AP® STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 4

Intent of Question

The primary goals of this question were to assess a student's ability to (1) design an experiment to compare two treatments and (2) identify the associated potential Type I and Type II errors and decide which of these two would be more serious.

Solution

Part (a):

Approach 1: Paired Design

Each subject will receive both treatments, with a suitable length of time between treatments. The order of the treatments will be randomly assigned to the subjects. For example, for each patient flip a coin to determine which treatment will be administered first. Measure diastolic blood pressure, then have the subject sit quietly for 20 minutes in either a noise-free environment or in a room where soothing music is played, depending on which treatment was selected at random (based on the coin flip). At the end of the 20 minutes, measure diastolic blood pressure again and compute its change (*after − before)*. After a suitable period of time, repeat with the other treatment.

When the data have been collected, the difference (*music − noise-free*) in the change in diastolic blood pressure will be computed for each subject, and then a paired *t*-test will be run to see if the mean difference is significantly different from zero.

Approach 2: Matched Pairs Design

Measure diastolic blood pressure for each of the 100 subjects and then form 50 pairs based on these readings by pairing the two with the highest diastolic blood pressure, then the two with the next highest, and so on. For each pair, toss a coin to determine which member of the pair will be assigned to group 1, and then assign the other member of the pair to group 2. For group 1, measure diastolic blood pressure, then have the subjects sit quietly in a noise-free environment for 20 minutes, and then measure diastolic blood pressure again and compute its change (*after − before)*. For group 2, the plan is the same, except that they will sit for 20 minutes in a room where soothing music is played between blood pressure measurements.

When the data have been collected, the difference (*music − noise-free*) in the change in diastolic blood pressure will be computed for each pair, and then a paired *t*-test will be run to see if the mean difference is significantly different from zero.

Approach 3: Completely Randomized Design (This is not as good a choice as the two previous approaches, but because of the large number of subjects available for each treatment group, it is considered an acceptable solution.)

Assign the 100 patients numbers from 00 to 99. From a random number table, select 50 unique numbers; the patients with the selected values will form group 1; the remaining 50 patients will form group 2. For group 1, measure diastolic blood pressure, then have the subjects sit quietly in a noise-free environment for 20 minutes, and then measure diastolic blood pressure again. For group 2, the plan is the same, except that they will sit for 20 minutes in a room where soothing music is played between blood pressure measurements.

Question 4 (continued)

When the data have been collected, the change in diastolic blood pressure will be computed for each subject, and then a two-sample *t*-test will be run to see if there is a significant difference between the mean change attributable to *music* and the mean change attributable to a *noise-free* environment.

Part (b):

Type I error: Concluding that soothing music does reduce mean diastolic blood pressure compared to sitting quietly, when in fact it does not. The consequence of this type of error is that the clinic will offer music therapy when it is not effective.

Type II error: Soothing music does reduce diastolic blood pressure compared to sitting quietly, but we fail to detect this and conclude that it does not. The consequence of this type of error is that the clinic will choose not to offer music therapy when it would have been effective.

Which type of error is more serious? A case can be made for either type of error, and the student can take either side as long as a reasonable justification is given. For example, the student can say a Type I error is more serious because it will cost the clinic money with no benefit, or the student can say that a Type II error is more serious because the clinic will miss an opportunity to improve the health and well-being of its patients.

Scoring

Part (a) is divided into two sections: section 1 is the randomization, and section 2 is the experimental runs. Section 1 is scored as essentially correct (E), partially correct (P), or incorrect (I). Section 2 is scored as essentially correct (E) or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the design includes an appropriate randomization of patients to treatment groups or randomization of the order of treatments in the paired design. The description of the randomization should be sufficiently clear that it could be duplicated by the reader.

Partially correct (P) if the student states that the patients should be randomized to treatment groups or that there should be randomization of the order of treatment in the paired design but does not specify how this randomization is to be accomplished *OR* if there are flaws in the randomization *OR* if the randomization could not be duplicated by the reader.

Incorrect (I) otherwise.

Note: It is acceptable to first block by variables such as gender or age if the student then correctly uses one of the above approaches.

AP® STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 4 (continued)

Section 2 is scored as follows:

Essentially correct (E) if treatments are applied and blood pressures are measured after treatment. (Although it would be better if the student suggests measuring blood pressure before and after treatment, it is sufficient that the student measures blood pressure only after treatment; however, measuring blood pressure both before and after treatment is considered a plus. Additionally, a statement that a comparison of the two groups will be made based on the change in blood pressure is not required but is also considered a plus.)

Incorrect (I) otherwise.

Part (b) is divided into two sections: section 1 consists of the identification of the errors, and section 2 consists of the consequences. Each section is scored as essentially correct (E), partially correct (P), or incorrect (I).

