

Module 8

Setup

Start by loading some packages (we already installed) that we want to use. The # denotes a comment - anything after a # on a line is ignored by R.

```
library(readxl) # reads excel files
library(lubridate) # working with dates
library(janitor) # frequency tables with percentages
library(gt) # nicely formatted tables for output
library(gtsummary) # summary tables of multiple variables
library(knitr) # for some output tables
library(tidyverse) # makes some data manipulation tasks
easier
library(epiR) # association tests
library(pwr) # power calculations
```

Import the data. The data files here are in the current working directory (where R starts to look for files in the file system). We can read directly from an Excel file without saving to CSV first.

```
uganda <- read_excel("Uganda Malaria HIV Data.xlsx",
sheet=1)
```

Previous Clean-Up

Some clean-up and variable creation steps done first in Module 6:

```
uganda$death <- ifelse(is.na(uganda$dod), 0, 1)
uganda$malaria <- ifelse(is.na(uganda$mal_dx_dt), 0, 1)
uganda$hiv <- ifelse(is.na(uganda$hiv_dx_dt), 0, 1)
uganda$log_hiv_rna <- log10(uganda$hiv_rna)
uganda$age <- floor(time_length(difftime(uganda$entry_dt,
uganda$dob), "years"))
uganda$bmi <- uganda$weight/((uganda$height/39.370)^2) # 1
meter is 39.370 inches
uganda$bmi_cat <- NA
uganda$bmi_cat[uganda$bmi < 18.5] <- "underweight"
```

```
uganda$bmi_cat[uganda$bmi >= 18.5 & uganda$bmi < 25] <-
"normal"
uganda$bmi_cat[uganda$bmi >= 25 & uganda$bmi < 30] <-
"overweight"
uganda$bmi_cat[uganda$bmi >= 30] <- "obese"
uganda$bmi_cat <- factor(uganda$bmi_cat,
levels=c("underweight", "normal",
"overweight", "obese"))
```

Malaria Incidence Rate

Compute the malaria incidence rate for the cohort. To do this, you need the number of events and follow-up time.

First, calculate the follow-up time from entry (`entry_dt`) to malaria diagnosis (`mal_dx_dt`), death (`dod`), or end of the study (we'll use "2013-12-31" as the end of study date). The `coalesce()` function will select the first value that isn't missing.

```
uganda$mal_time <- coalesce(uganda$mal_dx_dt, uganda$dod,
ymd("2013-12-31")) - uganda$entry_dt
```

`uganda$mal_time` is in days. Convert this to person-years and sum it up across the study. This will be the denominator: the total amount of time people were observed in the study.

```
mal_py <- sum(as.numeric(uganda$mal_time)/365.25)
mal_py
## [1] 1004.939
```

Now get the total number of malaria cases:

```
mal_events <- sum(uganda$malaria)
mal_events
## [1] 66
```

To compute the incidence rate, take the number of events (malaria cases) divided by the person years. Then multiply by 1000 to get the incidence rate per 1000 person years. The 1000 is to get the number to a scale that is easier to work with/interpret. The value is the number of cases per 1000 person years.

```
mal_incidence <- (mal_events/mal_py)*1000
mal_incidence
```

```
## [1] 65.67562
```

HIV Incidence Rate

Get crude HIV incidence rate for the cohort. Similar procedure as above. Again, this is in cases per 1000 person-years.

```
uganda$hiv_time <- coalesce(uganda$hiv_dx_dt, uganda$dod,  
  ymd("2013-12-31")) - uganda$entry_dt  
hiv_py <- sum(as.numeric(uganda$hiv_time)/365.25)  
hiv_events <- sum(uganda$hiv)  
hiv_incidence <- (hiv_events/hiv_py)*1000  
hiv_incidence  
## [1] 49.43099
```

Death Incidence Rate

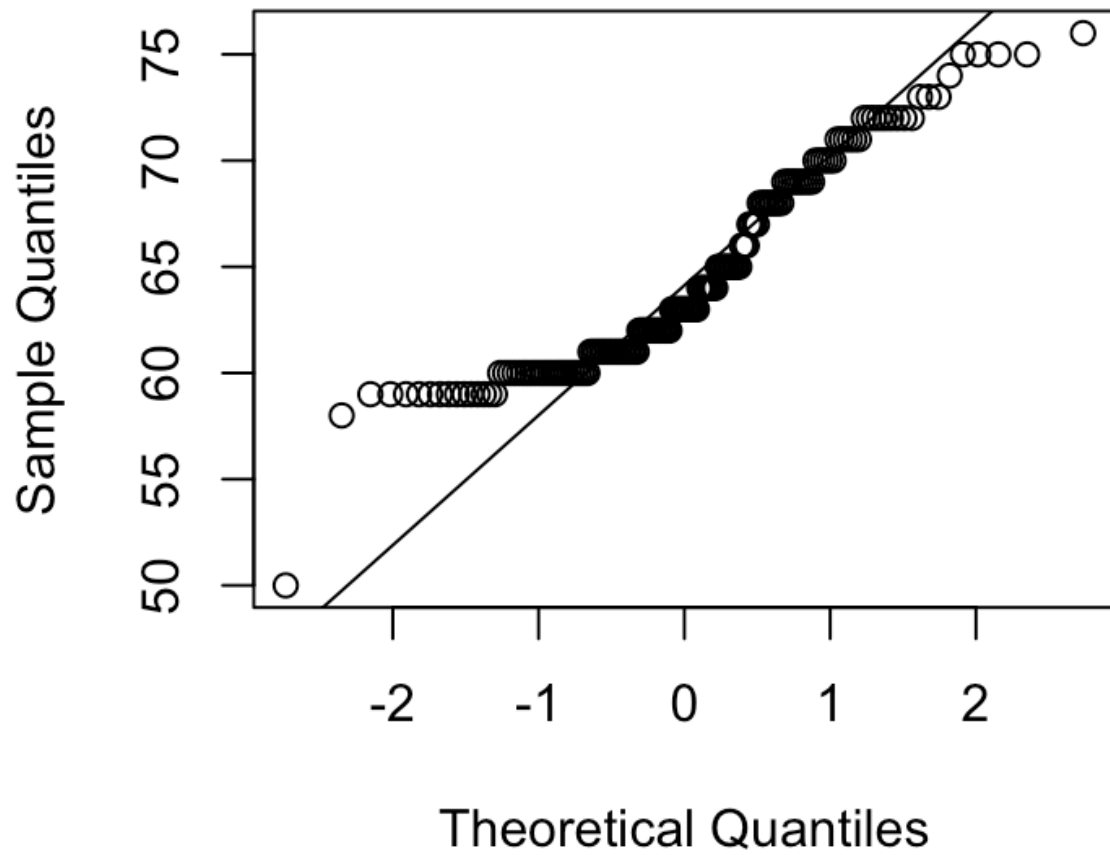
```
uganda$death_time <- coalesce(uganda$dod,  
  ymd("2013-12-31")) - uganda$entry_dt  
death_py <- sum(as.numeric(uganda$death_time)/365.25)  
death_events <- sum(uganda$death)  
death_incidence <- (death_events/death_py)*1000  
death_incidence  
## [1] 17.53649
```

QQ Plots

Check how closely distributed variables are to a normal distribution.

