Impact of Zinc Treatment on Immune Response in HIV-positive Children

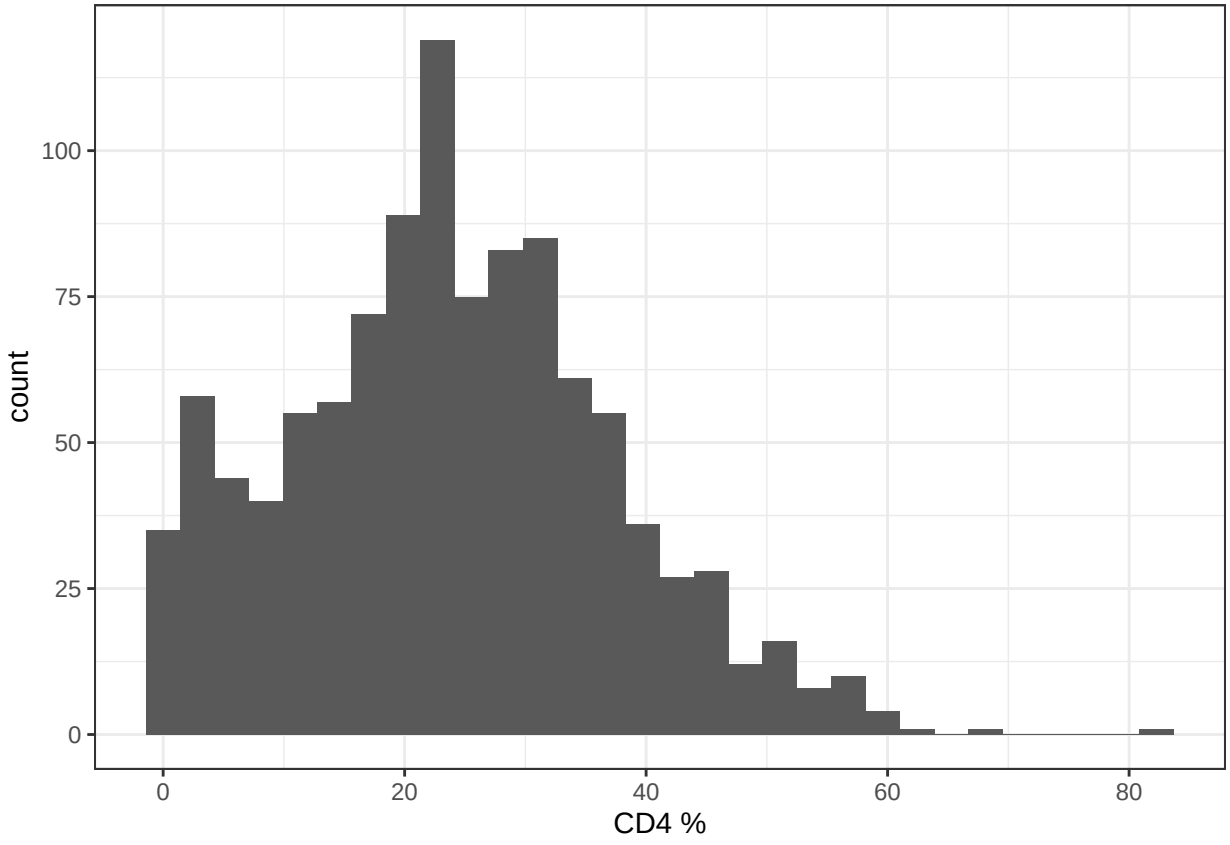
Alexey Arkhangelskiy

12/08/2022

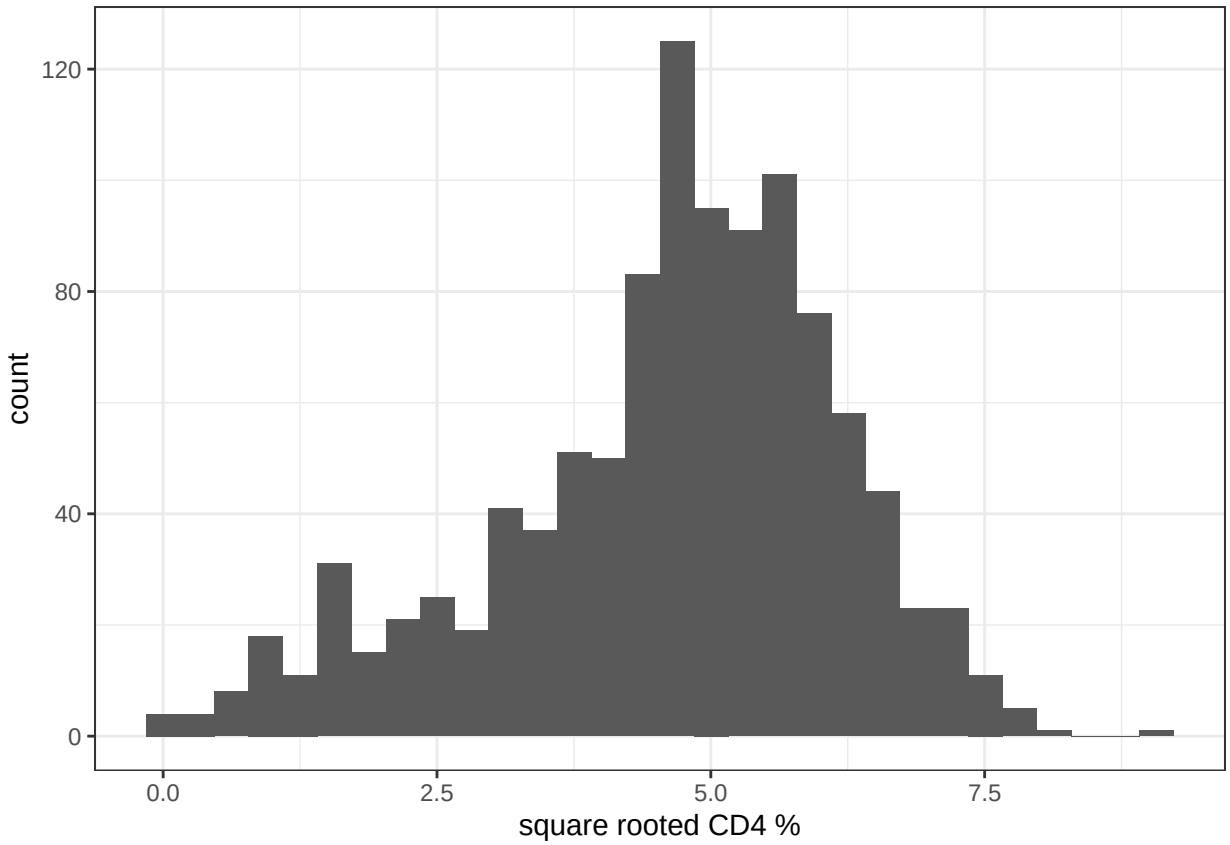
Question 1

a)

Let's plot a histogram of the CD4 percentages.

