Final Project

Parker Lambert

2023-11-16

Name: Parker Lambert

Assignment: Final Homework

Course: QBS119

Due: 11/17/23

Correlates of Phthalates Exposure

Libraries

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(vtable)
```

```
## Loading required package: kableExtra
```

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'
```

```
##
## Attaching package: 'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##     recode
##
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(ggplot2)
```

Load in Data

```
df <- read.csv("nhanes2009_phthalates.csv")
df |>
  dim()#head(3)
```

```
## [1] 2819   14
```

Removed Patients

Only include adults 21 years of age or older.

Make sure to report where/why we lose participants (e.g., how many participants are excluded because they are missing one or more of the variables of interest). It is critical to report why participants were excluded for reproducibility.

Exclude anyone missing the primary dependent and independent variables of interest – urinary metabolites, BMI, and diet data – do not make a "missing" category. Specifically, the analytical sample should be limited to those with data on each urinary metabolite, BMI, and each diet measure. That way the inference from any one analysis applies to same overall sample. Make sure the sociodemographic summaries also reflect that same analytic dataset.

```
# remove patients under the age of 21
df <- df |>
  filter(RIDAGEYR>=21)
# remove anyone missing urinary metabolites
df <- df |>
  filter(!is.na(URXMBP))
df <- df |>
  filter(!is.na(URXMZP))
# remove anyone missing BMI
df <- df |>
  filter(!is.na(BMXBMI))
df <- df |>
```

```
  filter(!is.na(BMXBMICAT))
# remove anyone missing diet data which is also as 9999
# First convert all instances of 9999 to na
df <- df |>
  mutate(DBD900 = replace(DBD900, DBD900 == 9999, NA)) |>
  mutate(DBD905 = replace(DBD905, DBD905 == 9999, NA)) |>
  mutate(DBD910 = replace(DBD910, DBD910 == 9999, NA))
df <- df |>
  filter(!is.na(DBD900)) |>
  filter(!is.na(DBD905)) |>
  filter(!is.na(DBD910))
# check dimentions
df |>
  dim()
```

```
## [1] 1843   14
```

We were asked to remove all patients under the age of 21 from our dataset this resulted in the rows in the dataset to reduce from 2819 down to 1931. After removing all patients with NA values in URXMBP we were left with 1908 rows. Then we removed all patients with NA values in URXMZP leaving us with 1863 rows. When we removed rows holding NA in our two BMI variables no extra rows were removed. Finally we need to remove all missing values from our 3 diet columns. In these columns NA could be in as NA or as 9999 so first we have to convert the 9999 to NA then remove NA. This dropped our row count down to 1843.

Format Factor Variables

First we need to change how BMXBMICAT is stored so that we can convert it to factor. **ALSO THE DATA DICTIONARY IS ACTUAL INCORRECT FOR BMXBMICAT IT SAYS THERE IS A GROUP 5 BUT THIS IS NOT THE CASE IT SKIPPS GROUP 2.**

```
# store BMXBMICAT as a single integer
df <- df |>
  mutate(BMXBMICAT = as.numeric(str_extract(BMXBMICAT, "\\d+")))
```

Now we can follow the data dictionary and create our factors/ catagories.

```
df <- df |>
  mutate(RIAGENDR = factor(RIAGENDR, levels = c(1,2), labels = c("Male",  "Female"))) |>
  mutate(RIDRETH1 = factor(RIDRETH1, levels = c(1, 2, 3, 4, 5),
                           labels = c("Mexican American",
                                      "Other Hispanic",
                                      "Non-Hispanic White",
                                      "Non-Hispanic Black",
                                      "Other Race - Including Multi-Racial"))) |>
  mutate(DMDEDUC2 = factor(DMDEDUC2, levels = c(1, 2, 3, 4, 5, 7, 9),
                           labels = c("Less Than 9th Grade",
                                      "Some high school, no diploma",
                                      "High School Grad/GED or Equivalent",
                                      "Some College or AA degree",
                                      "College Graduate or above",
                                      "Refused",
```

```
                                        "Don't Know"))) |>
  mutate(INDHHIN2 = factor(INDHHIN2, levels = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 77, 99),
                           labels = c("$ 0 to $ 4,999",
                                      "$ 5,000 to $ 9,999",
                                      "$10,000 to $14,999",
                                      "$15,000 to $19,999",
                                      "$20,000 to $24,999",
                                      "$25,000 to $34,999",
                                      "$35,000 to $44,999",
                                      "$45,000 to $54,999",
                                      "$55,000 to $64,999",
                                      "$65,000 to $74,999",
                                      "Over $20,000",
                                      "Under $20,000",
                                      "$75,000 to $99,999",
                                      "$100,000 and Over",
                                      "Refused",
                                      "Don't know"))) |>
  mutate(BMXBMICAT = factor(BMXBMICAT, levels = c(1, 0, 2, 3, 4),
                                labels = c("Underweight: BMI <18 kg/m2",
                                           "Healthy weight: BMI 18 to <25 kg/m2",
                                           "With overweight: BMI 25 to <30 kg/m2",
                                           "With class 1 obesity: BMI 30 to <35 kg/m2",
                                           "With class 2 or higher obesity: BMI >=35 kg/m2")))
```

Summarize the Dataset

Numerical Summary

For the final I am going to implement one of my favorite packages from undergrad vtable. This allows us to get a summary of all our variables in one table.

```
df |>
  st()
```

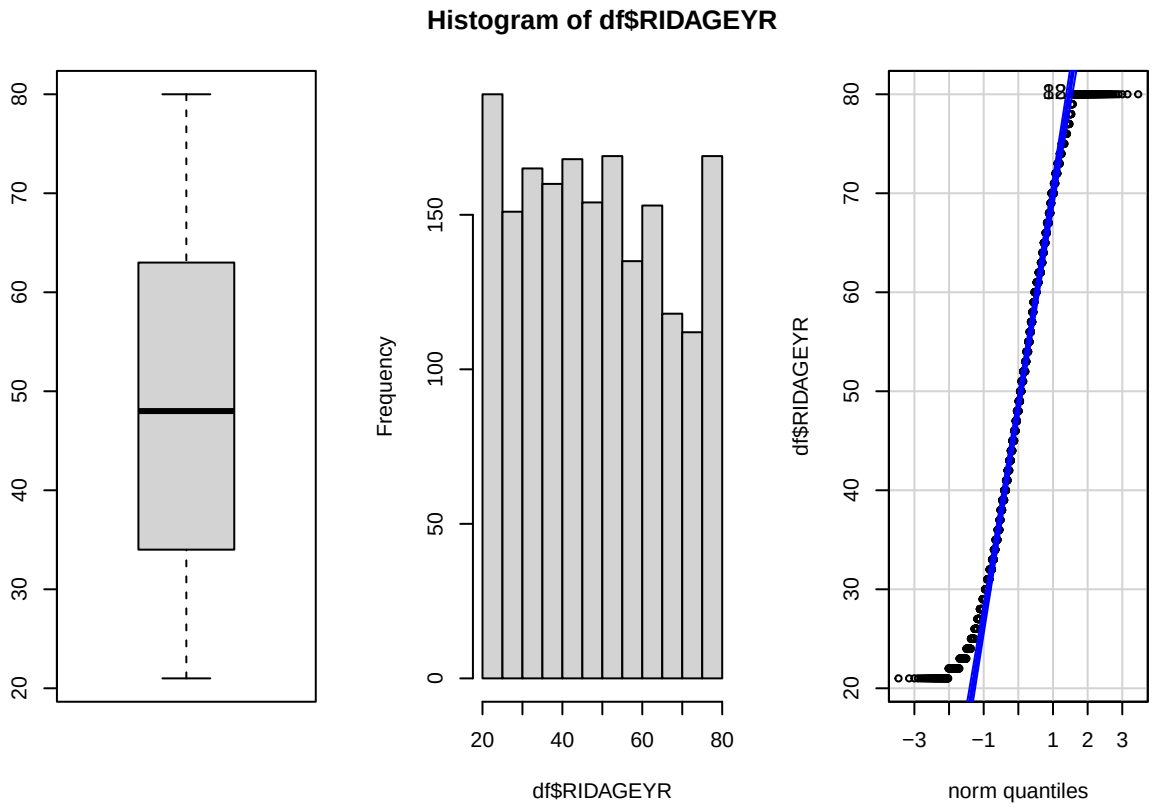
Visual Summaries

Continuous Variables Lets first look at **Age**

```
par(mfrow=c(1,3))
boxplot(df$RIDAGEYR)
hist(df$RIDAGEYR)
qqPlot(df$RIDAGEYR)
```

Table 1: Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
SEQN	1843	56871	3056	51629	54203	59526	62157
RIAGENDR	1843						
... Male	921	50%					
... Female	922	50%					
RIDAGEYR	1843	49	18	21	34	63	80
RIDRETH1	1843						
... Mexican American	329	18%					
... Other Hispanic	183	10%					
... Non-Hispanic White	901	49%					
... Non-Hispanic Black	328	18%					
... Other Race - Including Multi-Racial	102	6%					
DMDEDUC2	1843						
... Less Than 9th Grade	235	13%					
... Some high school, no diploma	294	16%					
... High School Grad/GED or Equivalent	436	24%					
... Some College or AA degree	489	27%					
... College Graduate or above	386	21%					
... Refused	3	0%					
... Don't Know	0	0%					
INDHHIN2	1834						
... $ 0 to $ 4,999	46	3%					
... $ 5,000 to $ 9,999	69	4%					
... $10,000 to $14,999	146	8%					
... $15,000 to $19,999	132	7%					
... $20,000 to $24,999	127	7%					
... $25,000 to $34,999	225	12%					
... $35,000 to $44,999	170	9%					
... $45,000 to $54,999	134	7%					
... $55,000 to $64,999	103	6%					
... $65,000 to $74,999	74	4%					
... Over $20,000	94	5%					
... Under $20,000	15	1%					
... $75,000 to $99,999	163	9%					
... $100,000 and Over	246	13%					
... Refused	49	3%					
... Don't know	41	2%					
INDFMPIR	1665	2.4	1.7	0	1	4.1	5
URXMBP	1843	13	19	0.11	3.1	15	261
URXMZP	1843	13	23	0.15	2.8	14	415
DBD900	1843	3.5	3.8	0	1	4	39
DBD905	1843	1.6	5.1	0	0	1	90
DBD910	1843	17	21	0	2	18	126
BMXBMI	1843	29	6.6	16	25	33	85
BMXBMICAT	1843						
... Underweight: BMI <18 kg/m2	15	1%					
... Healthy weight: BMI 18 to <25 kg/m2	487	26%					
... With overweight: BMI 25 to <30 kg/m2	649	35%					
... With class 1 obesity: BMI 30 to <35 kg/m2	389	21%					
... With class 2 or higher obesity: BMI >=35 kg/m2	303	16%					

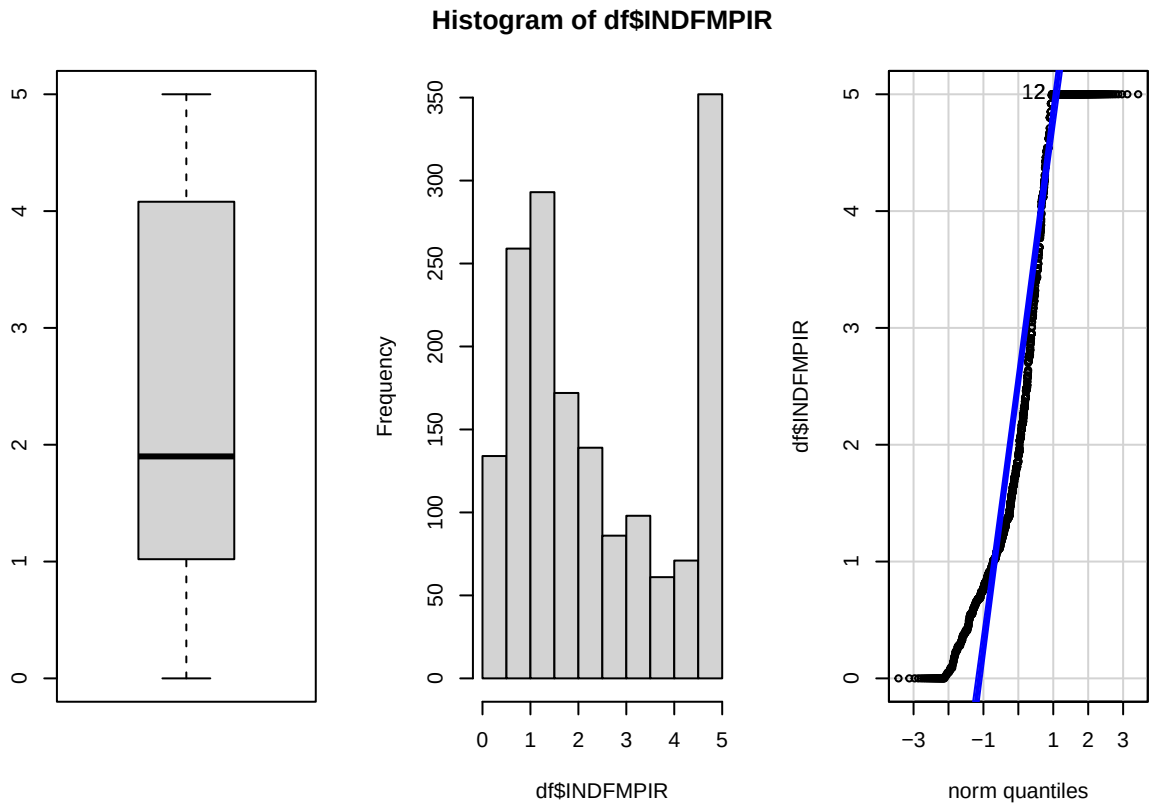



```
## [1] 18 82
```

From the Visual Summary above we can see that the distribution is roughly uniform.