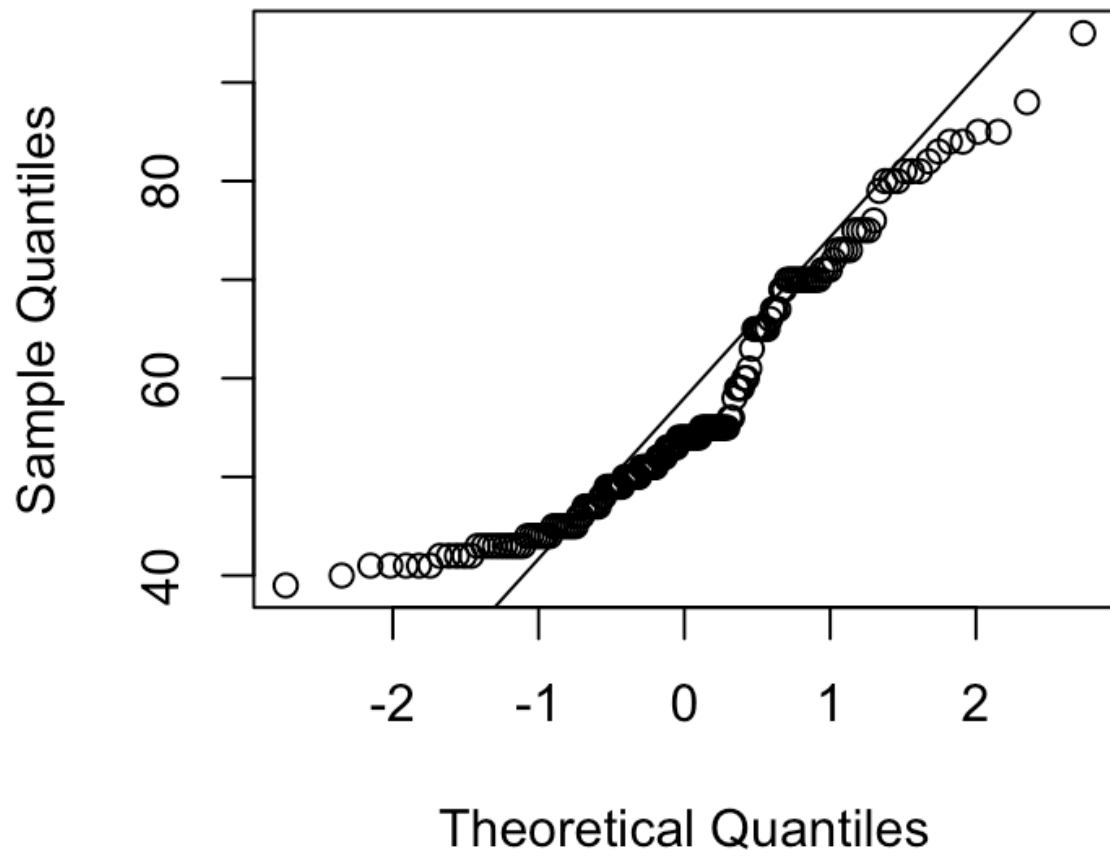
```
qqnorm(uganda$height, main="QQ Height")  
qqline(uganda$height)
```

QQ Height



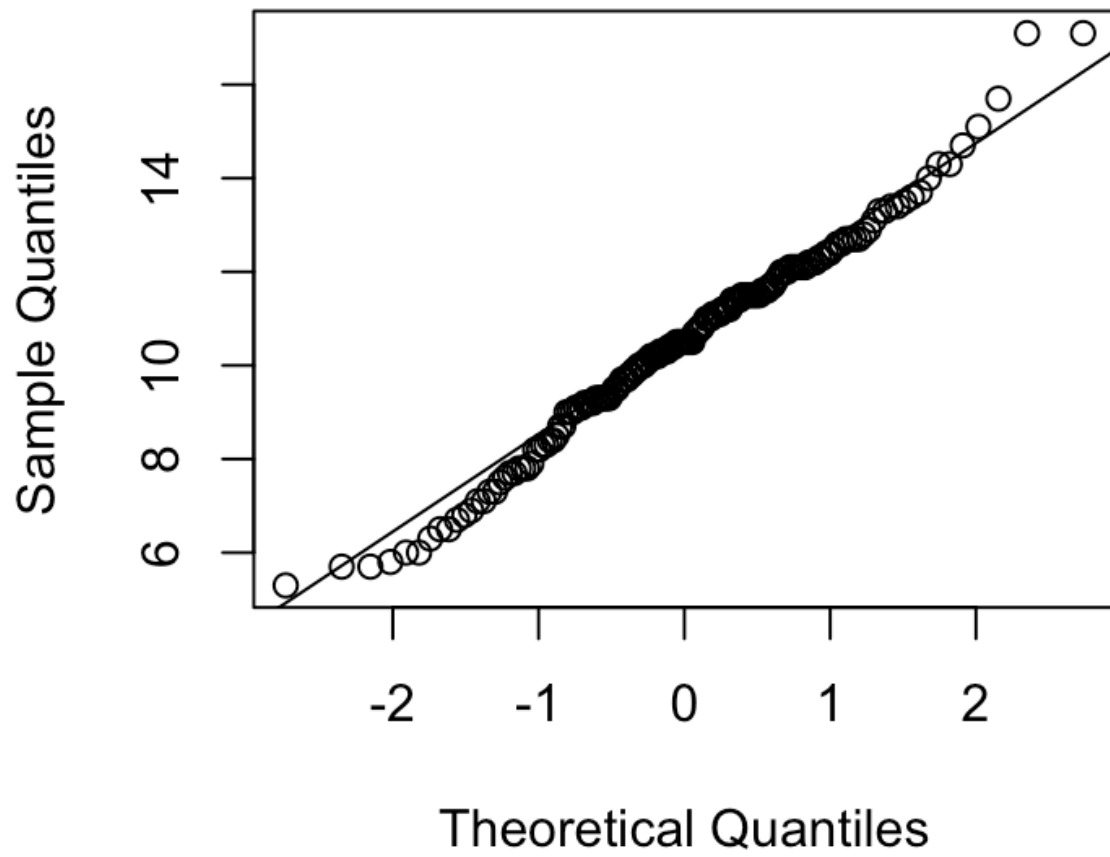
```
qqnorm(uganda$weight,main="QQ Weight")  
qqline(uganda$weight)
```