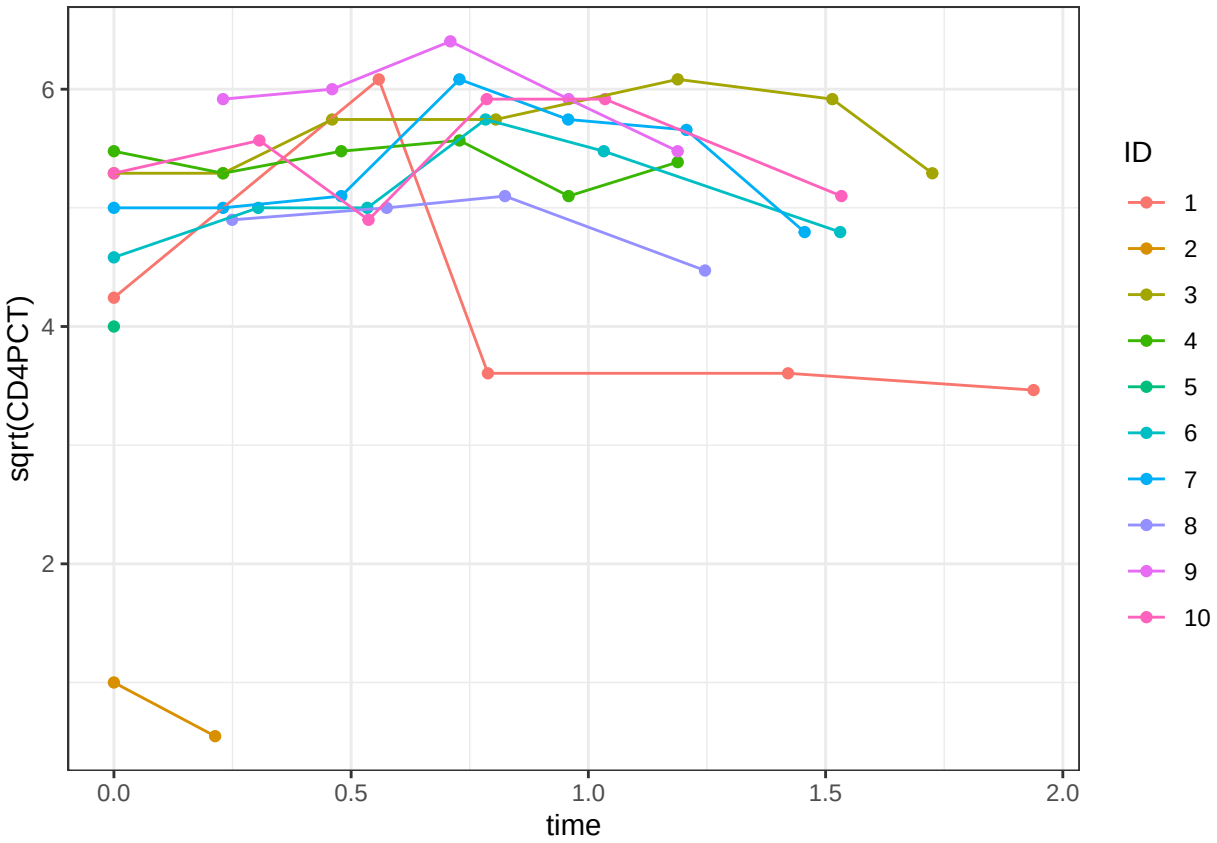


The issue here is that if we want to use a normal model, having that left-truncated data near the y-axis will inevitably cause issues down the line. A good transformation is the square root transformation, which leaves the distribution reasonable symmetric.



b)

Let's plot the CD4 percentages over time.



Part c)

```
simple_model <- lmer(data=cd4, sqrt(CD4PCT) ~ time + (1|newpid))
summary(simple_model)$varcor
```

```
##  Groups   Name        Std.Dev.
##  newpid   (Intercept) 1.3989
##  Residual             0.7725
```

```
summary(simple_model)$coefficients
```

```
##               Estimate Std. Error       df   t value      Pr(>|t|)
## (Intercept)  4.7634086 0.09647959 289.0009 49.372189 7.032875e-143
## time        -0.3660932 0.05398921 853.9097 -6.780858  2.228714e-11
```

Part d)

```
complex_model <- lmer(data=cd4, sqrt(CD4PCT) ~ time + baseage + treatmnt + (1|newpid))
summary(complex_model)$varcor
```

```
##  Groups   Name        Std.Dev.
##  newpid   (Intercept) 1.37466
##  Residual             0.77258
```

```
summary(complex_model)$coefficients
```

```
##               Estimate Std. Error       df    t value     Pr(>|t|)
## (Intercept)  4.9060558 0.31683637 247.0373 15.4845092 2.945757e-38
## time        -0.3621573 0.05398705 854.0961 -6.7082246 3.583000e-11
## baseage     -0.1194538 0.04000131 245.4081 -2.9862464 3.110165e-03
## treatmnt     0.1800822 0.18261614 243.0341  0.9861244 3.250528e-01
```

The interesting thing about these predictors is that they describe the GROUP of observations. In other words, they are constant for each child. In the Radon example, we had a predictor (basement) that described each house. But imagine if we had variables that described each county.

