate* containing the date of the earliest visit for which a subject has a row in the data set, and provide summary statistics to show that each subject only has one row in the dataset for that date. The following Stata code can be used to generate *mindate*:

```
egen mindate=min(intVisitDate), by(Subject)
```

Answer: No subject was missing a visit date for any visit, and each of the 117 subjects had only one row in which the visit date was equal to that subject's earliest visit, as determined by tabling the subject identifiers under the restriction that the visit date was the earliest visit.

An aside: I had you use the earliest visit date to identify a unique visit for each subject, because I am sure that everyone by definition has an earliest visit. However, when dealing with human subjects (not to mention human investigators), we cannot always count on having all the measurements we were supposed to have at all visits. In this data set, a total of 5 patients (2 placebo, 1 dose 0.25, 2 dose 0.5) had a visit labeled Visit 0 occurring after their start date (2, 2, 9, 11, and 24 days after randomization). A further 3 patients (2 placebo, 1 dose 0.25) had no Visit 0 listed, and their earliest visit date corresponded to Visit 4 occurring 35, 40 or 43 days after randomization.

It is not good that patients might miss their protocol specified visits, but it does in fact happen. And having missed, for instance, a week 4 visit, you can never go back and get that information. In this homework, we are just treating the earliest visit for these subjects as if it is a good estimate of those patients' baseline hearing levels. This would need to be made clear in the Materials and Methods of any paper describing the results of the study.

But the take home message: Always try to guess how people might have messed up data collection and look to see if they did.

c. Now, since we know that every individual has a single row corresponding to *mindate*, when we desire statistics on each patient, we could obtain summary statistics just for rows corresponding to *intVisitDate==mindate*. Describe the distribution of the number of measurements made on each subject. Provide descriptive statistics that allow us to compare the number of measurements per patient by treatment group. What might be the scientific importance of any differences between treatment groups? What might be the statistical ramifications of any differences? Are there differences that concern you?

Answer: Table 1 provides descriptive statistics for the number of visits for subjects in each dose group. While the groups have similar average number of visits, the highest dose group had 2 subjects with only 1 visit and 1 subject with 8 visits. Subjects in other dose groups had 2 to 7 visits, thus there is a slightly higher standard deviation for the number of visits in the high dose group. It is possible that any tendency toward lower numbers of visits in the high dose group reflects toxicities (either audiologic or other) that led to a subject dropping out of the study, and that any tendency toward higher numbers of visits in the high dose group reflects more intense audiologic monitoring of patients who are having symptoms of inner ear toxicity.

Table 1: Descriptive statistics for the number of audiology visits for each subject by dose.

	N	Mean (SD)	Mdn (IQR)	(Min, Max)
Dose 0	39	4.33 (1.68)	5 (3, 6)	(2, 7)
Dose 0.25	37	4.19 (1.61)	4 (3, 6)	(2, 7)
Dose 0.5	41	4.07 (1.81)	4 (2, 6)	(1, 8)
Combined	117	4.20 (1.69)	4 (3, 6)	(1, 8)

Statistically, we worry that more frequent measurements might lead to a greater chance to observe more extreme measurements even when there was no true tendency toward greater hearing damage. Furthermore, if the frequency of measurements was dictated by the results of the measurements themselves, this could potentially be biasing: Abnormally good measurements for a patient might be accepted as evidence of no toxicity, but higher measurements are verified with repeat measurements. Those repeat measurements are subject to variability, and thus we are in essence trying to get some random lows.

d. (There is nothing to answer in this part, it is purely informational.) An alternative approach to find a unique row for each patient is to use the "tag" function in "egen", which will tag a unique row for each Subject. The following Stata code can be used to generate variable *tag*, and then obtain descriptive statistics for *nR4000* on a per patient basis:

```
egen tag=tag(Subject)
tabstats nR4000 if tag, stat(n mean … max)
```

2. Generate a variable *mR4000* reflecting the average of all hearing measurements made for each individual (both before and after randomization).

a. Provide summary statistics for both *R4000* and *L4000* and *mR4000* and *mL4000* for the three treatment groups using all available data in the data set. What scientific question could be addressed using these descriptive statistics?