Now lets look at **Ratio of Family income to Poverty**. This is capped at 5 so lets keep that in mind when looking at the data.

```
par(mfrow=c(1,3))
boxplot(df$INDFMPIR)
hist(df$INDFMPIR)
qqPlot(df$INDFMPIR)
```

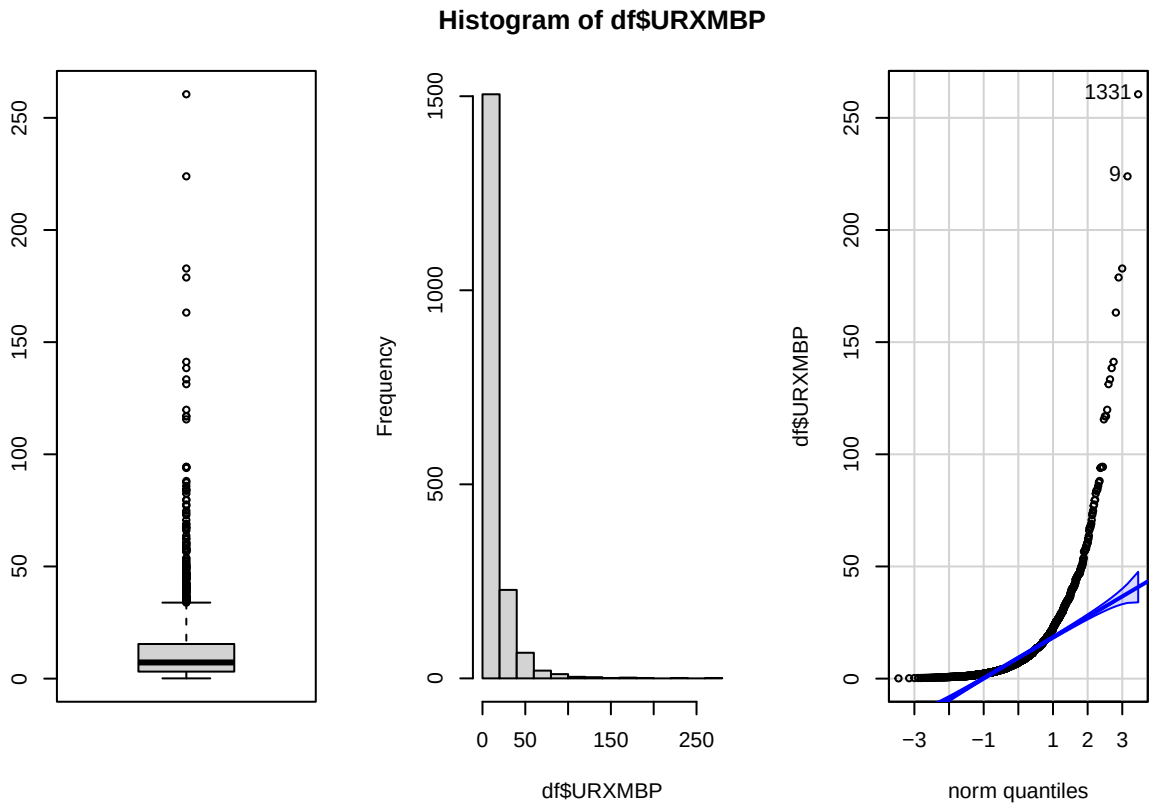



```
## [1]  2 12
```

From the Visual Summary above we can see that the distribution is roughly skewed positive however since all measures above 5 are converted to 5 we see a spike for five.

Mono-n-butyl phthalate (ng/mL)

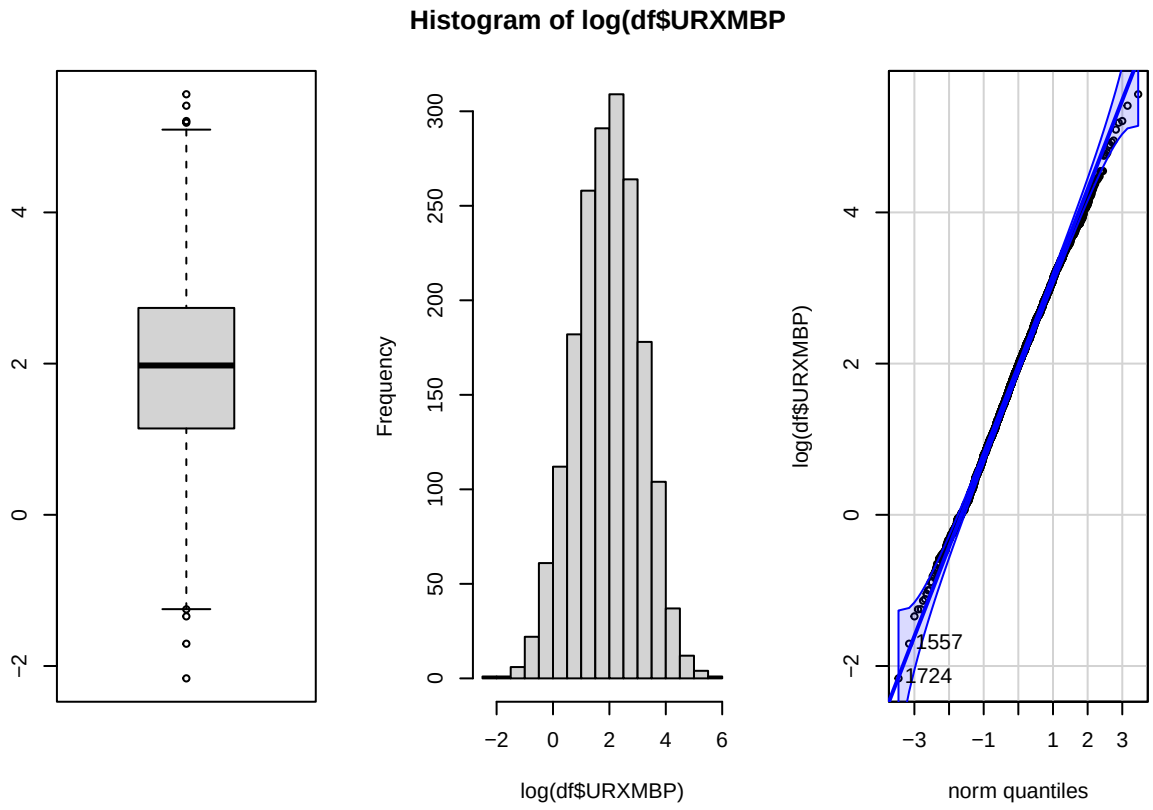
```
par(mfrow=c(1,3))
boxplot(df$URXMBP)
hist(df$URXMBP)
qqPlot(df$URXMBP)
```




```
## [1] 1331    9
```

From the Visual Summary above we can see that the distribution is skewed positive which we can now log transform to test if it is normal on a log scale.

```
par(mfrow=c(1,3))
boxplot(log(df$URXMBP))
hist(log(df$URXMBP))
qqPlot(log(df$URXMBP))
```

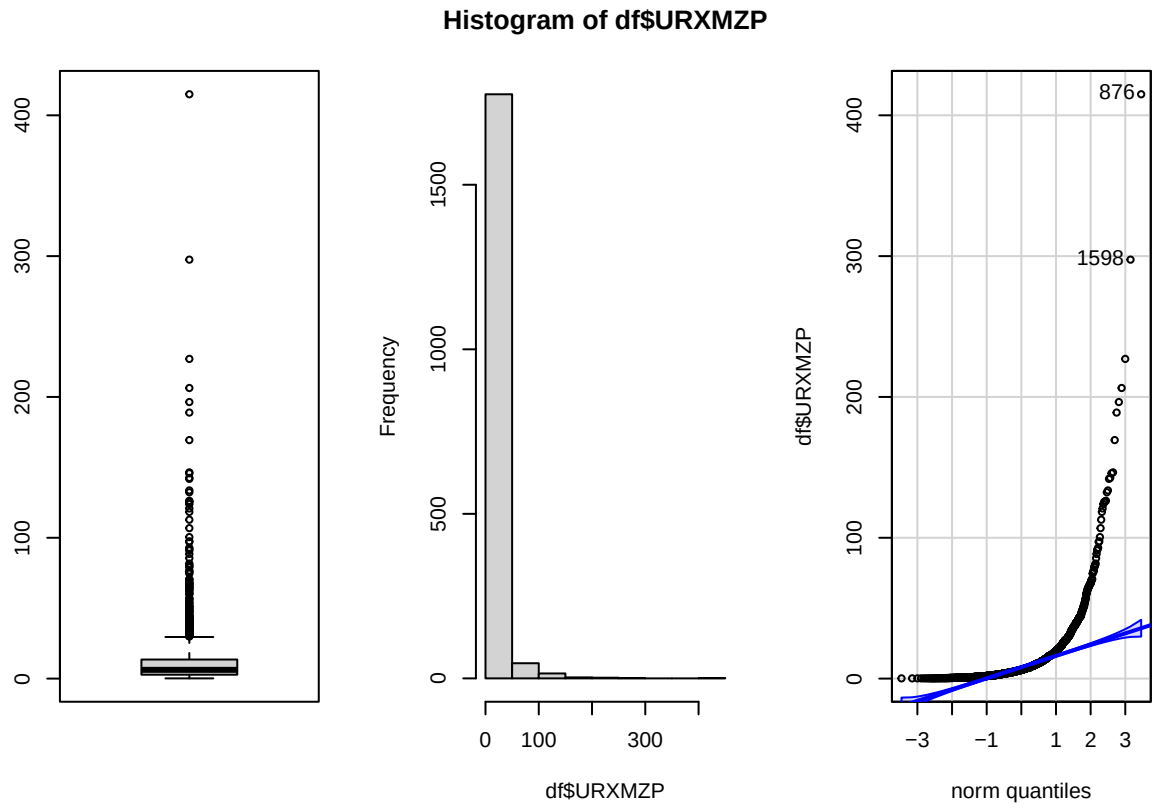



```
## [1] 1724 1557
```

From the Visual Summary above we can see that the distribution when log transformed is normal.

Mono-benzyl phthalate (ng/mL)

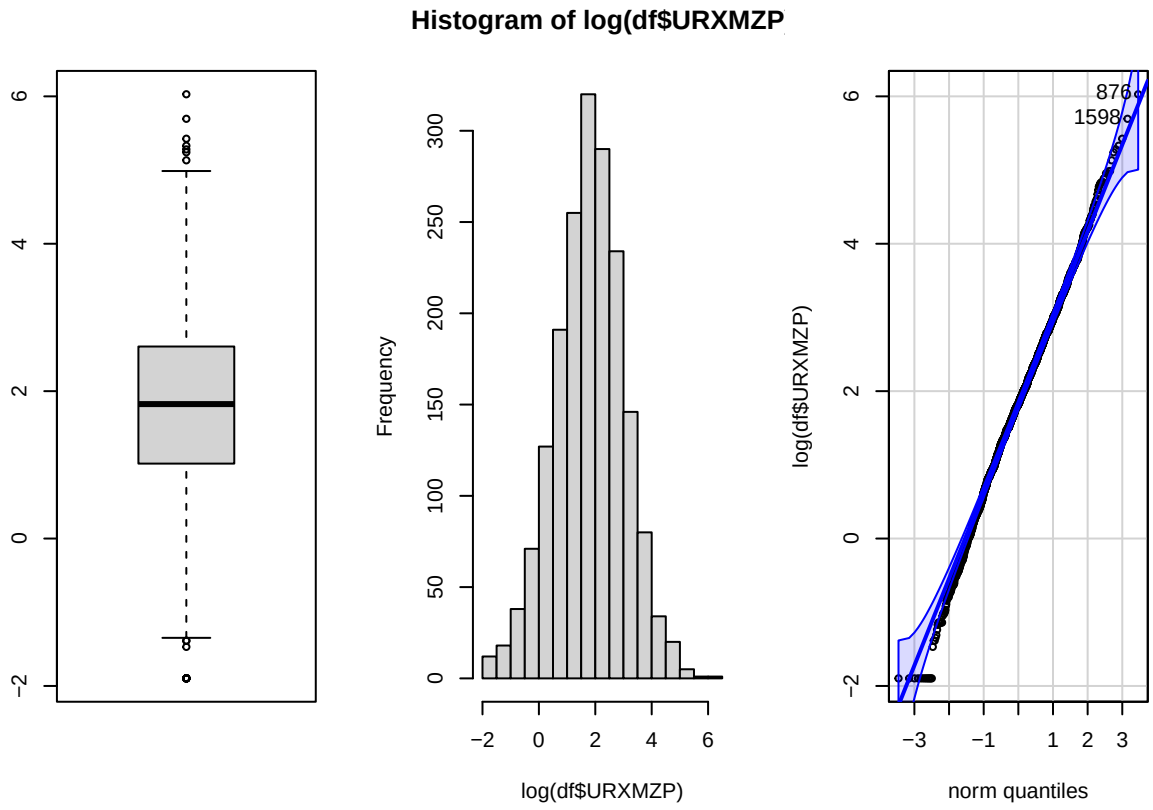
```
par(mfrow=c(1,3))
boxplot(df$URXMZP)
hist(df$URXMZP)
qqPlot(df$URXMZP)
```




```
## [1]  876 1598
```

From the Visual Summary above we can see that the distribution is skewed positive which we can now log transform to test if it is normal on a log scale.

```
par(mfrow=c(1,3))
boxplot(log(df$URXMZP))
hist(log(df$URXMZP))
qqPlot(log(df$URXMZP))
```

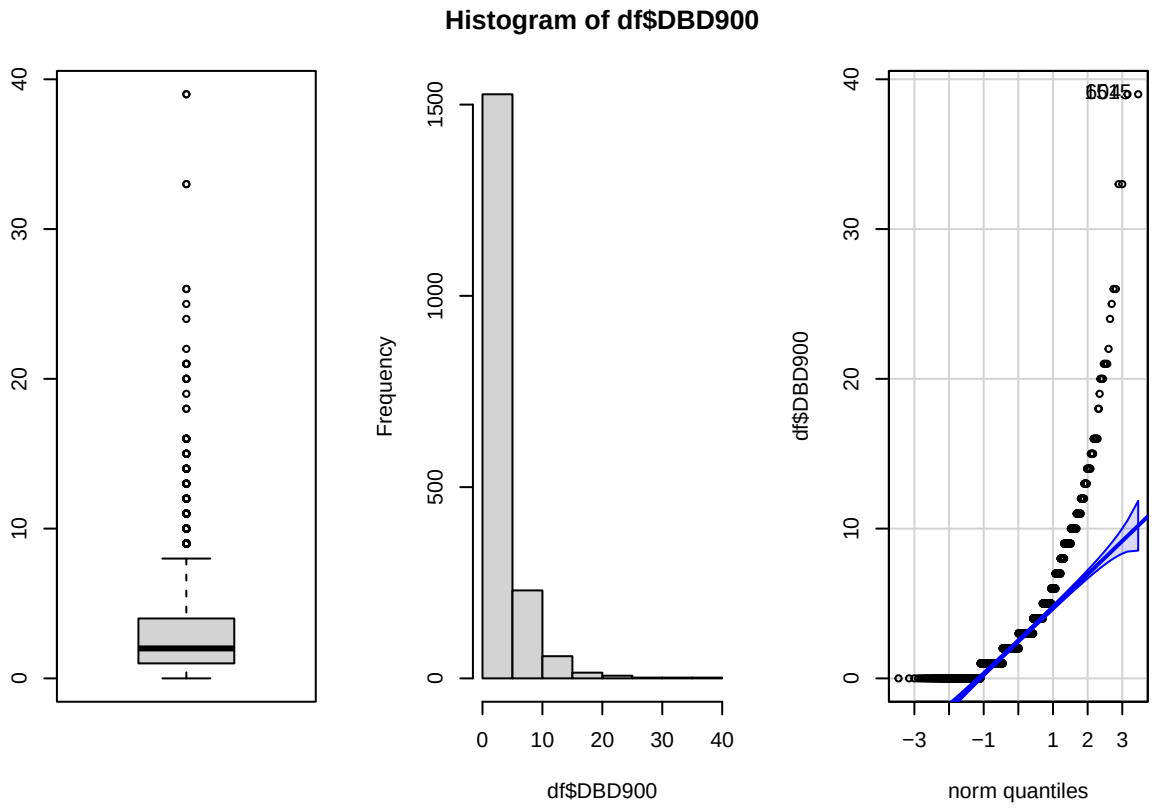



```
## [1]  876 1598
```

From the Visual Summary above we can see that the distribution when log transformed is normal.

