

# An analysis of my baby's daily life

**Peifen Lu 18008550**

Taking care of a baby can be one of the most difficult things. First-time parents or even experienced parents can be nervous and sometimes have hard times dealing with baby stuff. For me, I find it helpful to record the feeding, poops, sleeps etc so that I know whether there is anything I need to worry about. To understand my baby's daily patterns, I decide to do an analysis on it. Hopefully, this portfolio can also be helpful to other parents and people who want to have babies. This project presents a statistical analysis of my baby's daily life (including feeds, poops and sleeps) from 1st March to 30th April. I recorded 60 days of my baby's daily life.



Figure 1: My happy baby - Jacob

The dataset used has 60 rows and 13 columns. Each row corresponds to one day record. Here are some explanation of the data columns:

column[1]: day (I recorded 60 days from 1st March to 30th April)

column[2]: type means type of formula

column[3]: ml means total ml of formula he took per day

column[4]: feeds means number of feeds per day

column[5]: poops means number of poops per day

column[6]: colour means colour of poop

column[7]: shape means shape of the poop

column[8]: sleep means total sleep hours

column[9]: nightslee means sleeping hours at night (from 7 pm to 7 am)

column[10]: daysleep means sleeping hours in the daytime (from 7 am to 7 pm)

column[11]: naps: number of naps (including day and night)

column[12]: carer means the person who took care of my baby

column[13]: weather

```
babydata <- (read.table("babydata.txt", header=TRUE))
babydata
```

```
##   day    type   ml feeds poops colour shape sleep nightslee daysleep naps
## 1  1 aptamil 885     5     0     0 15.27  11.17   4.10     6
## 2  2 aptamil 910     5     2 yellow watery 13.97  10.97   3.00     7
## 3  3 aptamil 1080    6     2 yellow watery 13.83  10.08   3.75     7
## 4  4 aptamil 935     6     2 yellow normal 14.38  10.88   3.50     5
## 5  5 aptamil 960     6     1 yellow normal 14.58  11.17   3.42     6
## 6  6 aptamil 1090    6     1 yellow normal 14.28  10.72   3.57     6
## 7  7 aptamil 830     5     2 yellow normal 12.58  9.42    3.17     7
## 8  8 aptamil 920     5     1 yellow watery 13.25  9.75    3.50     6
## 9  9 aptamil 1110    6     1 yellow normal 14.48  11.00   3.48     6
## 10 10 aptamil 1085    6     1 yellow normal 14.00  10.50   3.50     4
## 11 11 aptamil 990     5     1 yellow normal 13.65  11.05   2.60     6
## 12 12 aptamil 1190    6     1 yellow watery 12.65  11.25   1.40     6
## 13 13 aptamil 1085    6     0     0     0 13.30  10.92   2.38     7
## 14 14 aptamil 830     6     2 yellow normal 14.07  11.00   3.07     6
## 15 15 aptamil 940     5     1 yellow normal 14.32  11.15   3.17     6
## 16 16 aptamil 910     5     1 yellow normal 12.32  8.50    3.82     7
## 17 17 aptamil 910     5     1 yellow watery 14.75  11.58   3.17     6
## 18 18 aptamil 1160    6     1 yellow watery 14.23  9.92    4.32     6
## 19 19 aptamil 930     5     1 yellow normal 15.15  11.73   3.42     7
## 20 20 aptamil 1120    6     1 yellow normal 14.15  11.00   3.15     8
## 21 21 aptamil 930     5     1 yellow normal 13.67  11.17   2.50     7
## 22 22 aptamil 1160    6     2 yellow normal 13.00  10.00   3.00     6
## 23 23 aptamil 1020    6     3 yellow watery 12.97  9.48    3.48     7
## 24 24 aptamil 940     5     1 yellow watery 14.90  10.67   4.23     6
## 25 25 aptamil 1180    6     1 yellow watery 12.50  8.50    4.00     7
## 26 26 aptamil 1010    5     2 yellow watery 13.13  10.00   3.13     8
## 27 27 aptamil 1100    6     1 yellow watery 13.83  11.25   2.58     7
## 28 28 aptamil 1145    7     0     0     0 14.20  10.83   3.37     7
## 29 29 aptamil 1070    6     1 yellow watery 13.05  8.45    4.60     7
## 30 30 aptamil 1050    6     2 yellow watery 12.23  8.50    3.73     7
## 31 31 neocate 710     5     1 green normal 15.40  10.00   5.40     5
## 32 32 neocate 660     5     1 green watery 13.82  11.08   2.73     8
## 33 33 neocate 780     6     2 green watery 13.23  10.75   2.48     5
## 34 34 neocate 750     6     1 yellow normal 14.42  11.17   3.25     5
## 35 35 neocate 770     5     1 yellow watery 14.75  11.25   3.50     6
## 36 36 neocate 990     8     1 green watery 12.00  8.75    3.25     9
## 37 37 neocate 840     6     3 green watery 14.33  10.25   4.08     6
## 38 38 neocate 910     7     2 yellow watery 11.33  9.00    2.33     8
## 39 39 neocate 700     5     2 yellow watery 14.37  11.50   2.87     6
## 40 40 neocate 830     7     0     0     0 12.78  10.78   2.00     8
## 41 41 neocate 850     6     1 green watery 12.83  10.50   2.33     7
## 42 42 neocate 850     6     1 yellow watery 14.83  11.00   3.83     6
## 43 43 neocate 850     6     1 yellow watery 12.77  10.10   2.67     7
```

```

## 44 44 neocate 900    6    2 green normal 13.35    9.00    4.35    8
## 45 45 neocate 910    6    2 green normal 14.98   11.42    3.57    5
## 46 46 neocate 960    6    4 yellow watery 13.33   8.33    5.00    8
## 47 47 neocate 1060   7    1 yellow watery 14.37   9.92    4.45    8
## 48 48 neocate 910    7    2 yellow watery 11.68   8.13    3.55    6
## 49 49 neocate 940    6    1 yellow normal 14.37  11.12    3.25    6
## 50 50 neocate 890    6    2 yellow watery 13.80  11.00    2.80    6
## 51 51 neocate 950    6    2 yellow normal 12.25   9.35    2.90    6
## 52 52 neocate 980    6    1 yellow watery 14.42  10.75    3.67    6
## 53 53 neocate 980    6    0      0      0 14.08  10.67    3.42    5
## 54 54 neocate 1130   6    2 green watery 14.15  10.75    3.40    6
## 55 55 neocate 990    5    1 yellow watery 13.18  10.85    2.33    5
## 56 56 neocate 1000   5    3 yellow watery 15.58  13.50    2.08    4
## 57 57 neocate 1010   5    1 yellow watery 12.58   9.75    2.83    6
## 58 58 neocate 1180   6    1 yellow normal 13.32   8.65    4.67    7
## 59 59 neocate 900    5    2 green watery 14.67  10.25    4.42    4
## 60 60 neocate 900    5    1      0      0 13.00  10.00    3.00    5
##           carer weather
## 1       me   rainy
## 2       me   rainy
## 3       me   rainy
## 4   husband   fine
## 5   husband   fine
## 6       me   rainy
## 7       me   fine
## 8       me   fine
## 9   husband   fine
## 10      me   fine
## 11  husband   rainy
## 12  husband   rainy
## 13  husband   fine
## 14  husband   fine
## 15      me   fine
## 16      me   fine
## 17  husband   fine
## 18  husband   fine
## 19      me   fine
## 20  husband   fine
## 21      me   fine
## 22  husband   fine
## 23      me   fine
## 24      me   fine
## 25  husband   fine
## 26      me   fine
## 27  husband   fine
## 28      me   fine
## 29  husband   rainy
## 30      me   rainy
## 31      me   fine
## 32      me   rainy
## 33  husband   rainy
## 34  husband   fine
## 35  husband   rainy
## 36  husband   rainy

```

```

## 37      me   fine
## 38 husband  fine
## 39 husband rainy
## 40 husband  fine
## 41 husband  fine
## 42 husband rainy
## 43      me   fine
## 44 husband  fine
## 45      me   rainy
## 46 husband  fine
## 47      me   fine
## 48      me   fine
## 49 husband rainy
## 50 husband  fine
## 51      me   fine
## 52 husband  fine
## 53      me   fine
## 54 husband  fine
## 55 husband  fine
## 56      me   fine
## 57      me   fine
## 58      me   fine
## 59 husband  fine
## 60      me   rainy

```

### First Analysis-t.test

My baby was taking aptamil in March and I changed his formula to neocate in April. The first analysis I do is to find out whether he is in favour of any type of formula.

