***REVIEW HW #1 Free Response Solutions***

***DATA ANALYSIS* Free Response HW Solutions and Scoring Guidelines**

Question 1: (1998 #1)

**4 Complete Response**

Standard symbols are acceptable without explanation, but non-standard symbols must be defined.

The question asks for exact results, but a student can receive full credit for CAREFULLY explaining

what might happen in a simulation.

(a) The mean of the sampling distribution is equal to the mean of the population.

(b) The standard deviation of the sampling distribution is equal to the standard deviation of the population divided by the square root of the sample size.

OR

Clearly states that the standard deviation of the sampling distribution decreases as n increases.

(c) The equivalent of the following two statements must be included:

1. The sampling distribution is skewed for small sample sizes. (A statement that does not use the term skewed but says the distribution will be non-normal is OK.)

2. The shape of the sampling distribution gets more and more normal-like (bell shaped) as the sample size increases.

**3 Substantial Response**

States both (a) and (b) correctly; has a weak, but correct, response on part (c). A weak response for (c) would include correctly one of the two statements in (c) above, but not both.

OR

States either (a) or (b) correctly and gives a complete and correct response to (c).

**2 Developing Response**

States both of (a) and (b) correctly but gives an incorrect response to (c).

OR

Gives a correct response to either (a) or (b), but gives a weak response to (c).

OR

Gives an incorrect response to both (a) and (b), but gives a complete and correct response to (c).

**1 Minimal Response**

States one of (a) or (b) correctly.

OR

Has substantive errors in (a) and (b), but gives a weak response to (c).

Notes:

Some students appear to have confused the sampling distribution with the histogram for a particular sample. There were a number of papers that had responses containing \the sample mean is close to" or \gets close to the population mean as n increases," or other rewordings of the law of large numbers. These statements, while true, do not answer the questions posed. These incorrectly worded responses may surround the correct formula in parts (a) and (b). If the written response is irrelevant but does not contradict the formula, credit can be awarded; if the written response directly contradicts the formula, credit should not be given.

Question 2 (1998 #2)

**4 Complete Response**

(a) Shows a correct histogram (or bar graph) of the number of defective items produced per day, as shown below. The vertical scale could be either frequency or relative frequency (8=20 = :40).

(insert graph)

(b) The most obvious fact from the histogram is that the number of defective items has a symmetric, mound-shaped, bell-shaped or approximately normal distribution. Any one of these terms (or phrases that convey the same notion) will be considered as correct. The term normal distribution is not acceptable for a complete response. Other facts that might be reported include the mean of 3 and/or some measure of spread, such as the standard deviation of approximately 1.1, but numerical summaries are not required, whereas some statement on shape is required.

(c) The most obvious fact from the scatterplot is that the number of defects per day has a decreasing, linear trend across the days. This point should not be missed even if other correct information, such as the larger spread for the later days, is given. Statements such as negative correlation, negative association and negative slope are acceptable as correct answers here.

**3 Substantial Response**

(a) Draws a histogram that is essentially correct, but may have a minor error such as an error in counting or scaling.

(b) Makes a nearly correct statement about the shape of the histogram. Stating that the number of defects is normally distributed without qualification will reduce the score to a maximum of 3.

(c) Makes a correct, but perhaps weak or incomplete, statement on the decreasing trend over time in the scatterplot.

**2 Developing Response**

Draws a histogram nearly correctly and has a nearly correct description of the shape of the histogram.

(The term normal distribution is acceptable here.) Makes no correct statement about essential features of the scatterplot.

OR

Draws a histogram incorrectly but then makes correct statements based on that drawing. Makes a correct statement about a feature of the scatterplot that reflects some understanding of the graphical image portrayed.

OR

Draws a histogram nearly correctly but interprets it incorrectly. Makes a correct statement about

a feature of the scatterplot that reflects some understanding of the graphical image portrayed.

**1 Minimal Response**

Has a partially correct response in at least one the three parts.