QQ Weight



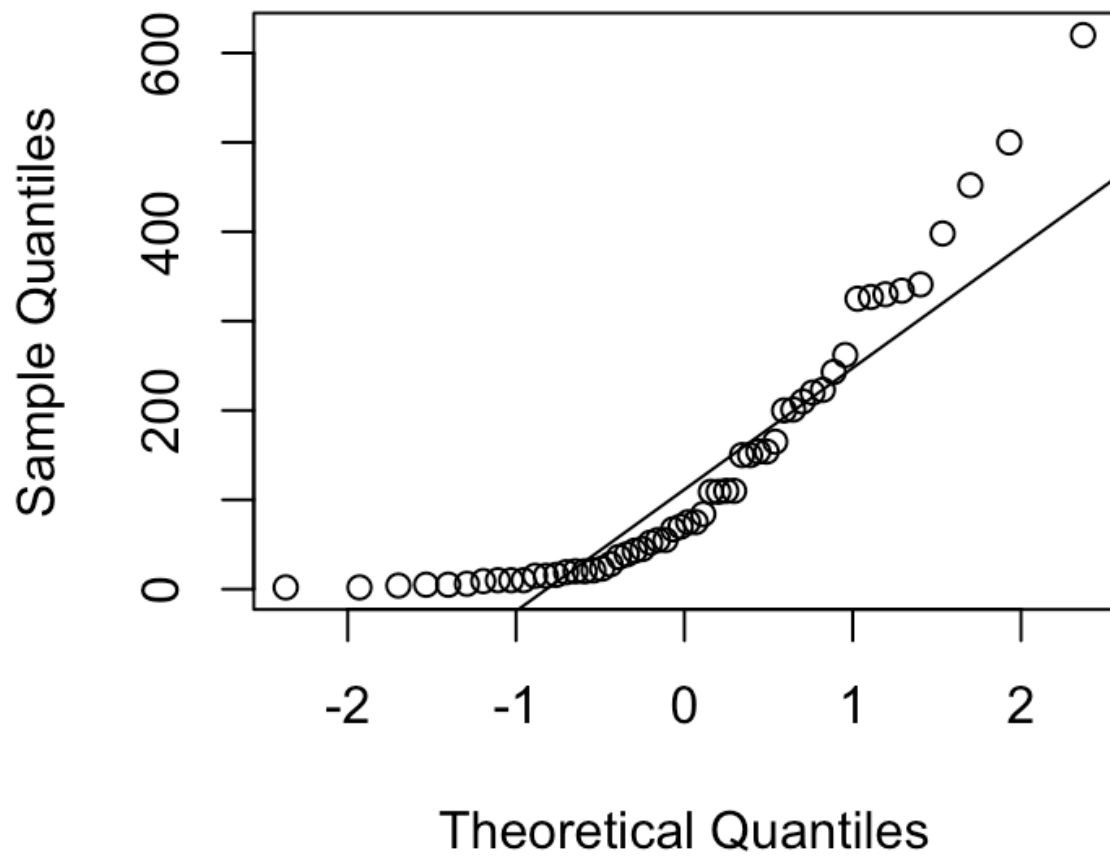
```
qqnorm(uganda$hgb,main="QQ Hgb")  
qqline(uganda$hgb)
```

QQ Hgb



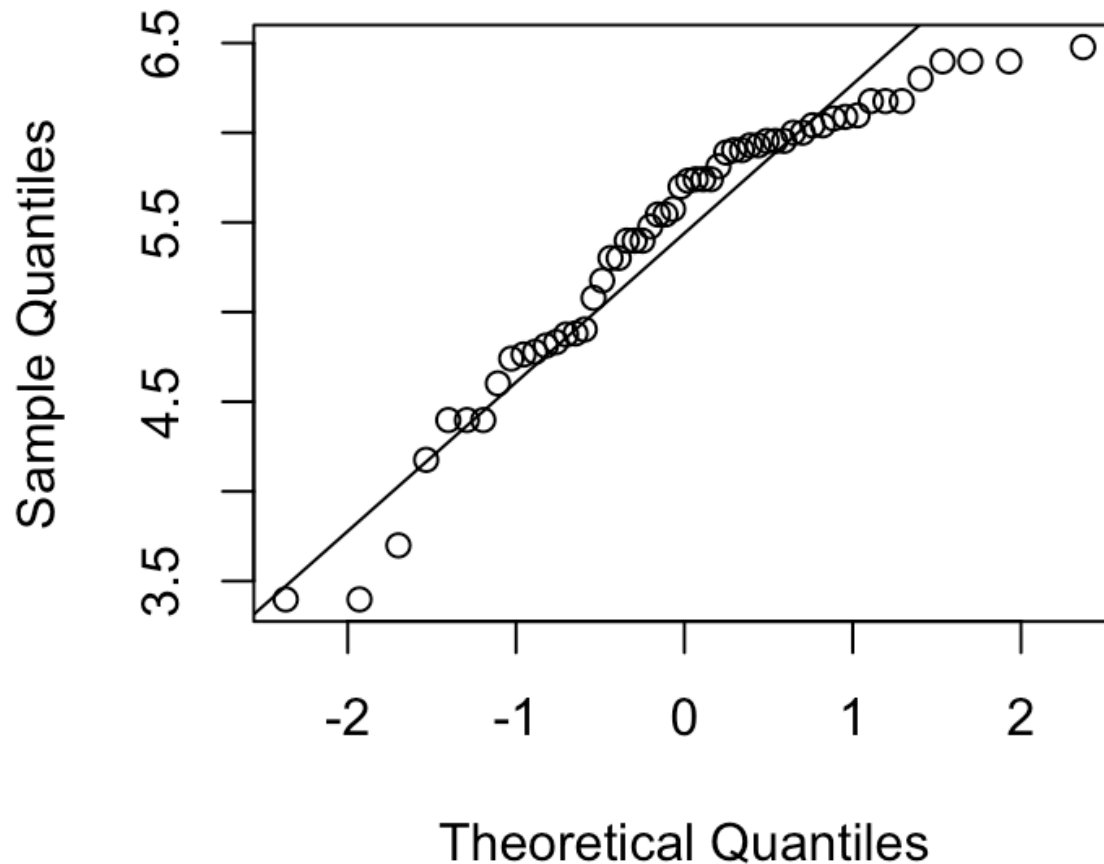
```
qqnorm(uganda$cd4,main="QQ CD4")  
qqline(uganda$cd4)
```

QQ CD4



```
qqnorm(uganda$log_hiv_rna,main="QQ HIV RNA")  
qqline(uganda$log_hiv_rna)
```