The subset data below is the ml he took when he was taking aptamil and ml he took when he was taking neocate.

```

formula_aptamil <- babydata[1:30,3]
formula_aptamil

```

```

## [1] 885 910 1080 935 960 1090 830 920 1110 1085 990 1190 1085 830 940
## [16] 910 910 1160 930 1120 930 1160 1020 940 1180 1010 1100 1145 1070 1050

```

```

formula_neocate <- babydata[31:60,3]
formula_neocate

```

```

## [1] 710 660 780 750 770 990 840 910 700 830 850 850 850 900 910
## [16] 960 1060 910 940 890 950 980 980 1130 990 1000 1010 1180 900 900

```

Null hypothesis: There is no difference in ml before and after changing formula

```
t.test(formula_aptamil,formula_neocate)
```

```

##
## Welch Two Sample t-test
##
```

```

## data: formula_aptamil and formula_neocate
## t = 3.8528, df = 57.21, p-value = 0.0002977
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##      54.35379 171.97954
## sample estimates:
## mean of x mean of y
## 1015.8333 902.6667

```

p-value is less than 5%. So there is strong evidence that there is a difference in total ml taken by my baby in these two formula.

### Second Analysis-Fisher's Exact Test

Second analysis: I will classify my baby's poop into yellow, green, and normal or watery. I want to know whether there is a colour difference in shape.

Here is the data:

```
head(babydata)
```

```

##   day    type    ml feeds poops colour  shape sleep nightslee daysleep naps
## 1  1 aptamil 885     5     0     0 15.27 11.17    4.10     6
## 2  2 aptamil 910     5     2 yellow watery 13.97 10.97    3.00     7
## 3  3 aptamil 1080    6     2 yellow watery 13.83 10.08    3.75     7
## 4  4 aptamil 935     6     2 yellow normal 14.38 10.88    3.50     5
## 5  5 aptamil 960     6     1 yellow normal 14.58 11.17    3.42     6
## 6  6 aptamil 1090    6     1 yellow normal 14.28 10.72    3.57     6
##   carer weather
## 1      me   rainy
## 2      me   rainy
## 3      me   rainy
## 4 husband   fine
## 5 husband   fine
## 6      me   rainy

```

```
contingency <- table(babydata[, "colour"], babydata[, "shape"])
contingency
```

```

##
##          0 normal watery
## 0       6     0     0
## green   0     3     7
## yellow  0    18    26

```

Deleting the records where there is no poop for the day

```

M <- matrix(c(3,18,7,26),2,2)

dimnames(M) <- list(colour=c("yellow","green"),shape=c("normal","watery"))
M

```

```

##      shape
## colour normal watery
##   yellow      3      7
##   green       18     26

```

As I am not sure whether yellow poop will have more normal shape or watery shape, this is a two-sided test.

Null Hypothesis: Shape is independent of colour.

```
fisher.test(M)
```

```

##
## Fisher's Exact Test for Count Data
##
## data: M
## p-value = 0.723
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.09185349 3.20983586
## sample estimates:
## odds ratio
## 0.6243692

```

pvalue is greater 5%. So there is insufficient evidence to reject the null and it is reasonable to suppose that shape is independent of colour.

### Third Analysis-Fisher's Exact Test

Third analysis: I am interested in knowing whether number of poops differ by the amount of ml my baby takes.

```

a <- sum(as.numeric(babydata[, "ml"])>=900 & as.numeric(babydata[, "poops"])>=2)
b <- sum(as.numeric(babydata[, "ml"])<900 & as.numeric(babydata[, "poops"])>=2)
c <- sum(as.numeric(babydata[, "ml"])>=900 & as.numeric(babydata[, "poops"])<2)
d <- sum(as.numeric(babydata[, "ml"])<900 & as.numeric(babydata[, "poops"])<2)
M <- matrix(c(a,b,c,d), 2, 2)
dimnames(M) <- list(ml=c(">=900", "<900"), poops=c(">=2", "<2"))
M

```

```

##      poops
## ml    >=2 <2
##   >=900 16 29
##   <900  6  9

```

I might expect eating more is more likely to poop more. Therefore, it is a one-sided test.

Null Hypothesis: the amount of milk taken does not differ the number of poops

```
fisher.test(M, alternative = "greater")
```

```

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

```

```

## 
## data: M
## p-value = 0.7343
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
## 0.2604627      Inf
## sample estimates:
## odds ratio
## 0.8302435

```

pvalue is greater than 5%. So there is insufficient evidence to reject the null and it is reasonable to suppose that the amount of milk taken does not differ the number of poops.

#### Fourth Analysis-Fisher's Exact Test

I worry about the shape of poops differs by the type of formula taken. I decide to investigate whether it is or not by using another fisher.test. My hypothesis is that the shape of poops is independent of type of formula taken.

```

contingency <- table(babydata[, "type"], babydata[, "shape"])
contingency

```

```

## 
##          0 normal watery
## aptamil   3     14     13
## neocate   3      7     20

```

Deleting the records where there is no poop for the day

```

M <- matrix(c(14, 7, 13, 20), 2, 2)

dimnames(M) <- list(formula=c("aptamil", "neocate"), shape=c("normal", "watery"))
M

```

```

##           shape
## formula  normal watery
## aptamil    14     13
## neocate     7     20

```

This is a two-sided test.

```

fisher.test(M)

## 
## Fisher's Exact Test for Count Data
## 
## data: M
## p-value = 0.09291
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.8579746 11.4612496
## sample estimates:
## odds ratio
## 3.010775

```

pvalue is greater than 5%. So there is insufficient evidence to reject the null and it is reasonable to suppose that the type of formula does not differ the shape of poops.

### Fifth Analysis-Fisher's Exact Test

Next Analysis: I am interested in whether the weather will differ the amount of ml taken. I expect my baby is more likely to take more milk if the weather is nicer. I will use a fisher.test to perform the analysis.

```
a <- sum(as.numeric(babydata[, "ml"])>=900 & babydata[, "weather"]=="fine")
b <- sum(as.numeric(babydata[, "ml"])<900 & babydata[, "weather"]=="fine")
c <- sum(as.numeric(babydata[, "ml"])>=900 & babydata[, "weather"]=="rainy")
d <- sum(as.numeric(babydata[, "ml"])<900 & babydata[, "weather"]=="rainy")

M <- matrix(c(a,b,c,d),2,2)
dimnames(M) <- list(ml=c(">=900", "<900"), weather=c("fine", "rainy"))
M
```

```
##           weather
## ml      fine  rainy
##   >=900    34    11
##   <900     9     6
```

Null Hypothesis: weather does not differ the amount of milk taken

```
fisher.test(M, alternative = "greater")

##
##  Fisher's Exact Test for Count Data
##
## data:  M
## p-value = 0.2021
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.5936296      Inf
## sample estimates:
## odds ratio
##  2.033865
```

pvalue is greater than 5%. So there is insufficient evidence to reject the null and it is reasonable to suppose that weather does not differ the amount of milk taken.

### Sixth Analysis-Pearson's chi-square Test

I am now test whether My husband and I are equally sharing the responsibility taking care of our baby. I will test it using chi-square:

```
sum(babydata[, "carer"]=="me")
```

```
## [1] 29
```

```
sum(babydata[, "carer"]=="husband")
```

```
## [1] 31
```

```
o <- c(me=29, husband=31)
o
```

```
##      me husband
##      29      31
```

The first step is to form a null hypothesis  $H_0$ :  $\text{Prob(me)} = \text{Prob(husband)} = 1/2$ , where  $\text{Prob(me)}$  is the probability of me taking care of our baby.  $\text{Prob(husband)}$  is the probability of my husband taking care of our baby. If the null is true, we would expect to observe that each of us spent 30 nights with our baby for the last 60 days.

```
e <- c(me=30, husband=30)
e
```

```
##      me husband
##      30      30
```

The chi-square statistic B:

```
B <- sum((o-e)^2/e)
B
```

```
## [1] 0.06666667
```

After that, the p-value is:

```
pchisq(B,df=2-1,lower.tail = FALSE)
```

```
## [1] 0.7962534
```

The pvalue is greater than 5%, so we fail to reject the null and it is reasonable to suppose that each of us has the same probability of taking care of our baby.