Question 3 (1999 #4)

1. Because it is stated as a normal distribution, we can use the Empirical Rule. 95% of the data lies within 2 standard deviation of the mean, so according to control plan A, the proportion of batteries that will be rejected is 2.5% (100-95 =5, so there’s 5% left for BOTH sides. Plan A rejects if they are BELOW the mean, so we need 5/2 = 2.5%)

or

P(defective) = P(z < -2) = 0.0228 (from z-table)

1. P(at least 1 of 2 randomly selected batteries will be rejected by Plan A) =

1 – P(none defective)

1 – (.95 \* .95) = .04938 (if using the Empirical Rule from above)

1 – (.9772 \* .9772) =0 .04498

1. Proportion of batteries rejected by Plan B:

More than 1.5 IQR below Q3…using the standard normal curve:

Look up 0.25 on the INSIE of the z-chart to find the z-value that corresponds. Z = 0.67

IQR = Q3-Q1 = .67 – (-.67) = 1.34

1.5IQR = 1.5(1.34) = 2.01

Q3 – 2.01 = -.67 – 2.01 = -2.68

P(z < -2.68) = 0.0037

Question 4 (2000 #1)

1. Draw two histograms side by side or on the same axis with two colors.
2. The distribution for middle-aged men is approximately symmetric, centered around 5.5 whereas the young adult men are skewed to the left, centered around 6.5 which is HIGHER than the middle aged men. There is quite a bit of variability in both distributions. In general, there were more young men with flexibility ratings at the high end of the scale and fewer at the low end than the middle aged men.

To get an “E” on part a), you need to do graphs on the SAME scale with labels and scales.

To get an “E” on part b), you need to give similarities and differences about the distributions in reference to center, shape and spread. Use the “-er” words. A “P” is talking about 2 out of 3 of center, shape and spread.

***DESIGN Free Response HW Answers and Scoring***

Question #1 (2007 #2 Dogs)

1. A control group gives the researchers a comparison group to be used to evaluate the effectiveness of the treatments. The control group allows the impact of the normal aging process on joint and hip health to be measured with appropriate response variables. The effects of glucosamine and chondroitin can be assessed by COMPARING the responses for these two treatment groups with those for the control group.
2. Put all 300 dogs’ names on a piece of paper and put them in a hat. Shake them up. Randomly choose 100 names, those dogs go to the glucosamine group. Pick 100 more names and they get condroitin and the rest are the control group.
3. The key question is which variable has the strongest association with joint and hip health. You can have chosen either variable as long as you said you picked it because it would have the most effect on the hip health of the dogs. For example: Different breeds of dogs are difference sizes, and size can effect hip health. So I think breed of dog would have MORE of an effect on hip health than clinic so I would block by breed of dog.

To get an “E” on part a, you had to say that the control group would allow you to COMPARE the effects of no drug to the effects of the glucosamine and condroitin AND you had to have CONTEXT (more than just “dog”) A “P” for having COMPARE or CONTEXT.

To get an “E” on part b), you had to have a correct method of an SRS AND you had to tell the reader which treatments the dogs would be going to. It didn’t matter what order you said, but they had to be named (glucosamine, condroitin, control)

To get an “E” on part c), you had to mention BOTH clinics and breed and say which ONE you would pick and relate to the fact that it would have the strongest effect on hip health. Or, you could have said which one you would block by because it would create the most homogeneous groups.

Question #2 (1999 #3 Dentist)

1. It is an observational study because there are no treatments imposed.
2. Two variables are confounded if their effects on the number of new cavities cannot be distinguished from one another. The students must mention not only that the confounding variables may affect the outcome but that they have differential effects within the two groups. For example, confounding would occur if patients who eat an apple a day differ from those who eat less than one apple a week on some variable that is related to dental health. For example, people that eat an apple a day might be more healthier in general and brush their teeth more than those that do not eat an apple a day. NOTE: There are many examples of confounding variables that can be cited.
3. No, because it is not an experiment and cause and effect relationships cannot be drawn from an observational study.