Section 1 is scored as follows:

Essentially correct (E) if the response gives correct descriptions of Type I and Type II errors in context.

Partially correct (P) if the response gives a correct description of one type of error in context and the description of the other type of error has a minor flaw.

Incorrect (I) otherwise.

Section 2 is scored as follows:

Essentially correct if:

(1) The response describes the consequences for each of Type I and Type II errors.
AND
(2) It states which type of error is more serious and gives a reason to support the selection made.

Partially correct if only one of (1) and (2) is correctly stated.

Incorrect (I) otherwise.

Notes:
- If Type I and Type II errors are reversed but the description of the errors and the consequences are correct for this reversal, give credit for part (b), section 2, but not for part (b), section 1.
- If only one type of error and its consequences are described, give credit for one section of part (b).

AP® STATISTICS
2008 SCORING GUIDELINES (Form B)

Question 4 (continued)

Each essentially correct response is worth 1 point; each partially correct response is worth ½ point.

4 **Complete Response**

3 **Substantial Response**

2 **Developing Response**

1 **Minimal Response**

If a response is between two scores (for example, 2½ points), use a holistic approach to determine whether to score up or down, depending on the strength of the response and communication.

4A1

4. A researcher wants to conduct a study to test whether listening to soothing music for 20 minutes helps to reduce diastolic blood pressure in patients with high blood pressure, compared to simply sitting quietly in a noise-free environment for 20 minutes. One hundred patients with high blood pressure at a large medical clinic are available to participate in this study. n=100.

(a) Propose a design for this study to compare these two treatments.

Randomly assign each patient a number from 1 to 100. Find 100 identical small pieces of paper and write one number, from 1 to 100, on each piece of paper. Throw the papers into a hat, shake the hat to make sure all papers are mixed together, and take out 50 pieces of paper from that hat. People whose assigned number correspond to the numbers on the papers drawn out from the hat will be the treatment group, Those 50 the other 50 will be the control group. Measure diastolic blood pressure in all 100 patients and record down the data. Place 2 groups (control and treatment) into 2 identical rooms (with identical temprature, humidity etc). Make sure people in one room will not be able to hear the sound from another room. Let the treatment group sit in room 1 and listen to soothing music for 20 mins. At the same time, the control group sit in room 2 without any distraction. Measure the diastolic blood pressure of all patients right after the 20 mins experimental period. Compare the data recorded.

GO ON TO THE NEXT PAGE.

4A2

(b) The null hypothesis for this study is that there is no difference in the mean reduction of diastolic blood pressure for the two treatments and the alternative hypothesis is that the mean reduction in diastolic blood pressure is greater for the music treatment. If the null hypothesis is rejected, the clinic will offer this music therapy as a free service to their patients with high blood pressure. Describe Type I and Type II errors and the consequences of each in the context of this study, and discuss which one you think is more serious.

$H_0: \mu_c = \mu_t$
$H_a: \mu_c < \mu_t$

μ: reduction in diastolic blood pressure.

Type I: reject H_0 when it is true.
Clinic offer music therapy as a free service to their patients even though the therapy is actually ineffective

Type II: not reject H_0 when it is false
The clinic assume that music therapy is ineffective and doesn't offer the therapy even though the therapy is actually effective.

Type II error is more serious because the patients are denied the chance to reduce their blood pressure when there is an effective treatment. Whereas in type I error, listen to music will not do any harm to the patients.

GO ON TO THE NEXT PAGE.

4B1

4. A researcher wants to conduct a study to test whether listening to soothing music for 20 minutes helps to reduce diastolic blood pressure in patients with high blood pressure, compared to simply sitting quietly in a noise-free environment for 20 minutes. One hundred patients with high blood pressure at a large medical clinic are available to participate in this study.

(a) Propose a design for this study to compare these two treatments.

Randomly divide the patients into two groups. Measure blood pressure of each patient prior to experiment. 50 will listen to music for 20 minutes and the other 50 will sit in silence. Re measure blood pressure after 20 minutes.

Age and gender of patient may affect the outcome, as well as race.

Some extraneous factors include the personality, medical history, and patience of patient.

4B2

(b) The null hypothesis for this study is that there is no difference in the mean reduction of diastolic blood pressure for the two treatments and the alternative hypothesis is that the mean reduction in diastolic blood pressure is greater for the music treatment. If the null hypothesis is rejected, the clinic will offer this music therapy as a free service to their patients with high blood pressure. Describe Type I and Type II errors and the consequences of each in the context of this study, and discuss which one you think is more serious.

Type I error would involve rejecting the null hypothesis when it should not have been.

