Sixth Examination Study Guide

1. Know how to conduct the three kinds of tests that we discussed that use the X^2 test statistic: the *goodness-of-fit test*, the *test of independence*, and *McNemar's test*. This includes the statement of the null and alternative hypotheses, calculation of the test statistic, calculation of the p-value, and the decision.
2. Know how to compute expected counts for a goodness-of-fit test and a test of independence.
3. Know how to compute the X^2 test statistic using observed and expected counts.
4. Know how to compute the p-value using the χ^2 sampling distribution.
5. What does it mean to say that two variables are *independent*?
6. The test statistic
$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$
can *sometimes* be used instead of the X^2 test statistic for a goodness-of-fit test. What is the relationship between the z and X^2 test statistics? What is the limitation of the z test statistic?
7. The test statistic
$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}.$$
can *sometimes* be used instead of the X^2 test statistic for a test of independence. What is the relationship between the z and X^2 test statistics? What is the limitation of the z test statistic?
8. What is the purpose of a *mark-recapture* study?
9. How do you compute the Lincoln-Petersen estimator? What does it estimate?
10. Understand the importance of independence of inclusion/exclusion within the two samples in the context of a mark-recapture study, and how this assumption might be violated.
11. What are *direct sampling* and *inverse sampling* in the context of a mark-recapture study? Why is the distinction between these two kinds of sampling important?
12. Understand the purpose and calculation of Cramer's V.
13. What is *Simpson's paradox*, *Berkson's paradox*, the *ecological fallacy*, a *suppressor variable*, and a *spurious relationship*?

Formulas/expressions you should understand when and how to use.

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$$

$$\text{expected count} = \text{probability} \times n$$

$$\text{expected count} = \frac{R \times C}{T}$$

$$\text{df} = \text{number of categories} - 1$$

$$\text{df} = (r-1)(c-1)$$

$$\hat{N} = \frac{n_1 n_2}{m}$$

$$\text{standard error} = \sqrt{\frac{n_1 n_2 (n_1 - m)(n_2 - m)}{m^3}}$$

$$\text{standard error} = \sqrt{\frac{n_1^2 n_2 (n_2 - m)}{m^2 (m + 1)}}$$

$$V = \sqrt{\frac{X^2/n}{\min(r - 1, c - 1)}}$$

$$X^2 = \frac{(O_{bl} - O_{tr})^2}{O_{bl} + O_{tr}}$$