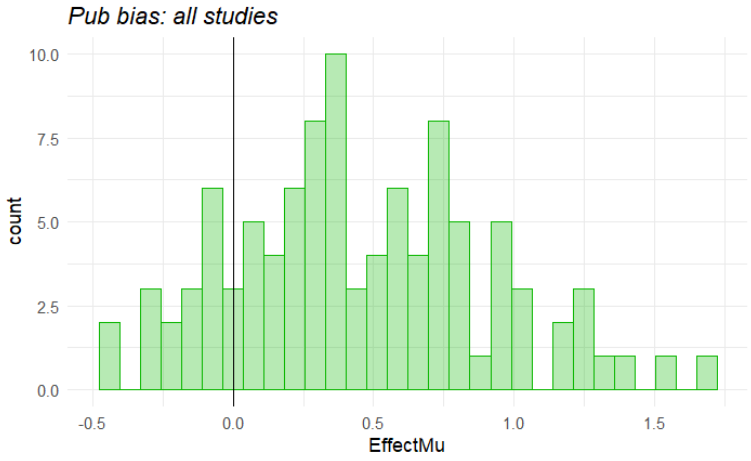
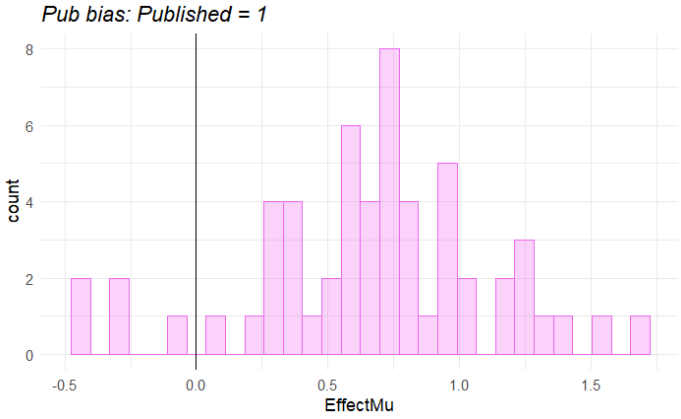
Ioana Luisa Forcas
au693527

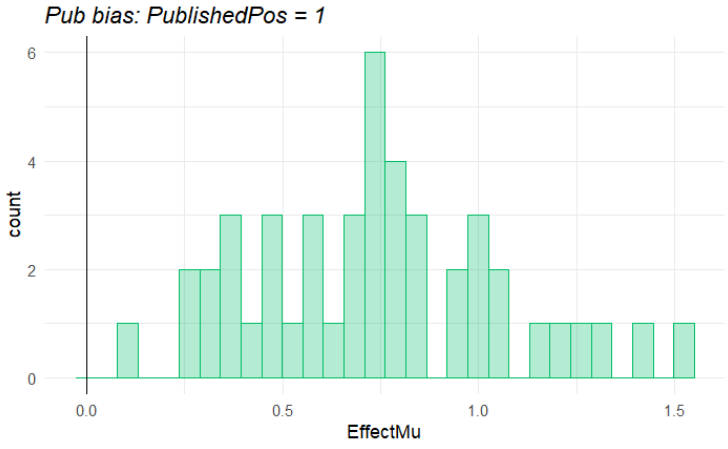
Assignment 2

Question 1

The following plots show the output of the simulated data according to the instructions. The first plot includes all the studies (effect mean size = 0), the second plot includes the published studies (the ones with significant results) and the third one is again with studies that were published (this time with both significant results and the expected outcome).







Ioana Luisa Forcas
au693527

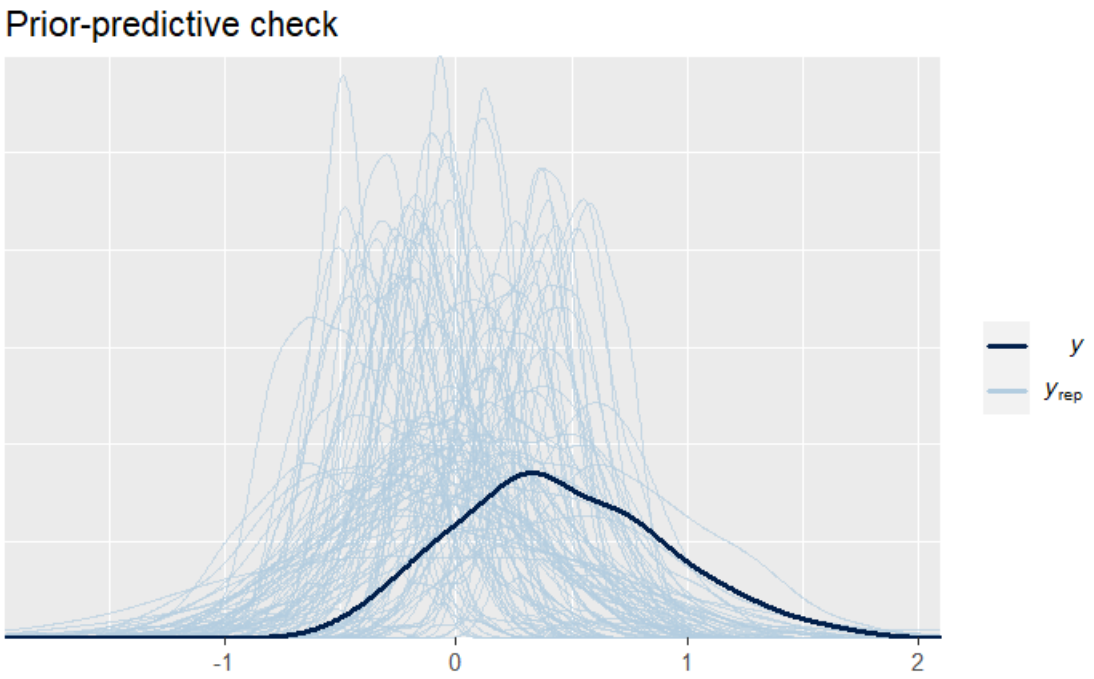
The distribution of impact sizes in the simulated data, where only the research with significant and anticipated findings are released, is shown in the last plot above.

