Stephanie Mecham
EPID 674.002

Exercise 4: Basic Hypothesis Testing with R

Problem 1.
Yes, we can conclude that there is a statistically significant association between urinary bpa and age category in our sample. The results of the Kruskal-Wallis rank sum test (which was performed as a non-parametric test for a continuous dependent variable and ordinal independent variable) show us that the p-value is < 2.2e-16, which is far below the commonly used alpha threshold of 0.05 to determine statistical significant associations.

Results:
Kruskal-Wallis rank sum test

data: ubpa by bpa$age_cat
Kruskal-Wallis chi-squared = 122.02, df = 2, p-value < 2.2e-16

Problem 2.
Yes, we can conclude that there is also a statistically significant association between urinary bpa and sex. The results of the Wilcoxon rank sum test (used as a non-parametric test for continuous dependent variables and dichotomous independent variables) resulted in a p-value of 1.977e-05, which is again below our alpha threshold of 0.05. Therefore we can reject the null hypothesis and conclude there is a statistically significant association between urinary bpa and sex.

Results:
Wilcoxon rank sum test with continuity correction

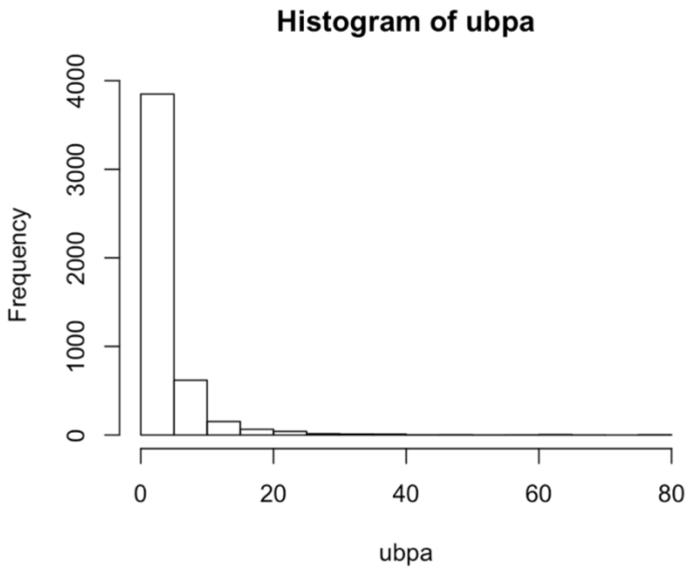
data: ubpa by gender
W = 3062300, p-value = 1.977e-05
alternative hypothesis: true location shift is not equal to 0

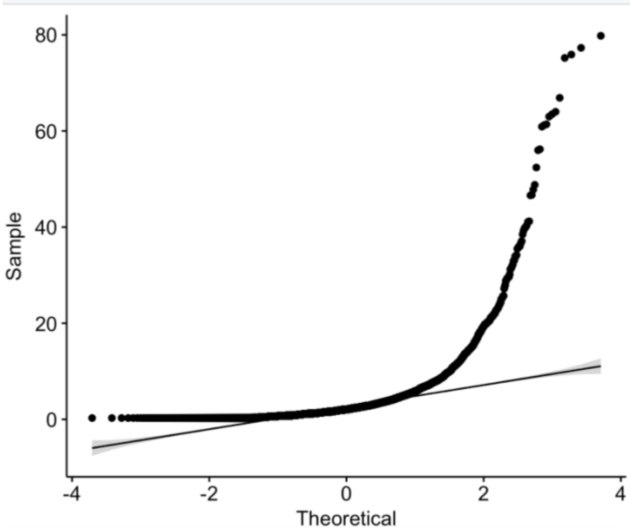
Problem 3.
The Kruskal-Wallis rank sum test we performed to test the association between urinary bpa and smoking status resulted in a p-value of = 0.004993. Since this is below our alpha level of 0.05, we can reject the null and state that there is a statistically significant association between these two variables in our sample.

Results:
Kruskal-Wallis rank sum test

data: ubpa by SMK
Kruskal-Wallis chi-squared = 10.599, df = 2, p-value = 0.004993

SIDENOTE – I chose to perform non-parametric tests because analysis of a histogram, qq-plot, and Shapiro-Wilk normality test showed that the ubpa distribution did not appear normal.





Shapiro-Wilk normality test

data: bpa$ubpa
W = 0.50874, p-value < 2.2e-16
Small p-value -> distribution is significantly different from a normal distribution.

