```{r echo = FALSE , include=FALSE}

knitr::opts\_chunk$set(echo = TRUE, warning = FALSE , message = FALSE)

library(tidyverse)

library(broom)

library(knitr)

library(brms)

library(tidybayes)

library(ggthemes)

library(coda)

```

## Introduction

In this post I will explore some basic approaches to analyze a simple simulated data example.

## Two sample continuous outcome

A depend variable Y and a indedependent factor X (two levels, A and B).

I standardize Y to get zero mean and unit standard deviation.

```{r A, include=T }

DAT <- tibble(N = 60 ,

X = rbinom(N, 1, 0.5),

Y = 0.5\*X + rnorm(N)) %>%

mutate(X = factor(X, labels = c("A", "B")) ,

Y = (Y-mean(Y, na.rm=T))/sd(Y, na.rm=T))

```

Describe the data (recall that Y has been standardized)

```{r B, include=T}

DESC <- DAT %>% group\_by(X) %>%

summarise( mean = mean(Y, na.rm=T) ,

sd = sd(Y, na.rm=T) ,

min=min(Y, na.rm=T),

max=max(Y, na.rm=T) ,

n = n()) %>%

kable(digits = 2)

DESC

```

Plot the data

```{r C, include=T}

PLT\_1 <- DAT %>% ggplot( aes(x=Y, color=X , fill=X )) +

# geom\_histogram(aes(y=..density..), position="identity", alpha=0.2 , binwidth= 0.333)+

geom\_density(alpha=0.2)+

scale\_color\_manual(values=c("red", "blue"))+

scale\_fill\_manual(values=c("red", "blue"))+

labs(x="Y", y = "Density")+

theme\_classic() +

geom\_rug() +

theme(legend.text = element\_text(size = 12, colour = "black", angle = 0)) +

theme(legend.position = c(0.8, 0.8)) +

theme(axis.text.x = element\_text(size=12), axis.text.y = element\_text(size=12))

```

```{r echo = FALSE , include=T , comment=NA}

PLT\_1

```

## Frequentist inference by a standard linear model (two-sample t-test):

```{r D, include=T}

m1 <- tidy(lm(Y~X , data = DAT), conf.int = T, conf.level = 0.95)[2,]

M <- signif(as.numeric(m1[, 2]) , digits = 2) ;

L<- signif(as.numeric(m1[, 6]) , digits = 2) ;

U <- signif(as.numeric(m1[, 7]), digits = 2) ;

UU\_ <- paste( U,"(95%CI:",L, ";",U, ")")

```

Mean difference and 95% compatibility intervals:

```{r echo = FALSE , include=T , comment=NA}

UU\_

```

## Bayesian inference:

Use the [\_rethinking\_ package](https://rdrr.io/github/rmcelreath/rethinking/)

```{r E, include=T , warnings=FALSE, results="hide" , message=FALSE, warning=FALSE, cache=0,eval=TRUE, error=FALSE, refresh =0 }

library(rstan)

library(rethinking)

options(mc.cores = parallel::detectCores())

rstan\_options(auto\_write = TRUE)

m12 <- ulam(

alist( Y ~ dnorm( mu , sigma ) ,

mu <- a + bA \* X ,

a ~ dnorm( 0 , 1 ) ,

bA ~ dnorm( 0 , 1 ) ,

sigma ~ dexp( 1 ) ) , data = DAT , chains=2, cores=1 , sample=TRUE )

m12\_ <- precis(m12,depth=1 , prob=0.95)

M <- signif(as.numeric(m12\_[2, 1]) , digits = 2) ;

L<- signif(as.numeric(m12\_[2, 3]) , digits = 2) ;

U <- signif(as.numeric(m12\_[2, 4]), digits = 2) ;

U\_ <- paste( U,"(95%CI:",L, ";",U, ")")

```

Posterior mean difference and 95% credible intervals:

```{r echo = FALSE , include=T , comment=NA}

U\_

```

Extract prior and posterior distribution and plot it.

```{r F, include=T }

post <- extract.samples(m12,n=1e4)

prior <- extract.prior(m12,n=1e4)

post\_plot <- dens(post$bA, main = "Posterior density")

prior\_plot <- dens(prior$bA , main = "Prior density (Draws from the MCMC procedure)")

```

```{r echo = FALSE , include=T , comment=NA}

prior\_plot

```

```{r echo = FALSE , include=T , comment=NA}

post\_plot

```

Put both prior and posterior densities in same graph with ggplot:

```{r G, include=T}

post\_ <- tibble(P = post$bA) %>%

mutate( type = c("Posterior") )

prior\_ <- tibble(P = prior$bA) %>%

mutate( type = c("Prior") )

por <- rbind(post\_ , prior\_)

PLT\_2 <- por %>% ggplot( aes(x=P, color=type , fill=type )) +

geom\_density(alpha=0.2)

```

```{r echo = FALSE , include=T , comment=NA}

PLT\_2

```

Above we used a normal and exponential prior for location and scale, respectively.

Let us use the [\_brms\_ package](https://cran.r-project.org/web/packages/brms/index.html).