Formula for the analysis:

f1 <- bf (EffectMu | se (EffectSigma) ~ 1 + (1 | Study))

Every study's impact size mean and standard error are different, proving that the result is actually a distribution rather than just a single point.

The plots for the prior-predictive checks:



I have not changed any of the initial priors I established because the range of result values seems to be plausible (within the order of magnitude). The model was then created for the real data:

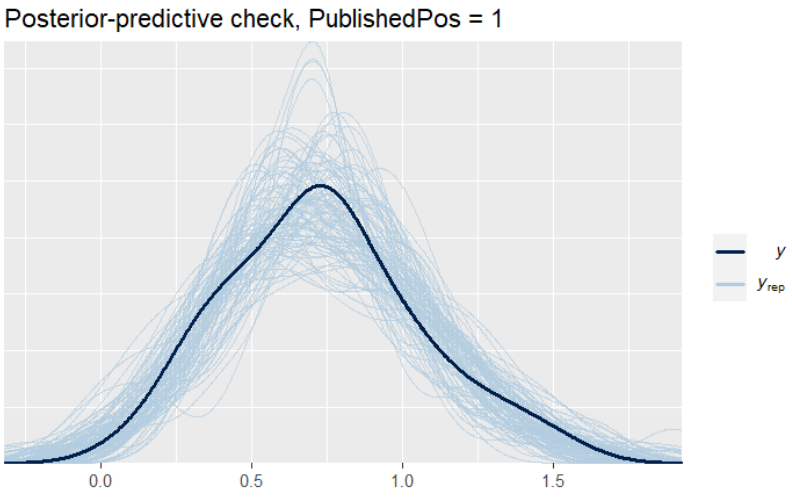
Ioana Luisa Forcas
au693527

In the first plot, it is shown that the model is rather effective in capturing the data.

Publication bias is considered in the second and third plots. The figure in the middle shows that research not published are being excluded in order to examine symmetric publication bias (no significant outcome). The last plot is a study of the asymmetric publication bias, which states that research with a substantial and anticipated outcome are published.

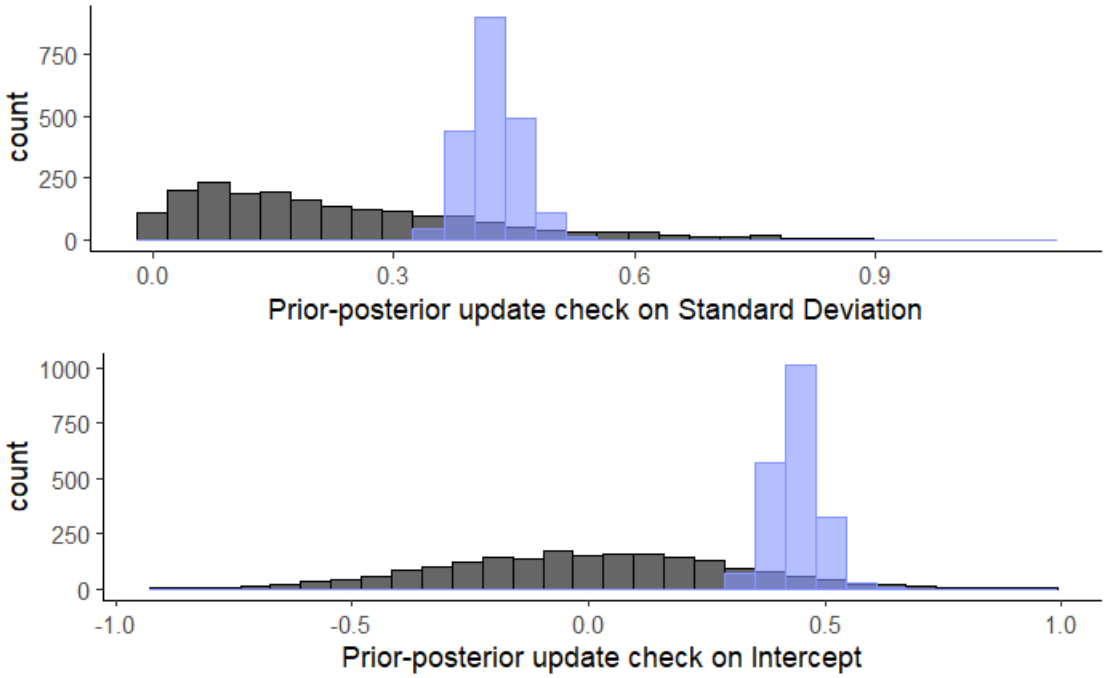
The model is attempting to "smooth-out" the 0-outcome because the publishing bias is not built into the model (the middle plot). The drawings (blue lines) in all three plots closely resemble the actual data (black line).