QQ HIV RNA



HIV RNA and CD4 Correlation

Spearman is a rank-based correlation more appropriate for the distribution of `hiv_rna`.

```
cor.test(uganda$hiv_rna, uganda$cd4,  
         use="complete.obs", # drops missing values  
         method="spearman")
```

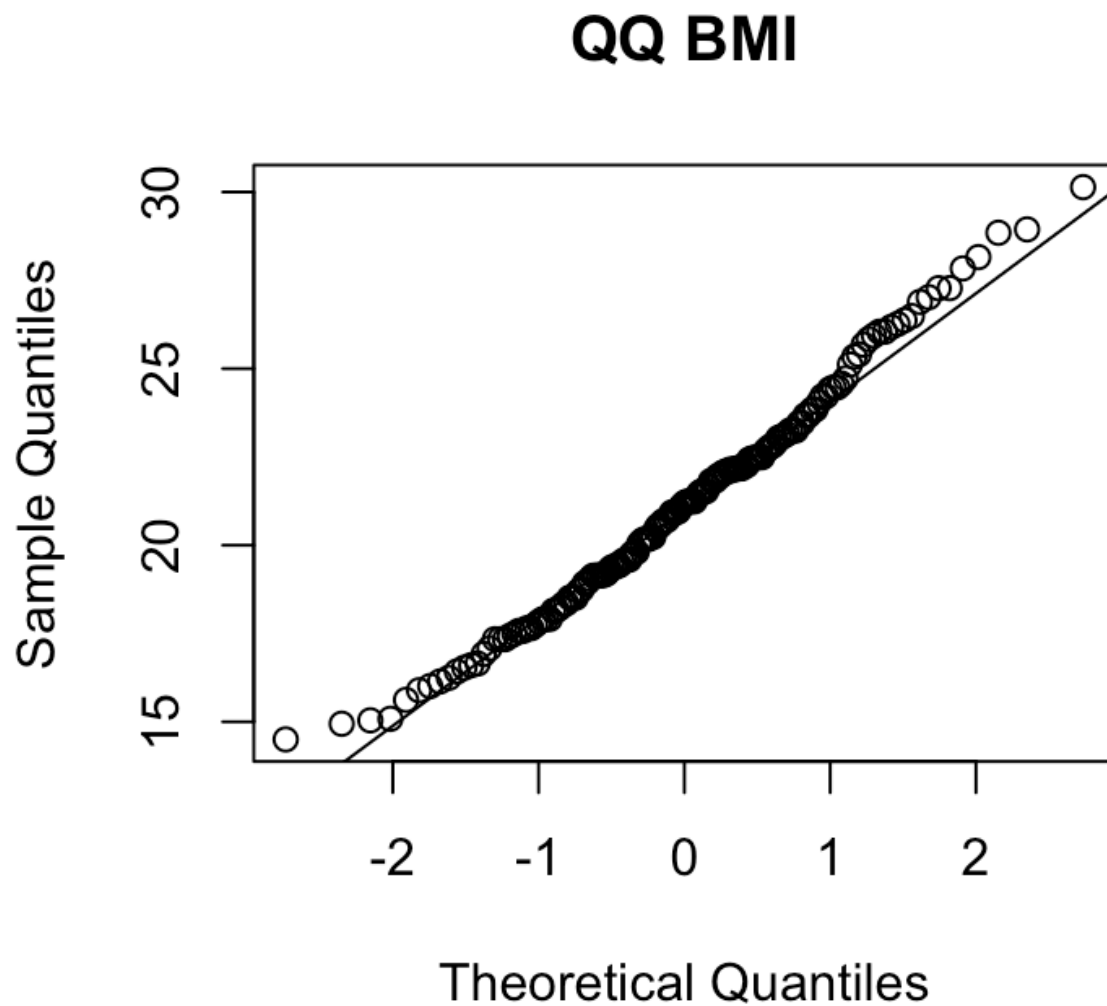
```
##  
## Spearman's rank correlation rho  
##
```



```
## data: uganda$hiv_rna and uganda$cd4
## S = 55635, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.9014109
```

Hemoglobin and BMI

```
qqnorm(uganda$bmi, main="QQ BMI")
qqline(uganda$bmi)
```



```
cor.test(uganda$hgb, uganda$bmi,
```

```

      use="complete.obs",
      method="pearson")
##
## Pearson's product-moment correlation
##
## data:  uganda$hgb and uganda$bmi
## t = -1.3984, df = 158, p-value = 0.164
## alternative hypothesis: true correlation is not equal to
0
## 95 percent confidence interval:
## -0.26124584  0.04536811
## sample estimates:
##          cor
## -0.110569

```

Condoms and HIV Infection

Is condom use associated with HIV infection?

Make an indicator for condom use – whether the person reports ever using condoms

```

uganda$condom <- ifelse(uganda$condom_use != "never",
"yes", "no")

```

Table of condom use and HIV (0=no, 1=yes):

```

uganda$condom <- factor(uganda$condom, levels=c("yes",
"no"))

```

```

uganda$hiv <- factor(uganda$hiv, levels=c(1,0),
labels=c("HIV+", "HIV-"))

```

```

tabyl(uganda, condom, hiv) %>% ## table
gt() # output formatting

```

condom	HIV+	HIV-
yes	15	94
no	41	10

```

epi.2by2(table(uganda$condom, uganda$hiv),
method="cohort.count", conf.level=0.95)

```

##		Outcome +	Outcome -	Total
Inc risk *		Odds		
## Exposed +		15	94	109
13.8	0.160			
## Exposed -		41	10	51
80.4	4.100			
## Total		56	104	160
35.0	0.538			

##

Point estimates and 95% CIs:

##

```

-----
-----
## Inc risk ratio                                0.17
(0.10, 0.28)
## Odds ratio                                    0.04
(0.02, 0.09)
## Attrib risk in the exposed *                  -66.63
(-79.30, -53.96)
## Attrib fraction in the exposed (%)            -484.18
(-852.72, -258.21)
## Attrib risk in the population *               -45.39
(-58.56, -32.23)
## Attrib fraction in the population (%)         -129.69
(-177.72, -89.97)

```

##

```

-----
-----
## Uncorrected chi2 test that OR = 1: chi2(1) = 67.802
Pr>chi2 = <0.001
## Fisher exact test that OR = 1: Pr>chi2 = <0.001
## Wald confidence limits
## CI: confidence interval
## * Outcomes per 100 population units
Calculate number needed to treat (NNT) using "Attrib risk in the exposed,"
which is per 100

```

```
condom_nnt <- 1/(66.63/100)
```

```
condom_nnt
```

```
## [1] 1.500825
```