Question #3 (1998 #3 Butterflies)

**4 Complete Response**

(a) A complete response should try to make the six treatment groups look as nearly alike as possible, except for the treatment, so as to balance out any characteristics (seen or unseen) that might affect chances of successful migration.

Some ways to do this are as follows:

-Randomly allocate the 3600 butteries to groups of 600 for each marking location. A correctly described method does not necessarily have to be practical. Although it would be difficult, the 3600 butteries could be each assigned a unique number and then a random number table could be used to select 600 for position A, and so on. A more practical randomization could be accomplished by putting 3600 slips of paper n a bag, 600 for each letter. As each buttery is captured, a slip is randomly selected from the bag (without replacement) to indicate the marking location for the buttery.

-Divide the 3600 butteries into groups smaller than 600 and then randomly assign an equal number of groups to each marking location.

-Systematically assign marking locations A through F to successive groups of 6 butteries until all 3600 have been assigned.

Note: Responses for (a) should be read carefully to assess understanding of randomness. For example, only indicating an allocation of 600 butteries per location is an incorrect response. Only stating that 600 butteries are randomly assigned to each marking location is an incomplete response.

(b) Generate six possible outcomes independently, each with probability 1/6, perhaps by tossing a balanced die. Assign one number to each location, such as 1 = A, 2 = B, and so on. Toss the die and assign the marking location sequentially to the 3600 butteries as they are caught. The order of assignment does not matter.

(c) The analysis requires a test of the null hypothesis that the proportions of butteries that successfully migrate are the same for all locations. Either a frequency table equivalent to the one shown below can be indicated as the proper way to arrange the data, or a written description of how the counts are tabulated and proportions compared is acceptable.

(Insert Table)

The student must indicate an appropriate test statistic. A chi-square test is anticipated. Multiple comparisons of proportions would also work, but this requires 15 simple comparisons. (The student need not indicate degrees of freedom or formulas. However, if these are included, they must be correct.)

Note: No credit is given for writing down \chi-square test" without an essentially correct method of assigning treatments to the butteries.

(d) In (a) the number launched is maxed at 600 for each marking location. The chi-square test for the equality of proportions is still an acceptable test, so the method of analysis need not be changed.

**3 Substantial Response**

Gives essentially correct methods for (a) and (b) but has an error in one of the methods of analysis in (c) or (d). A common error is to think that the chi-square test will not work when the sample sizes are random, part (c), and to try to concoct a new test here.

OR

Gives essentially correct methods for (a) and (b) and, in (c) and (d), indicates how to tally the counts and that the analysis requires a comparison of proportions (but does not name a specific test to analyze this comparison).

OR

Has an incomplete response for either (a) or (b), but indicates appropriate analysis in (c) and (d).

**2 Developing Response**

Gives an essentially correct method for one of (a) or (b) and a correct analysis that goes with it in (c) or (d). Has a substantive error in one of (a) or (b) and the related analysis.

OR

Gives an essentially correct method for (a) and (b), but (c) and (d) are incorrect or blank.

OR

Gives an essentially correct method for (a), but (b) is incorrect or missing, and (c) and (d) have incomplete analyses.

**1 Minimal Response**

Gives an essentially correct method for (b), but (a), (c), and (d) are incorrect or missing.

OR

Gives an essentially correct method for (a), but (b), (c), and (d) are incorrect or missing.

OR

Has an incomplete response for (a) or (b), and in (c) or (d) indicates how to tally the counts and that the analysis requires a comparison of proportions.

Question 4: 2000 #5 Diet and Exercise

1. Describes an experimental design that includes:
2. Random assignment of treatments
3. Identification of treatment groups as old drug and new drug
4. Indicates a comparison or measurement of cholesterol level.

