Suggested readings and practice problems fromOpenIntro Statistics, 3rd edition (a free online introductory statistics textbook co-authored by Dr. Cetinkaya-Rundel) for this week:

Suggested reading: Chapter 6, Sections 6.1 - 6.6

Suggested exercises: (End of chapter exercises from OpenIntro Statistics)

- Single proportion: 6.1, 6.3, 6.5, 6.9, 6.11, 6.13, 6.15, 6.19, 6.21
- Comparing two proportions: 6.23, 6.25, 6.27, 6.29, 6.31, 6.33, 6.35
- Inference for proportions via simulation: 6.51, 6.53, 6.55
- Comparing three or more proportions (Chi-square): 6.39, 6.41, 6.43, 6.45, 6.47

(Reminder: the solutions to the end of chapter exercises are at the end of the *OpenIntro Statistics*book)

Test yourself:

1. *Suppose 10% of Coursera students smoke. You collect many random samples of 100 Coursera students at a time, and calculate a sample proportion (p̂) for each sample, indicating the pro- portion of students in that sample who smoke. What would you expect the distribution of these p̂s to be? Describe its shape, center, and spread.*
2. *Suppose you want to construct a confidence interval with a margin of error no more than 4% for the proportion of Coursera students who smoke. How would your calculation of the required sample size change if you don't know anything about the smoking habits of Coursera students vs. if you have a reliable previous study estimating that about 10% of Coursera students smoke?*
3. *Suppose a 95% confidence interval for the difference between male and female Coursera students who smoke (male - female) is (-0.08,0.11). Interpret this interval, making sure to incorporate into your interpretation a comparative statement about the two sexes of Coursera students.*
4. *Does the above interval suggest a significant difference between the true proportions of smokers in the two groups?*
5. *Suppose you had a sample of 100 male Coursera students where 11 of them smoke, and a sample of 80 female Coursera students where 10 of them smoke. Calculate p^pool.*
6. *When and why do we use p^pool in calculation of the standard error for the difference between two sample proportions?*
7. *Explain the different hypothesis tests one could use when assessing the distribution of a categorical variable (e.g. smoking status) with only two levels (e.g. levels: smoker and non-smoker) vs. more than two levels (e.g. levels: heavy smoker, moderate smoker, occasional smoker, non-smoker).*
8. *Why is the p-value for chi-square tests always "one sided"?*
9. *What are the null and alternative hypotheses in the chi-square test of independence?*
10. *Suppose a chi-square test of independence between two categorical variables (one with 5, the other with 3 levels) yields a test statistic of χ^2 =14. What's the conclusion of the hypothesis test at 5% significance level?*
11. *Suppose you want to estimate the proportion of Coursera students who smoke. You collect a random sample of*

100 students, where only 8 of them smoke. Can you use theoretical methods (Z) to construct a confidence interval based on these data? If not, describe how you could calculate a 95% bootstrap confidence interval.

12. *Why is the p-value for chi-square tests always "one sided"?*