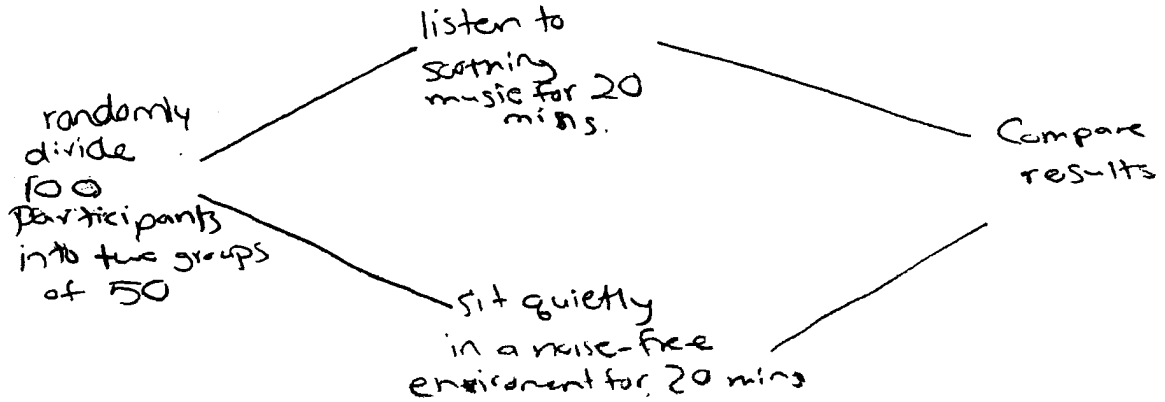
Type II error would be keeping the null hypothesis instead of rejecting it.

In this study, a Type I error would mean that the clinic would offer this unnecessary music service to their patients. Type II error would mean denying a service which could lower their blood pressure, perhaps save their lives, meaning a Type II error would be more serious.

4C1

4. A researcher wants to conduct a study to test whether listening to soothing music for 20 minutes helps to reduce diastolic blood pressure in patients with high blood pressure, compared to simply sitting quietly in a noise-free environment for 20 minutes. One hundred patients with high blood pressure at a large medical clinic are available to participate in this study.

   (a) Propose a design for this study to compare these two treatments.



4C2

(b) The null hypothesis for this study is that there is no difference in the mean reduction of diastolic blood pressure for the two treatments and the alternative hypothesis is that the mean reduction in diastolic blood pressure is greater for the music treatment. If the null hypothesis is rejected, the clinic will offer this music therapy as a free service to their patients with high blood pressure. Describe Type I and Type II errors and the consequences of each in the context of this study, and discuss which one you think is more serious.

Type I error: is when you reject the null when actually it was true.

Type 2 error: is when you accept the null when actually it was false.

In this certain proble type I error would be if they believe music therapy helps high blood pressure patients when actually it doesen't.

Type II would be it they believe there is no difference in the two treatments when in fact the music treatment actually reduces the mean diastolic blood pressure.

If the clinic makes Type I error they would lose money more than anything else. They would be giving music treatment when it doesn't have an effect. Another consequense that may occur sometimes is that belief of the patient that they are getting treatment may lead to the reduction of their bp anyways (kind of like a placebo). If the clinic makes a type two error they would be deprivating patients from a possible cure for their bp, which I think is worse. People wouldn't be getting the music therapy, which would have helped them reduce their bp.

AP® STATISTICS
2008 SCORING COMMENTARY (Form B)

Question 4

Sample: 4A
Score: 4

The completely randomized design described in part (a) is a good method of comparing the two treatments. The randomization of treatments to subjects is described in sufficient detail that it could be duplicated by the reader. Blood pressure is measured both before and after a treatment is applied, which is preferable, but not a required component of the design. Each section of part (a) was scored as essentially correct. In part (b) correct descriptions of Type I and Type II errors are given in context, and a reasonable determination is made as to which type is more serious. Correct terminology is used throughout. Each section of part (b) was scored as essentially correct. The entire answer was judged a complete response, based on all four sections.

Sample: 4B
Score: 3

The completely randomized design described in part (a) is acceptable, except that no method of randomization is included. The first section of part (a) was scored as partially correct and the second section as essentially correct. In part (b) the description of a Type II error is incomplete, as it is not clearly conditional: a Type II error is failing to reject the null hypothesis instead of rejecting it *when it is false*. The first section of part (b) was scored as partially correct and the second section as essentially correct. The overall answer was deemed a substantial response.

Sample: 4C
Score: 2

The description of the completely randomized design in part (a) does not include a method for the randomization, nor does it say when or if measurements of blood pressure are made. (The use of a diagram is not in itself a defect. A diagram can be judged essentially correct if all the necessary components are included.) The first section of part (a) was scored as partially correct and the second section as incorrect. Part (b) is well done except that "accept the null" hypothesis is not considered suitable language. The first section of part (b) was scored as partially correct and the second section as essentially correct. On the whole, this answer was considered a developing response.