(This is 100% acceptable in picture form as shown below)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Old drug |  |
| Volunteers | Random  Assignment |  | Measure and Compare Cholesterol Levels |
|  |  | New drug |  |

1. Describes an experimental design that includes:
2. Creating blocks based on level of exercise or cholesterol level, or creating blocks using age, diet, gender, or any other factor plausibly related to cholesterol level **with explanation** (that is, I blocked by gender because males and females’ cholesterol level may respond differently to the drug)
3. Random assignment of treatments within the blocks

(This is 100% acceptable in picture form as shown below)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Random assignment | Old drug | Measure and Compare Cholesterol Level | |
|  |  | Exercise | New Drug |
| Volunteers | Block |  |  |  |  |
|  |  | No exercise  Random assignment | Old drug | Measure and Compare Cholesterol Level | |
|  |  |  |  |
|  |  |  | New drug |  |  |

1. Clearly explains a double-blind experiment – neither the subjects (people receiving the treatments) nor those administering the drugs or monitoring the results know which of the two drugs are being used.

**Probability HW Free Response Answers and Scoring**

Question #1 (1998 #6)

**Solutions and Scoring**

**Section 1: Finding a probability**

(a) Using the normal density function, determines that P(7 < X < 9) = 0:38, which rounds to 0:4.

A \substantially correct" response must show some indication of how this value was determined.

For example, either

P(7 < X < 9) = P(􀀀0:5 < Z < 0:5) = 0:6915 􀀀 0:3085 = 0:3830

or

Normalcdf(7; 9; 8; 2) is fine.

If there is no explanation for the 0:38, this is considered only an “acceptable approach." Other acceptable approaches, but not complete responses, include answers where normal curve calculations or geometry is incorrect. Such acceptable approaches include answers of 1 􀀀 0:3830 = 0:6170 and 0:1915. Note that the answer 0:5(0:68) = 0:34 should receive no credit.

**Section 2: Setting up and running a simulation**

(b) Shows a listing of four of the ten digits designed to be equivalent to finding a pearl of the correct size (a \success") and states that the other digits represent finding no pearl. The value used in (b) must match the value calculated in (a). If no solution is given in (a), a value can be “made up" and the problem completed. It is OK to use a two digit simulation with p = 0:38.

Describes a correct method of sampling digits from a random digit table until two successes are found.

To get any credit for Section II, the student must describe a waiting time simulation. A correct assignment of digits in (b) is not sufficient to get any credit if the waiting time component is missing.

(c) Correctly identifies a success, based on the definition in part (b) and counts the number of digits it took to get two successes.

Correctly runs three trials (more is OK) and records the number of digits it took to get two successes. (Starting points for the trials will vary.)

If the assignment of digits in (b) is incorrect, the response should be considered an “acceptable approach" if a waiting time simulation is carried out correctly in part (c) according to the assignments of digits in part (b).

**Section 3: Interpreting results of a simulation**

(d) Identifies A as the correct distribution for P = 0:4 by noting that the probability of finding a pearl between 4 and 6.5 mm is smaller than the probability of finding a pearl between 7 and 9 mm, and it therefore takes longer to find two pearls between 4 and 6.5 than two pearls between 7 and 9. If the student computes the probability of finding a smaller pearl, this should be considered a bonus. (P(4 < X < 6:5) = 0:2039.) If the argue that a pearl between 7 and 9 is more likely because this interval contains the mean or is closer to the mean than [4; 6:5], this is not a major error but is considered a minus. If the student selects Distribution A with no reason or an incorrect reason, they get no credit on part (d). If the student selects a distribution based on how well it matches their simulation, this is an \acceptable approach."

Another correct response is to observe that the mean should be 2(1=0:38) = 5:26, which is closer to 5 than to 10.