Bed Nets

Is bed net use associated with malaria infection?

```
uganda$bed_net <- factor(uganda$bed_net, levels=c("y",  
"n"), labels=c("bed net", "no bed net"))  
uganda$malaria_labeled <- factor(uganda$malaria,  
levels=c(1, 0), labels=c("Malaria", "No Malaria"))  
tabyl(uganda, bed_net, malaria_labeled) %>%  
  gt()
```

bed_net	Malaria	No Malaria
bed net	12	64
no bed net	54	30

```
epi.2by2(table(uganda$bed_net, uganda$malaria_labeled),  
  method="cohort.count", conf.level=0.95)
```

```
##              Outcome +      Outcome -      Total  
Inc risk *      Odds  
## Exposed +          12          64          76  
15.8      0.188  
## Exposed -          54          30          84  
64.3      1.800  
## Total              66          94          160  
41.2      0.702  
##  
## Point estimates and 95% CIs:  
##
```

```
-----  
## Inc risk ratio              0.25  
(0.14, 0.42)  
## Odds ratio                  0.10  
(0.05, 0.22)  
## Attrib risk in the exposed *      -48.50  
(-61.62, -35.37)  
## Attrib fraction in the exposed (%)      -307.14  
(-600.85, -136.52)
```

```
## Attrib risk in the population * -23.04
(-35.81, -10.26)
## Attrib fraction in the population (%) -55.84
(-78.24, -36.26)
##
-----
-----
## Uncorrected chi2 test that OR = 1: chi2(1) = 38.722
Pr>chi2 = <0.001
## Fisher exact test that OR = 1: Pr>chi2 = <0.001
## Wald confidence limits
## CI: confidence interval
## * Outcomes per 100 population units
Calculate number needed to treat (NNT) using "Attrib risk in the exposed,"
which is per 100

bednet_nnt <- 1/(48.50/100)
bednet_nnt
## [1] 2.061856
```

Factors Associated with HIV

Use multiple logistic regression

```
# set category order
uganda$hiv <- ifelse(uganda$hiv == "HIV-", 0, 1)
uganda$bmi_cat <- factor(uganda$bmi_cat, levels=c("normal",
"underweight",
"overweight", "obese"))

hivlogreg <- glm(hiv ~ age + sex + village + marital_status
+
bmi_cat + sex_partners + condom_use,
data=uganda, family="binomial")
summary(hivlogreg) # console output
##
## Call:
## glm(formula = hiv ~ age + sex + village + marital_status
+ bmi_cat +
```

```
##      sex_partners + condom_use, family = "binomial", data
= uganda)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -1.93871   -0.03922   -0.00614    0.00859    2.50715
##
## Coefficients:
##
##              Estimate Std. Error z value
Pr(>|z|)
## (Intercept)          -1.527e+01  9.849e+00  -1.550
0.121085
## age                 -5.357e-04  1.336e-01  -0.004
0.996801
## sexMale              1.041e+00  2.317e+00   0.450
0.653041
## villageLori         -2.512e+00  4.485e+00  -0.560
0.575401
## villageNoko         -1.022e+00  3.183e+00  -0.321
0.748035
## villageOmugo        -1.711e+00  3.255e+00  -0.526
0.599075
## villageYumbe        -1.881e+00  3.911e+00  -0.481
0.630513
## marital_statusMarried 1.478e+00  6.670e+00   0.222
0.824673
## marital_statusSingle  1.884e+00  6.940e+00   0.271
0.786019
## marital_statusWidowed -1.278e-01  7.369e+00  -0.017
0.986160
## bmi_catunderweight  -2.189e+00  3.413e+00  -0.641
0.521345
## bmi_catoverweight     7.112e-01  4.516e+00   0.157
0.874865
## bmi_catobese         -6.859e+00  3.956e+03  -0.002
0.998617
## sex_partners         4.811e+00  1.450e+00   3.317
0.000909 ***
## condom_usenever      5.900e+00  2.924e+00   2.018
0.043640 *
```

```
## condom_usesometimes      2.758e+00  2.514e+00   1.097
0.272593
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 207.183  on 159  degrees of freedom
## Residual deviance:  14.307  on 144  degrees of freedom
## AIC: 46.307
##
## Number of Fisher Scoring iterations: 16
tbl_regression(hivlogreg, exponentiate=TRUE) ## adding
odds-ratio (OR) and formatting
```

Characteristic	OR ¹	95% CI ¹	p-value
age	1.00	0.75, 1.34	>0.9
sex			
Female	—	—	
Male	2.83	0.02, 530	0.7
village			
Kaya	—	—	
Lori	0.08	0.00, 533	0.6
Noko	0.36	0.00, 111	0.7
Omugo	0.18	0.00, 171	0.6
Yumbe	0.15	0.00, 114	0.6
marital_status			

Divorced	—	—	
Married	4.38	0.00, 4,728,042	0.8
Single	6.58	0.00, 14,734,727	0.8
Widowed	0.88	0.00, 1,164,008	>0.9
bmi_cat			
normal	—	—	
underweight	0.11	0.00, 29.2	0.5
overweight	2.04	0.00, 29,222	0.9
obese	0.00		>0.9
sex_partners	123	15.7, 6,540	<0.001
condom_use			
always	—	—	
never	365	5.90, 1,164,477	0.044
sometimes	15.8	0.20, 9,390	0.3

Factors Associated with Malaria

```
mallogreg <- glm(malaria ~ age + sex + village + bmi + hgb
+ bed_net_days,
                 data=uganda, family="binomial")
summary(mallogreg)
```



```
##
## Call:
## glm(formula = malaria ~ age + sex + village + bmi + hgb
+ bed_net_days,
##       family = "binomial", data = uganda)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1774  -0.3459  -0.1535   0.6982   2.1408
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.563802    2.210997  -0.255    0.7987
## age           0.041084    0.023001   1.786    0.0741 .
## sexMale       0.254277    0.552179   0.460    0.6452
## villageLori  -1.247920    0.773761  -1.613    0.1068
## villageNoko   0.934073    0.844093   1.107    0.2685
## villageOmugo -0.588816    0.686483  -0.858    0.3910
## villageYumbe -1.597151    0.666510  -2.396    0.0166 *
## bmi           0.022899    0.072528   0.316    0.7522
## hgb           0.008275    0.108197   0.076    0.9390
## bed_net_days -0.671748    0.116365  -5.773 7.8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 216.88  on 159  degrees of freedom
## Residual deviance: 122.62  on 150  degrees of freedom
## AIC: 142.62
##
## Number of Fisher Scoring iterations: 6
tbl_regression(mallogreg, exponentiate=TRUE)
```