Number of meals from a fast food or pizza place in the past week

```
par(mfrow=c(1,3))
boxplot(df$DBD900)
hist(df$DBD900)
qqPlot(df$DBD900)
```



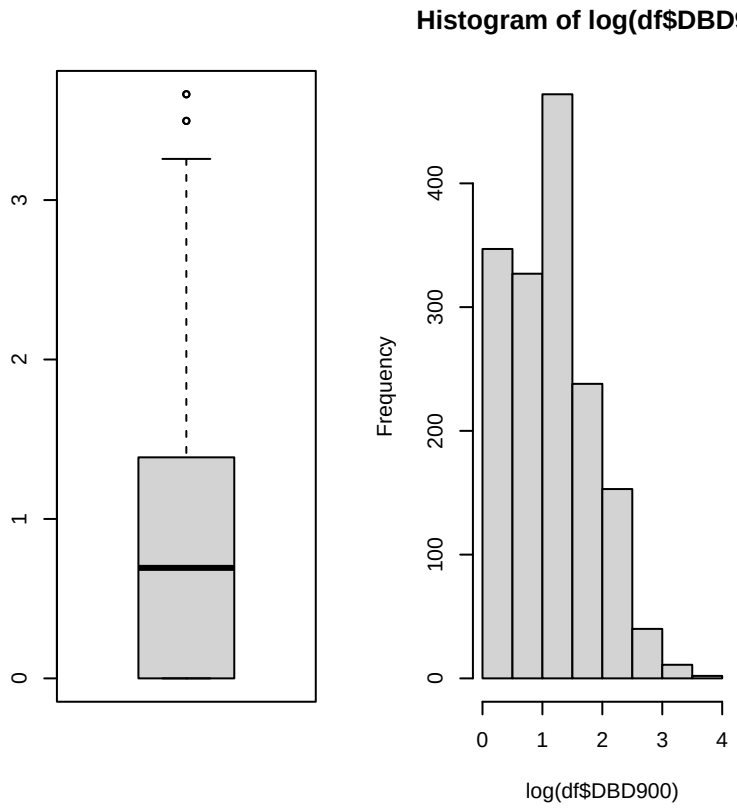

```
## [1]  604 1515
```

From the Visual Summary above we can see that the distribution is skewed positive which we can now log transform to test if it is normal on a log scale.

```
par(mfrow=c(1,3))
boxplot(log(df$DBD900))
```

```
## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out =
## z$out[z$group == : Outlier (-Inf) in boxplot 1 is not drawn
```

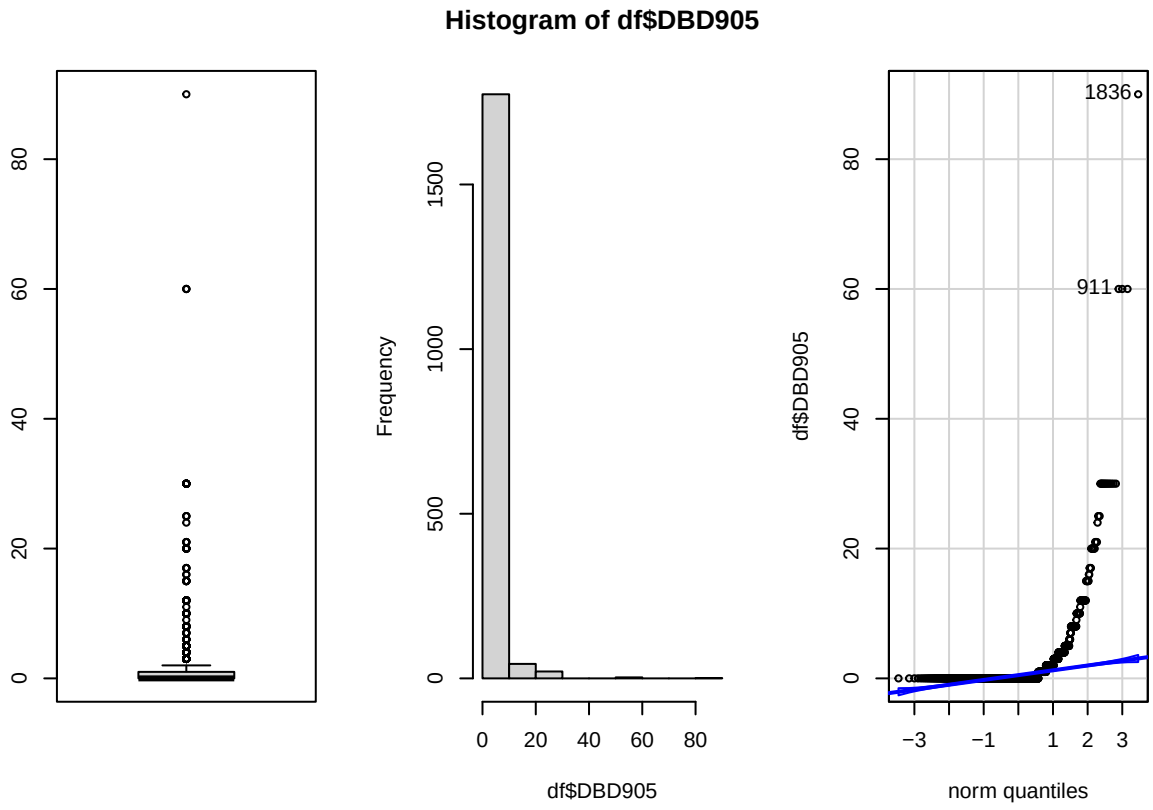
```
hist(log(df$DBD900))
#qqPlot(log(df$DBD900))
```



From the Visual Summary above we can see that the distribution when log transformed is still skewed postive.

Number of ready-to-eat foods in past 30 days

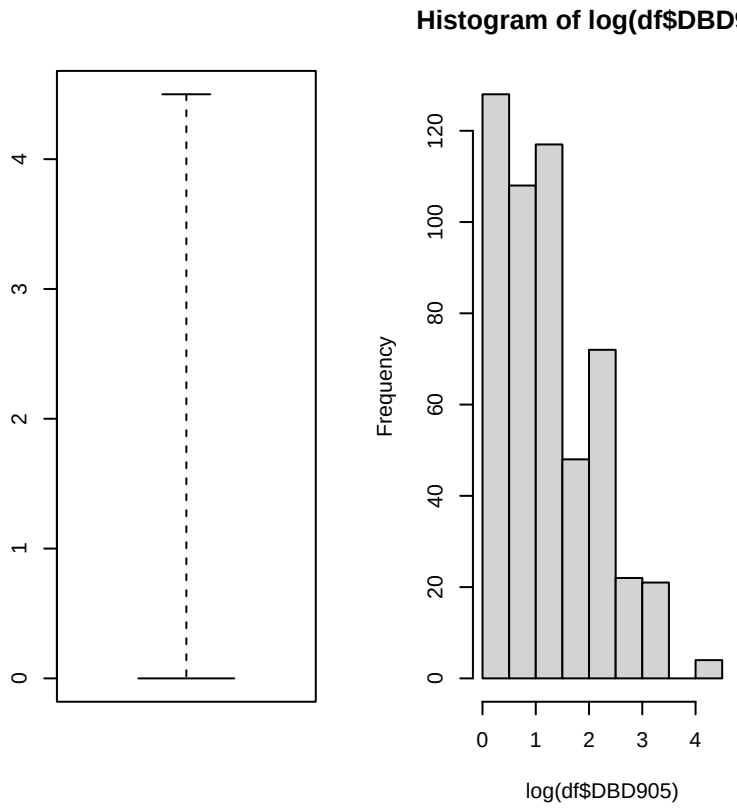
```
par(mfrow=c(1,3))
boxplot(df$DBD905)
hist(df$DBD905)
qqPlot(df$DBD905)
```




```
## [1] 1836  911
```

From the Visual Summary above we can see that the distribution is skewed positive which we can now log transform to test if it is normal on a log scale.

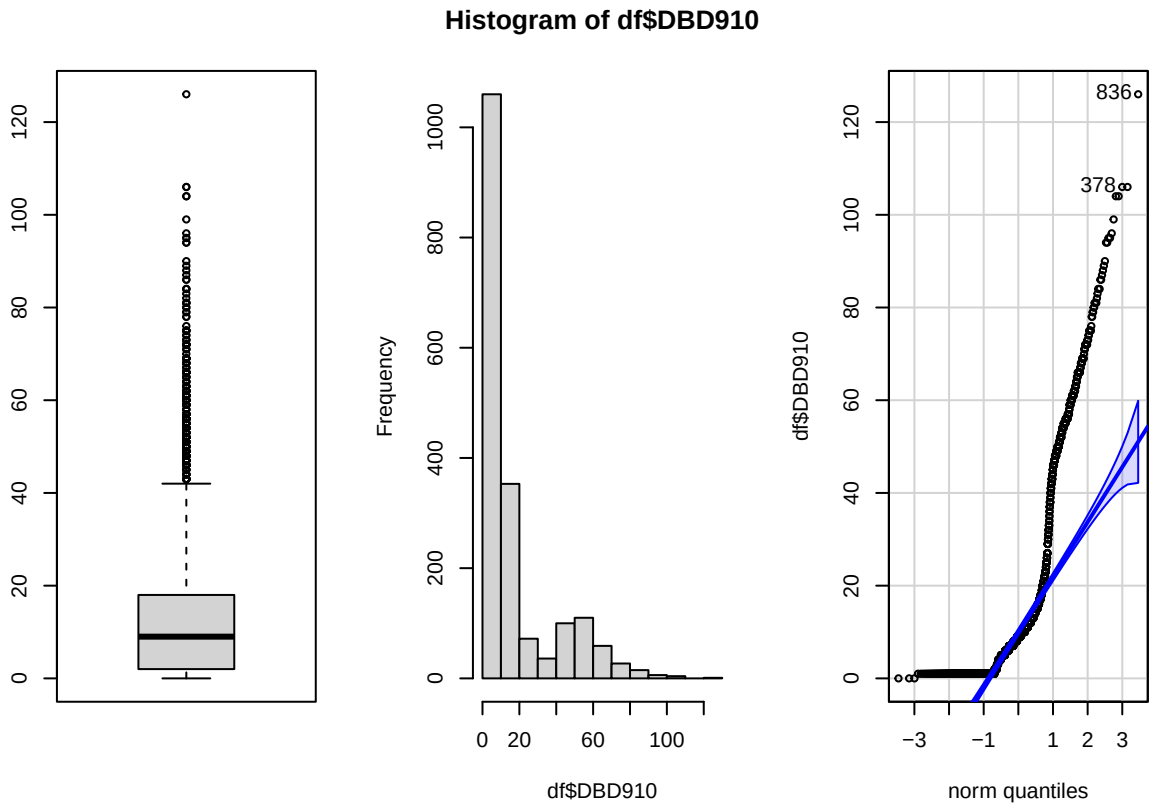
```
par(mfrow=c(1,3))
boxplot(log(df$DBD905))
hist(log(df$DBD905))
#qqPlot(log(df$DBD905))
```



From the Visual Summary above we can see that the distribution when log transformed is still skewed positive.

Number of frozen meals/pizza in past 30 days

```
par(mfrow=c(1,3))
boxplot(df$DBD910)
hist(df$DBD910)
qqPlot(df$DBD910)
```



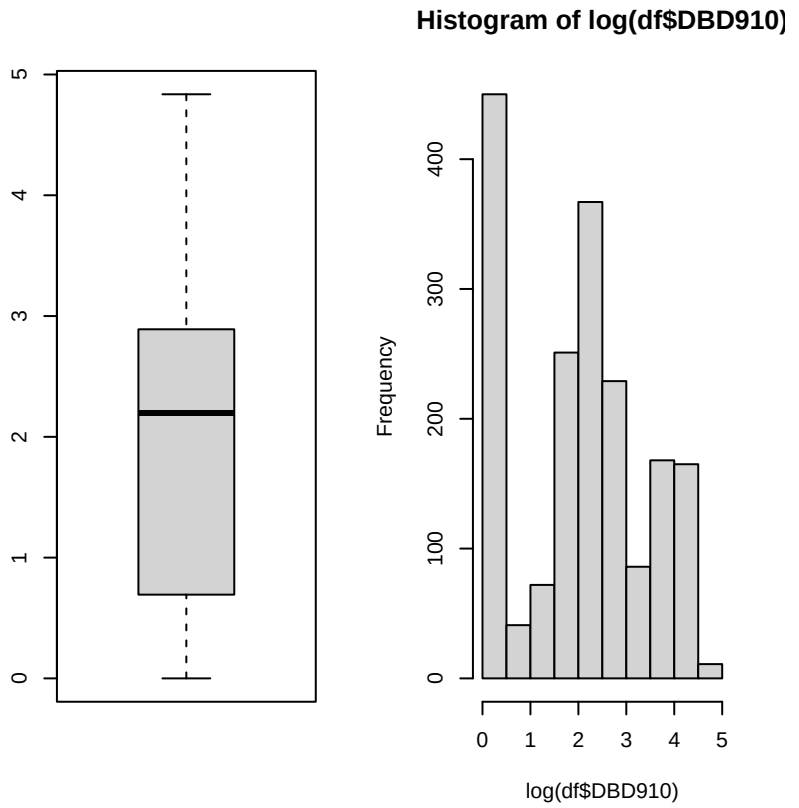

```
## [1] 836 378
```

From the Visual Summary above we can see that the distribution is skewed positive which we can now log transform to test if it is normal on a log scale.

```
par(mfrow=c(1,3))
boxplot(log(df$DBD910))
```

```
## Warning in bplt(at[i], wid = width[i], stats = z$stats[, i], out =
## z$out[z$group == : Outlier (-Inf) in boxplot 1 is not drawn
```

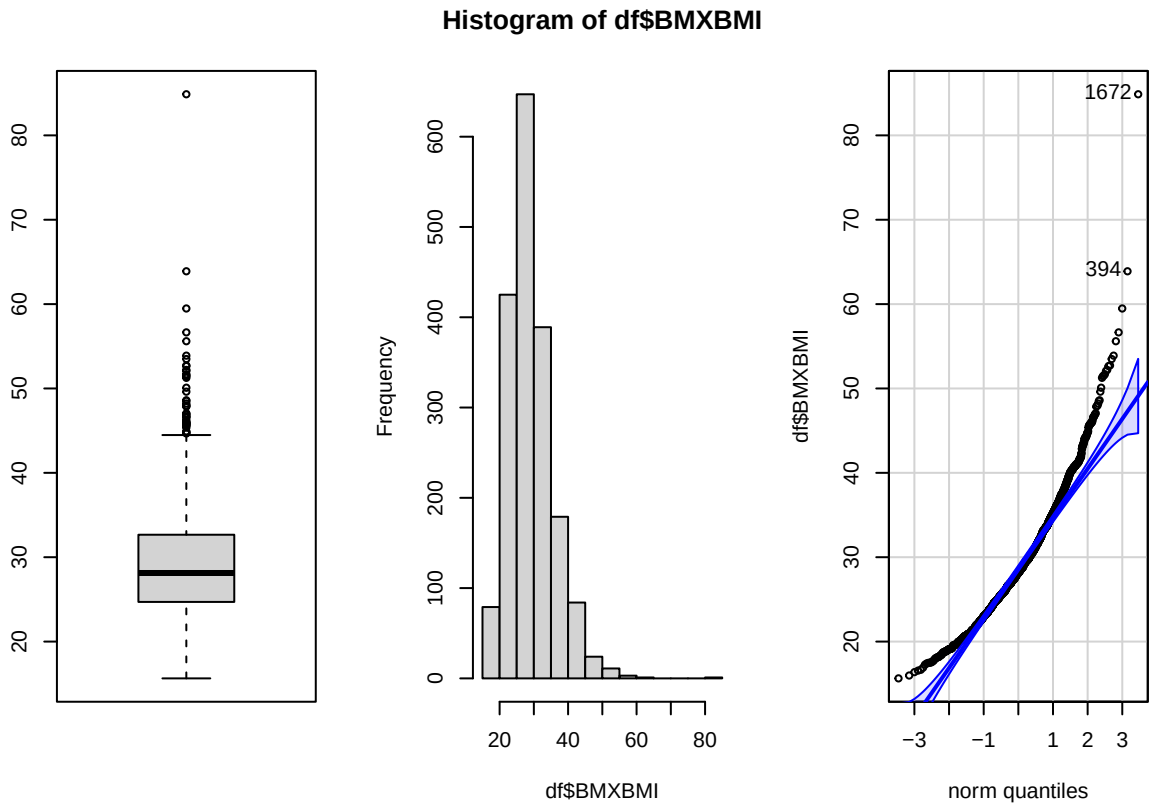
```
hist(log(df$DBD910))
#qqPlot(log(df$DBD910))
```



From the Visual Summary above we can see that the distribution when log transformed is still skewed positive.

