4. Baby walkers are seats hanging from frames that allow babies to sit upright with their legs dangling and feet touching the floor. Walkers have wheels on their legs that allow the infant to propel the walker around the house long before he or she can walk or even crawl. Typically, babies use walkers between the ages of 4 months and 11 months.

   Because most walkers have tray tables in front that block babies' views of their feet, child psychologists have begun to question whether walkers affect infants' cognitive development. One study compared mental skills of a random sample of those who used walkers with a random sample of those who never used walkers. Mental skill scores averaged 113 for 54 babies who used walkers (standard deviation of 12) and 123 for 55 babies who did not use walkers (standard deviation of 15).

   (a) Is there evidence that the mean mental skill score of babies who use walkers is different from the mean mental skill score of babies who do not use walkers? Explain your answer.

   (b) Suppose that a study using this design found a statistically significant result. Would it be reasonable to conclude that using a walker causes a change in mean mental skill score? Explain your answer.

5. High cholesterol level in people can be reduced by exercise or by drug treatment. A pharmaceutical company has developed a new cholesterol-reducing drug. Researchers would like to compare its effects to the effects of the cholesterol-reducing drug that is currently available on the market. Volunteers who have a history of high cholesterol and who are currently not on medication will be recruited to participate in a study.

   (a) Explain how you would carry out a completely randomized experiment for the study.

   (b) Describe an experimental design that would improve the design in (a) by incorporating blocking.

   (c) Can the experimental design in (b) be carried out in a double blind manner? Explain.

Question 4

Solution

a. **part 1**: States a correct pair of hypotheses

H_o: $\mu_W = \mu_N$ $\qquad$ H_o: $\mu_W - \mu_N = 0$

OR

H_a: $\mu_W \neq \mu_N$ $\qquad$ H_a: $\mu_W - \mu_N \neq 0$

where μ_W is the mean mental skill score for babies who used walkers and μ_N is the mean for those who did not. Nonstandard notation must be explained. Hypotheses about statistics (e.g. $\bar{x}$ or $\hat{p}$) are unacceptable.

part 2: Identifies a correct test (by name or by formula), and **checks** appropriate assumptions.

Note: Problem states that samples are random samples, so this does not need to be addressed in the assumptions.

Independent samples t test. Assumptions: large sample or normal population distributions. Check: OK, because, for example, n_1&n_2>30.

OR

Pooled t test. Assumptions: large samples or normal population distributions, equal population standard deviations. Checks: OK because, for example, n_1&n_2>30 and $s_1 \approx s_2$.

OR

Independent samples z test. Assumptions: large samples. Check: OK because, for example, n_1&n_2>30.

part 3: Correct mechanics, including value of test statistic, df (if appropriate), and P-value or rejection region (except for minor arithmetic errors)

- For independent samples t test:

$$t = \frac{\bar{x}_W - \bar{x}_N}{\sqrt{\frac{s_W^2}{n_W} + \frac{s_N^2}{n_N}}} = \frac{113 - 123}{\sqrt{\frac{12^2}{54} + \frac{15^2}{55}}} = \frac{-10}{\sqrt{6.7576}} = -3.8468$$

(Calculator: t = -3.846843677)
df = 102.828 (OK to use 102), P-value = .0002

OR conservative df = 54 - 1 = 53, P-value = 2(.00016)=.00032
OR using tables (for either df) P-value < 2(.0005) = .001

- For pooled t test: s_p = 13.597, t = -3.839, df = 107, P-value = .0002 (or < .001 from tables)

- For independent samples z test, z = -3.8468, P-value = .0001 (or <2(.0002) = .0004 from tables)

Question 4 continued

part 4: Stating a correct conclusion in the context of the problem, using the result of the statistical test (i.e., **linking the conclusion to the result of the hypothesis test**).

Reject the null hypothesis because P-value is less than stated α (or because P-value is very small, or because test statistic falls in the rejection region). There is convincing evidence that the mean mental score of babies who used walkers is different from the mean score for babies who did not use walkers.

If both an α and a P-value are given, the linkage is implied. If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the P-value or explaining how the conclusion follows from the P-value.

If the P-value in part 3 is incorrect but the conclusion is consistent with the computed P-value, part 4 can be considered as correct.

NOTE: A confidence interval approach will earn full credit for

- correct hypotheses at outset or, implicitly, in conclusion,
- correct procedure (by name or formula) and assumptions checked,
- correct mechanics, including specification of a (reasonable) confidence level, degrees of freedom specified (if appropriate)
  - 2-sample t interval, unpooled, 95%, df=102 or 53: (−15.2, −4.8)
  - 2-sample t interval, pooled, 95%, df=107: (−15.2, −4.8)
  - 2-sample z interval, 95%: (−15.1, −4.9)
- correct conclusion in context: "Since 0 is not in the 95% confidence interval, there is a significant difference between the mean mental skill scores of babies with walkers and babies without at the α=.05 level of significance."

part (b): No. This was an observational study, and a causal relationship can not be inferred from an observational study.

- It ***is*** sufficient to say any of:
  - "no; observational study" (or "no; not controlled experiment").
  - "no; no randomization in grouping" or "no; parents choose which babies use walkers".
  - "no" and then cite a plausible confounding variable **and** indicate how it is confounded with the formation of the groups.
- It is ***not*** sufficient to either:
  - merely mention lurking and/or confounding variables without indicating how they are confounded with the formation of the groups.
  - mention a causal factor which is a treatment "side effect", e.g. that walkers may contain plastics which are toxic to children.

Question 4 continued

Scoring

Part (a) is evaluated based on the four parts of the test. Each part must be COMPLETELY correct (except for minor arithmetical errors in part 3) to consider the part correct.

Part (b) is either correct or incorrect. If a student just answers "no" without giving a correct explanation that relates to the design of the study, part (b) is incorrect.

Note: *A 1-sided test can earn, at most, a score of 3.*

4 Complete Response

All four parts of the hypothesis test in part (a) correct and part (b) correct. (4-E)
(4 parts correct in (a) -- Correct in (b))

3 Substantial Response

All four parts of the hypothesis test in part (a) correct and part (b) incorrect (4-I)
OR
Three parts of the hypothesis test in part (a) correct and part (b) correct. (3-E)

2 Developing Response

Two parts of the hypothesis test in part (a) correct and part (b) correct. (2-E)
OR
Three parts of the hypothesis test in part (a) correct and part (b) incorrect. (3-I)

Note: *For papers judged a 2 because a one-tailed test is done and assumptions are not checked, exceptionally strong answers to the rest of the problem can be used to score the paper a "holistic" 3.*

1 Minimal Response

Two parts of the hypothesis test in part (a) correct and part (b) incorrect (2-I)
OR
None or one part of the test in part (a) correct and part (b) correct.
(0-E or 1-E)

0 No credit
Note that a 1-I earns a score of 0.