(e) The best estimate of the expected number of pearls is the sample mean of 5.16, for distribution A. This can be arrived at by direct computation using the formula

x =P x \_ f 100 = 5:16

by estimating the balance point of the distribution, or by using the formula 2(1=0:38) = 5:26. [The standard deviation is 2.56 in case anyone uses it.] A computational mistake using the right formula is okay, if the answer is reasonable. If no indication is given of how the answer was determined, this is an \acceptable approach,"

but not a \complete response." However, if no work was shown in part (a) as well, don't count o\_ again.

If a student uses the median of 4 or the mode of 3 as the best measure of center with justification, this is OK. Without justification, it is an \acceptable approach."

If the student selects distribution B in part (d), the mean should be 9.36, the median 8, and the mode 7 (and the standard deviation is 5.27).

**4 Complete Response**

The response is substantially correct for all three sections, I, II, and III.

**3 Substantial Response**

The response is substantially correct for two of the three sections, I, II, and III, or the response is substantially correct for one of the three sections and nearly correct for each of the other two.

**2 Developing Response**

The response is substantially correct for one of the three sections, I, II, and III, and shows an acceptable approach in at least one of the other sections.

**1 Minimal Response**

The response is substantially correct for one of the three sections, I, II, and III, or shows an acceptable approach in at least one of the three sections.

Question 2: (1999 #5)



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **DIE A** | | | | | |
|  |  | 0 | 0 | 9 | 9 | 9 | 9 |
| **DIE B** | 3 | B | B | A | A | A | A |
| 3 | B | B | A | A | A | A |
| 3 | B | B | A | A | A | A |
| 3 | B | B | A | A | A | A |
| 11 | B | B | B | B | B | B |
| 11 | B | B | B | B | B | B |

P(A wins) = 16/36 = 4/9

P(B wins) = 20/36 = 5/9

Choose Die B because the probability of winning is higher (5/9 compared to 4/9 for Die A).

1. Let X be the number of tokens the player using die B should receive.

|  |  |  |
| --- | --- | --- |
| Tokens received | 45 | X |
| Prob of winning | 4/9 | 5/9 |

For the game to be fair we need 45(4/9) = X(5/9)

Solving this equation for X gives X=36. Player B should receive 36 tokens.

Question #3 (2001 #2)

1. For Machine A, the total 3-year cost =

This cost is fixed

For Machine B: Let X = number of repairs in one year

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X | 0 | 1 | 2 | 3 |
| P(X) | .5 | .25 | .15 | .1 |

E(X) = .85. Multiply this by 3 for the three years.

Expected number of repairs in 3 years =

Expected cost for repairs in 3 years = 2.55($200) = $510.

Expected 3-year total cost: $10,500 + $510 = **$11,010**

**Choice:** Choose B because it has a lower expected (or average) cost. (A has a fixed cost that is $790 ($11,800-$11,010) higher than the expected cost of B.

The solution should include the following four elements:

1. Correct calculation of 3-year cost for A
2. Correct calculation of a relevant expected **value** for B (expected number of repairs per year or per 3 years or expected cost of repairs per year or per 3 years). Calculation of expected value must be shown.
3. Correct calculation of expected **total cost** for B.
4. Choice of B with a *complete* *and coherent* explanation that is based on student’s prior calculations for A and B.

“*Complete and coherent”* means that:

* Costs for A and B are compared

AND

* B’s cost has been indicated as “expected” or “average” or “mean” or “estimated” or “approximate” or “predicted” etc.

Question 4 (1997 #3)

Either by formula oy by constructing a table or tree diagram, demonstrates understanding of the three pieces of information given in the problem.