Answer: Table 2a provides descriptive statistics for the hearing thresholds for every row in the data file. We obtain the same mean whether we take the mean of all visits in the data set, or whether we first compute subject specific means, and then replace each visit's measurement with the subject specific mean. The standard deviation, however, is less when we use the subject specific means, because we have removed the "within subject variability". Similarly, the minima and maxima are less extreme for the subject specific mean measurements.

There is no particularly interesting scientific question that could be answered by this data analysis:

- **Subjects are not weighted equally, so we are not gaining information on patients.**
- **While you might imagine that an audiologist would be interested in the range of measurements that would typically have to be sampled (and thus they might want to weight more heavily those subjects who are measured more frequently), the protocolized RCT setting may not generalize to a typical clinic population's schedule of measurements and severity of hearing impairment.**

Table 2a: Descriptive statistics for the hearing threshold (in dB) at 4000 Hz when computed for each row in the data file. Statistics are computed for right and left ears separately within dose groups.

		N	Mean (SD)	Mdn (IQR)	(Min, Max)
Dose 0	**Each visit: Right**	169	42.1 (25.9)	35.0 (20.0, 65.0)	(0.0, 95.0)
	Each visit: Left	169	46.4 (25.2)	45.0 (25.0, 65.0)	(0.0, 95.0)
	Subject mean: Right	169	42.1 (25.7)	35.8 (20.0, 65.8)	(6.7, 90.8)
	Subject mean: Left	169	46.4 (25.0)	48.0 (25.0, 64.2)	(4.2, 92.5)
Dose 0.25	**Each visit: Right**	155	43.8 (25.7)	40.0 (25.0, 60.0)	(5.0, 110.0)
	Each visit: Left	155	46.4 (25.2)	40.0 (25.0, 60.0)	(10.0, 105.0)
	Subject mean: Right	155	43.8 (25.3)	40.0 (27.0, 57.5)	(7.0, 107.0)
	Subject mean: Left	155	46.4 (24.9)	40.0 (29.0, 58.8)	(10.0, 99.0)
Dose 0.5	**Each visit: Right**	167	41.2 (22.7)	35.0 (20.0, 60.0)	(-5.0, 85.0)
	Each visit: Left	167	43.7 (21.9)	40.0 (25.0, 60.0)	(5.0, 85.0)
	Subject mean: Right	167	41.2 (22.2)	35.8 (22.5, 60.6)	(2.5, 84.3)
	Subject mean: Left	167	43.7 (21.4)	44.2 (25.8, 65.0)	(11.0, 83.0)
Combined	**Each visit: Right**	491	42.4 (24.7)	35.0 (20.0, 60.0)	(-5.0, 110.0)
	Each visit: Left	491	45.4 (24.1)	45.0 (25.0, 65.0)	(0.0, 105.0)
	Subject mean: Right	491	42.4 (24.4)	36.7 (23.3, 60.6)	(2.5, 107.0)
	Subject mean: Left	491	45.4 (23.8)	42.5 (25.8, 64.2)	(4.2, 99.0)

b. Provide summary statistics for *mR4000* and *mL4000* for the three treatment groups when each patient is represented only once. What scientific question could be addressed using these descriptive statistics?

Answer: Table 2b provides descriptive statistics for the hearing thresholds for every subject in the data set. Statistics are now presented for the earliest visit by each patient, as well as for the subject specific mean measurement across all visits. We now obtain different

statistics according to whether we are taking only the earliest measurement for each subject or whether we are taking the average value for each subject.

While we would expect the patient specific means to be less variable than a single measurement on each individual (again owing to damping the within subject variability), the fact that the measurements made at different times may be affected by aging, drug effects, other hearing damage, etc., it is difficult to predict whether the added measurements would add more precision by reducing "random error", or add more variability by including other "systematic" variability due to other effects on hearing.

While not necessarily the optimal choice, these statistics might be expected to provide some information about drug effects on hearing. (Even though they often include the baseline values for each subject, any post randomization differences would be expected to affect the overall means.)

Table 2b: Descriptive statistics for the hearing threshold (in dB) at 4000 Hz when computed for each subject in the data file. Statistics are computed for the hearing level at each subject's earliest visit as well as the subject specific mean of all visits. Statistics are presented for right and left ears separately within dose groups.