Use default priors (what is a \_default\_ prior?)

```{r H, include=T}

library(brms)

```

```{r I, include=T , warnings=FALSE, results="hide"}

m13 <- brm(formula = Y ~ X,

data = DAT,

seed = 123)

```

Look what was the \_default\_ priors:

```{r J, include=T}

prior\_summary(m13)

```

Use same priors as was used in the first model (m12):

```{r K, include=T , warnings=FALSE, results="hide"}

get\_prior(Y ~ X, data = DAT)

priors2 <- c(set\_prior("normal(0, 1)", class = "Intercept"),

set\_prior("normal(0, 1)", class = "b", coef = "XB"),

set\_prior("exponential(1)", class = "sigma"))

m14 <- brm(formula = Y ~ X,

data = DAT,

seed = 123,

prior = priors2)

```

Did they model actually use the priors I specified?

```{r L, include=T}

prior\_summary(m14)

```

```{r M, include=T}

resA <- fixef(m13)

resB <- fixef(m14)[2,]

M2 <- signif(as.numeric(fixef(m14)[2, 1]) , digits = 2) ;

L2<- signif(as.numeric(fixef(m14)[2, 3]) , digits = 2) ;

U2 <- signif(as.numeric(fixef(m14)[2, 4]), digits = 2) ;

U\_2 <- paste( M2,"(95%CI:",L2, ";",U2, ")")

```

So using the same priors, we now get:

```{r N, include=T}

U\_2

```

Lets say I have a firm prior belief that the mean difference between A and B is close to one.

Let the prior reflect this belief and then estimate the model.

```{r P, include=T , warnings=FALSE, results="hide"}

priors3 <- c(set\_prior("normal(0, 1)", class = "Intercept"),

set\_prior("normal(1, 0.25)", class = "b", coef = "XB"),

set\_prior("exponential(1)", class = "sigma"))

m15 <- brm(formula = Y ~ X,

data = DAT,

seed = 123,

prior = priors3)

```

Plotting the posterior distribution using the [\_tidybayes\_](http://mjskay.github.io/tidybayes/index.html/) package

```{r Q , include=T}

PPP <- posterior\_samples(m15) %>%

ggplot(aes(x = b\_XB , y = 0)) +

stat\_halfeye(fill = "orange2", .width = .95) +

scale\_y\_continuous(NULL, breaks = NULL) +

theme\_fivethirtyeight() +

theme(axis.text.x = element\_text( size = 15))

```

```{r echo = FALSE , include=T , comment=NA}

PPP

```

Extract the posterior distributions from the three models (default priors, mean zero prior and mean one prior).

I took the code below from [\_here\_ ](https://www.rensvandeschoot.com/tutorials/r-linear-regression-bayesian-using-brms/).

```{r R, include=T}

posterior1.2.3 <- bind\_rows("Default priors" = as\_tibble(as.mcmc(m13, pars = "XB", fixed = F ,combine\_chains = TRUE)),

"Mean Zero prior" = as\_tibble(as.mcmc(m14, pars = "XB", fixed = F ,combine\_chains = TRUE)),

"Mean one prior" = as\_tibble(as.mcmc(m15, pars = "XB", fixed = F ,combine\_chains = TRUE)),

.id = "id1")

prior1.2.3 <- bind\_rows("Default priors" = enframe(runif(10000, min = -2.1, max = 2.1)),

"Mean Zero prior" = enframe(rnorm(10000, mean=0, sd=1)),

"Mean one prior" = enframe(rnorm(10000, mean=1, sd=0.25)),

.id = "id1") %>%

rename(b\_XB = value)

priors.posterior <- bind\_rows("posterior" = posterior1.2.3, "prior" = prior1.2.3, .id = "id2")

```

Plot the data. Dotted lines are priors and solid lines are psoterior. Default flat ("improper", that is a prior that is does not fulfil the requiremtents of a probability distribution) prior and Mean zero prior give very similar posterior. For the mean one prior you see that the posterior is inbetween the mean of one and the frequentist estimate of 0.53.

```{r S , include=T}

PLT\_3 <- ggplot(data = priors.posterior,

mapping = aes(x = b\_XB,

fill = id1,

colour = id2,

linetype = id2,

alpha = id2)) +

geom\_density(size=1) +

scale\_x\_continuous(limits=c(-2.5, 2.5)) +

scale\_colour\_manual(name = 'Posterior/Prior',

values = c("black","red"),

labels = c("posterior", "prior")) +

scale\_linetype\_manual(name = 'Posterior/Prior',

values = c("solid","dotted"),

labels = c("posterior", "prior")) +

scale\_alpha\_discrete(name ='Posterior/Prior',

range = c(.7,.3),

labels = c("posterior", "prior")) +

scale\_fill\_manual(name = "Densities",

values = c("darkred","blue", "green" )) +

labs(title = expression("Influence of (Informative) Priors on" ~ beta[X]),

subtitle = "3 different densities of priors and posteriors")

```

```{r echo = FALSE , include=T , comment=NA}

PLT\_3

```