P(positive|no disease) = .05

P(negative| disease) = .003

P(disease) = .02

1. Correctly reworks that conditional information to answer the unconditional probability as

P(positive) = **.0689**

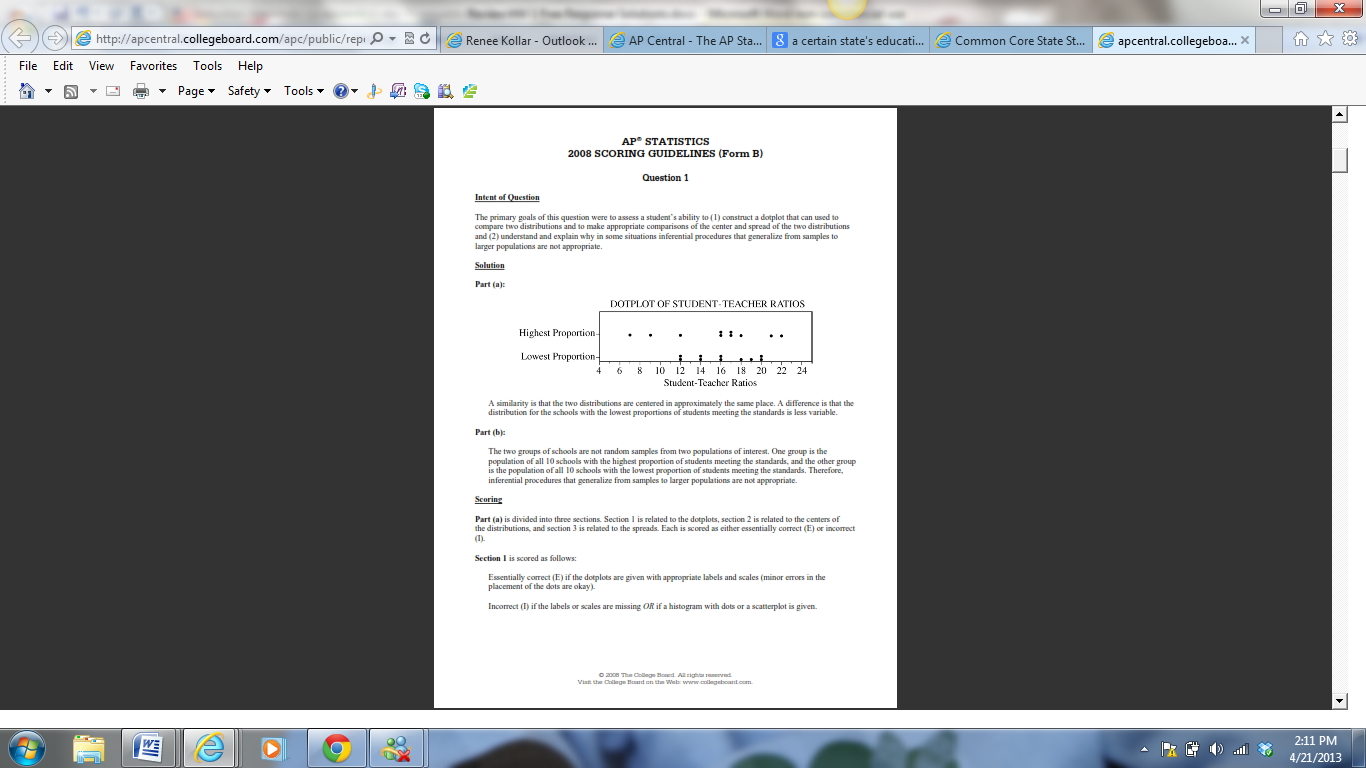
1. Correctly reworks that conditional information to answer the conditional probability as

P(no disease|positive) = =

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Disease** |  | **Test**  + |  |  |
| + |  |  | **.997** | **.01994** |
|  | **.02** | - |  |  |
| - |  |  | **.003** | **.00006** |
|  |  | + |  |  |
|  | **.98** | - | **.05** | **.049** |
|  |  |  |  |  |
|  |  |  | **.95** | **.931** |
|  |  |  |  |  |

**Inference FR Answers and Solutions**

**Question #1 (1998B #1)**



1. The two groups of schools are not random samples from the two populations of interest, therefore inferential procedures (test or confidence interval) are not appropriate. These are groups from the highest ten and lowest 10 schools…not a random sample.