		N	Mean (SD)	Mdn (IQR)	(Min, Max)
Dose 0	**First visit: Right**	39	43.2 (26.0)	35.0 (20.0, 65.0)	(5.0, 90.0)
	First visit: Left	39	44.9 (25.2)	45.0 (20.0, 65.0)	(5.0, 90.0)
	Subject mean: Right	39	43.6 (26.3)	36.7 (22.5, 65.8)	(6.7, 90.8)
	Subject mean: Left	39	46.2 (25.5)	48.0 (22.5, 64.2)	(4.2, 92.5)
Dose 0.25	**First visit: Right**	37	41.4 (24.1)	35.0 (25.0, 50.0)	(5.0, 105.0)
	First visit: Left	37	43.9 (23.4)	40.0 (25.0, 55.0)	(10.0, 105.0)
	Subject mean: Right	37	41.5 (24.5)	37.5 (25.0, 50.0)	(7.0, 107.0)
	Subject mean: Left	37	44.0 (22.9)	40.0 (29.0, 55.0)	(10.0, 99.0)
Dose 0.5	**First visit: Right**	41	35.6 (22.3)	30.0 (20.0, 55.0)	(5.0, 85.0)
	First visit: Left	41	39.5 (21.0)	35.0 (20.0, 55.0)	(10.0, 85.0)
	Subject mean: Right	41	37.5 (21.9)	31.7 (20.8, 57.5)	(2.5, 84.3)
	Subject mean: Left	41	41.2 (20.5)	40.0 (21.7, 55.0)	(11.0, 83.0)
Combined	**First visit: Right**	117	40.0 (24.2)	35.0 (20.0, 55.0)	(5.0, 105.0)
	First visit: Left	117	42.7 (23.2)	40.0 (20.0, 55.0)	(5.0, 105.0)
	Subject mean: Right	117	40.8 (24.2)	35.8 (22.5, 58.3)	(2.5, 107.0)
	Subject mean: Left	117	43.7 (22.9)	41.7 (25.0, 57.5)	(4.2, 99.0)

3. In problem 2, you took the mean of all hearing measurements for an individual—both before and after randomization. The following code will create a variable *mtrtR4000* which will be the mean of hearing measurements made post randomization. (Note the need to ensure that the first row for each patient, or the "tagged" case if you use that approach, will not have a missing value for *mtrtR4000*.):

```
egen grbg=mean(R4000) if postRand==1, by(Subject)
egen mtrtR4000=mean(grbg), by(Subject)
```

Table 3: Descriptive statistics for individual specific hearing levels when averaged over all measurements made after randomization or averaged over all measurements made while

taking study drug. Statistics are provided for both the absolute measurements made as well as for the change from the earliest measurement for each subject.