## Seventh Analysis-Pearson's chi-square Test

The number of poops dropped per day seems to follow Poisson distribution. I would do a test on whether the number of poops is Poisson. My null hypothesis is that the number of poops in a day follows Poisson distribution. The alternative hypothesis is that it does not follow Poisson distribution.

```
o <- table(as.numeric(babydata[, "poops"]))
o
```

```
##
##   0   1   2   3   4
##   5 33 18  3  1
```

Thus, we have 5 days with zero poops, 33 days with 1 poop, 18 days with 2 poops, 3 days with 3 poops and 1 day with 4 poops. The mean number of poops is:

```
mean(as.numeric(babydata[, "poops"]))
```

```
## [1] 1.366667
```

Thus, we can model the number of poops dropped per day as a Poisson distribution with lambda = 1.366667. I will classify day as having 0,1,2,>=3 poops. The probabilities of 0-2 poops is

```
dpois(0:2,lambda = mean(as.numeric(babydata[, "poops"])))
```

```
## [1] 0.2549554 0.3484390 0.2381000
```

The probability of having 3 or more will be

```
1-sum(dpois(0:2,lambda = mean(as.numeric(babydata[, "poops"]))))
```

```
## [1] 0.1585056
```

Therefore, the probabilities of 0,1,2,>=3 poops is

```
probs <- c(dpois(0:2,lambda = mean(as.numeric(babydata[, "poops"]))),  
           1-sum(dpois(0:2,lambda = mean(as.numeric(babydata[, "poops"])))))  
probs
```

```
## [1] 0.2549554 0.3484390 0.2381000 0.1585056
```

To verify:

```
sum(probs)
```

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

The expected number of day with 0,1,2,>=3 poops is:

```
e <- length(as.numeric(babydata[, "poops"]))*probs  
e
```

```
## [1] 15.297324 20.906342 14.286001 9.510333
```

The chi-square test is:

```
o
```

```
##  
## 0 1 2 3 4  
## 5 33 18 3 1
```

```
o <- c("0"=5,"1"=33,"2"=18,">=3"=4)
o
```

```
##    0    1    2 >=3
##    5   33   18    4
```

```
B <- sum((o-e)^2/e)
B
```

```
## [1] 18.08565
```

```
pchisq(B,df=4-2,lower.tail = FALSE)
```

```
## [1] 0.0001182361
```

pvalue is less than 5%. We can reject the null. There is strong evidence that the observations(number of poops) are not consistent with a Poisson distribution.

### Another Analysis-Linear Regression on sleeps



Figure 2: My baby is sleeping.

As parents, it is very important to know their babies' routines and getting to know whether they are having a good sleeping routine. I found some healthlines mention that as the days go by, babies are more likely to have less amount of daytime sleep. I am interested in whether my baby's daytime sleep is dependent on the day.

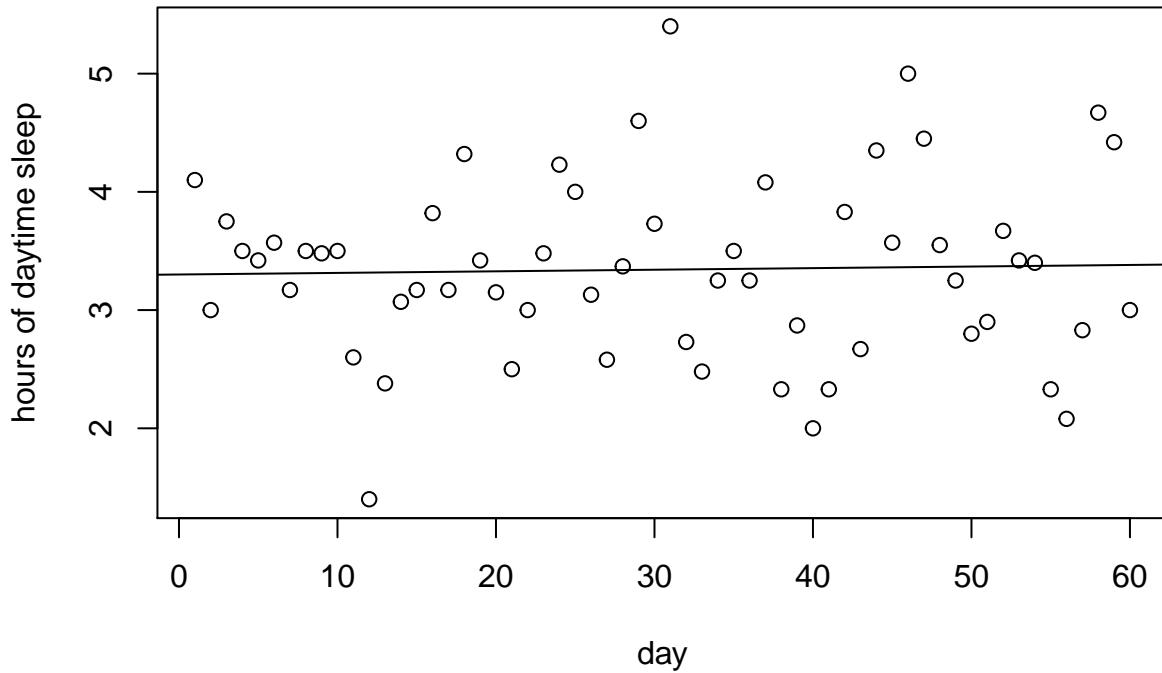
My Null Hypothesis \$H\_0\$: \$ My baby's daytime sleep is independent of the day

```
day <- as.numeric(babydata[, "day"])

daysleep <- as.numeric(babydata[, "daysleep"])

plot(daysleep~day,xlab="day",ylab = "hours of daytime sleep")

abline(lm(daysleep~day))
```



Is the regression line significant?

```
summary(lm(daysleep~day))
```

```
##
## Call:
## lm(formula = daysleep ~ day)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.9172 -0.4995  0.0289  0.4027  2.0573
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.301051  0.201262 16.402 <2e-16 ***
## day         0.001343  0.005738  0.234    0.816
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7698 on 58 degrees of freedom
## Multiple R-squared:  0.000943, Adjusted R-squared: -0.01628
## F-statistic: 0.05474 on 1 and 58 DF, p-value: 0.8158
```

pvalue is greater than 5%, so we fail to reject the null. So, there is no strong evidence that my baby's daytime sleep hours are dependent on the day.

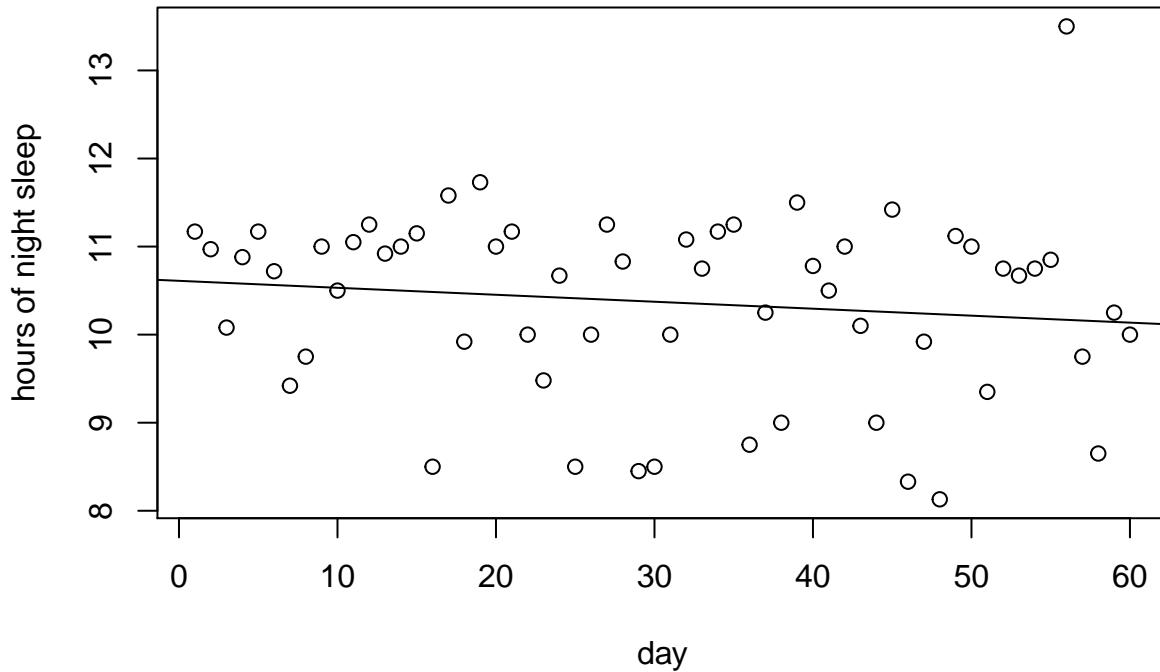
I want to do another analysis on the nightsleep. My Null Hypothesis \$H\_0\$: My baby's night sleep is independent of the day

```
day <- as.numeric(babydata[, "day"])

nightsleep <- as.numeric(babydata[, "nightsleep"])

plot(nightsleep-day, xlab="day", ylab = "hours of night sleep")

abline(lm(nightsleep~day))
```



