

PubHealthStats Midterm, S03

- 📎 Coverage: Article by Gould, article by Jolson, risk ratios, odds ratios, data management.
- 📎 Time limit: 1.25 hours.
- 📎 For T/F questions, circle T or F. If false, change one or two words in the statement to make it true.

Part A (Closed Book)

1. The article by Stephen Jay Gould claimed that median survival time provided little useful information about the author's particular survival. What type of statistic might have proved more useful?
2. [M/C]: The article by Jolson et al. (1992) discussed an increased risk of toxic reactions in users of a generic drug. One of the problems in interpreting the results of the study was the potential for clinicians to be "looking harder" for cases taking the generic drug ("diagnostic bias"). Diagnostic bias is a form of:

(a) Imprecision (b) Information bias (c) Selection bias (d) Confounding
3. [M/C]: Another problem with the Jolson study was the small number of cases studied (11 exposed cases and 3 non-exposed cases). What type of error is associated with small sample size?

(a) Imprecision (b) Information bias (c) Selection bias (d) Confounding
4. [M/C]: Jolson and co-workers (1992) were concerned with the kidney function of patients. If the group receiving the generic drug has poorer kidney function than the other group, then the higher risk of toxicity could be explained by this factor rather than the generic drug. This type of bias is called:

(a) Imprecision (b) Information bias (c) Selection bias (d) Confounding
5. What is the primary benefit of randomization?
6. Blinding is the process by which we keep data collectors in the dark about the exposure status of individuals. What is the primary benefit of blinding?
7. What is the primary distinction between cross-sectional and longitudinal measurements?
8. Are prevalence studies cross-sectional or longitudinal?
9. How does "processing error" differ from "measurement error"?
10. Describe a method used to avoid or minimize or detect data entry errors.
11. The old joke goes "the two types of computer users are those who have lost data and those who are going to lose data." *Besides* from using logical filenames and storing data in known (media) locations, what is the most sure-fire way to minimize data loss?

12. Besides helping to set up the data entry screen, what function does the .qes file serve?
13. What type of *EpiData* file actually contains the data? (Identify the three letter file extension that is used to identify this type of file.)
14. When you export data from *EpiData* to SPSS, an .sps file and a .txt file are created. What function does the .sps file serve?
15. *Fill in the blank:* A risk ratio of 1.5 suggests a _____% *increase* in risk compared with the non-exposed group.
16. *Fill in the blank:* A risk ratio of 1.5 suggest that the risk in the exposed group is _____ times that of the non-exposed group.
17. [T/F]: A broad confidence interval provides evidence that the estimate is biased.
18. [T/F]: It is unethical to conduct a placebo controlled trial when an effective treatment is available.
19. [T/F]: According to generally accepted benchmarks, a p value of .0047 provides statistical evidence against the null hypothesis.
20. [T/F]: According to Fisher's flexible method of significance testing, a p value of .07 provides no statistical evidence against the null hypothesis.
21. [T/F]: : Epidemiologists will occasionally use a prevalence ratio to infer a risk ratio (while making certain unstated assumptions about the duration of illness in the groups, unchanging nature of the exposure status, and so on).

Part B (Procedure Notebook)

22. A statistician calculates a chi-square statistics (with one degree of freedom) of 4.71. Draw the chi-square probability distribution for this statistic and shade the area corresponding to the p value region. Then provide a landmark (or landmarks) on the distribution that help quantify the area under the curve corresponding to the p value region. Report the p value. [4 pts]
23. What field type (variable indicator) does *EpiData* use to create an upper-case variable that is one character in length?
24. What field type (variable indicator) does *EpiData* use to create a date field in American format?
25. Write the text you would use to have *EpiData* create a variable called “age” that is 3 bytes long.
26. What is the name of the type of file or data document that contains information about the structure, date, labels and content variables in a data set. Please be more specific than simply saying a .not file, as there are different types of .not files?
27. In requesting a cross-tabulation for the purpose of risk estimation using SPSS, which variable is identified as the *column variable*, the “exposure variable” or the “disease variable”?

The remainder of the questions address the following cross-tabulation:

	D+	D-	Total
E+	9	53	62
E-	10	156	166
Total	19	209	228

Calculate the:

28. The incidence in the exposed group.
29. The incidence in the non-exposed group.
30. The risk ratio.
31. A 95% confidence interval for the risk ratio. Show intermediate calculations. [4 pts]
32. Interpret the above statistics. [4 pts]
33. Calculate a χ^2 test for the data. [4 pts]