		N (Msng)	Mean (SD)	Mdn (IQR)	(Min, Max)
	Mean Hearing Threshold After Randomization				
Dose 0	**Right**	39 (0)	43.7 (26.4)	37.0 (20.0, 66.0)	(7.0, 91.0)
	Left	39 (0)	46.7 (25.6)	48.0 (25.0, 65.8)	(3.0, 93.0)
Dose 0.25	**Right**	37 (0)	41.8 (24.9)	40.0 (25.0, 55.0)	(6.3, 107.5)
	Left	37 (0)	44.0 (22.8)	40.0 (28.3, 55.0)	(10.0, 97.5)
Dose 0.5	**Right**	39 (2)	37.0 (21.9)	31.3 (20.0, 58.3)	(-5.0, 84.2)
	Left	39 (2)	40.5 (20.5)	35.0 (22.0, 52.5)	(11.3, 82.5)
Combined	**Right**	115 (2)	40.8 (24.4)	36.7 (20.0, 60.0)	(-5.0, 107.5)
	Left	115 (2)	43.8 (23.0)	41.0 (25.0, 60.0)	(3.0, 97.5)
	Mean Difference in Hearing Threshold After Randomization				
Dose 0	**Right**	39 (0)	0.49 (4.87)	0.00 (-2.50, 5.00)	(-12.5, 10.0)
	Left	39 (0)	1.83 (4.04)	2.00 (0.00, 5.00)	(-7.0, 11.0)
Dose 0.25	**Right**	37 (0)	0.42 (6.21)	0.00 (-3.33, 4.00)	(-13.3, 20.0)
	Left	37 (0)	0.12 (4.45)	0.00 (-1.00, 3.00)	(-15.0, 8.3)
Dose 0.5	**Right**	39 (2)	2.65 (6.45)	2.50 (0.00, 5.00)	(-15.0, 17.9)
	Left	39 (2)	2.06 (7.00)	1.67 (-2.50, 6.00)	(-10.0, 25.0)
Combined	**Right**	115 (2)	1.20 (5.92)	1.00 (-1.67, 5.00)	(-15.0, 20.0)
	Left	115 (2)	1.36 (5.36)	1.25 (-1.00, 5.00)	(-15.0, 25.0)
	Mean Hearing Threshold on Drug				
Dose 0	**Right**	39 (0)	43.6 (26.2)	37.0 (20.0, 66.3)	(7.0, 91.0)
	Left	39 (0)	46.6 (25.5)	48.0 (25.0, 65.8)	(2.5, 93.0)
Dose 0.25	**Right**	37 (0)	41.8 (24.9)	40.0 (25.0, 55.0)	(6.7, 107.5)
	Left	37 (0)	44.0 (22.8)	40.0 (28.3, 55.0)	(10.0, 97.5)
Dose 0.5	**Right**	38 (3)	37.1 (22.2)	31.9 (20.0, 58.3)	(-5.0, 84.2)
	Left	38 (3)	40.7 (20.8)	37.5 (22.0, 52.5)	(10.0, 82.5)
Combined	**Right**	114 (3)	40.8 (24.4)	36.7 (20.0, 60.0)	(-5.0, 107.5)
	Left	114 (3)	43.8 (23.1)	41.5 (25.0, 60.0)	(2.5, 97.5)
	Mean Difference in Hearing Threshold on Drug				
Dose 0	**Right**	39 (0)	0.36 (4.68)	0.00 (-2.50, 5.00)	(-12.5, 10.0)
	Left	39 (0)	1.73 (4.01)	2.00 (0.00, 5.00)	(-7.5, 11.3)
Dose 0.25	**Right**	37 (0)	0.43 (6.20)	0.00 (-3.33, 4.00)	(-13.3, 20.0)
	Left	37 (0)	0.10 (4.43)	0.00 (0.00, 3.00)	(-15.0, 8.3)
Dose 0.5	**Right**	38 (3)	2.41 (6.36)	1.67 (0.00, 5.00)	(-15.0, 17.0)
	Left	38 (3)	2.06 (7.01)	1.83 (-2.50, 6.00)	(-10.0, 24.0)
Combined	**Right**	114 (3)	1.07 (5.81)	1.00 (-1.67, 5.00)	(-15.0, 20.0)
	Left	114 (3)	1.31 (5.34)	0.92 (-1.00, 5.00)	(-15.0, 24.0)

a. Provide descriptive statistics which compare the treatment groups with respect to the patient specific mean hearing levels post randomization. Based on these statistics, do you worry about any outliers in the data? Explain.

Answer: Table 3 (top section) contains descriptive statistics on the patient specific average hearing thresholds across post randomization visits. Note that two subjects in the highest

dose group have no post randomization measurements, and an additional subject only has post randomization measurements made while no longer taking study drug.

In all cases, the means and medians appear fairly similar and based on the fact that the medians appear to be approximately the mid point of the interquartile range, the central part of the distribution appears to be fairly symmetric. There is somewhat of a tendency for the maximum to be further from the mean and median than the minimum is from those centerpoints, but this is not too striking. (There is certainly a lot of variability in the measurements across people.)

(Note that because negative measurements are in fact possible with these logarithmic scale measurements, we cannot easily use comparisons between the mean and standard deviation to look for outliers.)

b. Provide descriptive statistics which compare the treatment groups with respect to the difference between the patient specific mean hearing level post randomization and the patient's hearing level at randomization. (Note that for the case representing *mindate*, the difference *mtrtR4000 – R4000* is the value we are interested in.)

Answer: Table 3 (second section) contains descriptive statistics on the patient specific average change in hearing thresholds across post randomization visits. These measurements are markedly less variable than the absolute measurements, which argues for high correlation among measurements made on the same subjects. There is no clear toward markedly higher changes in hearing thresholds with higher doses. This is certainly true for the average changes, which at 1 – 2 dB change is substantially what might be noticed by an individual patient (we only measure hearing thresholds in units of 5 dB.) The maximal changes of +25dB are a bit more than the minimal changes of -15 dB, but there do not appear to be striking differences by dose.

c. Create a new variable *mdrgR4000* representing the mean hearing level for each patient while taking study drug (experimental or placebo), and repeat parts (a) and (b) for this measure of treatment outcome.