Is the regression line significant?

```
summary(lm(nightsleep~day))

##
## Call:
## lm(formula = nightsleep ~ day)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1012 -0.5176  0.3377  0.6618  3.3321
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```

## (Intercept) 10.611006   0.272510  38.938   <2e-16 ***
## day        -0.007913   0.007770  -1.018    0.313
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 1.042 on 58 degrees of freedom
## Multiple R-squared:  0.01757, Adjusted R-squared:  0.0006298
## F-statistic: 1.037 on 1 and 58 DF, p-value: 0.3127

```

pvalue is greater than 5%, so we fail to reject the null. So, there is no strong evidence that my baby's night time sleep hours are dependent on the day.

I will do another linear regression on the total sleep hours of a function of day My Null Hypothesis \$H\_0\$: \$ My baby's total sleep is independent of the day

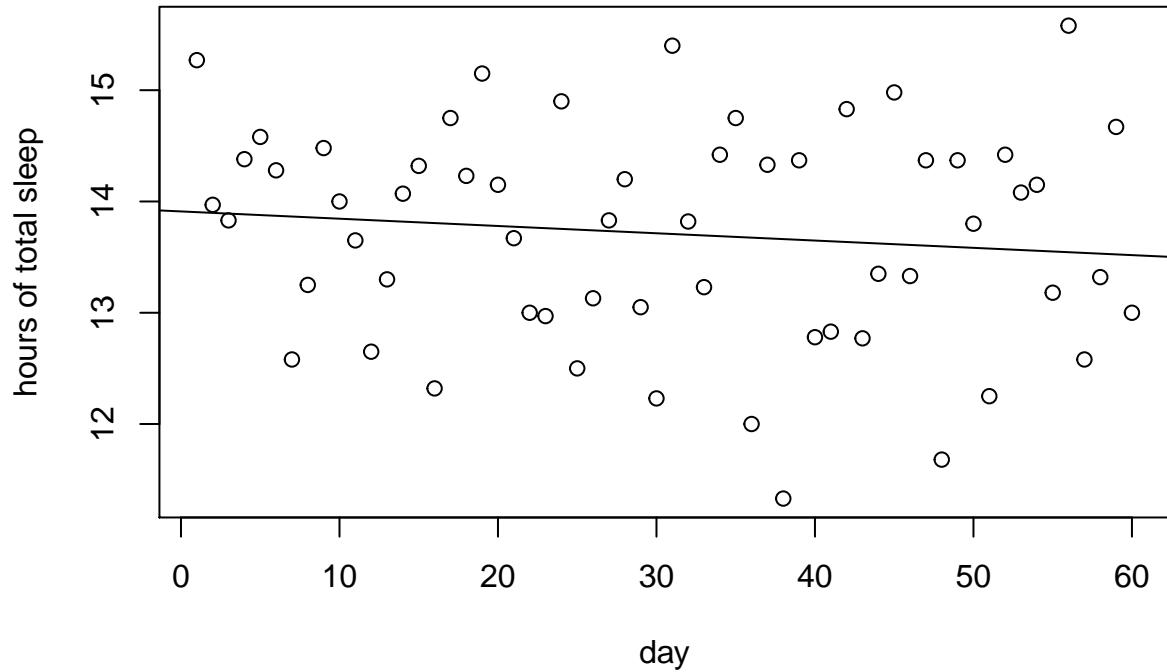
```

day <- as.numeric(babydata[, "day"])

totalsleep <- as.numeric(babydata[, "sleep"])

plot(totalsleep~day,xlab="day",ylab = "hours of total sleep")
abline(lm(totalsleep~day))

```



Is the regression line significant?

```

summary(lm(totalsleep~day))

##
## Call:
## lm(formula = totalsleep ~ day)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -2.3319 -0.6948  0.1075  0.7052  2.0359 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 13.910661  0.252972 54.989 <2e-16 ***
## day        -0.006546  0.007213 -0.908   0.368    
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 0.9675 on 58 degrees of freedom
## Multiple R-squared:  0.014, Adjusted R-squared:  -0.002996 
## F-statistic: 0.8238 on 1 and 58 DF, p-value: 0.3678

```

pvalue is greater than 5%, so we fail to reject the null. The coefficients are not statistically significant. So, there is no strong evidence that there is a change in my baby's total sleep hours with day.

### Another Analysis-Linear Regression on feeds

I would like to know whether my baby's total amount of milk increases as the day goes by. Because I changed his formula in day 31, I will only take day 1-30 for first analysis to rule out the cause of changing formula.

```

ml_aptamil <- as.numeric(babydata[1:30,3])

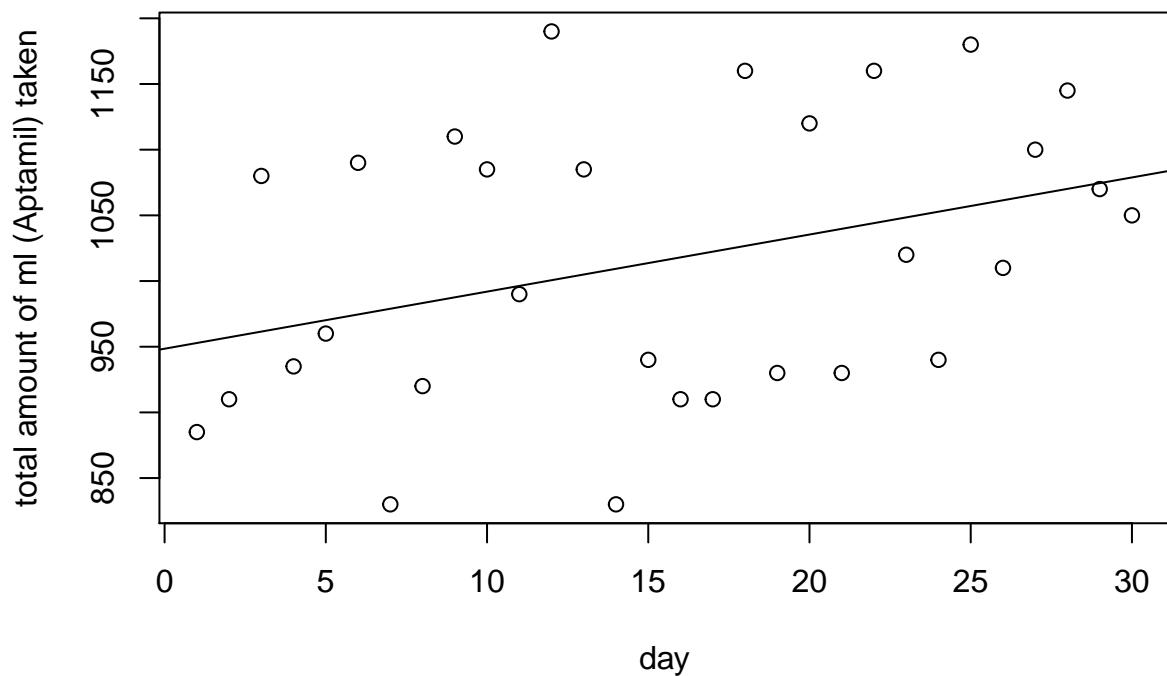
day <- seq_along(ml_aptamil)

plot(ml_aptamil~day,xlab="day",ylab = "total amount of ml (Aptamil) taken")
abline(lm(ml_aptamil~day))