Problem 4.
Compared to the first quartile, the second quartile had an odds ratio of 1.02 (0.79-1.3), the third quartile had an odds ratio of 1 (0.78-1.29), and the fourth had an odds ratio of 0.98 (0.76,

1.25). These odds ratios are not particularly striking. None of the quartiles had an odds ratio greater than 0.02 away from the reference group. Using the odds ratios alone, I wold reasonably conclude that there is not a statistically significant association between urinary bpa and type 2 diabetes in the U.S. The Chi-squared and Fisher's square test corroborates this conclusion, with p-values well above the standard alpha level of 0.05.

Results:

```
          bpa$quantfac
T2DM         First Quantile Fourth Quantile Second Quantile
  0          1089        1044        1058
  1           154         144         152

  Odds ratio    1           0.98        1.02
  lower 95% CI              0.76        0.79
  upper 95% CI              1.25        1.3
          bpa$quantfac
T2DM         Third Quantile
  0          1001
  1           142

  Odds ratio    1
  lower 95% CI  0.78
  upper 95% CI  1.29
```

Chi-squared = 0.112 , 3 d.f., P value = 0.99
Fisher's exact test (2-sided) P value = 0.99

CODE:

```
> #Exercise 4 | EPID 674.002 | Stephanie Mecham
>
>attach(bpa)
>
>#Assessing normality of urinary bpa
>hist(ubpa) #Histogram
>install.packages("ggpubr")
>library("dplyr")
>library("ggpubr")
> ggqqplot(bpa$ubpa) #QQ Plot
> shapiro.test(bpa$ubpa) #Shapiro-Wilk normality test

        Shapiro-Wilk normality test
```

```
data:  bpa$ubpa
W = 0.50874, p-value < 2.2e-16
>
>#Creating age categories
>
>#Problem One:
> bpa$age_cat <- as.factor(bpa$age_cat)
> kruskal.test (ubpa ~ bpa$age_cat)

	Kruskal-Wallis rank sum test

data:  ubpa by bpa$age_cat
Kruskal-Wallis chi-squared = 122.02, df = 2, p-value < 2.2e-16
>
>
>#Problem Two
> wilcox.test(ubpa ~ gender)

	Wilcoxon rank sum test with continuity correction

data:  ubpa by gender
W = 3062300, p-value = 1.977e-05
alternative hypothesis: true location shift is not equal to 0
>
>#Problem Three
> kruskal.test(ubpa ~ SMK)

	Kruskal-Wallis rank sum test

data:  ubpa by SMK
Kruskal-Wallis chi-squared = 10.599, df = 2, p-value = 0.004993
>
>
>#Problem Four
>
>#Creating binary T2DM variable:
>diabetic <- ifelse(a1c >= 6.5, 1, 0)
> diabetic_meds <- ifelse (dmmed == 1,1,0)
> T2DM <- ifelse (!is.na(diabetic) & (diabetic==1) | (!is.na(diabetic_meds) & (diabetic_meds
==1)) ,1, 0)
>
>#Creating quartiles:
> quantile(ubpa)
   0%   25%   50%   75%  100%
```

```
 0.28  1.00  2.10  4.10 79.80
> bpa$quant[ubpa <= 1.00] <- "First Quantile"
> bpa$quant[ubpa > 1.00 & ubpa <= 2.10] <- "Second Quantile"
> bpa$quant[ubpa > 2.10 & ubpa <= 4.10] <- "Third Quantile"
> bpa$quant[ubpa > 4.10] <- "Fourth Quantile"
>
>
>#Calculating number of cases per quartile
> sum(T2DM==1 & bpa$quant=="First Quantile")
[1] 154
> sum(T2DM==1 & bpa$quant=="Second Quantile")
[1] 152
> sum(T2DM==1 & bpa$quant=="Third Quantile")
[1] 142
> sum(T2DM==1 & bpa$quant=="Fourth Quantile")
[1] 144
> sum(T2DM==1)
[1] 592
>
>#Calculating odds ratios
> bpa$quantfac <- factor(bpa$quant) #converting quant to a factor
> bpa$quantfac <- relevel(bpa$quantfac, ref = "First Quantile") #Setting reference group
> cc(T2DM, bpa$quantfac) #Calculating odds ratio

            bpa$quantfac
T2DM         First Quantile Fourth Quantile Second Quantile
  0           1089          1044           1058
  1           154           144            152

  Odds ratio     1           0.98           1.02
  lower 95% CI               0.76           0.79
  upper 95% CI               1.25           1.3
            bpa$quantfac
T2DM         Third Quantile
  0           1001
  1           142

  Odds ratio     1
  lower 95% CI   0.78
  upper 95% CI   1.29

Chi-squared = 0.112 , 3 d.f., P value = 0.99
Fisher's exact test (2-sided) P value = 0.99
>
```

```
>
>#End of code
```