Answer: Table 3 (third section) contains descriptive statistics on the patient specific average absolute hearing thresholds across post randomization visits while the patient is still taking study drug. The fourth section of Table 3 contains statistics on change from baseline. These statistics differ little from those computed on all post randomization visits.

d. Which of these analyses are scientifically useful in assessing the effect of study drug on hearing levels? Why? What are their relative advantages and disadvantages?

Answer: Given the apparent high correlation of measurements made on the same subject, we would have more precision when we adjust for the baseline measurements in some way. Because this is a safety endpoint, it is highly important that we have the most precision possible to detect any ill effects of the drug. (A sponsor with a financial conflict of interest might be motivated to try to downplay such safety concerns by performing less precise analyses.)

We are hampered in this study due to the 8 subject for whom no pre-treatment audiology levels are available. However, to the extent that we might be comfortable with no short term toxic effects of therapy, using the measurements made soon after randomization might be preferable to ignoring those patients entirely. ***(Taking the difference is not the most precise way to adjust for baseline—a regression model adjusting for baseline is. However, given the high correlation within subjects, taking the difference is probably getting the majority of precision possible in this case.)***

Now, choosing between all post randomization variables and those made on drug is often difficult.

- **Certainly treatments can have long term toxic effects, and ignoring the later measurements made after stopping study drug might miss these.**
- **But it is also possible that any toxic effects might be transient and/or reversible, and by including measurements made off study drug might decrease our ability to see an effect.**

With safety data, we do not usually has as high a burden of proof to establish harm. We just want to find those results that are suggestive of needing additional study. Hence, for a safety endpoint we might do both of these analyses (and more like in problem 4 below). ***(For an efficacy endpoint, we would not be happy with the multiple comparisons' inflation of the inferential type I error that we shall be talking about in coming weeks.)***

4. Now suppose we consider a treatment outcome based on the worst hearing level for each patient, instead of the mean. The following code will create a variable *maxtrtR4000* which will be the maximum of hearing threshold observed post randomization. (Note the need to ensure that the first row for each patient, or the "tagged" case if you use that approach, will not have a missing value for *maxtrtR4000*.):

```
egen grbg=max(R4000) if postRand==1, by(Subject)
egen mastrtR4000=mean(grbg), by(Subject)
```

Table 4: Descriptive statistics for individual specific maximal hearing thresholds across all measurements made after randomization or all measurements made while taking study drug. Statistics are provided for both the absolute measurements made as well as for the change from the earliest measurement for each subject.