Now for BMI (BMXBMI)

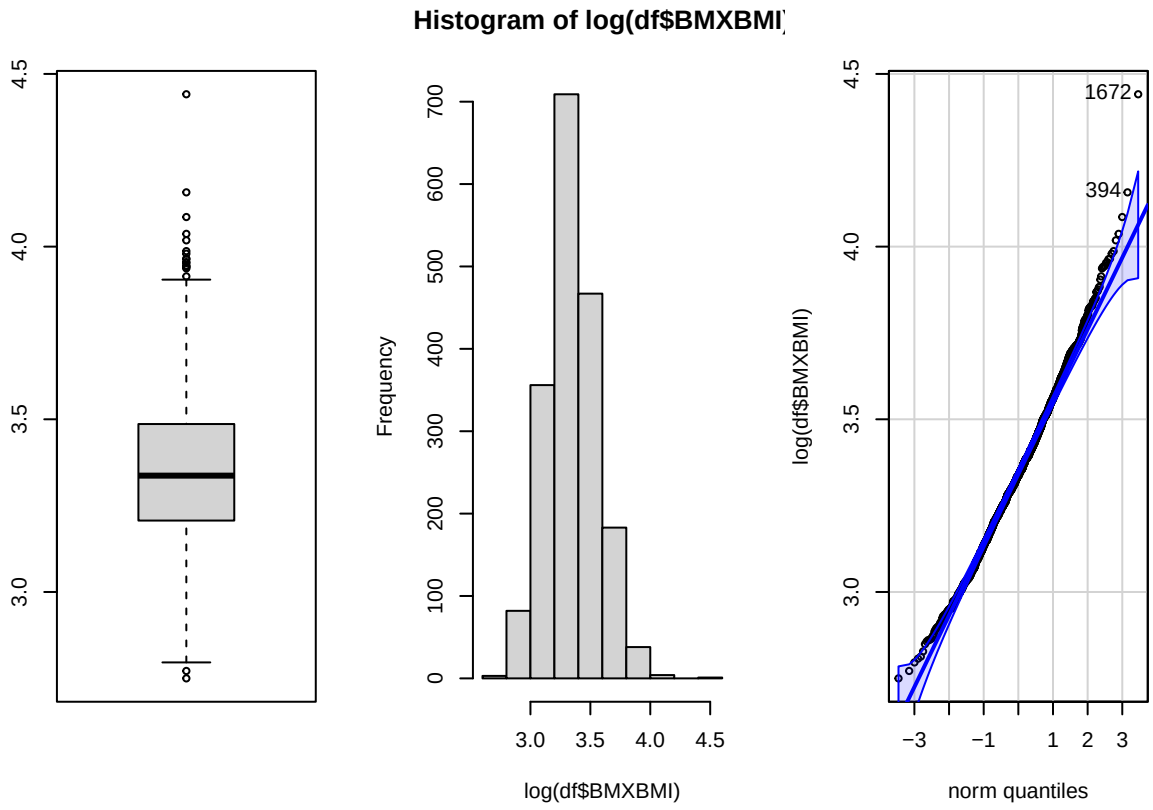
```
par(mfrow=c(1,3))
boxplot(df$BMXBMI)
hist(df$BMXBMI)
qqPlot(df$BMXBMI)
```




```
## [1] 1672  394
```

From the graph above we can see it the data exhibits a positive skew so next we will try a log transformation.

```
par(mfrow=c(1,3))
boxplot(log(df$BMXBMI))
hist(log(df$BMXBMI))
qqPlot(log(df$BMXBMI))
```




```
## [1] 1672  394
```

Now the log BMXBMI data appears normal.

Research Question 1

Are urinary concentrations of phthalate metabolites associated with BMI among adults in the US? (BMI is being used as a proxy for adiposity)

a)

Is that association linear? Compare the two phthalate metabolites to BMI when BMI is
treated as continuous and when BMI when treated as categorical. Report a standardized effect size between each metabolite and BMI for when BMI is continuous. However, the PI will need to know what model may be better to use as the final model for each metabolite – BMI as continuous or BMI as categorical? Do the results support one model over the other?

Continuous BMI vs our two Phthalate's Mono-n-butyl phthalate (ng/mL)

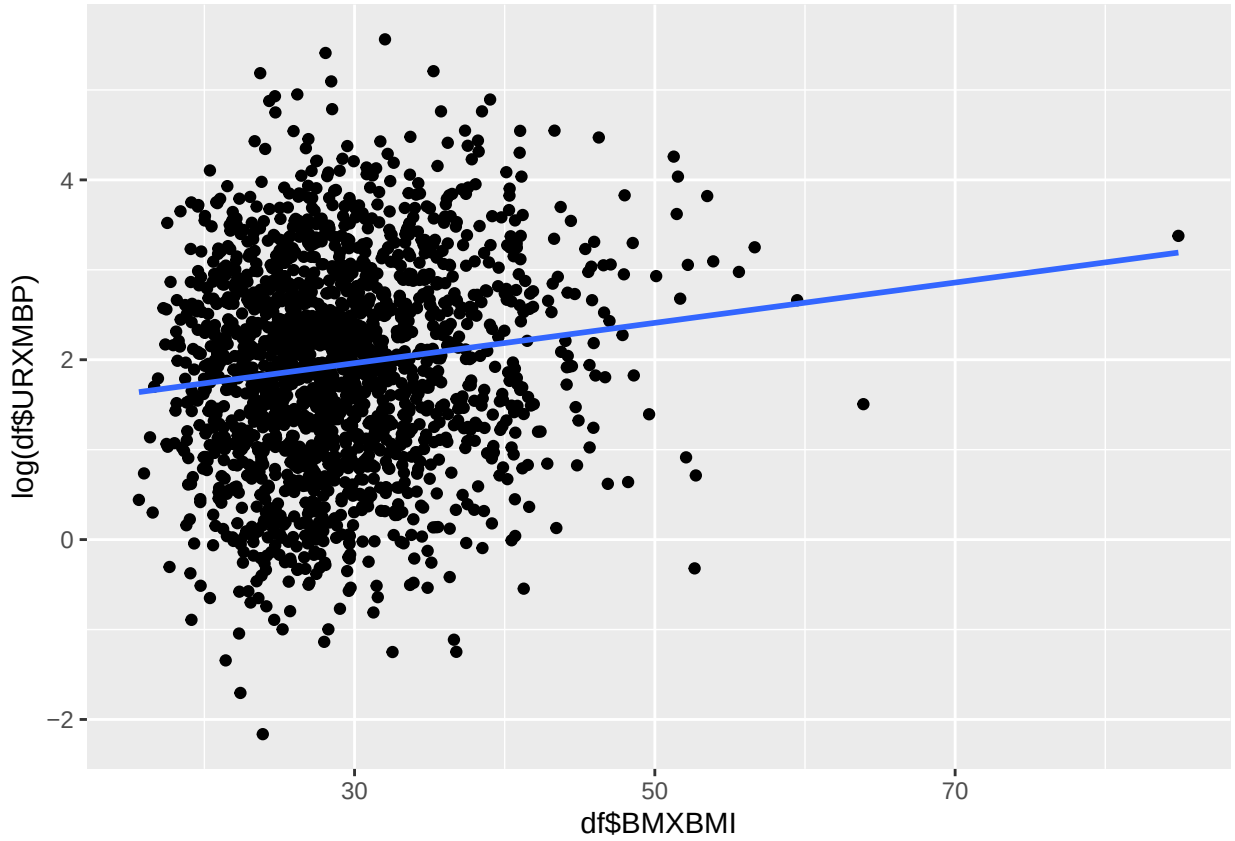
From our summary above we know that it would not be proper to do this test without a log transform on the URXMBP variable.

```r
lm(log(df$URXMBP) ~ df$BMXBMI) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMBP) ~ df$BMXBMI)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9884 -0.7889  0.0515  0.8121  3.5552
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.289618   0.120580  10.695  < 2e-16 ***
## df$BMXBMI   0.022410   0.004041   5.546 3.35e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.138 on 1841 degrees of freedom
## Multiple R-squared:  0.01643,    Adjusted R-squared:  0.0159
## F-statistic: 30.75 on 1 and 1841 DF,  p-value: 3.354e-08
```

```r
ggplot(df, aes(df$BMXBMI,log(df$URXMBP))) +
  geom_point() +
  geom_smooth(method = "lm",se=F)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



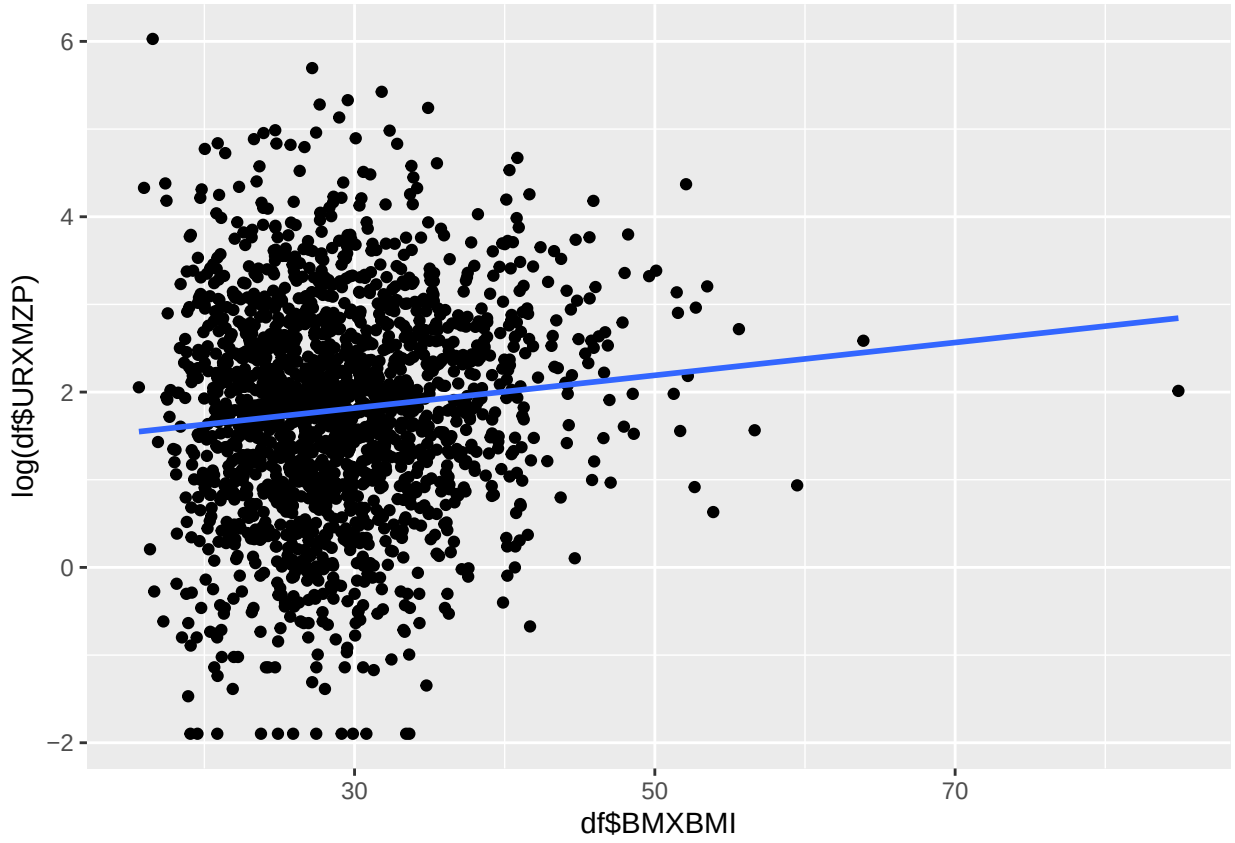
Mono-benzyl phthalate (ng/mL)

```
lm(log(df$URXMZP) ~ df$BMXBMI) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMZP) ~ df$BMXBMI)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7827 -0.7706  0.0320  0.7917  4.4620
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.256480   0.128954   9.744  < 2e-16 ***
## df$BMXBMI   0.018696   0.004322   4.326  1.6e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.217 on 1841 degrees of freedom
## Multiple R-squared:  0.01006,    Adjusted R-squared:  0.009526
## F-statistic: 18.72 on 1 and 1841 DF,  p-value: 1.599e-05
```

```
ggplot(df, aes(df$BMXBMI,log(df$URXMZP))) +
  geom_point() +
  geom_smooth(method = "lm",se=F)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



The summaries from the two linear models tell us that there is a statistically significant linear association between the two Phthalate's and continuous BMI. For every one increase in Mono-n-butyl phthalate (ng/mL) aka URXMBP the BMI prediction increases by 0.022. For every one increase in Mono-benzyl phthalate (ng/mL) aka URXMZP the BMI prediction increases by 0.0186. Also the Intercept for both models is statistically significant.