Characteristic	OR ¹	95% CI ¹	p-value
age	1.04	1.00, 1.09	0.074
sex			

Female	—	—	
Male	1.29	0.44, 3.91	0.6
village			
Kaya	—	—	
Lori	0.29	0.06, 1.29	0.11
Noko	2.54	0.51, 14.9	0.3
Omugo	0.55	0.14, 2.12	0.4
Yumbe	0.20	0.05, 0.72	0.017
bmi	1.02	0.89, 1.18	0.8
hgb	1.01	0.81, 1.25	>0.9
bed_net_days	0.51	0.39, 0.62	<0.001

Power Analysis

Using Power Analysis determine samples size for 50% decrease incidence. Determine effect size (h) based on 15% baseline then reduce to 7.5% (50% reduction). Use the equation from the website:

```
h50 <- 2*asin(sqrt(0.075))-2*asin(sqrt(0.15))
h50
```

```
## [1] -0.2405878
```

Now calculate the sample size needed (the “n”) for that h value, with power .8 and a 0.05 significance level:

```
pwr.2p.test(h=h50,
             sig.level=0.05,
             power=0.80,
             alternative="less")
```

```
##
##      Difference of proportion power calculation for
binomial distribution (arcsine transformation)
##
##              h = -0.2405878
##              n = 213.6244
##      sig.level = 0.05
##      power = 0.8
##      alternative = less
##
## NOTE: same sample sizes
Repeat for a 25% reduction (to 11.25%) and a 75% reduction (to 3.75%)
```

```
h25 <- 2*asin(sqrt(0.1125))-2*asin(sqrt(0.15))
pwr.2p.test(h=h25,
            sig.level=0.05,
            power=0.80,
            alternative="less")
```

```
##
##      Difference of proportion power calculation for
binomial distribution (arcsine transformation)
##
##              h = -0.1113176
##              n = 997.862
##      sig.level = 0.05
##      power = 0.8
##      alternative = less
##
```

```
## NOTE: same sample sizes
h75 <- 2*asin(sqrt(0.0375))-2*asin(sqrt(0.15))
pwr.2p.test(h=h75,
            sig.level=0.05,
            power=0.80,
            alternative="less")
```

```
##
##      Difference of proportion power calculation for
binomial distribution (arcsine transformation)
##
##              h = -0.4056381
##              n = 75.14856
```

```
##          sig.level = 0.05
##          power = 0.8
##    alternative = less
##
## NOTE: same sample sizes
```

RCT

Import the data on the randomized clinical trial. It's in a special R file type:

```
load("ugandarct.RData")
```

This loads a data frame called `ugandarct` into the R environment.

Baseline Characteristics

In the RCT, are baseline characteristics balanced?

Categorical variables

Marital Status

```
maritalrct <- table(ugandarct$marital_status,
ugandarct$RCT_treat)
kable(maritalrct) %>% kableExtra::kable_styling(full_width
= FALSE)
```

	0	1
Divorced	2	1
Married	14	16
Single	9	5
Widowed	0	3

```
fisher.test(maritalrct)
```

```
##
```

```
## Fisher's Exact Test for Count Data
```

```
##
```

```
## data:  maritalrct
```

```
## p-value = 0.2241
## alternative hypothesis: two.sided
Sex
```

```
sexrct <- table(ugandarct$sex, ugandarct$RCT_treat)
kable(sexrct) %>% kableExtra::kable_styling(full_width =
FALSE)
```

	0	1
Female	16	19
Male	9	6

```
fisher.test(sexrct)
##
## Fisher's Exact Test for Count Data
##
## data: sexrct
## p-value = 0.538
## alternative hypothesis: true odds ratio is not equal to
1
## 95 percent confidence interval:
## 0.1343541 2.2463926
## sample estimates:
## odds ratio
## 0.5679574
```

Village

```
villagerct <- table(ugandarct$village, ugandarct$RCT_treat)
kable(villagerct) %>% kableExtra::kable_styling(full_width
= FALSE)
```

	0	1
Kaya	7	7
Lori	3	2
Noko	9	6
Omugo	3	5

Yumbe	3	5
-------	---	---

```
fisher.test(villagerct)
##
## Fisher's Exact Test for Count Data
##
## data: villagerct
## p-value = 0.7903
## alternative hypothesis: two.sided
BMI category
```

```
bmicatrct <- table(ugandarct$bmi_cat, ugandarct$RCT_treat)
kable(bmicatrct) %>% kableExtra::kable_styling(full_width =
FALSE)
```

	0	1
normal	17	17
overweight	3	2
underweight	5	6

```
fisher.test(bmicatrct)
##
## Fisher's Exact Test for Count Data
##
## data: bmicatrct
## p-value = 1
## alternative hypothesis: two.sided
```

Continuous variables

Age

```
t.test(age~RCT_treat, data=ugandarct)
##
## Welch Two Sample t-test
##
## data: age by RCT_treat
## t = 1.3449, df = 47.656, p-value = 0.185
## alternative hypothesis: true difference in means between
group 0 and group 1 is not equal to 0
```

```
## 95 percent confidence interval:
## -1.941349  9.781349
## sample estimates:
## mean in group 0 mean in group 1
##          40.20          36.28
```

Height

```
t.test(height~RCT_treat, data=ugandarct)
##
##  Welch Two Sample t-test
##
## data:  height by RCT_treat
## t = 0.85, df = 44.38, p-value = 0.3999
## alternative hypothesis: true difference in means between
group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -1.480087  3.640087
## sample estimates:
## mean in group 0 mean in group 1
##          64.88          63.80
```

Weight

```
t.test(weight~RCT_treat, data=ugandarct)
##
##  Welch Two Sample t-test
##
## data:  weight by RCT_treat
## t = 1.5393, df = 43.984, p-value = 0.1309
## alternative hypothesis: true difference in means between
group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -1.447357 10.807357
## sample estimates:
## mean in group 0 mean in group 1
##          58.96          54.28
```

BMI

```
t.test(bmi~RCT_treat, data=ugandarct)
##
##  Welch Two Sample t-test
##
```

```
## data:  bmi by RCT_treat
## t = 1.1437, df = 46.806, p-value = 0.2585
## alternative hypothesis: true difference in means between
group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.721548  2.622604
## sample estimates:
## mean in group 0 mean in group 1
##      21.60741      20.65688
```

Hemoglobin

```
t.test(hgb~RCT_treat, data=ugandarct)
##
##  Welch Two Sample t-test
##
## data:  hgb by RCT_treat
## t = 0.63873, df = 42.67, p-value = 0.5264
## alternative hypothesis: true difference in means between
group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.7423647  1.4303647
## sample estimates:
## mean in group 0 mean in group 1
##      10.548      10.204
```

Parasitemia

```
t.test(parasitemia~RCT_treat, data=ugandarct)
##
##  Welch Two Sample t-test
##
## data:  parasitemia by RCT_treat
## t = 0, df = 47.853, p-value = 1
## alternative hypothesis: true difference in means between
group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -1.843801  1.843801
## sample estimates:
## mean in group 0 mean in group 1
##      6.48      6.48
```