		N (Msng)	Mean (SD)	Mdn (IQR)	(Min, Max)
	Mean Hearing Threshold After Randomization				
Dose 0	**Right**	39 (0)	45.9 (26.3)	45.0 (20.0, 70.0)	(10.0, 95.0)
	Left	39 (0)	49.1 (26.2)	50.0 (25.0, 70.0)	(5.0, 95.0)
Dose 0.25	**Right**	37 (0)	44.6 (25.9)	40.0 (25.0, 60.0)	(10.0, 110.0)
	Left	37 (0)	46.8 (23.2)	45.0 (30.0, 60.0)	(10.0, 100.0)
Dose 0.5	**Right**	39 (2)	39.6 (23.5)	35.0 (20.0, 60.0)	(-5.0, 85.0)
	Left	39 (2)	43.5 (21.6)	40.0 (25.0, 55.0)	(15.0, 85.0)
Combined	**Right**	115 (2)	43.3 (25.2)	40.0 (25.0, 60.0)	(-5.0, 110.0)
	Left	115 (2)	46.4 (23.6)	45.0 (30.0, 60.0)	(5.0, 100.0)
	Mean Difference in Hearing Threshold After Randomization				
Dose 0	**Right**	39 (0)	2.69 (5.94)	5.00 (0.00, 5.00)	(-10.0, 15.0)
	Left	39 (0)	4.23 (4.66)	5.00 (0.00, 5.00)	(-5.0, 15.0)
Dose 0.25	**Right**	37 (0)	3.24 (7.29)	0.00 (0.00, 5.00)	(-10.0, 25.0)
	Left	37 (0)	2.84 (4.94)	0.00 (0.00, 5.00)	(-10.0, 15.0)
Dose 0.5	**Right**	39 (2)	5.26 (8.58)	5.00 (0.00, 10.00)	(-15.0, 40.0)
	Left	39 (2)	5.00 (9.03)	5.00 (0.00, 10.00)	(-10.0, 40.0)
Combined	**Right**	115 (2)	3.74 (7.37)	5.00 (0.00, 5.00)	(-15.0, 40.0)
	Left	115 (2)	4.04 (6.55)	5.00 (0.00, 5.00)	(-10.0, 40.0)
	Mean Hearing Threshold on Drug				
Dose 0	**Right**	39 (0)	45.5 (25.8)	45.0 (20.0, 70.0)	(10.0, 95.0)
	Left	39 (0)	48.8 (25.9)	50.0 (25.0, 70.0)	(5.0, 95.0)
Dose 0.25	**Right**	37 (0)	44.6 (25.9)	40.0 (25.0, 60.0)	(10.0, 110.0)
	Left	37 (0)	46.6 (23.2)	45.0 (30.0, 60.0)	(10.0, 100.0)
Dose 0.5	**Right**	38 (3)	39.9 (23.8)	37.5 (20.0, 60.0)	(-5.0, 85.0)
	Left	38 (3)	43.8 (21.8)	40.0 (25.0, 55.0)	(15.0, 85.0)
Combined	**Right**	114 (3)	43.3 (25.1)	40.0 (25.0, 60.0)	(-5.0, 110.0)
	Left	114 (3)	46.4 (23.6)	45.0 (30.0, 60.0)	(5.0, 100.0)
	Mean Difference in Hearing Threshold on Drug				
Dose 0	**Right**	39 (0)	2.31 (5.60)	0.00 (0.00, 5.00)	(-10.0, 15.0)
	Left	39 (0)	3.97 (4.61)	5.00 (0.00, 5.00)	(-5.0, 15.0)
Dose 0.25	**Right**	37 (0)	3.24 (7.29)	0.00 (0.00, 5.00)	(-10.0, 25.0)
	Left	37 (0)	2.70 (4.80)	0.00 (0.00, 5.00)	(-10.0, 15.0)
Dose 0.5	**Right**	38 (3)	5.13 (8.66)	5.00 (0.00, 10.00)	(-15.0, 40.0)
	Left	38 (3)	5.13 (9.12)	5.00 (0.00, 10.00)	(-10.0, 40.0)
Combined	**Right**	114 (3)	3.55 (7.31)	5.00 (0.00, 5.00)	(-15.0, 40.0)
	Left	114 (3)	3.95 (6.53)	5.00 (0.00, 5.00)	(-10.0, 40.0)

a. Provide descriptive statistics which compare the treatment groups with respect to the patient specific worst hearing levels post randomization. Based on these statistics, do you worry about any outliers in the data? Explain.

Answer: Table 4 (top section) contains descriptive statistics on the patient specific maximal hearing thresholds across post randomization visits. Note again that two subjects in the highest dose group have no post randomization measurements, and an additional subject only has post randomization measurements made while no longer taking study drug.

Again, there are no striking examples of outliers evidenced from these descriptive statistics.

b. Provide descriptive statistics which compare the treatment groups with respect to the difference between the patient specific worst hearing levels post randomization and the patient's hearing levels at randomization. (Note that for the case representing *mindate*, the difference *maxtrtR4000 – R4000* is the value we are interested in.)

Answer: Table 4 (second section) contains descriptive statistics on the patient specific maximal hearing thresholds adjusted for baseline values. There is a slight suggestion that the highest dose group might have a greater tendency toward greater hearing loss, though the average loss is not at a level that would necessarily be judged clinically important.

Of note some subjects in the highest dose group were measured to have as much a 40 dB change from baseline, which if persistent would certainly meet criteria for clinical importance. If the number of subjects experiencing such a profound hearing loss is relatively small, it may not affect the dose group mean enough to yield a clinically important population decline, but the individual toxicities would still be of concern.

c. Create a new variable *maxdrgR4000* representing the worst hearing levels for each patient while taking study drug (experimental or placebo), and repeat parts (a) and (b) for this measure of treatment outcome.

