Biostat 140.623
Practice Quiz 1 B

The questions pertain to the following scenario: A survey of **adolescent teens** in Baltimore was performed to investigate the association between **HIV status (+ or -), gender** and **health care costs (in dollars)**.

1) Suppose that the investigators were interested in testing the null hypothesis that **mean health care costs** were equal for HIV+ and HIV- adolescents. In order to calculate a sample size, which one of the following quantities is **not needed**? (Circle only one response.)

   a) α
   b) $1-\beta$
   c) Δ = a meaningful hypothesized difference in mean health care costs by HIV status.
   d) The estimated variance of health care costs in HIV+ and the estimated variance of health care costs in HIV- adolescents.
   e) The estimated proportion of males who are HIV+ and the estimated proportion of females who are HIV+.

   Since health care costs in dollars is a continuous variable, assumptions must be made about the magnitude of a meaningful difference in mean costs between the two groups, the magnitude of the variance, and α and β.

2) Which of the following is the best definition for **α in problem 1**? (Circle only one response.)

   a) The probability of falsely detecting a difference in mean health care costs between HIV status groups.
   b) The probability of falsely concluding no difference in mean health care cost between HIV status groups.
   c) The probability of falsely detecting a difference in the prevalence of HIV between the health care cost groups.
   d) The probability of falsely concluding no difference in the prevalence of HIV between the health care cost groups.
   e) The probability of correctly detecting a difference in mean health care costs between HIV status groups.

   α is the probability of a Type I error. If the null hypothesis is no difference in mean costs between HIV status groups, then α is the probability of falsely rejecting a true null hypothesis (concluding that there is a difference in mean health care costs between HIV status groups when there truly is no difference).

3) Suppose that after performing their survey and collecting their data, the investigators performed a statistical test and concluded that there was no statistically significant difference in mean health care costs between HIV status groups. Which of the following could lead to this result? (Circle only one response. Circle one of the responses from a) through d) if you believe it is the ONLY correct response.)

   a) The statistical power of the test was too large to detect a statistically significant difference in mean health care costs between HIV status groups.
      FALSE, low power leads to failure to detect a difference.
   b) There is truly no difference in mean health care costs between HIV status groups.
      POSSIBLE
   c) The sample size may be too small to detect a statistically significant difference in mean health care costs between HIV status groups.
      POSSIBLE
   d) A Type I error has occurred.
      FALSE, can only occur when a difference is found.
   e) Either b) or c).
   f) Either a) or d).

4) What sample size in each group would be needed to estimate the difference in **HIV prevalence** between males and females to within ± 5%? Assume **α=0.05** and **equal sample sizes by gender**. Also assume that the prevalence of HIV+ by gender is **unknown** prior to conducting the study. Estimate the largest (most conservative) sample size. (Circle only one response.)

   a) 20
   b) 200
   c) 384
   d) 768
   e) 1536

Set d= 0.05 = $1.96\sqrt{\frac{.5(.5)}{n}+\frac{.5(.5)}{n}}$ **with p_1=q_1=0.5 and p_2=q_2=0.5 for the most conservative estimate.**

Solving for n:

$$(0.05)^2 = 1.96^2\left(\frac{.5(.5)}{n}+\frac{.5(.5)}{n}\right)$$

$$n = \frac{1.96^2(2)(.25)}{(0.05)^2} = 768.3$$

5) How would the **width of the confidence interval** from problem 1) change if the sample size for each group was quadrupled (multiplied by a factor of 4)? (Circle only one response.)

   a) The width of the confidence interval would be multiplied by a factor of 4.
   b) The width of the confidence interval would be multiplied by a factor of ¼
   c) The width of the confidence interval would be multiplied by a factor of ½.
   d) The width of the confidence interval would be multiplied by a factor of 2.
   e) There would be no change in the width of the confidence interval.

The change in the width of the CI depends on the change in the SE that results from the quadrupling of sample size in this example. When the sample size is multiplied by 4, then

$$\sqrt{\frac{p_1q_1}{4n}+\frac{p_2q_2}{4n}}=\sqrt{\frac{1}{4}\left(\frac{p_1q_1}{n}+\frac{p_2q_2}{n}\right)}=\frac{1}{2}\sqrt{\frac{p_1q_1}{n}+\frac{p_2q_2}{n}}=\frac{1}{2}\textbf{old SE}$$

The old width of the CI was 2(1.96)SE.

The new width of the resulting CI is $\mathbf{2(1.96)\frac{1}{2}\textbf{old SE}}$ **.**

Thus, the width of the old CI is multiplied by a factor of ½ when the sample size in each group is quadrupled.