Scoring:

“E” if the dotplots are given with appropriate labels and scales

“E” if the response states that the two distributions have approximately the same center OR if they say the centers are *slightly* different

“E” if the response states that the spread of the distributions differ

“E” if part b) states that the data are not samples from some larger population OR that they are not random samples but instead are those in the highest and lowest proportions of students meeting the standard.

Question 2 (1998B #3)

1. Using the formula for a 95% confidence interval for a margin of error of 2, solve for *n* in the following:

n = 138.3, so use n = 139

Since the cost is $100 per person, it would cost $100\*139 = $13,900 which exceeds the amount budgeted for the study. Sufficient funds are not available.

1. To be within 2 feet of the true mean with 95% confidence, 139 observations are required. The budget of $12000 only allows 120 observations to be taken. Therefore the company will not be able to meet the regulatory agency’s requirements with the allocated budget.

OR

Given the fact that the $12000 will allo for 120 observations, the margin of error for this would be 2.15 which exceeds the requirement of 2. Therefore the company will not be able to meet the regulatory agency’s requirements with the allocated budget.

**Question #3(1998B #4)**

1. 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, each patient could flip a coin to determine which test goes first. Measure diastolic blood pressure, then have the subject sit quietly for 20 minutes in either a noise-free environment or a room where soothing music is played, depending on which treatment was selected at random. At the end of 20 min, measure the diastolic blood pressure again and compute the change (after – before). After a suitable period of time, repeat with the other treatment.

When the data have been collected, the difference (music – music free) in the change in the diastolic blood pressure will be compared for each subject and then a paired t-test will be run to see if the mean difference is significantly different from 0.

Approach 2: Matched Pairs Design

Measure the diastolic blood pressure of the 100 subjects and then form 50 pairs based on these readings by pairing the two with the highest diastolic blood pressure, the next highest, etc. For each pair, toss a coin (or another random assignment method) to determine which individual gets the noise free and which gets the soothing music. Leave the subjects in their respective environments for 20 minutes and measure their diastolic blood pressure after the 20 minutes. Compute the change for each individual (after – before).

When the data are collected, the difference (music – noise-free) in the change in diastolic blood pressure will be compared for each pair and then a paired t-test will be run to see if the mean difference is significantly different from 0.

Approach 3: Completely randomized design (in a picture is fine)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | 50 noise free |  |
| 100 patients | Random  Assignment |  | Measure and Compare Diastolic Blood pressure |
|  |  | 50 soothing music |  |

Random assignment can be done by putting all 100 patients names in a hat. Shake it up. The first 50 chosen form the noise free group, the others form the soothing music group.

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.

1. 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 to not offer music therapy when it would have been effective.

Which type is more serious? A case can be made for either 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 a student can say a Type II Error is more serious because the clinic will miss the opportunity to improve the health and well being of patients.

**Question #4 (1998B #6)**

1. The trait that distinguishes the two groups in the scatterplot is dominant foot (left or right). All of the patients in the upper left hand cluster represent measurements from individuals whole dominant foot is the right foot while all of the points in the lower right cluster represent measurements from individuals whose dominant foot is the left foot.
2. There is a positive linear relationship between swelling in the dominant foot and swelling in the nondominant foot. Swelling in the dominant foot tends to be greater than in the nondominant foot.
3. Component 1: States a correct pair of hypotheses:

, where

mean difference in swelling for (dominant foot - nondominant foot)

Component 2: Identifies a correct test by name or by formula and checks the conditions

Conditions:

1. Random sample, given in problem
2. Normal? Must graph the DIFFERENCES in one graph (boxplot and look for symmetry or normal probability plot and look for linearity)

Component 3: Perform correct mechanics, including the test statistic and p-value

t = 10.68, df = 11, p-value is approximately 0

Component 4: Draws an appropriate conclusion in context with linkage.

Because the P-value is so small (or less than a stated ) we reject Ho. There is convincing evidence that the mean swelling is different for dominant and nondominant feet for women with MN.

1. To be continued