```



Figure 3: My baby is drinking milk-Jacob



Is the regression line significant?

```
summary(lm(ml_aptamil~day))

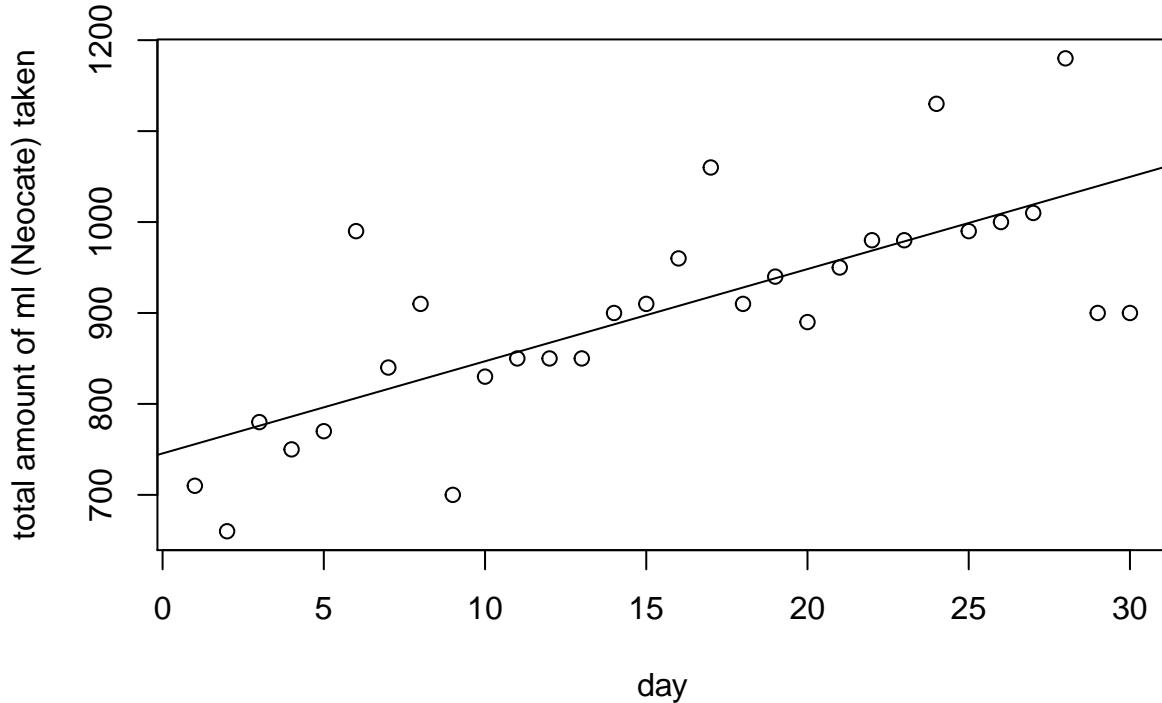
##
## Call:
## lm(formula = ml_aptamil ~ day)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -179.31   -72.20   -19.31    90.95   189.38 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 948.471    38.028  24.941 <2e-16 ***
## day          4.346     2.142   2.029   0.0521 .  
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
##
## Residual standard error: 101.6 on 28 degrees of freedom
## Multiple R-squared:  0.1282, Adjusted R-squared:  0.09703 
## F-statistic: 4.116 on 1 and 28 DF,  p-value: 0.05208
```

The two-sided pvalue is not significant. However, the one-sided value is. In this case, the one-sided test is appropriate because it is reasonable to assume that babies take more food as they grow bigger. Therefore, we can reject the null and conclude that there is evidence that the total amount of milk taken is dependent on the day.

```
ml_neocate <- as.numeric(babydata[31:60,3])

day <- seq_along(ml_neocate)

plot(ml_neocate~day,xlab="day",ylab = "total amount of ml (Neocate) taken")
abline(lm(ml_neocate~day))
```



Is the regression line significant?

```
summary(lm(ml_neocate~day))
```

```
##
## Call:
## lm(formula = ml_neocate ~ day)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -149.634  -27.056   -8.684  12.503  183.623
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 745.563    30.725  24.266 < 2e-16 ***
## day         10.136     1.731   5.856 2.69e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 82.05 on 28 degrees of freedom
## Multiple R-squared:  0.5505, Adjusted R-squared:  0.5345
## F-statistic: 34.3 on 1 and 28 DF, p-value: 2.695e-06
```

pvalue is less than 5% which is significant. Therefore, we can reject the null and conclude that there is evidence that there is a positive change in the total amount of milk taken with time.

## Another Analysis-Logistic Regression on milk taken

I am seek to estimate the colour of poops as a function of the amount of milk taken. I am wondering whether the colour is dependent of the amount of milk taken. My null hypothesis is that the colour of poops is independent of the amount of milk taken per day (ml).

```
colour <- babydata$colour

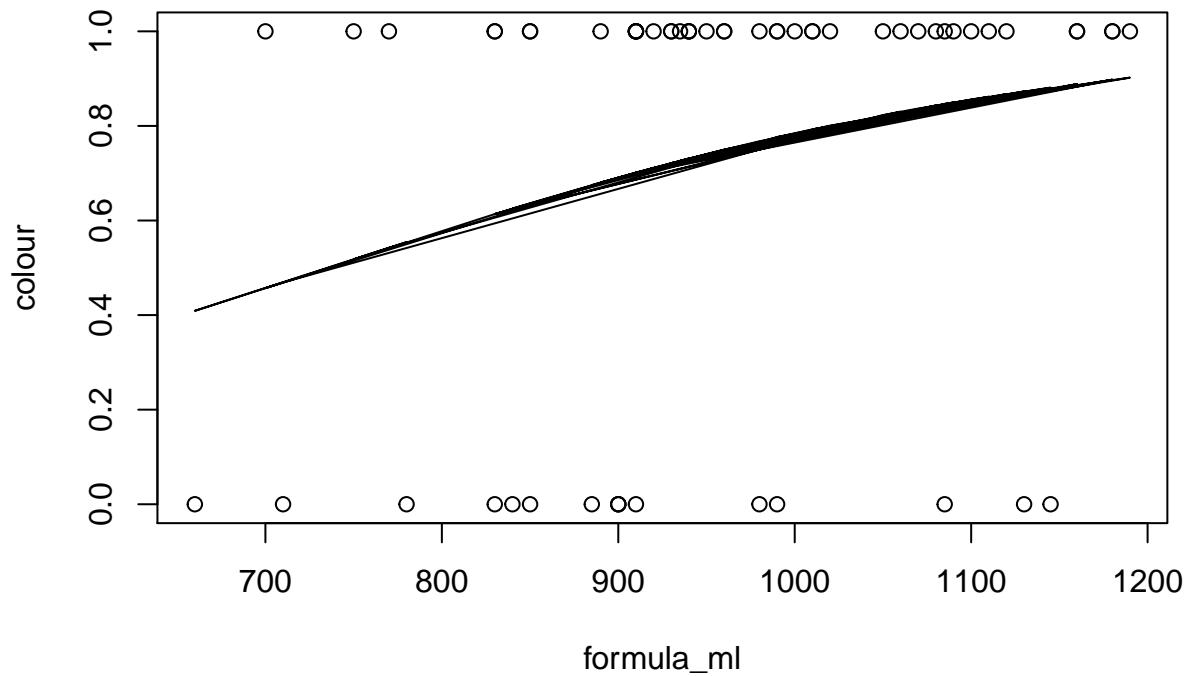
for(i in 1:60){
  if(colour[i]=="yellow"){
    colour[i] <- 1
  }else{
    colour[i] <- 0
  }
}
colour <- as.numeric(colour)

formula_ml <- babydata[1:60,3]

L0 <- function(p)(log(p/(1-p)))
pr <- function(L0){exp(L0)/(1+exp(L0))}
summary(glm(colour~formula_ml,family = "binomial"))

##
## Call:
## glm(formula = colour ~ formula_ml, family = "binomial")
##
## Deviance Residuals:
##      Min        1Q     Median        3Q       Max 
## -2.0631   -1.0506    0.6454    0.8083    1.2515 
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept) -3.592299  2.382615 -1.508   0.1316    
## formula_ml  0.004885  0.002549  1.917   0.0552 .  
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 69.590  on 59  degrees of freedom
## Residual deviance: 65.547  on 58  degrees of freedom
## AIC: 69.547
##
## Number of Fisher Scoring iterations: 4

plot(colour~formula_ml)
points(formula_ml,pr(-3.592299+0.004885*formula_ml),type = "l")
```