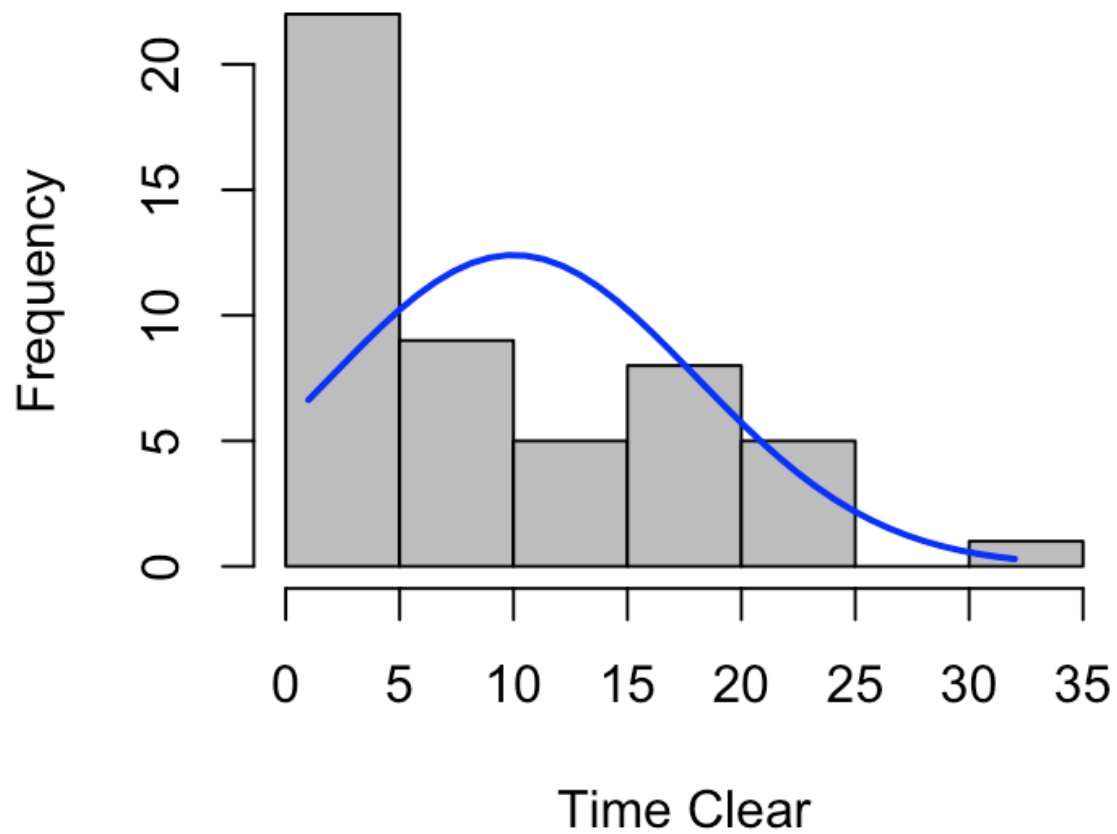

Parasitemia

Is there a difference in time to clearance of parasitemia?

Is the outcome normally distributed? Create histogram with curve and QQ plot

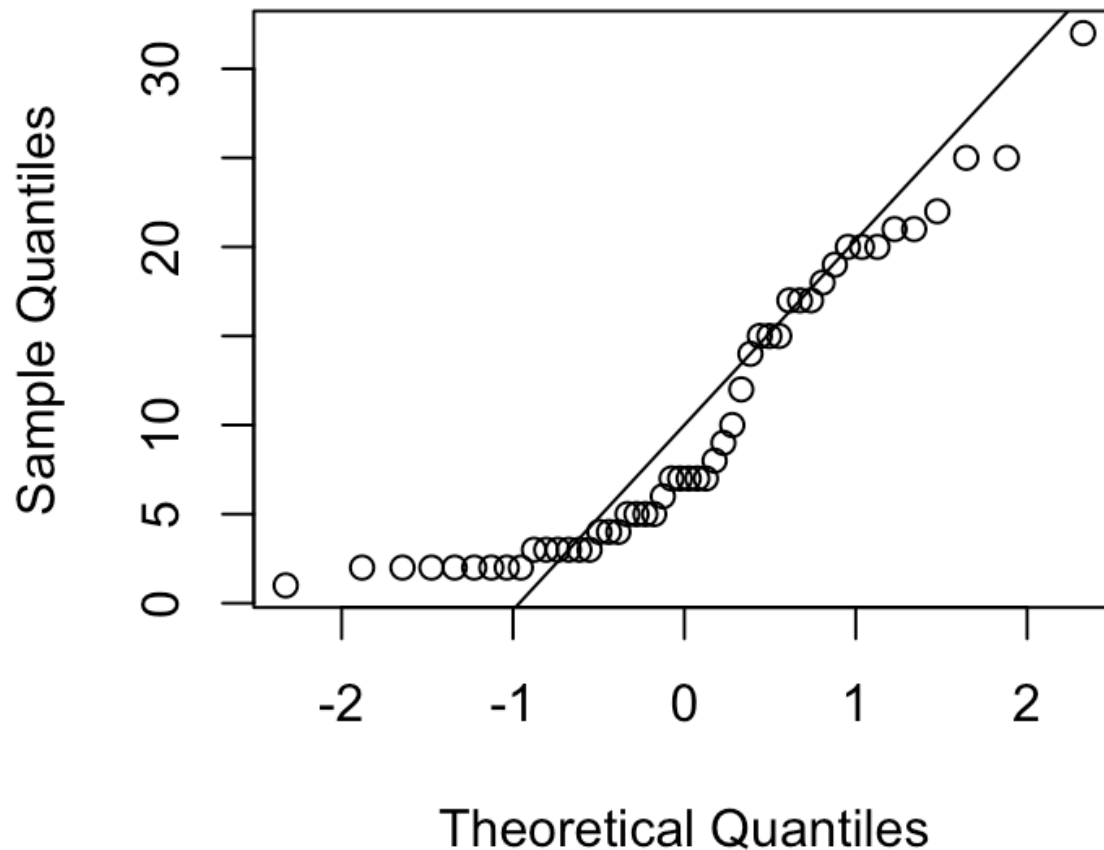
```
h <- hist(ugandarct$time_clear,
          main="Histogram of Clearance Outcome",
          xlab="Time Clear",
          col="gray")
xfit <- seq(min(ugandarct$time_clear),
            max(ugandarct$time_clear), length=40)
yfit <- dnorm(xfit, mean=mean(ugandarct$time_clear),
              sd=sd(ugandarct$time_clear))
yfit <- yfit*diff(h$mids[1:2])*length(ugandarct$time_clear)
lines(xfit, yfit, col="blue", lwd=2)
```

