Exercise 11.1

To study the effect of prolonged inhalation of cadmium, a researcher randomized a group of 20 dogs into two groups: "experimental" (`group=1`) and "control" (`group=1`). The 10 experimental animals were exposed to cadmium oxide while the 10 control animals were not exposed. At the end of the experiment, the level of hemoglobin (`hemo`) was determined for each of the 20 dogs.

A Wilcoxon rank sum test was performed to compare the two groups; Stata output is below.

```
. ranksum hemo, by(group)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

       group |      obs    rank sum    expected
-------------+---------------------------------
           0 |       10       137.5         105
           1 |       10        72.5         105
-------------+---------------------------------
    combined |       20         210         210

unadjusted variance      175.00
adjustment for ties       -1.05
                     ----------
adjusted variance        173.95

Ho: hemo(group==0) = hemo(group==1)
             z =   2.464
    Prob > |z| =   0.0137
```

(questions on next page)

(a) What are the null and alternative hypotheses for this test?

H_0 : the distribution of hemoglobin is the same for dogs exposed to cadmium oxide and dogs who were not exposed
H_a : the values of hemoglobin are systematically larger for one group

(b) What is the value of the test statistic, its distribution under the null, and the p-value for this test?

test statistic z = 2.464
$z \sim N(0,1)$ under H_0
p-value = 0.0137

(c) Write a one-sentence conclusion (in context) based on the results of this test. Assume $\alpha = 0.05$.

There is evidence that the values of hemoglobin are systematically larger in one group of dogs.

(d) The largest value of hemoglobin in the data set is 18. Suppose this were a typo, and this one value was really 32. How would the result of the Wilcoxon rank sum test change? (In other words, if we replace the largest value of 18 with the value 32, what happens to the test result?)

There would be no change – only the **ranks** *are used in the calculations, so since this will remain the largest value, its rank won't change (will remain rank=20), so the test result will be unchanged.*

Exercise 11.2

We have permeability constants of the human chorioamnion (a placental membrane) taken at two time points: 40 weeks gestational age (GA), and between 12 to 26 weeks GA. Note that these measurements were all on different pregnant women (not the same women measured twice). A nonparametric test was used to determine if there is different permeability of the human chorioamnion by GA. Write a one sentence conclusion of the results of this test, including reference to a p-value. Assume $\alpha = 0.05$.

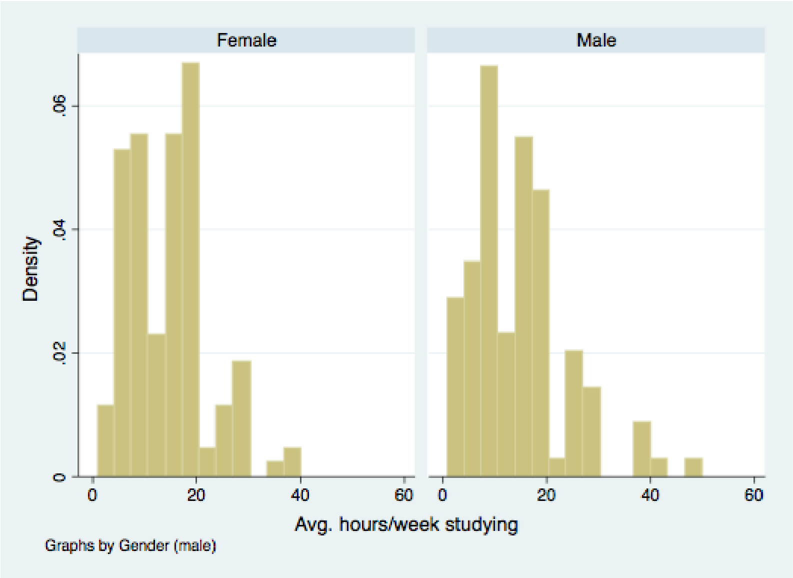
```
. ranksum perm, by(ga)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

          ga |      obs    rank sum    expected
-------------+---------------------------------
    40 weeks |       10          90          80
 12-26 weeks |        5          30          40
-------------+---------------------------------
    combined |       15         120         120

unadjusted variance       66.67
adjustment for ties        0.00
                     ----------
adjusted variance         66.67

Ho: perm(ga==40 weeks) = perm(ga==12-26 weeks)
             z =   1.225
    Prob > |z| =   0.2207
```

There is no evidence that the distribution of permeability of the human chorioamnion is different at 40 weeks versus 12-26 weeks (p=0.2207).

Exercise 11.3

A survey of a random sample of students at the University of New Hampshire was conducted. School administrators would like to use these data to test whether there are differences between male (`gender=1`) and female (`gender=0`) students in the number of hours they report studying per week (`study`).