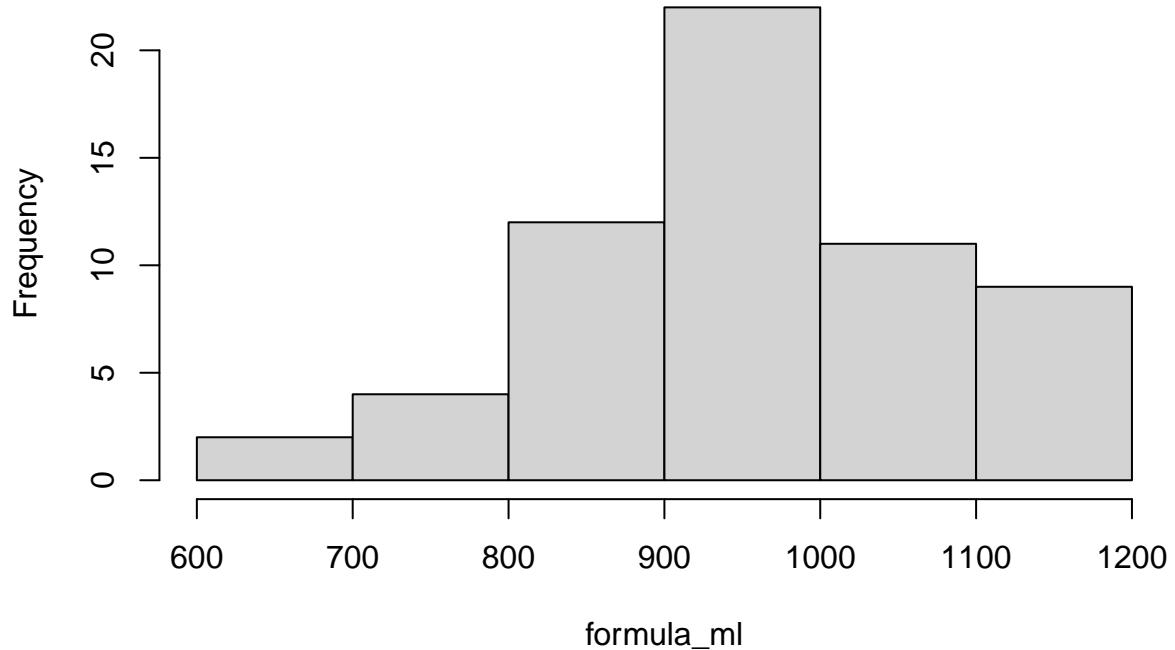
The pvalue is 0.0552 which is greater than 5%. It is not significant. There is no evidence that colour of poops is dependent on the amount of milk taken per day (ml).

### Another Analysis

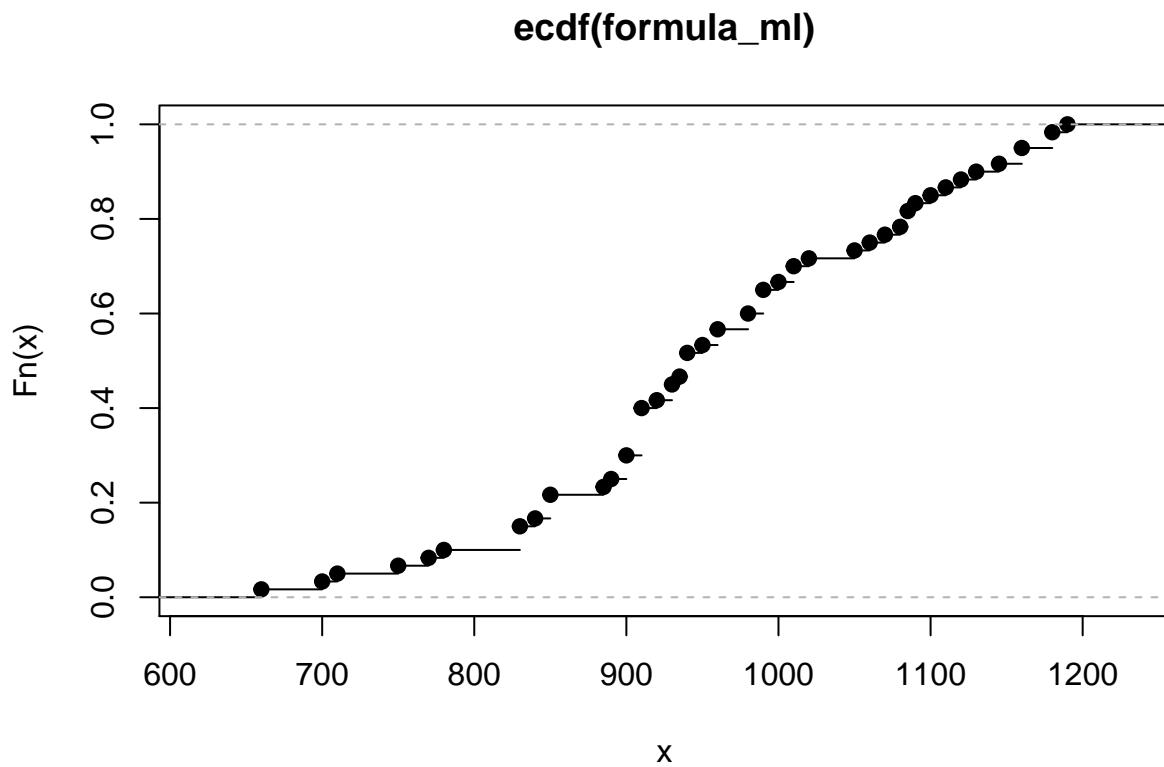
I would like to investigate whether the total amount of milk taken is Gaussian. My null hypothesis is that the total amount of milk taken (ml) is Gaussian. I will use quantile methods to investigate.

```
formula_ml <- babydata$ml
hist(formula_ml)
```