Histogram of Clearance Outcome



```
qqnorm(ugandarct$time_clear,main="QQ Time Clear")  
qqline(ugandarct$time_clear)
```

QQ Time Clear



Compare descriptive stats and analyze continuous outcome with non-parametric test since the variable is not distributed normally

```
ugandarct$RCT_treat_txt <- ifelse(ugandarct$RCT_treat > 0,  
  "Drug", "Placebo")
```

```
tapply(ugandarct$time_clear, ugandarct$RCT_treat_txt,  
  summary)
```

```
## $Drug
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      1.00   2.00   3.00   3.56   5.00   8.00
```

```
##
```

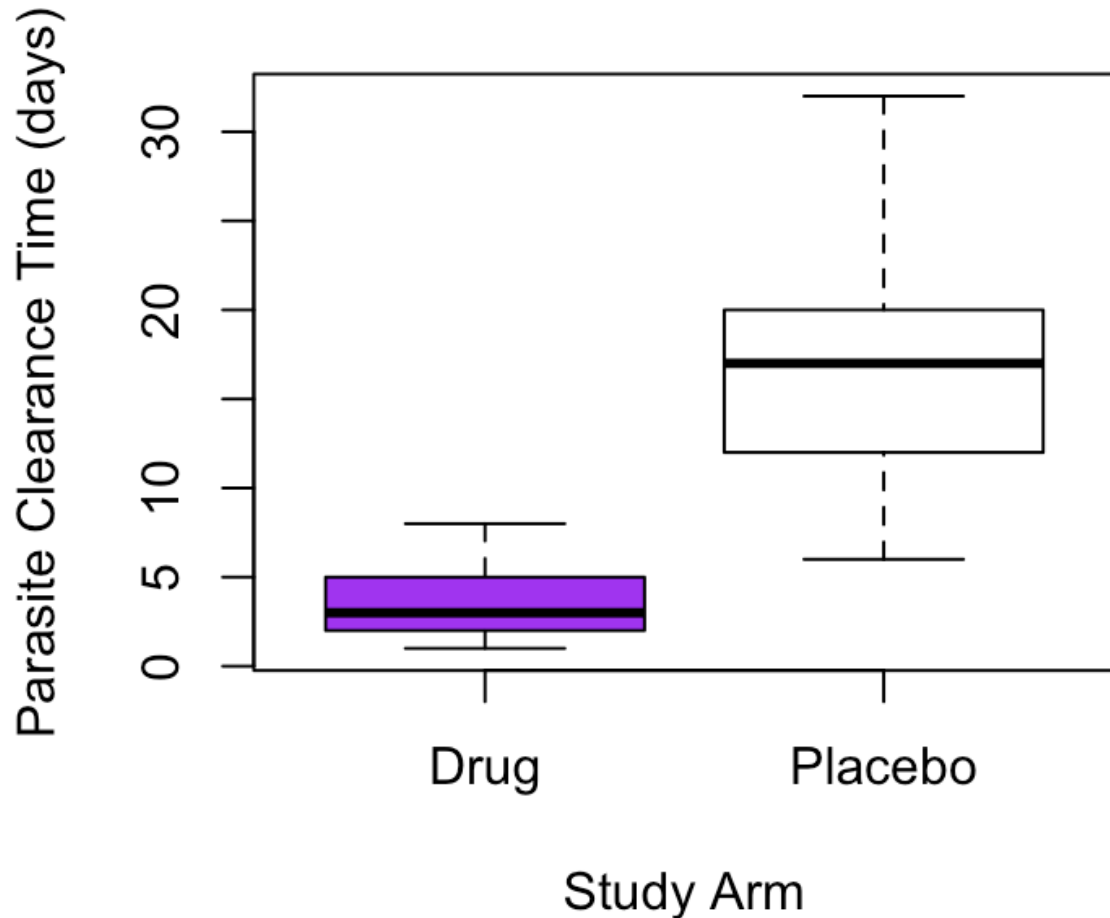
```
## $Placebo
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      6.00   12.00   17.00   16.44   20.00   32.00
wilcox.test(time_clear~RCT_treat, data=ugandarct)
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  time_clear by RCT_treat
## W = 616, p-value = 3.574e-09
## alternative hypothesis: true location shift is not equal
to 0
```

Create a visualization of the difference above

```
boxplot(time_clear~RCT_treat_txt,
        data=ugandarct,
        col=(c("purple", "white")),
        main="Malaria RCT Results",
        xlab="Study Arm",
        ylab="Parasite Clearance Time (days)")
```

Malaria RCT Results



Clearance Time

Linear regression for clearance time outcome with treatment and other important factors: age, bmi, parasitemia

```
malrctlinreg <- lm(time_clear ~ RCT_treat + age + bmi +  
parasitemia,  
                    data=ugandarct)  
summary(malrctlinreg)  
##  
## Call:
```

```
## lm(formula = time_clear ~ RCT_treat + age + bmi +
parasitemia,
##      data = ugandarct)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -9.8067 -1.8934 -0.3632  2.3936 15.1952
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.99969    5.97909   1.672   0.101
## RCT_treat  -12.36302    1.39249  -8.878 1.88e-11 ***
## age           0.08441    0.06687   1.262   0.213
## bmi           0.19577    0.23460   0.834   0.408
## parasitemia  -0.18259    0.21268  -0.858   0.395
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
##
## Residual standard error: 4.772 on 45 degrees of freedom
## Multiple R-squared:  0.6765, Adjusted R-squared:  0.6477
## F-statistic: 23.53 on 4 and 45 DF,  p-value: 1.521e-10
```

Power

What was the power of your randomized clinical trial?

What was n in each group?

```
tabyl(ugandarct, RCT_treat_txt) %>% gt()
```

RCT_treat_txt	n	percent
Drug	25	0.5
Placebo	25	0.5

Calculate effect size (d) from equation on Quick-R website, first calculate standard deviation of time to clear

```
d <- (mean(ugandarct$time_clear[ugandarct$RCT_treat_txt ==
"Placebo"]) -
      mean(ugandarct$time_clear[ugandarct$RCT_treat_txt
== "Drug"])) /
sd(ugandarct$time_clear)
d
```

```
## [1] 1.601848
```

Calculate power with pwr.t.test since equal size groups and d from above

```
pwr.t.test(n=25,
           d=d, # value computed above
           sig.level = 0.05)

##
##      Two-sample t test power calculation
##
##              n = 25
##              d = 1.601848
##      sig.level = 0.05
##      power = 0.9998328
##      alternative = two.sided
##
## NOTE: n is number in *each* group
```