(a) A histogram of the hours studied, separated by sex, was produced and is below. Explain why we might want to use a nonparametric test to compare the two groups instead of a parametric one, and state which nonparametric test would be appropriate to use.



Distribution is clearly non-normal in each group, thus the normality assumption required for a t-test or linear regression is violated. The appropriate nonparametric test would be a Wilcoxon rank sum test.

(b) The nonparametric test was run in Stata, output is below. Draw a conclusion from the test (in context), referencing a p-value. Assume $\alpha = 0.05$.

```
Two-sample Wilcoxon rank-sum (Mann-Whitney) test

      gender |      obs    rank sum    expected
-------------+---------------------------------
      Female |      133     16232.5       15960
        Male |      106     12447.5       12720
-------------+---------------------------------
    combined |      239       28680       28680

unadjusted variance   281960.00
adjustment for ties    -3783.87
                     ----------
adjusted variance     278176.13

Ho: study(gender==Female) = study(gender==Male)
             z =   0.517
    Prob > |z| =   0.6054
```

We do not have evidence that the distribution of hours studied is different for male and female students (p=0.6054).

(c) If we are willing to assume that the shape of the distributions is the same in both groups, what is another way we could state our conclusion?

There is no evidence that the median hours studied is different for male and female students (p=0.6054).

Exercise 11.4

A survey of a random sample of students at the University of New Hampshire was conducted. School administrators would like to use these data to test whether there are differences among students based on year in school (`year`) in the number of hours they report studying per week (`study`). A nonparametric test was used to test this; Stata output is below. Write a one sentence conclusion of the results of this test, including reference to a p-value. Assume $\alpha = 0.05$.

```
. kwallis study, by(year)

Kruskal-Wallis equality-of-populations rank test

  +----------------------------+
  |      year | Obs | Rank Sum |
  |-----------+-----+----------|
  |  Freshman |  39 |  5150.00 |
  | Sophomore |  64 |  7011.50 |
  |    Junior |  73 |  9137.00 |
  |    Senior |  63 |  7381.50 |
  +----------------------------+

chi-squared =     3.159 with 3 d.f.
probability =     0.3678

chi-squared with ties =     3.202 with 3 d.f.
probability =     0.3615
```

There is no evidence that the distribution of hours studied is not the same for freshman, sophomores, juniors, and seniors.

Exercise 11.5

A survey of a random sample of students at the University of New Hampshire was conducted. Information was collected on the amount of alcohol students consumed. The result was a 33-point drinking scale score, where a higher score means more alcohol consumption. We are interested in whether there are differences in alcohol consumption among the years in school (`year`: 1=Freshman, 2=Sophomore, 3=Junior, 4=Senior). However, we are concerned that the distribution of the drinking scores might not be normal within each group, so we use a Kruskal-Wallis test. Stata output is below.

```
. kwallis drink, by(year)

Kruskal-Wallis equality-of-populations rank test

  +---------------------------+
  |      year | Obs | Rank Sum |
  |-----------+-----+----------|
  |  Freshman |  40 |  4914.00 |
  | Sophomore |  65 |  9341.50 |
  |    Junior |  75 |  9300.50 |
  |    Senior |  63 |  6090.00 |
  +---------------------------+

chi-squared =    14.453 with 3 d.f.
probability =     0.0023

chi-squared with ties =    14.490 with 3 d.f.
probability =     0.0023
```

(questions next page)

(a) What are the null and alternative hypotheses for this test?

H_0 : the distribution of drinking score is the same for all four years in school
H_a : the values of drinking score are systematically larger for one group based on year in school

(b) What is the value of the test statistic, its distribution under the null, and the p-value for this test?

test statistic H = 14.490 (taking the "adjusted for ties" value)
$H \sim \chi^2(3)$ under H_0
p-value = 0.0023

(c) Write a one-sentence conclusion (in context) based on the results of this test. Assume $\alpha = 0.05$.

There is evidence that the values of drinking score are systematically larger in at least one group of students (freshman, sophomores, juniors, seniors).

(d) The median drinking scores for each group are: freshman=19.0, sophomores=21.2, juniors=19.5, seniors=16.7. Based on this and the test result, can we conclude that seniors have significantly lower drinking scores than the other three groups? Why or why not?

*No – we can only conclude that *some* group is different, not which one (and not any directionality).*

(e) Suppose that drinking scores really **are** normally distributed (in the population) for each group of students. Name two disadvantages of using the nonparametric Kruskal-Wallis test instead of a parametric test like ANOVA.

– K-W test may be less powerful than ANOVA
– Hypotheses for the K-W test ("distributions") are more awkward than those for ANOVA ("means")
– We can't easily adjust for other predictors that might be important (e.g., confounders)