Answer: Table 4 (third and fourth sections) contains descriptive statistics on the patient specific maximal hearing thresholds across post randomization visits while still taking study drug. As with question 3, there is not much difference between the analyses made across all post randomization visits and those made only while on study drug.

d. Which of these analyses are scientifically useful in assessing the effect of study drug on hearing levels? Why? What are their relative advantages and disadvantages?

Answer: A drug might have many different effects on hearing over time:

- **The drug might cause an immediate permanent loss of hearing.**
- **The drug might cause a delayed, but permanent loss of hearing.**
- **The drug might cause a transient loss of hearing that lasts for a variable time period in different patients and then returns to baseline levels even while treatment continues.**

- **The drug might cause a reversible loss of hearing (either immediately or after some delay) while taking the study drug, but that hearing returns to baseline after stopping the drug.**
- **The drug might have no effect whatsoever on hearing.**
- **The drug might act differently in different people, with there being some subsets of people for each of the above patterns.**

No single statistic is going to capture all of these patterns. But we might expect

- **Looking at the maximum will find both transient and permanent effects whenever they occur.**
- **Looking at the mean will tend to downweight effect that are only temporary and accentuate effects that are long lasting.**
- **Looking at the final measurement made while on study drug will ignore effects that are temporary (unless patients drop off for other reasons).**
- **Looking only at measurements made after stopping study drug will be looking for permanent effects (i.e., irreversible).**
- **Looking at means or proportion above thresholds at each visit time may help us characterize population average trends, though it will smear out transient effects that happen at different times in different people or that happen according to cumulative dose received.**

In this homework, I just wanted you to think a little bit about the difficulties in summarizing measurements made over time in subjects, especially when the schedule of measurements differs across subjects.

e. What additional problem might be posed by using the maximum rather than the mean as was used in problem 3?

<u>Answer:</u> Differences in the number and timing of measurements can lead to spurious identification of trends in minima and maxima, because sample size alone affects the distribution of sample minima and sample maxima. Even if the underlying distribution of measurements is the same in all groups, groups having more measurements are expected to have lower lows and higher highs..

Questions for Biost 514 <u>only</u>:

5. In the above questions I (hopefully) seem preoccupied with the multiple measurements made on each individual. The question is whether this perseveration is just pre-senile dementia or whether it is justified.
    a. Show that if my true interest is differences among patients, the sample mean computed using all observations without regard to patient can be either a biased or unbiased estimate of the mean of the patient specific values. (*Hint: Consider the setting in which my sampling is "balanced" (equal numbers of measurements on each subject) versus the setting in which my sampling is "unbalanced" (some subjects might have more measurements than others).)*
    b. Show that if my true interest is differences among patients, the sample variance computed using all observations without regard to patient can be either a biased

or unbiased estimate of the variance of the patient specific values. (*Hint: Again, consider the balanced and unbalanced settings.)*

c. Show that if you obtain an unbiased estimate of the population variance, your estimate of the population standard deviation is biased. (*Note that this result is unrelated to whether or not you have repeated measurements on the subjects.)*

d. Later in the course, we will be interested in making statistical inference about the population means. Central to much of that inference will be the standard error of the mean, which for sample means computed from independent data is the standard deviation divided by the sample size. Explain how multiple measurements made on the same subject might complicate the computation of the standard error of the mean even when the repeated sampling of patients is balanced.

6. We often compute a normalized Z score even when the data might not be normally distributed. In such an instance we are measuring a variable in units of standard deviations. This is still useful due to Chebyshev's inequality: For random variable X having mean μ and variance $\sigma^2 > 0$, and arbitrary $k \geq 1$,

$$\Pr\left(\left|\frac{X-\mu}{\sigma}\right| \geq k\right) \leq \frac{1}{k^2}$$

The question then arises as to how conservative the bound provided by Chebyshev's inequality is.

a. Show that Chebyshev's inequality can be extremely conservative for $k > 1$ by finding a nondegenerate distribution (i.e., with a positive variance) that has 100% of its data within 1 standard deviation of the mean.

b. Show that Chebyshev's inequality is not always conservative, because there is a distribution that meets the bound exactly. That is, under what conditions will exactly $100/k^2$% of the data lie at least k standard deviations away from its mean for a specified value of k?

(Hint: Consider a distribution that takes on values -1 with probability p, 1 with probability p, and 0 with probability 1-2p.)