Practice Exam #1 - Sketch Solution

Ryan Miller

1

A) True - larger samples lead to less sampling variability (which is what SE describes)
B) False - sampling bias pertains to how cases are selected, not the number of cases sampled (see The Literary Digest and the 1936 election example)
C) True - achieving a greater level of confidence requires the interval to be wider
D) True - correlation is symmetric, but there are two regression lines for given pair of quantitative variables (see the Resting Metabolism example)
E) True - randomization is used to achieve balanced groups (so no variables will satisfy the definition of confounding)
F) False - it's possible for a convenience sample to be representative

2

A) Contingency table (4)
B) Standard deviation (7)
C) Correlation coefficient (3)
D) Side-by-side boxplots (6)
E) Median (1)
F) Standard error (5)

3-A

The cases are the 40 individuals suffering from insomnia that participated in the study.

3-B

The variables are "group" (binary categorical), which describes whether or not they received therapy, and "sleep improvement", which appears to be reported as categorical.

3-C

40 rows and 2 columns (or 3 if an identifier was used for each case)

3-D

The explanatory variable is "group" and the response variable is "sleep improvement". There is an association between these variables because a much higher proportion of the therapy group experienced sleep improvement (ie: there's a different in conditional proportions across groups).

3-E

A stacked bar chart would be most appropriate.

4-A

There's a weak, positive, linear relationship between AFQT percentile scores and income.

4-B

The regression line is:
$$\widehat{Income} = 21181.657 + 518.682 * AFQT$$

So the predicted income for an AFQT of 50 is:

$$21181.657 + 518.682 * 50 = 47115.76$$

4-C

The sample correlation between these variables is $r = 0.308$. And from the bootstrap distribution, $SE = 0.017$, so the 95% CI is:
$$0.308 \pm 2 * 0.017 = (0.274, 0.342)$$

4-D

Sampling bias is not a viable explanation because this is a representative sample (cases were randomly chosen from the target population). Sampling variability is not a viable explanation because the 95% CI suggests the only plausible values for the population-level relationship are weak, positive correlation coefficients.