These are called group level predictors, and they should effect our group-level variance, but not our individual level variance. In our case, the effect is very small (sd went from 1.399 to 1.375), but definitely noticeable. Note that our individual (residual) level variance barely changed at all.

Part e)

On average, patients who received the treatment had a CD4 percentage that was 0.18 higher on the square root scale than those who did not receive the treatment. This result was not statistically significant when using a t-test with corrected degrees of freedom. We don't have enough evidence to conclude that the treatment is effective.

Part f)

Need information on this child:

```
newpid9 <- cd4 %>% filter(newpid==9) %>% select(baseage, treatmnt) %>% colMeans
```

```
alpha9 <- coef(complex_model)$newpid[9,1]
betas <- fixef(complex_model)
sigma_eps <- summary(complex_model)$sigma

n.sims <- 1000
set.seed(12314112)

newpid9_new <- rnorm(n.sims, alpha9 + betas[2]*1.4 + betas[3]*newpid9[1], sigma_eps)

pi <- quantile(newpid9_new, probs=c(0.025, 0.975))
```

The 95% prediction interval is (15.58 % to 48.26 %), with a point estimate of 29.34 %

Part g)

```r
baseages <- cd4 %>% group_by(newpid) %>% summarise(age = mean(baseage))

sigma_alpha <- summary(complex_model)$varcor$newpid[1] %>% sqrt()
mu_alpha <- betas[1]

alpha_new <- rnorm(n.sims, mu_alpha, sigma_alpha)

new_newpid <- rnorm(n.sims, alpha_new + betas[2]*1 + betas[3]*mean(baseages$age), sigma_eps)

pi2 <- quantile(new_newpid, probs = c(0.025, 0.5, 0.975))^2
```

The prediction interval for the new child is (1.13% to 54.02%), with a point estimate of 16.95%.

Had we used the model from c), our interval would be wider because we are more uncertain about the child's intercept. I.e there would be more variation in α_{new}.