Categorical BMI vs our two Phthalate's Mono-n-butyl phthalate (ng/mL)

```
lm(log(df$URXMBP) ~ df$BMXBMICAT) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMBP) ~ df$BMXBMICAT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9609 -0.7854  0.0528  0.8151  3.5528
```

```
##
## Coefficients:
##                                                          Estimate Std. Error
## (Intercept)                                                1.5834     0.2940
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2            0.2140     0.2985
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2           0.2981     0.2974
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2      0.4263     0.2996
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2 0.6508     0.3012
##                                                          t value Pr(>|t|)
## (Intercept)                                                5.386 8.14e-08 ***
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2            0.717   0.4735
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2           1.002   0.3163
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2      1.423   0.1549
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2 2.161   0.0309 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.139 on 1838 degrees of freedom
## Multiple R-squared:  0.01737,    Adjusted R-squared:  0.01523
## F-statistic: 8.124 on 4 and 1838 DF,  p-value: 1.718e-06
```

Mono-benzyl phthalate (ng/mL)

```
lm(log(df$URXMZP) ~ df$BMXBMICAT) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMZP) ~ df$BMXBMICAT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6724 -0.7689  0.0460  0.7707  4.0063
##
## Coefficients:
##                                                          Estimate Std. Error
## (Intercept)                                               2.2378     0.3138
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2           -0.4625     0.3186
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2          -0.5486     0.3174
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2     -0.4756     0.3198
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2 -0.1295     0.3215
##                                                          t value Pr(>|t|)
## (Intercept)                                               7.130 1.43e-12 ***
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2           -1.451   0.1468
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2          -1.728   0.0841 .
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2     -1.487   0.1372
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2 -0.403   0.6872
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.215 on 1838 degrees of freedom
## Multiple R-squared:  0.01469,    Adjusted R-squared:  0.01255
## F-statistic: 6.853 on 4 and 1838 DF,  p-value: 1.792e-05
```

For both our categorical BMI models above the intercepts are statistically significant. For our first model we also have one statistically significant association this is for the significantly obese group. When a person is in the significantly obese group it is the prediction for Phthalats increases by 0.65. The second model has no other statistically significant attributes other than the intercept.

The results support the association between the continuous variables more than the categorical. However the R squared for both is low. Both of these models account for less than 2% of the variation in the phthalate.

b)

Does gender modify the association between BMI and urinary metabolites? (Gender is the proxy for circulating sex hormones. Admittedly, NHANES captured self-reported gender only and this is binary; this variable may more reflect biological sex

i) Use BMI as categorical for this analysis. Do not use BMI as continuous for the effect modification analysis (in practice we can do that, but for this assignment we won't).

Mono-n-butyl phthalate (ng/mL)

```
lm(log(df$URXMBP) ~ df$BMXBMICAT : df$RIAGENDR) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMBP) ~ df$BMXBMICAT:df$RIAGENDR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0046 -0.7936  0.0459  0.7853  3.5732
##
## Coefficients: (1 not defined because of singularities)
##                                                                          Estimate
## (Intercept)                                                               2.50248
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale                   -0.14171
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale          -0.75460
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale         -0.66438
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale    -0.63637
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale -0.65032
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale                 -1.03863
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale        -0.66136
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale       -0.56374
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale  -0.34529
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale       NA
##                                                                          Std. Error
## (Intercept)                                                                 0.08468
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale                      0.80337
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale             0.11300
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale            0.10310
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale       0.11684
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale  0.13184
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale                    0.32459
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale           0.11000
```

```
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale              0.10830
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale         0.11756
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale         NA
##                                                                                t value
## (Intercept)                                                                     29.551
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale                          -0.176
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale                 -6.678
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale                -6.444
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale           -5.447
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale      -4.933
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale                        -3.200
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale               -6.013
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale              -5.205
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale         -2.937
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale       NA
##                                                                               Pr(>|t|)
## (Intercept)                                                                    < 2e-16
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale                         0.86000
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale               3.21e-11
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale              1.49e-10
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale         5.82e-08
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale    8.85e-07
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale                       0.00140
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale             2.20e-09
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale            2.16e-07
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale        0.00335
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale       NA
##
## (Intercept)                                                                    ***
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale                ***
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale               ***
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale          ***
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale     ***
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale                       **
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale              ***
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale             ***
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale        **
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.13 on 1833 degrees of freedom
## Multiple R-squared:  0.03525,    Adjusted R-squared:  0.03051
## F-statistic: 7.441 on 9 and 1833 DF,  p-value: 9.624e-11
```

From above we can see that Gender absolutely has an interaction with BMI by category. The linear model went from having only one statistically significant interaction to all but one interaction being statistically significant.

Mono-benzyl phthalate (ng/mL)

```
lm(log(df$URXMZP) ~ df$BMXBMICAT : df$RIAGENDR) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMZP) ~ df$BMXBMICAT:df$RIAGENDR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7969 -0.7714  0.0427  0.7825  3.9364
##
## Coefficients: (1 not defined because of singularities)
##                                                                          Estimate
## (Intercept)                                                               2.21794
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale                   -0.57109
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale          -0.31816
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale         -0.45893
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale    -0.38016
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale -0.26584
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale                  0.11072
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale        -0.55226
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale       -0.62084
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale  -0.53341
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale       NA
##                                                                          Std. Error
## (Intercept)                                                                 0.09091
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale                      0.86242
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale             0.12131
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale            0.11068
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale       0.12542
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale  0.14154
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale                    0.34845
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale           0.11808
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale          0.11627
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale     0.12620
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale       NA
##                                                                          t value
## (Intercept)                                                               24.398
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale                    -0.662
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale           -2.623
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale          -4.146
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale     -3.031
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale -1.878
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale                   0.318
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale         -4.677
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale        -5.340
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale   -4.227
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale      NA
##                                                                          Pr(>|t|)
## (Intercept)                                                               < 2e-16
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale                    0.50793
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale           0.00880
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale         3.53e-05
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale     0.00247
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale  0.06050
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale                  0.75072
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale        3.13e-06
```

```
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale           1.05e-07
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale      2.49e-05
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale       NA
##
## (Intercept)                                                                  ***
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRMale
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRMale              **
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRMale             ***
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRMale        **
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRMale   .
## df$BMXBMICATUnderweight: BMI <18 kg/m2:df$RIAGENDRFemale
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2:df$RIAGENDRFemale            ***
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2:df$RIAGENDRFemale           ***
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2:df$RIAGENDRFemale      ***
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2:df$RIAGENDRFemale
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.213 on 1833 degrees of freedom
## Multiple R-squared:  0.02162,    Adjusted R-squared:  0.01682
## F-statistic: 4.501 on 9 and 1833 DF,  p-value: 7.146e-06
```

From above we can see that Gender absolutely has an interaction with BMI by category. The linear model went from having only one statistically significant interaction to all but one interaction being statistically significant.

Research Question 2

Are urinary concentrations of phthalate metabolites associated with a) past week intake of fast food/pizza, b) past 30-day intake of ready-to-eat foods, or c) past 30-day intake of frozen meals/frozen pizza among adults in the US?

a)

Are those associations linear? Compare the two metabolites to each dietary intake measure with dietary intake treated as continuous and also as categorical.

Mono-n-butyl phthalate (ng/mL)

We still need to log transform to be normal

by past week intake of fast food/pizza

Continuous

```
lm(log(df$URXMBP) ~ df$DBD900) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMBP) ~ df$DBD900)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0872 -0.8033  0.0321  0.7991  3.6178
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.923768   0.036116  53.266   <2e-16 ***
## df$DBD900   0.005261   0.007033   0.748    0.455
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.148 on 1841 degrees of freedom
## Multiple R-squared:  0.0003038,  Adjusted R-squared:  -0.0002392
## F-statistic: 0.5594 on 1 and 1841 DF,  p-value: 0.4546
```

The association is not statistically significant.

Categorical

```
lm(log(df$URXMBP) ~ as.factor(df$DBD900)) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMBP) ~ as.factor(df$DBD900))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1278 -0.7900  0.0288  0.8026  3.6637
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.96436    0.07218  27.214   <2e-16 ***
## as.factor(df$DBD900)1  -0.10876    0.09492  -1.146    0.252
## as.factor(df$DBD900)2   0.05899    0.09613   0.614    0.540
## as.factor(df$DBD900)3  -0.06185    0.09815  -0.630    0.529
## as.factor(df$DBD900)4  -0.06539    0.11308  -0.578    0.563
## as.factor(df$DBD900)5   0.04773    0.12453   0.383    0.702
## as.factor(df$DBD900)6  -0.10482    0.17211  -0.609    0.543
## as.factor(df$DBD900)7   0.15530    0.16956   0.916    0.360
## as.factor(df$DBD900)8  -0.26569    0.20972  -1.267    0.205
## as.factor(df$DBD900)9   0.16471    0.17482   0.942    0.346
## as.factor(df$DBD900)10 -0.23405    0.20972  -1.116    0.265
## as.factor(df$DBD900)11  0.19786    0.26669   0.742    0.458
## as.factor(df$DBD900)12  0.22925    0.32651   0.702    0.483
## as.factor(df$DBD900)13  0.18810    0.38946   0.483    0.629
## as.factor(df$DBD900)14 -0.38730    0.37018  -1.046    0.296
## as.factor(df$DBD900)15  0.65999    0.47425   1.392    0.164
## as.factor(df$DBD900)16 -0.30428    0.41230  -0.738    0.461
## as.factor(df$DBD900)18 -0.73738    0.81506  -0.905    0.366
## as.factor(df$DBD900)19  0.54381    1.15040   0.473    0.636
## as.factor(df$DBD900)20 -0.56180    0.57859  -0.971    0.332
## as.factor(df$DBD900)21  0.45578    0.57859   0.788    0.431
## as.factor(df$DBD900)22 -1.25090    1.15040  -1.087    0.277
```

```
## as.factor(df$DBD900)24  1.18819    1.15040   1.033    0.302
## as.factor(df$DBD900)25 -0.34491    1.15040  -0.300    0.764
## as.factor(df$DBD900)26 -0.80679    0.81506  -0.990    0.322
## as.factor(df$DBD900)33  0.74849    0.81506   0.918    0.359
## as.factor(df$DBD900)39  0.89866    0.81506   1.103    0.270
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.148 on 1816 degrees of freedom
## Multiple R-squared:  0.01293,    Adjusted R-squared:  -0.001203
## F-statistic: 0.9149 on 26 and 1816 DF,  p-value: 0.5881
```

By level the association at any level is not statistically significant.

by past 30-day intake of ready-to-eat foods

Continuous

```
lm(log(df$URXMBP) ~ df$DBD905) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMBP) ~ df$DBD905)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1034 -0.8006  0.0307  0.7948  3.6129
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.949698   0.028020  69.582   <2e-16 ***
## df$DBD905   -0.004878   0.005280  -0.924    0.356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.147 on 1841 degrees of freedom
## Multiple R-squared:  0.0004633,  Adjusted R-squared:  -7.967e-05
## F-statistic: 0.8533 on 1 and 1841 DF,  p-value: 0.3558
```

The association is not statistically significant.

Categorical

```
lm(log(df$URXMBP) ~ as.factor(df$DBD905)) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMBP) ~ as.factor(df$DBD905))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1308 -0.7958  0.0245  0.7848  3.5953
```

```
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              1.967e+00  3.157e-02  62.315   <2e-16 ***
## as.factor(df$DBD905)1  -1.482e-01  1.063e-01  -1.395   0.1633
## as.factor(df$DBD905)2   7.554e-05  1.149e-01   0.001   0.9995
## as.factor(df$DBD905)3   1.786e-02  1.567e-01   0.114   0.9092
## as.factor(df$DBD905)4  -3.059e-01  1.504e-01  -2.035   0.0420 *
## as.factor(df$DBD905)5   9.639e-02  1.940e-01   0.497   0.6193
## as.factor(df$DBD905)6  -1.314e-01  4.352e-01  -0.302   0.7627
## as.factor(df$DBD905)7   2.921e-01  5.145e-01   0.568   0.5703
## as.factor(df$DBD905)8  -3.693e-02  2.054e-01  -0.180   0.8574
## as.factor(df$DBD905)9   2.852e-01  1.149e+00   0.248   0.8040
## as.factor(df$DBD905)10 -1.518e-01  2.803e-01  -0.542   0.5881
## as.factor(df$DBD905)11  8.518e-01  1.149e+00   0.742   0.4585
## as.factor(df$DBD905)12 -2.618e-01  2.526e-01  -1.036   0.3001
## as.factor(df$DBD905)15  2.864e-01  4.072e-01   0.703   0.4820
## as.factor(df$DBD905)16 -1.568e+00  8.126e-01  -1.930   0.0538 .
## as.factor(df$DBD905)17 -5.809e-01  5.750e-01  -1.010   0.3126
## as.factor(df$DBD905)20 -3.948e-01  4.072e-01  -0.969   0.3324
## as.factor(df$DBD905)21 -6.836e-01  5.750e-01  -1.189   0.2347
## as.factor(df$DBD905)24  5.760e-01  1.149e+00   0.501   0.6162
## as.factor(df$DBD905)25  5.476e-02  6.637e-01   0.082   0.9343
## as.factor(df$DBD905)30  1.941e-01  3.200e-01   0.606   0.5443
## as.factor(df$DBD905)60  4.873e-01  6.637e-01   0.734   0.4629
## as.factor(df$DBD905)90 -1.693e+00  1.149e+00  -1.474   0.1407
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.148 on 1820 degrees of freedom
## Multiple R-squared:  0.01046,    Adjusted R-squared:  -0.001503
## F-statistic: 0.8744 on 22 and 1820 DF,  p-value: 0.6305
```

By level the association at any level is not statistically significant besides level 4.

by past 30-day intake of frozen meals/frozen pizza

Continuous

```
lm(log(df$URXMBP) ~ df$DBD910) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMBP) ~ df$DBD910)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0881 -0.8015  0.0330  0.7904  3.6251
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.923599   0.034333  56.028   <2e-16 ***
## df$DBD910   0.001074   0.001262   0.851    0.395
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.148 on 1841 degrees of freedom
## Multiple R-squared:  0.0003931,  Adjusted R-squared:  -0.0001498
## F-statistic: 0.7241 on 1 and 1841 DF,  p-value: 0.3949
```

The association is not statistically significant.