**Histogram of formula\_ml**



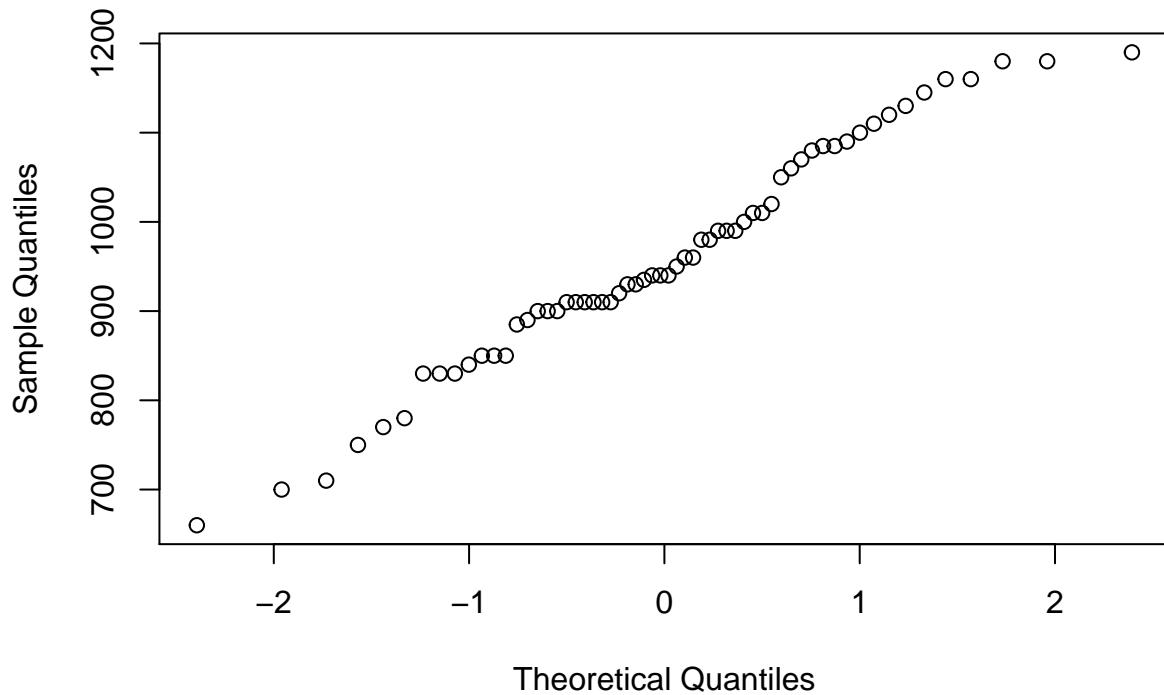
```
plot(ecdf(formula_ml))
```



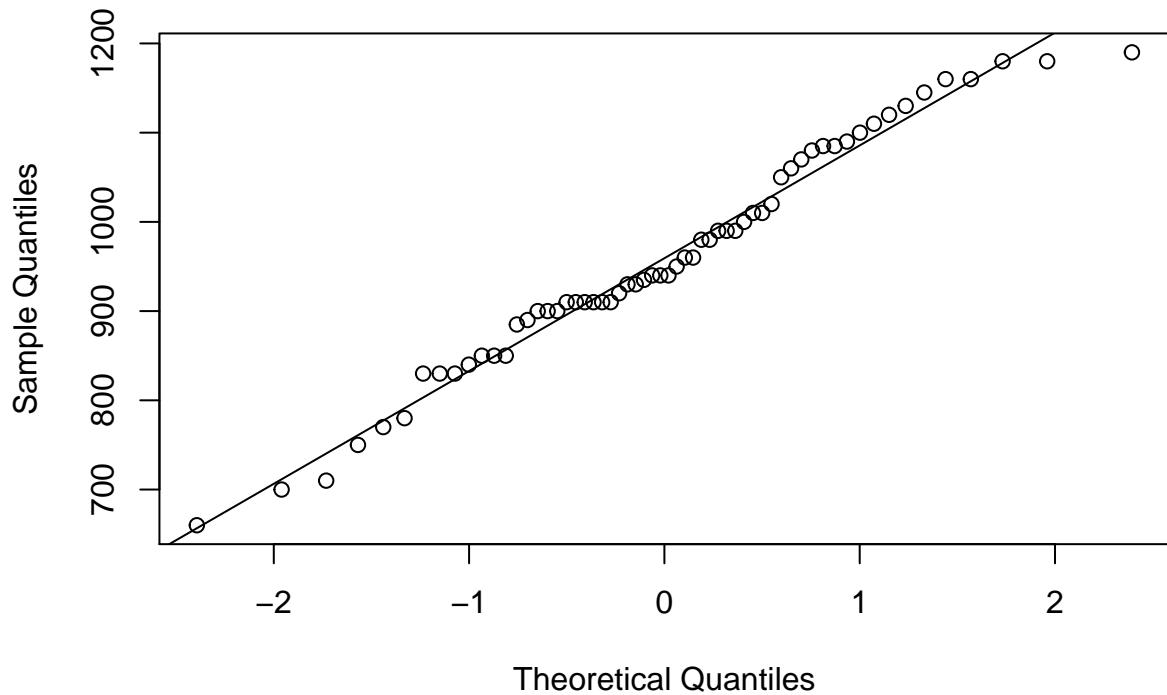
As we can see from the histogram, the data seems to follow Gaussian distribution. And then I conduct the Empirical Cumulative Distribution Function. I will perform `qqnorm()` to detect non-normality.

```
plot(qqnorm(formula_ml), ylab = "Sample Quantiles", xlab = "Theoretical Quantiles")
```

### Normal Q-Q Plot



```
abline(mean(formula_m1),sd(formula_m1))
```



As we can see from the diagram above, the data fall approximately on a straight line. It means that it is normally distributed. I will perform a Shapiro-Wilk test to detect whether the data is normally distributed as well.

```
shapiro.test(formula_ml)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data: formula_ml  
## W = 0.97744, p-value = 0.33
```

From the Shapiro.test above, the pvalue is 0.33 which is greater than 5%. We can not reject the null and can conclude that it is reasonably that the data-the total amount of milk taken is Gaussian distribution.

### Another Analysis

I would like to investigate whether there is any significant difference between the ml of milk taken while I am the carer and the ml of milk taken while my husband is the carer.

My null hypothesis: There is no difference between the two data sets, the total ml of milk taken while I am the carer and total ml of milk taken while my husband is the carer.

Let's investigate:

To get the total amount of ml taken while Jacob is taken care by me:

```

carer_me_ml <- babydata[which(babydata$carer=='me'),]
carer_me_ml <- carer_me_ml$ml
carer_me_ml

## [1] 885 910 1080 1090 830 920 1085 940 910 930 930 1020 940 1010 1145
## [16] 1050 710 660 840 850 910 1060 910 950 980 1000 1010 1180 900

```

To get the total amount of ml taken while Jacob is taken care by my husband:

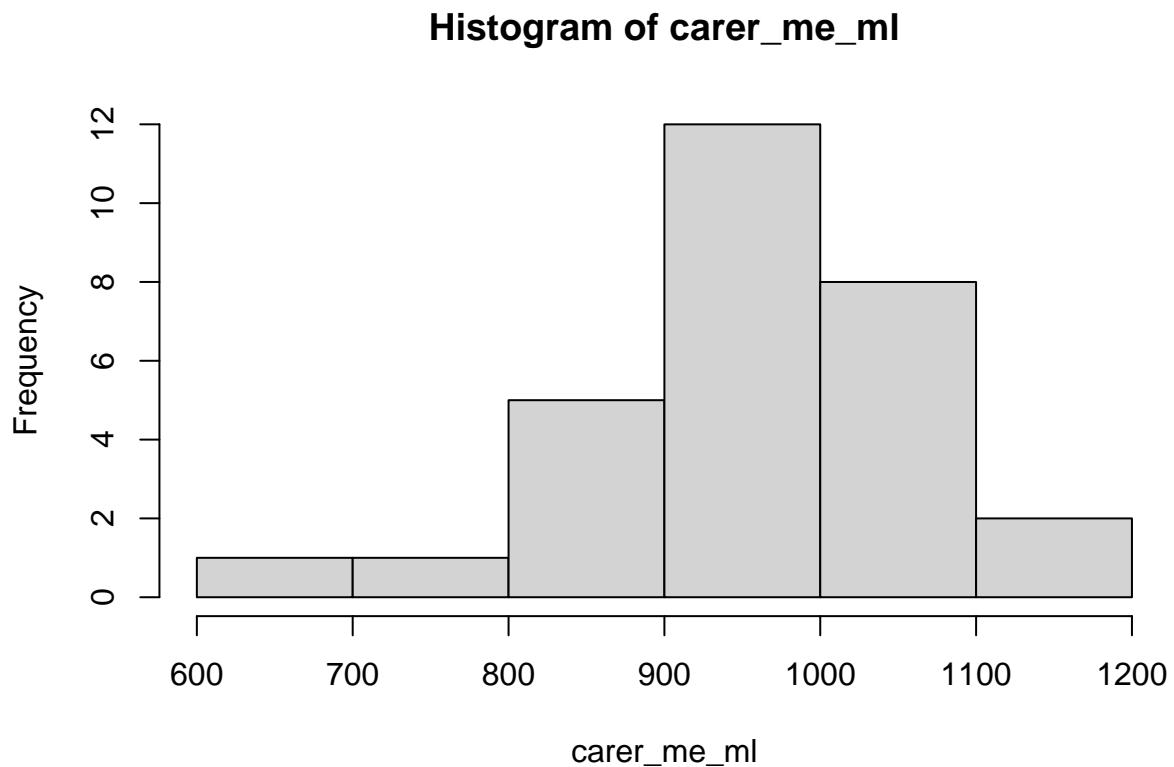
```

carer_husband_ml <- babydata[which(babydata$carer=='husband'),]
carer_husband_ml <- carer_husband_ml$ml
carer_husband_ml

## [1] 935 960 1110 990 1190 1085 830 910 1160 1120 1160 1180 1100 1070 780
## [16] 750 770 990 910 700 830 850 850 900 960 940 890 980 1130 990
## [31] 900

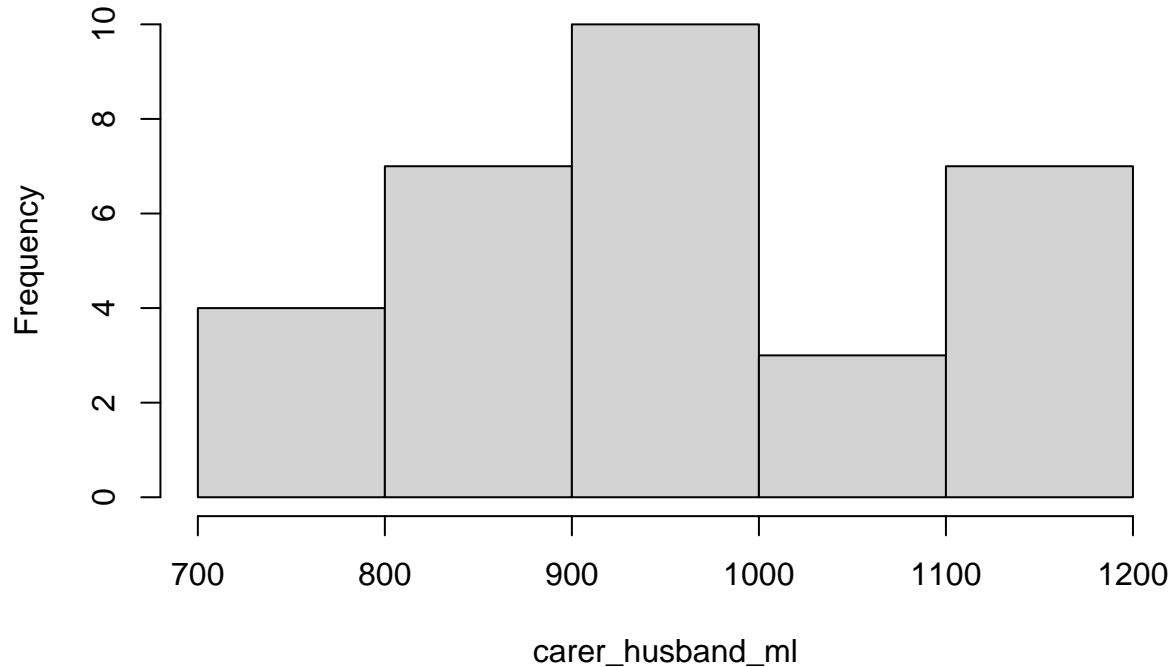
hist(carer_me_ml)

