****KEY 267 Exam 1 KEY****

Grading: Each asterisk (*) represents one point.

Do *not* share this key with anyone without written permission.

- Permitted: Procedure notebook, statistical tables, WinPepi, calculator, SPSS.
- Show work when applicable.
- Present your work in an accurate and organized manner.
- Time limit: 1½ hours.

1. Of 33 cases, 19 are male. Describe the results. Then test the hypothesis of "no sex preference" for the disease (i.e., equal distribution by gender). Show all work and hypothesis testing steps. Summarize your findings in concise terms. [10 points]

* 57.6% of cases were male
** H_0: $p = 0.5$ [two-sided alternative is implied]
**** $z_{\text{stat}} = \frac{\hat{p} - p_0}{SE_{\hat{p}}} = \frac{0.5758 - 0.5}{0.08704} = 0.87$
* $P = .3843$ (two-tailed)
** The gender preference was *not* statistically significant.

Notes:

- $SE_p = \sqrt{\frac{p_0 q_0}{n}} = \sqrt{\frac{.5 \cdot .5}{33}} = 0.08704$
- One-sided $P = 0.19$
- With continuity correction, $z_{\text{stat}} = 0.70$ and $P = .4839$
- $np_0q_0 = 33 \cdot .5 \cdot .5 = 8.25$ (allowing use of the z test)
- Rounding errors lose one-eighth to one-quarter.
- Exact test two-tailed $P = 0.487$. With mid-P correction $P = 0.392$.

2. A cohort study identifies 48 cases in 414 exposed individuals. Among 912 nonexposed individuals, there were 55 such incidents. Calculate the relative risk associated with the exposure. Include a 95% confidence interval for the *RR*. Show work. Interpret your results. [10 pts]

*** *RR-hat* = 0.1159 / 0.06031 = 1.923
**** 95% CI = $e^{\ln 1.923 \pm 1.96 \cdot 0.188} = e^{0.6539 \pm 0.3696} = e^{0.2843,\ 1.0235} = 1.33$ to 2.78
*** Interpretation: 92% increase in risk with exposure ("almost a doubling of risk"), with 95% confidence the RR parameter is between 1.33 and 2.78

3. Three ethnic groups demonstrate the cross-tabulated results shown below. Calculate the prevalence of the condition within each group. Then test the association for statistical significance. Show all work and hypothesis testing steps. Summarize your results. [10 pts]

Group	+	–	Total
1	6	615	621
2	12	933	945
3	4	123	127
Total	22	1671	1693

* p-hat$_1$ = 6 /621 = 0.009662 (1.0%)
* p-hat$_2$ = 12 /945= 0.012698 (1.3%)
* p-hat$_3$ = 4 /127 = 0.031496 (3.1%)

* Expected frequencies

8.070	612.930
12.280	932.720
1.650	125.350

* $(O - E)^2 / E$

0.531	0.007
.006	0.000
3.347	.044

* Chi-square = 0.531 + 0.007 + 0.006 + 0.000 + 3.345 + 0.044 = **3.93**
* df = (3-1)(2-1) = 2
* $P = 0.14$ ($0.10 < P < 0.15$)
* Prevalence varied from 3.1% (group 3) to 1.0% (or some other reasonable statement).
* Observed differences were not statistically significant ($P = 0.14$).

Note: Chi-square, Yates = 2.414, df = 2, $P = 0.299$

Problems involving no or minimal calculation

4. When do you need to do a Fisher's exact test? [1 pt]

* For testing R-by-C tables when more than 20% of cells have an expected frequency that is less than 5 and no cell has an expected value less than 1.

5. What distinguishes an incidence proportion from a prevalence proportion? [1 pt]

* The proportion represents an incidence when the count of successes is based on onsets over time. Otherwise, we are dealing with a prevalence.

6. Which will produce a larger *P*-value when testing H_0: $p = p_0$, a regular z statistic or continuity-corrected z statistic? [1 pt]

* The continuity corrected z statistic will produce the larger *P*-value. It is the more conservative test.

7. Interpret an *RR* of 1.25. Be specific. [2 pt]

** An *RR* of 1.25 represents a 25% increase in risk relative to baseline. This is equivalent to saying that the exposed group has 1.25 times the risk as the nonexposed group.

8. A 95% confidence interval for an *RR* is 1.08 to 2.08. A different 95% confidence interval for an *RR* is 1.08 to 1.68. The study designs were similar. Which study had the larger sample? [1 pt]

* Clearly, the second study.

9. Name the three general types of biases in observational studies. [3 pts]

*Selection bias, *information bias (or misclassification), *confounding

10. A chi-square statistic with 1 *df* is 2.943. What is the z statistic for these data?

*sqrt(2.943) = 1.72

After the exam, please check the course calendar for the week's assignment.