Categorical

```
lm(log(df$URXMBP) ~ as.factor(df$DBD910)) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMBP) ~ as.factor(df$DBD910))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0301 -0.7506  0.0290  0.7810  3.5893
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.80705    0.66014   2.737  0.00626 **
## as.factor(df$DBD910)1     0.05965    0.66233   0.090  0.92826
## as.factor(df$DBD910)2     0.33218    0.68386   0.486  0.62721
## as.factor(df$DBD910)3    -0.23293    0.70371  -0.331  0.74068
## as.factor(df$DBD910)4     0.18963    0.67965   0.279  0.78027
## as.factor(df$DBD910)5     0.09520    0.67240   0.142  0.88743
## as.factor(df$DBD910)6     0.05013    0.67271   0.075  0.94061
## as.factor(df$DBD910)7     0.26897    0.67070   0.401  0.68845
## as.factor(df$DBD910)8     0.11262    0.67271   0.167  0.86707
## as.factor(df$DBD910)9     0.04002    0.67287   0.059  0.95258
## as.factor(df$DBD910)10    0.05428    0.67130   0.081  0.93557
## as.factor(df$DBD910)11    0.18639    0.67567   0.276  0.78269
## as.factor(df$DBD910)12    0.18720    0.67618   0.277  0.78192
## as.factor(df$DBD910)13    0.16624    0.67965   0.245  0.80680
## as.factor(df$DBD910)14    0.12060    0.68785   0.175  0.86084
## as.factor(df$DBD910)15    0.31579    0.68386   0.462  0.64430
## as.factor(df$DBD910)16    0.65646    0.69585   0.943  0.34561
## as.factor(df$DBD910)17    0.06284    0.69862   0.090  0.92834
## as.factor(df$DBD910)18    0.28102    0.70572   0.398  0.69052
## as.factor(df$DBD910)19    0.51742    0.73806   0.701  0.48336
## as.factor(df$DBD910)20    0.32606    0.71303   0.457  0.64752
## as.factor(df$DBD910)21   -0.28351    0.73236  -0.387  0.69872
## as.factor(df$DBD910)22    0.27485    0.73806   0.372  0.70965
## as.factor(df$DBD910)23    0.35500    0.73806   0.481  0.63058
## as.factor(df$DBD910)24    0.06560    0.83502   0.079  0.93739
## as.factor(df$DBD910)25    0.72182    0.76226   0.947  0.34380
## as.factor(df$DBD910)26   -0.23479    0.83502  -0.281  0.77861
## as.factor(df$DBD910)27   -0.05944    0.80850  -0.074  0.94140
## as.factor(df$DBD910)29   -0.53788    0.80850  -0.665  0.50596
## as.factor(df$DBD910)30    0.50458    0.87328   0.578  0.56347
```

```
## as.factor(df$DBD910)31   1.09180    0.87328   1.250  0.21138
## as.factor(df$DBD910)32  -0.48439    0.83502  -0.580  0.56192
## as.factor(df$DBD910)33   0.05183    0.83502   0.062  0.95051
## as.factor(df$DBD910)34  -0.51206    1.04377  -0.491  0.62378
## as.factor(df$DBD910)35  -0.81086    1.04377  -0.777  0.43735
## as.factor(df$DBD910)36   0.80298    1.04377   0.769  0.44182
## as.factor(df$DBD910)37  -0.46443    0.93358  -0.497  0.61891
## as.factor(df$DBD910)38   0.15374    0.78902   0.195  0.84554
## as.factor(df$DBD910)39  -0.42231    1.04377  -0.405  0.68582
## as.factor(df$DBD910)40   0.52261    0.87328   0.598  0.54962
## as.factor(df$DBD910)41  -0.22731    0.80850  -0.281  0.77863
## as.factor(df$DBD910)42   0.04547    0.80850   0.056  0.95516
## as.factor(df$DBD910)43   0.68486    0.76226   0.898  0.36906
## as.factor(df$DBD910)44  -0.35066    0.83502  -0.420  0.67457
## as.factor(df$DBD910)45   1.39754    0.87328   1.600  0.10970
## as.factor(df$DBD910)46  -0.31620    0.73236  -0.432  0.66597
## as.factor(df$DBD910)47   0.50324    0.74474   0.676  0.49930
## as.factor(df$DBD910)48   0.82952    0.71937   1.153  0.24902
## as.factor(df$DBD910)49  -0.92428    0.72744  -1.271  0.20404
## as.factor(df$DBD910)50   0.35192    0.71937   0.489  0.62476
## as.factor(df$DBD910)51   0.14332    0.75267   0.190  0.84901
## as.factor(df$DBD910)52   0.34056    0.72314   0.471  0.63774
## as.factor(df$DBD910)53   0.52559    0.77408   0.679  0.49724
## as.factor(df$DBD910)54   0.52178    0.73236   0.712  0.47627
## as.factor(df$DBD910)55  -0.07794    0.71602  -0.109  0.91334
## as.factor(df$DBD910)56  -0.18277    0.71602  -0.255  0.79855
## as.factor(df$DBD910)57  -0.69659    0.75267  -0.925  0.35484
## as.factor(df$DBD910)58   0.32730    0.93358   0.351  0.72594
## as.factor(df$DBD910)59   0.47779    0.77408   0.617  0.53716
## as.factor(df$DBD910)60   0.70902    0.76226   0.930  0.35242
## as.factor(df$DBD910)61  -0.26871    0.74474  -0.361  0.71829
## as.factor(df$DBD910)62   0.57190    0.77408   0.739  0.46012
## as.factor(df$DBD910)63   0.59487    0.83502   0.712  0.47631
## as.factor(df$DBD910)64   0.23935    0.87328   0.274  0.78406
## as.factor(df$DBD910)65   0.37567    0.87328   0.430  0.66712
## as.factor(df$DBD910)66   0.36192    0.74474   0.486  0.62705
## as.factor(df$DBD910)67   0.29421    0.93358   0.315  0.75269
## as.factor(df$DBD910)68   0.82524    0.83502   0.988  0.32315
## as.factor(df$DBD910)69   0.04590    0.78902   0.058  0.95361
## as.factor(df$DBD910)70  -2.15736    1.32028  -1.634  0.10243
## as.factor(df$DBD910)71  -0.26898    0.93358  -0.288  0.77329
## as.factor(df$DBD910)72  -0.20065    0.78902  -0.254  0.79929
## as.factor(df$DBD910)73   0.21874    0.93358   0.234  0.81478
## as.factor(df$DBD910)74   1.55804    1.04377   1.493  0.13569
## as.factor(df$DBD910)75   0.50909    0.83502   0.610  0.54216
## as.factor(df$DBD910)76  -0.35662    1.32028  -0.270  0.78711
## as.factor(df$DBD910)78   0.37737    1.04377   0.362  0.71774
## as.factor(df$DBD910)79   0.51938    1.04377   0.498  0.61883
## as.factor(df$DBD910)80   0.40383    1.04377   0.387  0.69888
## as.factor(df$DBD910)81  -0.30866    0.87328  -0.353  0.72380
## as.factor(df$DBD910)82   0.10607    1.32028   0.080  0.93597
## as.factor(df$DBD910)83   1.01206    1.32028   0.767  0.44345
## as.factor(df$DBD910)84   0.59809    0.93358   0.641  0.52184
## as.factor(df$DBD910)86  -0.98898    1.04377  -0.948  0.34351
```

```
## as.factor(df$DBD910)87   0.95206    1.32028   0.721  0.47094
## as.factor(df$DBD910)88   0.07079    1.32028   0.054  0.95725
## as.factor(df$DBD910)89  -0.14121    1.32028  -0.107  0.91484
## as.factor(df$DBD910)90   2.95477    1.32028   2.238  0.02535 *
## as.factor(df$DBD910)94  -0.20408    1.04377  -0.196  0.84501
## as.factor(df$DBD910)95  -1.10205    1.04377  -1.056  0.29119
## as.factor(df$DBD910)96   0.99982    1.32028   0.757  0.44898
## as.factor(df$DBD910)99  -0.41698    1.32028  -0.316  0.75217
## as.factor(df$DBD910)104 -0.29692    1.04377  -0.284  0.77608
## as.factor(df$DBD910)106 -0.35332    1.04377  -0.339  0.73503
## as.factor(df$DBD910)126  0.24886    1.32028   0.188  0.85051
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.143 on 1748 degrees of freedom
## Multiple R-squared:  0.05773,    Adjusted R-squared:  0.007055
## F-statistic: 1.139 on 94 and 1748 DF,  p-value: 0.1757
```

By level the association at any level is not statistically significant.

Mono-benzyl phthalate (ng/mL)

We still need to log transform to be normal

by past week intake of fast food/pizza

Continuous

```
lm(log(df$URXMZP) ~ df$DBD900) |>
  summary()
```

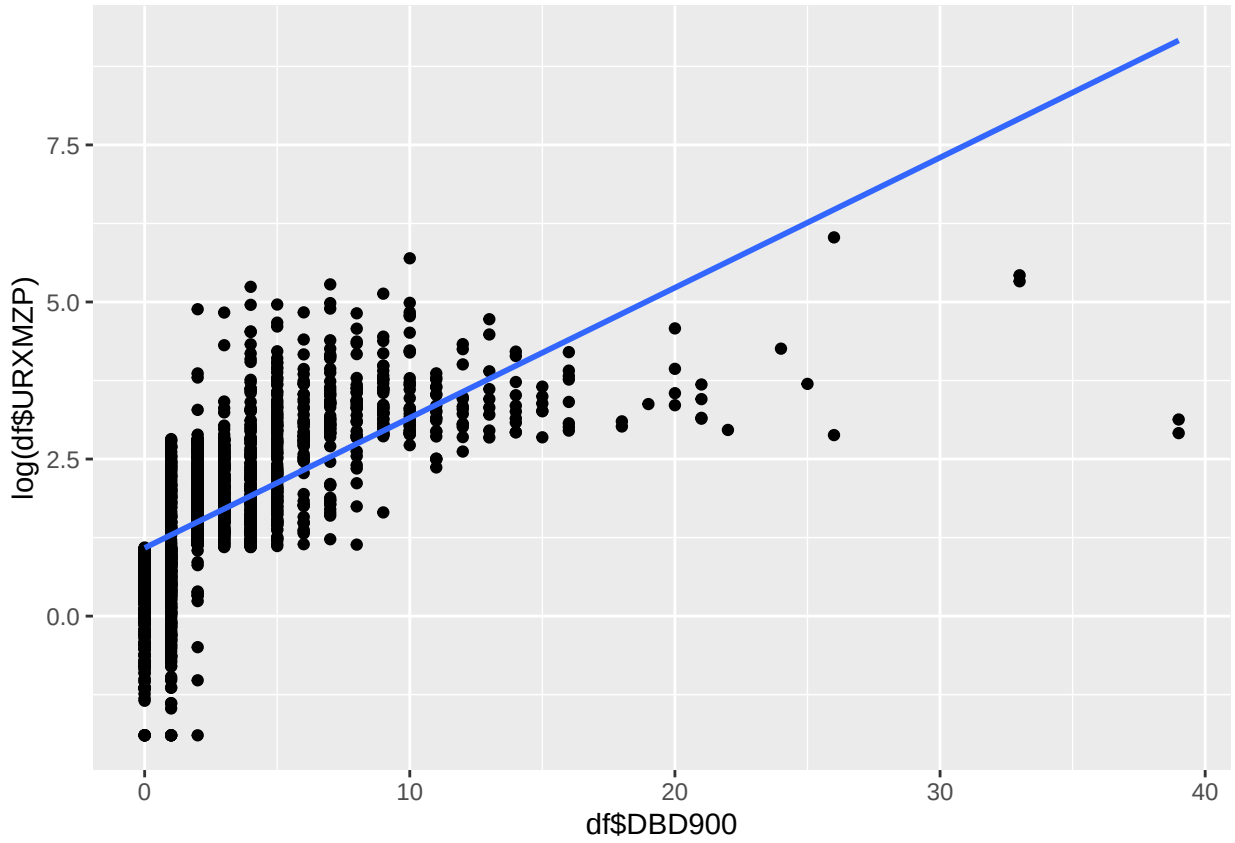
```
##
## Call:
## lm(formula = log(df$URXMZP) ~ df$DBD900)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.2496 -0.4956  0.0024  0.6331  3.3852
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.085565   0.029469   36.84   <2e-16 ***
## df$DBD900   0.207100   0.005739   36.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9364 on 1841 degrees of freedom
## Multiple R-squared:  0.4143, Adjusted R-squared:  0.414
## F-statistic:  1302 on 1 and 1841 DF,  p-value: < 2.2e-16
```

The association is statistically significant for every one increase in fast food intake the Mono-benzyl phthalate increases by 0.207!

For fun I want to graph it.

```
ggplot(df, aes(df$DBD900,log(df$URXMZP))) +
  geom_point() +
  geom_smooth(method = "lm",se=F)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Categorical

```
lm(log(df$URXMZP) ~ as.factor(df$DBD900)) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMZP) ~ as.factor(df$DBD900))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8064 -0.4721 -0.0026  0.4769  3.0987
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.24160    0.04783   5.051 4.83e-07 ***
## as.factor(df$DBD900)1     0.65059    0.06289  10.344  < 2e-16 ***
## as.factor(df$DBD900)2     1.66771    0.06370  26.180  < 2e-16 ***
```

```
## as.factor(df$DBD900)3   1.78269    0.06504  27.410  < 2e-16 ***
## as.factor(df$DBD900)4   1.90098    0.07493  25.371  < 2e-16 ***
## as.factor(df$DBD900)5   2.24704    0.08252  27.230  < 2e-16 ***
## as.factor(df$DBD900)6   2.55031    0.11405  22.362  < 2e-16 ***
## as.factor(df$DBD900)7   2.87462    0.11235  25.585  < 2e-16 ***
## as.factor(df$DBD900)8   2.98910    0.13897  21.510  < 2e-16 ***
## as.factor(df$DBD900)9   3.11261    0.11584  26.870  < 2e-16 ***
## as.factor(df$DBD900)10  3.34824    0.13897  24.094  < 2e-16 ***
## as.factor(df$DBD900)11  2.94183    0.17671  16.647  < 2e-16 ***
## as.factor(df$DBD900)12  3.14034    0.21636  14.515  < 2e-16 ***
## as.factor(df$DBD900)13  3.36925    0.25807  13.056  < 2e-16 ***
## as.factor(df$DBD900)14  3.18614    0.24529  12.989  < 2e-16 ***
## as.factor(df$DBD900)15  3.07609    0.31425   9.789  < 2e-16 ***
## as.factor(df$DBD900)16  3.27568    0.27320  11.990  < 2e-16 ***
## as.factor(df$DBD900)18  2.81768    0.54008   5.217 2.03e-07 ***
## as.factor(df$DBD900)19  3.13359    0.76229   4.111 4.12e-05 ***
## as.factor(df$DBD900)20  3.61364    0.38339   9.425  < 2e-16 ***
## as.factor(df$DBD900)21  3.11818    0.38339   8.133 7.67e-16 ***
## as.factor(df$DBD900)22  2.72161    0.76229   3.570 0.000366 ***
## as.factor(df$DBD900)24  4.01557    0.76229   5.268 1.55e-07 ***
## as.factor(df$DBD900)25  3.45450    0.76229   4.532 6.23e-06 ***
## as.factor(df$DBD900)26  4.21319    0.54008   7.801 1.03e-14 ***
## as.factor(df$DBD900)33  5.13571    0.54008   9.509  < 2e-16 ***
## as.factor(df$DBD900)39  2.77976    0.54008   5.147 2.93e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7608 on 1816 degrees of freedom
## Multiple R-squared:  0.6186, Adjusted R-squared:  0.6131
## F-statistic: 113.3 on 26 and 1816 DF,  p-value: < 2.2e-16
```

Above we can see that there is a statistically significant association at every factor level of fast food intake.

by past 30-day intake of ready-to-eat foods

Continuous