```



```
hist(carer_husband_ml)
```

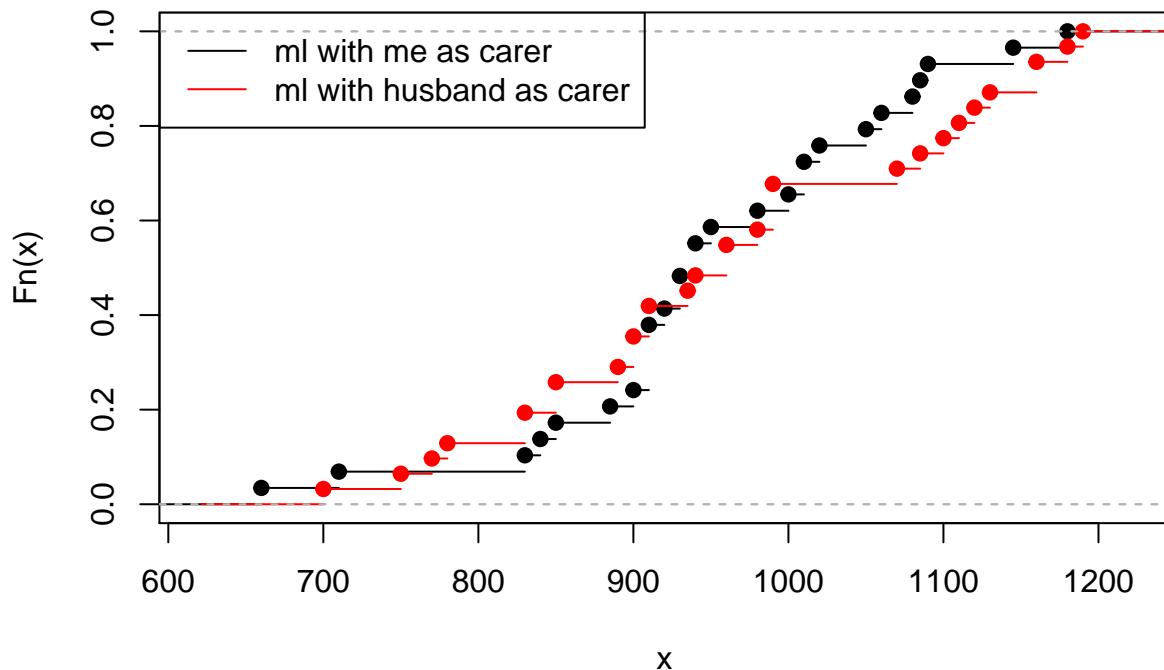
## Histogram of carer\_husband\_ml



To plot their empirical cumulative distribution function on the same axes:

```
plot(ecdf(carer_me_ml))
plot(ecdf(carer_husband_ml), add=TRUE, col='red')
legend("topleft", col=c("black", "red"), legend=c("ml with me as carer", "ml with husband as carer"), lty=1)
```

### ecdf(carer\_me\_ml)



The two ECDFs show a difference. We will use Kolmogorov Smirnov test to figure out whether there is statistically significant.

```
ks.test(carer_me_ml,carer_husband_ml)

## Warning in ks.test(carer_me_ml, carer_husband_ml): cannot compute exact p-value
## with ties

##
## Two-sample Kolmogorov-Smirnov test
##
## data: carer_me_ml and carer_husband_ml
## D = 0.1891, p-value = 0.6576
## alternative hypothesis: two-sided
```

pvalue is 0.6576 which is greater than 5%. The difference is not statistically significant. We fail to reject the null and can conclude that the two data sets are drawn from same distribution.

The Student t-test is also conducted:

```
t.test(carer_me_ml,carer_husband_ml)

##
## Welch Two Sample t-test
##
## data: carer_me_ml and carer_husband_ml
```

```

## t = -0.37389, df = 57.431, p-value = 0.7099
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -77.72273 53.26222
## sample estimates:
## mean of x mean of y
## 952.9310 965.1613

```

pvalue from the t.test is 0.7099 which is greater than 5%. The difference in means are not statistically significant. We fail to reject the null and conclude that there is no statistically difference in the means between these two data sets.

## Conclusions

The dataset of babydata were analysed using statistical methods and a number of results have been obtained. There is a statistical difference in the total of ml taken before and after changing formula by using t.test. Through fisher.test, we can reasonably conclude that the shape of my baby's poop's is independent of the colour. The number of poops does not differ by the amount of ml taken. The weather does not differ the amount of ml taken. And it is reasonable to suppose that me and my husband shares the same probability of taking care of our baby by performing Pearson's chi-square test. We can also conclude that the number of poops my baby has per day is not consistent with Poisson distribution through Pearson chi-square test. By using linear regression, there is no evidence to suggest that the day does not have a statistically significant effect on my baby's daytime, nighttime sleep, and total sleep. There is strong evidence that there is a positive change in the total amount of milk (aptamil) taken and the total amount of milk (neocate) taken with day. The total amount of milk taken (ml) follows Gaussian distribution (Shapiro-Wilk test). By conducting the Kolmogorov-Smirnov test, we can find that there is not significant difference between the ml of milk taken while I am the carer and the ml of milk taken while my husband is. The Student t-test is also used to find out that the difference in means between the ml of milk taken while I being the carer and the ml of milk taken while my husband being the carer. The result shows the difference is not statistically significant.