WORKSHEET: CHI-SQUARED

Problem 1. In each of the following, suppose χ^2 is a random variable with a chi-square distribution with the indicated number of degrees of freedom. Use the R functions `pchisq` and `qchisq` to calculate the indicated quantity.

(a) df $= 5$. Calculate $P(\chi^2 \leq 10)$.

(b) df $= 10$. Calculate $P(\chi^2 \leq 10)$.

(c) df $= 15$. Calculate $P(\chi^2 \leq 10)$.

(d) df $= 15$. Calculate $P(\chi^2 \geq 10)$.

(e) df $= 15$. Calculate $P(5 \leq \chi^2 \leq 10)$.

(f) df $= 15$. Calculate the value of the observation that is larger than exactly 80% of observations.

(g) df $= 15$. Calculate the value of the observation that is less than exactly 80% of observations.

Problem 2. Determine if each of the following statements is true or false.

(a) The chi-square distribution is always right skewed.

(b) As df increases, the mean of the chi-square distribution increases.

(c) As df increases, the shape of the chi-square distribution becomes more skewed.

Problem 3. A survey asked 827 randomly sampled registered voters in California "Do you support? Or do you oppose? Drilling for oil and natural gas off the Coast of California? Or do you not know enough to say?" Below is the distribution of responses, separated based on whether or not the respondent graduated from college.

	College Grad	Not College Grad	Total
Support	154	132	286
Oppose	180	126	306
Don't Know	104	131	235
Total	438	389	827

Conduct a hypothesis test to determine if opinions on this issue are associated with whether or not an individual has graduated from college.

Problem 4. A study examined microhabitat factors associated with forage and bed sites of barking deer in Hainan Island, China. In this region, woods make up 4.8% of the land, cultivated grass plot makes up 14.7%, deciduous forests make up 39.6%, and "other" makes up the rest.

Of the 426 sites where the deer forage, 4 were categorized as woods, 16 as cultivated grassplot, 61 as deciduous forests, and the rest as "other." Conduct a hypothesis test to determine if barking deer prefer to forage in certain habitats over others.

Problem 5. Researchers conducted a study investigating the relationship between caffeinated coffee consumption and risk of depression in women. They collected data on 50,739 women free of depression symptoms at the start of the study in the year 1996, and these women were followed through 2006. The researchers used questionnaires to collect data on caffeinated coffee consumption, asked each individual about physician-diagnosed depression, and also asked about the use of antidepressants. The table below shows the distribution of incidences of depression by amount of caffeinated coffee consumption.

	$\leq$ 1 cup/week	2-6 cups/week	1 cup/day	2-3 cups/day	$\geq$ 4 cups/day	Total
Depression	670	373	905	546	95	2607
No Depression	11545	6244	16329	11726	2288	48132
Total	12215	6617	17234	12290	2383	50739

(a) Conduct a hypothesis test to determine if there is an association between coffee intake and clinical depression.

(b) One of the authors of this study was quoted on the NYTimes as saying it was "too early to recommend that women load up on extra coffee" based on just this study. Do you agree with this statement? Explain your reasoning.