```
lm(log(df$URXMZP) ~ df$DBD905) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMZP) ~ df$DBD905)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7264 -0.7857  0.0233  0.7983  4.2345
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.793602   0.029871  60.044   <2e-16 ***
## df$DBD905   0.004454   0.005629   0.791    0.429
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 1.223 on 1841 degrees of freedom
## Multiple R-squared:  0.0003399,  Adjusted R-squared:  -0.0002031
## F-statistic: 0.626 on 1 and 1841 DF,  p-value: 0.4289
```

The association is not statistically significant.

Categorical

```
lm(log(df$URXMZP) ~ as.factor(df$DBD905)) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMZP) ~ as.factor(df$DBD905))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8481 -0.7458  0.0404  0.7772  4.2280
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               1.80014    0.03358  53.606   <2e-16 ***
## as.factor(df$DBD905)1    -0.16396    0.11306  -1.450   0.1472
## as.factor(df$DBD905)2     0.04285    0.12224   0.351   0.7259
## as.factor(df$DBD905)3    -0.13312    0.16664  -0.799   0.4245
## as.factor(df$DBD905)4     0.11562    0.15996   0.723   0.4699
## as.factor(df$DBD905)5    -0.15106    0.20633  -0.732   0.4642
## as.factor(df$DBD905)6    -0.39579    0.46288  -0.855   0.3926
## as.factor(df$DBD905)7     0.62855    0.54728   1.148   0.2509
## as.factor(df$DBD905)8     0.15088    0.21852   0.690   0.4900
## as.factor(df$DBD905)9     1.03484    1.22191   0.847   0.3972
## as.factor(df$DBD905)10    0.47433    0.29814   1.591   0.1118
## as.factor(df$DBD905)11    1.73033    1.22191   1.416   0.1569
## as.factor(df$DBD905)12    0.49369    0.26865   1.838   0.0663 .
## as.factor(df$DBD905)15   -0.15462    0.43315  -0.357   0.7212
## as.factor(df$DBD905)16    2.21135    0.86435   2.558   0.0106 *
## as.factor(df$DBD905)17   -0.67783    0.61165  -1.108   0.2679
## as.factor(df$DBD905)20   -0.12994    0.43315  -0.300   0.7642
## as.factor(df$DBD905)21   -0.82935    0.61165  -1.356   0.1753
## as.factor(df$DBD905)24    1.46409    1.22191   1.198   0.2310
## as.factor(df$DBD905)25   -0.41555    0.70600  -0.589   0.5562
## as.factor(df$DBD905)30   -0.03240    0.34043  -0.095   0.9242
## as.factor(df$DBD905)60    0.23488    0.70600   0.333   0.7394
## as.factor(df$DBD905)90    0.20404    1.22191   0.167   0.8674
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.221 on 1820 degrees of freedom
## Multiple R-squared:  0.01475,    Adjusted R-squared:  0.002842
## F-statistic: 1.239 on 22 and 1820 DF,  p-value: 0.2039
```

Here there is one level at 16 ready to eat foods per week that has a statistically significant association. However all the others do not.

by past 30-day intake of frozen meals/frozen pizza

Continuous

```
lm(log(df$URXMZP) ~ df$DBD910) |>
  summary()
```

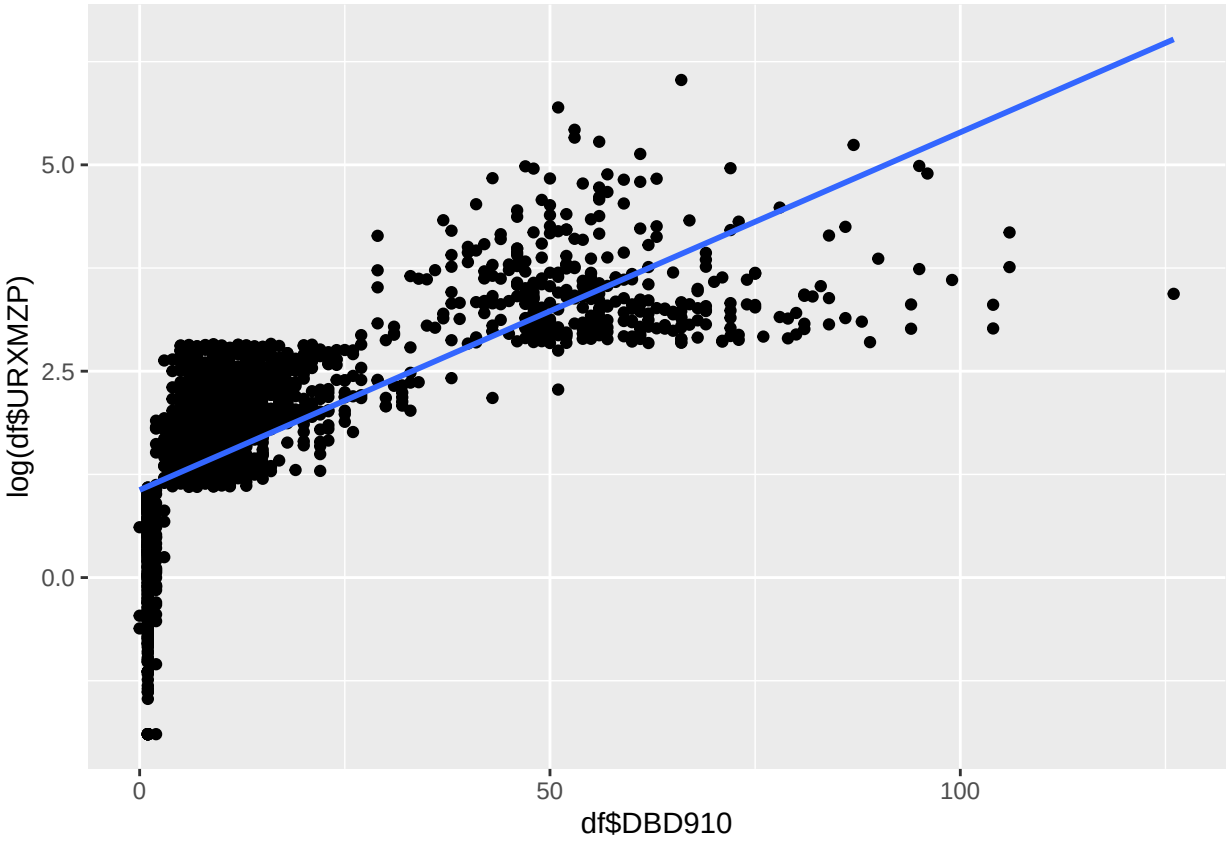
```
##
## Call:
## lm(formula = log(df$URXMZP) ~ df$DBD910)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.08323 -0.39231  0.06401  0.54161  2.42461
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.0610523  0.0241970   43.85   <2e-16 ***
## df$DBD910   0.0433285  0.0008896   48.71   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8088 on 1841 degrees of freedom
## Multiple R-squared:  0.5631, Adjusted R-squared:  0.5628
## F-statistic:  2372 on 1 and 1841 DF,  p-value: < 2.2e-16
```

The association is statistically significant for every one increase in frozen meal intake the Mono-benzyl phthalate increases by 0.0433!

For fun I want to graph it.

```
ggplot(df, aes(df$DBD910,log(df$URXMZP))) +
  geom_point() +
  geom_smooth(method = "lm",se=F)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Categorical

```
lm(log(df$URXMZP) ~ as.factor(df$DBD910)) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$URXMZP) ~ as.factor(df$DBD910))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.3564 -0.3400  0.0000  0.3771  2.7151
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -0.1562     0.3206  -0.487    0.626
## as.factor(df$DBD910)1      0.4366     0.3216   1.357    0.175
## as.factor(df$DBD910)2      0.6154     0.3321   1.853    0.064 .
## as.factor(df$DBD910)3      1.5621     0.3417   4.571 5.19e-06 ***
## as.factor(df$DBD910)4      1.7398     0.3300   5.271 1.52e-07 ***
## as.factor(df$DBD910)5      1.9546     0.3265   5.986 2.60e-09 ***
## as.factor(df$DBD910)6      1.9204     0.3267   5.879 4.94e-09 ***
## as.factor(df$DBD910)7      2.0600     0.3257   6.325 3.21e-10 ***
## as.factor(df$DBD910)8      1.9787     0.3267   6.057 1.69e-09 ***
## as.factor(df$DBD910)9      2.0624     0.3268   6.312 3.49e-10 ***
## as.factor(df$DBD910)10     2.0070     0.3260   6.157 9.19e-10 ***
```

```
## as.factor(df$DBD910)11    2.1507     0.3281   6.555 7.32e-11 ***
## as.factor(df$DBD910)12    2.2828     0.3284   6.952 5.06e-12 ***
## as.factor(df$DBD910)13    2.1897     0.3300   6.635 4.33e-11 ***
## as.factor(df$DBD910)14    2.3207     0.3340   6.948 5.22e-12 ***
## as.factor(df$DBD910)15    2.1868     0.3321   6.585 6.00e-11 ***
## as.factor(df$DBD910)16    2.3214     0.3379   6.870 8.90e-12 ***
## as.factor(df$DBD910)17    2.4113     0.3393   7.108 1.71e-12 ***
## as.factor(df$DBD910)18    2.4335     0.3427   7.101 1.80e-12 ***
## as.factor(df$DBD910)19    2.3885     0.3584   6.664 3.55e-11 ***
## as.factor(df$DBD910)20    2.3840     0.3463   6.885 8.02e-12 ***
## as.factor(df$DBD910)21    2.5704     0.3556   7.228 7.31e-13 ***
## as.factor(df$DBD910)22    2.1301     0.3584   5.943 3.37e-09 ***
## as.factor(df$DBD910)23    2.3733     0.3584   6.622 4.70e-11 ***
## as.factor(df$DBD910)24    2.7587     0.4055   6.803 1.40e-11 ***
## as.factor(df$DBD910)25    2.3834     0.3702   6.439 1.55e-10 ***
## as.factor(df$DBD910)26    2.5284     0.4055   6.236 5.63e-10 ***
## as.factor(df$DBD910)27    2.6692     0.3926   6.799 1.44e-11 ***
## as.factor(df$DBD910)29    3.3627     0.3926   8.565  < 2e-16 ***
## as.factor(df$DBD910)30    2.4572     0.4241   5.794 8.12e-09 ***
## as.factor(df$DBD910)31    2.9723     0.4241   7.009 3.42e-12 ***
## as.factor(df$DBD910)32    2.3495     0.4055   5.794 8.13e-09 ***
## as.factor(df$DBD910)33    2.8171     0.4055   6.947 5.23e-12 ***
## as.factor(df$DBD910)34    3.1496     0.5069   6.214 6.45e-10 ***
## as.factor(df$DBD910)35    3.4887     0.5069   6.883 8.14e-12 ***
## as.factor(df$DBD910)36    3.5315     0.5069   6.967 4.55e-12 ***
## as.factor(df$DBD910)37    3.7115     0.4534   8.187 5.12e-16 ***
## as.factor(df$DBD910)38    3.5767     0.3832   9.335  < 2e-16 ***
## as.factor(df$DBD910)39    3.3869     0.5069   6.682 3.15e-11 ***
## as.factor(df$DBD910)40    3.8062     0.4241   8.975  < 2e-16 ***
## as.factor(df$DBD910)41    3.5659     0.3926   9.083  < 2e-16 ***
## as.factor(df$DBD910)42    3.6952     0.3926   9.412  < 2e-16 ***
## as.factor(df$DBD910)43    3.5128     0.3702   9.490  < 2e-16 ***
## as.factor(df$DBD910)44    3.8193     0.4055   9.419  < 2e-16 ***
## as.factor(df$DBD910)45    3.6117     0.4241   8.517  < 2e-16 ***
## as.factor(df$DBD910)46    3.9360     0.3556  11.067  < 2e-16 ***
## as.factor(df$DBD910)47    3.6321     0.3616  10.043  < 2e-16 ***
## as.factor(df$DBD910)48    3.5551     0.3493  10.177  < 2e-16 ***
## as.factor(df$DBD910)49    3.5225     0.3532   9.972  < 2e-16 ***
## as.factor(df$DBD910)50    3.7028     0.3493  10.600  < 2e-16 ***
## as.factor(df$DBD910)51    3.6485     0.3655   9.982  < 2e-16 ***
## as.factor(df$DBD910)52    3.7132     0.3512  10.574  < 2e-16 ***
## as.factor(df$DBD910)53    4.1872     0.3759  11.139  < 2e-16 ***
## as.factor(df$DBD910)54    3.5512     0.3556   9.986  < 2e-16 ***
## as.factor(df$DBD910)55    3.5186     0.3477  10.120  < 2e-16 ***
## as.factor(df$DBD910)56    3.8822     0.3477  11.165  < 2e-16 ***
## as.factor(df$DBD910)57    3.7121     0.3655  10.156  < 2e-16 ***
## as.factor(df$DBD910)58    3.6772     0.4534   8.111 9.34e-16 ***
## as.factor(df$DBD910)59    3.8468     0.3759  10.234  < 2e-16 ***
## as.factor(df$DBD910)60    3.3477     0.3702   9.044  < 2e-16 ***
## as.factor(df$DBD910)61    3.6899     0.3616  10.203  < 2e-16 ***
## as.factor(df$DBD910)62    3.5189     0.3759   9.361  < 2e-16 ***
## as.factor(df$DBD910)63    4.0560     0.4055  10.003  < 2e-16 ***
## as.factor(df$DBD910)64    3.3000     0.4241   7.782 1.22e-14 ***
## as.factor(df$DBD910)65    3.4299     0.4241   8.088 1.12e-15 ***
```

```
## as.factor(df$DBD910)66    3.4692     0.3616   9.593  < 2e-16 ***
## as.factor(df$DBD910)67    3.7232     0.4534   8.213 4.16e-16 ***
## as.factor(df$DBD910)68    3.4071     0.4055   8.403  < 2e-16 ***
## as.factor(df$DBD910)69    3.6382     0.3832   9.495  < 2e-16 ***
## as.factor(df$DBD910)70    3.7388     0.6411   5.832 6.53e-09 ***
## as.factor(df$DBD910)71    3.2770     0.4534   7.228 7.27e-13 ***
## as.factor(df$DBD910)72    3.7025     0.3832   9.663  < 2e-16 ***
## as.factor(df$DBD910)73    3.5336     0.4534   7.794 1.10e-14 ***
## as.factor(df$DBD910)74    3.6148     0.5069   7.132 1.44e-12 ***
## as.factor(df$DBD910)75    3.6829     0.4055   9.083  < 2e-16 ***
## as.factor(df$DBD910)76    3.0755     0.6411   4.797 1.75e-06 ***
## as.factor(df$DBD910)78    3.9758     0.5069   7.844 7.54e-15 ***
## as.factor(df$DBD910)79    3.1740     0.5069   6.262 4.77e-10 ***
## as.factor(df$DBD910)80    3.2314     0.5069   6.375 2.33e-10 ***
## as.factor(df$DBD910)81    3.3882     0.4241   7.990 2.43e-15 ***
## as.factor(df$DBD910)82    3.5613     0.6411   5.555 3.21e-08 ***
## as.factor(df$DBD910)83    3.6866     0.6411   5.750 1.05e-08 ***
## as.factor(df$DBD910)84    3.6881     0.4534   8.135 7.72e-16 ***
## as.factor(df$DBD910)86    3.8523     0.5069   7.600 4.79e-14 ***
## as.factor(df$DBD910)87    5.3974     0.6411   8.419  < 2e-16 ***
## as.factor(df$DBD910)88    3.2562     0.6411   5.079 4.20e-07 ***
## as.factor(df$DBD910)89    3.0080     0.6411   4.692 2.92e-06 ***
## as.factor(df$DBD910)90    4.0192     0.6411   6.269 4.57e-10 ***
## as.factor(df$DBD910)94    3.3178     0.5069   6.546 7.76e-11 ***
## as.factor(df$DBD910)95    4.5181     0.5069   8.914  < 2e-16 ***
## as.factor(df$DBD910)96    5.0511     0.6411   7.878 5.78e-15 ***
## as.factor(df$DBD910)99    3.7614     0.6411   5.867 5.30e-09 ***
## as.factor(df$DBD910)104   3.3175     0.5069   6.545 7.79e-11 ***
## as.factor(df$DBD910)106   4.1272     0.5069   8.143 7.28e-16 ***
## as.factor(df$DBD910)126   3.5934     0.6411   5.605 2.42e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5552 on 1748 degrees of freedom
## Multiple R-squared:  0.8045, Adjusted R-squared:  0.7939
## F-statistic: 76.51 on 94 and 1748 DF,  p-value: < 2.2e-16
```

Above we can see that there is a statistically significant association at every factor level of fast food intake.

Dietary Intake and BMI

Finally, please examine the associations (with hypothesis testing) between each dietary intake measure and BMI. Include hypothesis testing. Please decide how to best complete those comparisons. Those findings will help me determine the next steps

Continuous BMI (BMXBMI)

by past week intake of fast food/pizza

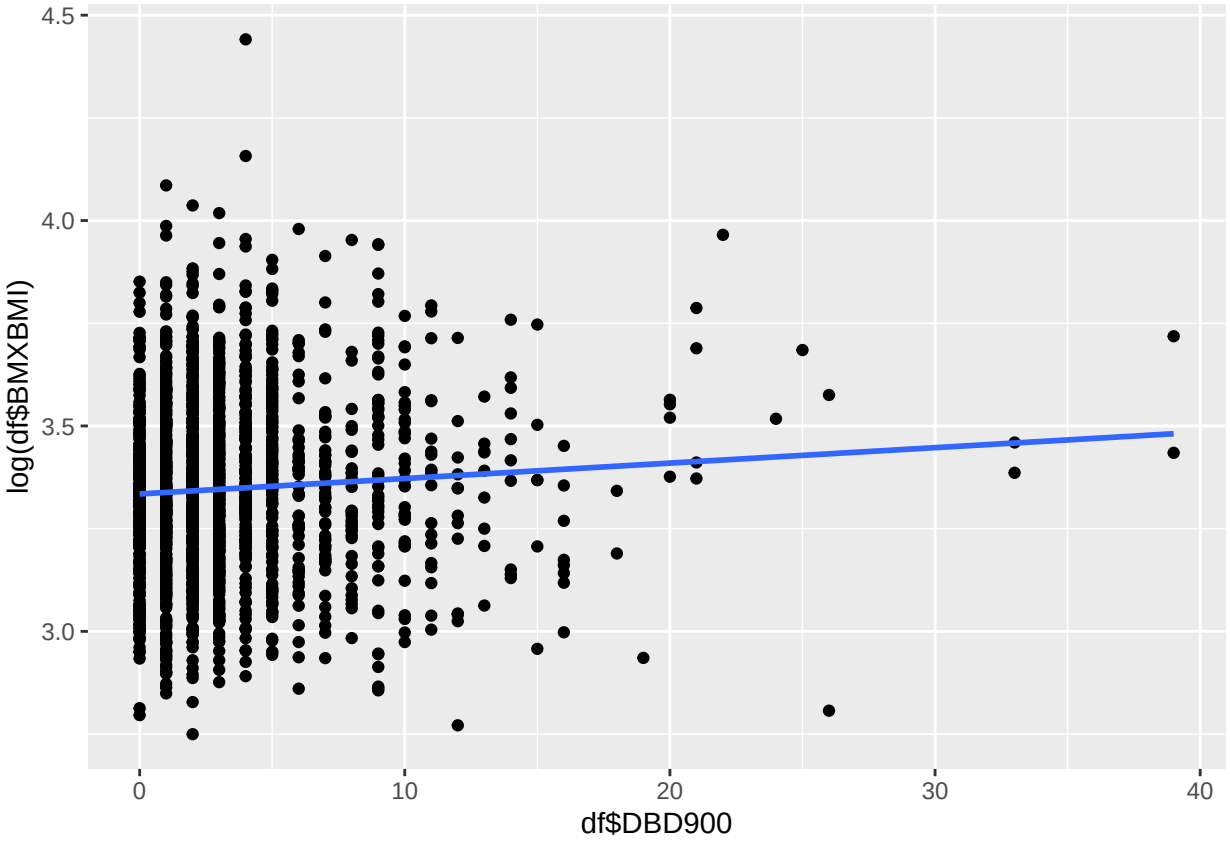
```
lm(log(df$BMXBMI) ~ df$DBD900) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$BMXBMI) ~ df$DBD900)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62523 -0.14298 -0.00894  0.13567  1.09159
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.334502   0.006739 494.799  < 2e-16 ***
## df$DBD900   0.003758   0.001312   2.864  0.00423 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2141 on 1841 degrees of freedom
## Multiple R-squared:  0.004435,   Adjusted R-squared:  0.003894
## F-statistic: 8.201 on 1 and 1841 DF,  p-value: 0.004234
```

The association is statistically significant for every one increase in fast food intake the BMXBMI increases by 0.003758!

```
ggplot(df, aes(df$DBD900,log(df$BMXBMI))) +
  geom_point() +
  geom_smooth(method = "lm",se=F)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



by past 30-day intake of ready-to-eat foods

```
lm(log(df$BMXBMI) ~ df$DBD905) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$BMXBMI) ~ df$DBD905)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59729 -0.14066 -0.01029  0.13755  1.09400
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.3471223  0.0052403 638.723   <2e-16 ***
## df$DBD905   0.0002242  0.0009876   0.227     0.82
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2146 on 1841 degrees of freedom
## Multiple R-squared:  2.8e-05,    Adjusted R-squared:  -0.0005152
## F-statistic: 0.05154 on 1 and 1841 DF,  p-value: 0.8204
```

The association is not statistically significant.

by past 30-day intake of frozen meals/frozen pizza

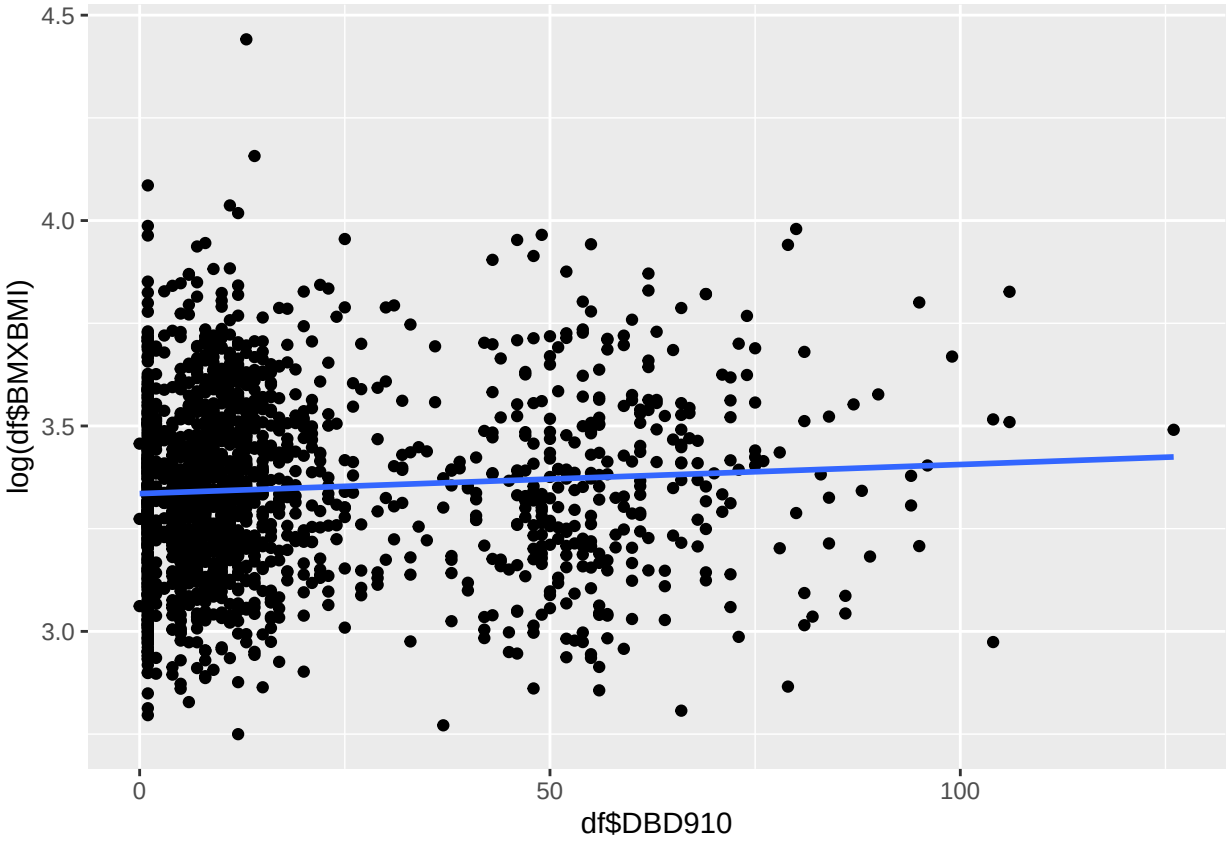
```r
lm(log(df$BMXBMI) ~ df$DBD910) |>
  summary()
```

```
##
## Call:
## lm(formula = log(df$BMXBMI) ~ df$DBD910)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59406 -0.14358 -0.00781  0.13482  1.09652
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.3353902  0.0064051 520.740  < 2e-16 ***
## df$DBD910   0.0007082  0.0002355   3.007  0.00267 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2141 on 1841 degrees of freedom
## Multiple R-squared:  0.004889,   Adjusted R-squared:  0.004349
## F-statistic: 9.045 on 1 and 1841 DF,  p-value: 0.00267
```

The association is statistically significant for every one increase in fast food intake the BMXBMI increases by 0.0007!

```r
ggplot(df, aes(df$DBD910,log(df$BMXBMI))) +
  geom_point() +
  geom_smooth(method = "lm",se=F)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Categorical BMI (BMXBMICAT)

by past week intake of fast food/pizza

```
lm(df$DBD900 ~ df$BMXBMICAT) |>
  summary()
```

```
##
## Call:
## lm(formula = df$DBD900 ~ df$BMXBMICAT)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.533 -2.312 -1.169  0.831 35.450
##
## Coefficients:
##                                                          Estimate Std. Error
## (Intercept)                                                5.5333     0.9782
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2           -2.2212     0.9931
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2          -2.3638     0.9894
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2     -1.9832     0.9969
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2 -1.4739     1.0021
##                                                          t value Pr(>|t|)
## (Intercept)                                                5.657 1.79e-08 ***
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2           -2.237   0.0254 *
```

```
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2            -2.389   0.0170 *
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2       -1.989   0.0468 *
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2  -1.471   0.1415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.789 on 1838 degrees of freedom
## Multiple R-squared:  0.009084,   Adjusted R-squared:  0.006927
## F-statistic: 4.212 on 4 and 1838 DF,  p-value: 0.002131
```

The model above found an association with every category besides the BMI greater than 35 group. This group had smaller sample size which could account for the larger standard error.

by past 30-day intake of ready-to-eat foods

```
lm(df$DBD905 ~ df$BMXBMICAT) |>
  summary()
```

```
##
## Call:
## lm(formula = df$DBD905 ~ df$BMXBMICAT)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.815 -1.728 -1.387 -0.630 88.272
##
## Coefficients:
##                                                             Estimate Std. Error
## (Intercept)                                                   1.2000     1.3080
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2               0.4304     1.3280
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2              0.1867     1.3231
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2        0.5275     1.3330
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2   0.6152     1.3400
##                                                             t value Pr(>|t|)
## (Intercept)                                                   0.917    0.359
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2               0.324    0.746
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2              0.141    0.888
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2        0.396    0.692
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2   0.459    0.646
##
## Residual standard error: 5.066 on 1838 degrees of freedom
## Multiple R-squared:  0.001114,   Adjusted R-squared:  -0.00106
## F-statistic: 0.5124 on 4 and 1838 DF,  p-value: 0.7267
```

None of the levels were statistically significant.

by past 30-day intake of frozen meals/frozen pizza

```
lm(df$DBD910 ~ df$BMXBMICAT) |>
  summary()
```

```
##
## Call:
```

```
## lm(formula = df$DBD910 ~ df$BMXBMICAT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.467 -13.844  -8.596   2.156 108.404
##
## Coefficients:
##                                                          Estimate Std. Error
## (Intercept)                                                23.467      5.447
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2            -6.458      5.531
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2           -8.622      5.510
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2      -5.870      5.551
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2 -2.519      5.581
##                                                          t value Pr(>|t|)
## (Intercept)                                                4.308 1.73e-05 ***
## df$BMXBMICATHealthy weight: BMI 18 to <25 kg/m2           -1.168    0.243
## df$BMXBMICATWith overweight: BMI 25 to <30 kg/m2          -1.565    0.118
## df$BMXBMICATWith class 1 obesity: BMI 30 to <35 kg/m2     -1.057    0.290
## df$BMXBMICATWith class 2 or higher obesity: BMI >=35 kg/m2 -0.451    0.652
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.1 on 1838 degrees of freedom
## Multiple R-squared:  0.01027,    Adjusted R-squared:  0.008121
## F-statistic:  4.77 on 4 and 1838 DF,  p-value: 0.0007884
```

None of the levels were statistically significant.

Summary

First we removed all instances that had missing data this lowered our number of rows to 1843. Then we numerically and visually summarized the data. For the first research question we found that there are is a statistically significant linear association between our continuous BMI BMXBMI and our two phthalates. However their R squared are both less than 2%. For part b we found that gender was a statistically significant interaction for the categorical BMI BMXBMICAT. For research question 2 we found that our second Phthalate was more promising, It showed a statistically significant linear association with fast food intake and frozen meal intake. Finally for our look at Dietary Intake and BMI we found that fast food intake had a statistically significant linear association with continuous BMI and categorical BMI. Also frozen meal intake had a statistically significant linear association with continuous BMI.