Posterior-predictive check



Ioana Luisa Forcas
au693527

Update checks:



The prior-posterior update check incorporates all research without consideration to publication bias. The posterior exhibits confidence and appears to have learnt from the priors.

Question 2

Description

The 50 research studies included in the data from Parola et al (2020) were published between 1977 and 2018.

Each research had, on average, 13 male and 11 female individuals from the control group, and 17 male and 7 female participants with schizophrenia. Participants with schizophrenia are 35.95 years old on average, with an 8.6 standard deviation. Participants in the control group had a median lifespan of 35.34 and a standard deviation of 9.

The fundamental frequency of pitch (F0), which is the focus variable for pitch variability, changes by 21.6 SD on average in the control group whereas by 26.8 SD in schizophrenia subjects.

Analysis

Ioana Luisa Forcas
au693527

The model from the first section of the assignment was utilized to examine the data (with adjusted variable names).

The results were translated to Cohen's D since various scales for the variables of interest might be used by different papers.

As a result, the pitch F0 mean and standard deviation values for the control and schizophrenia groups were modified.

Formula:

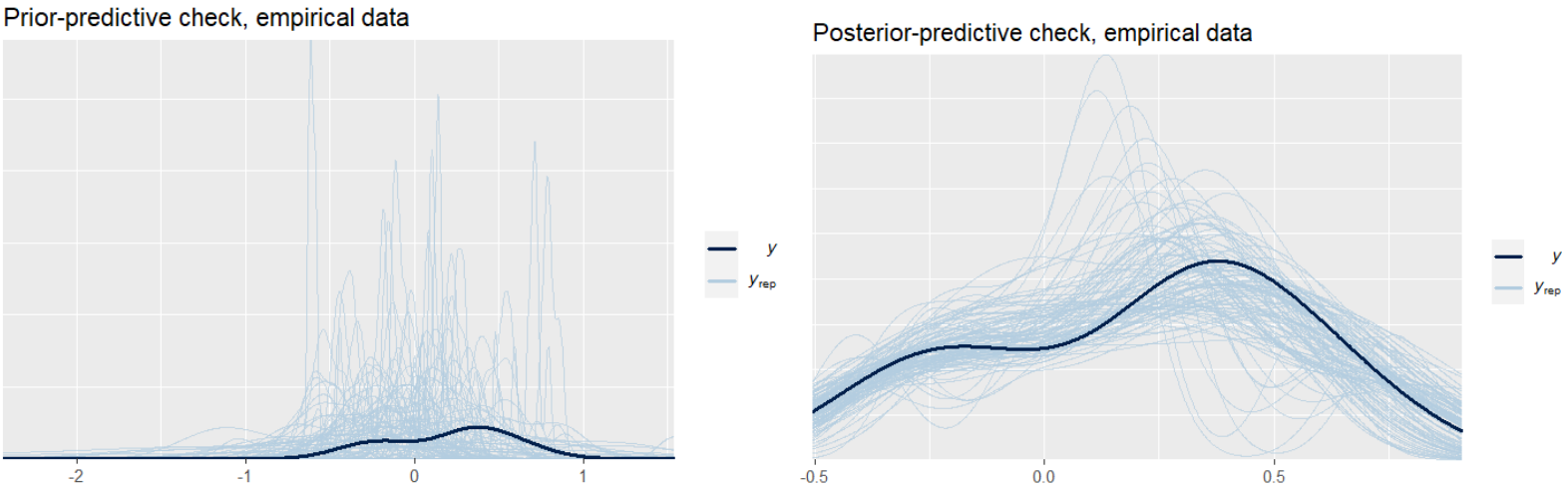
yi | se(vi) ~ 1 + (1 | StudyID)

-"yi" stands for the effect size

-"vi" for the Cohen's D scale-converted standard error of each research.

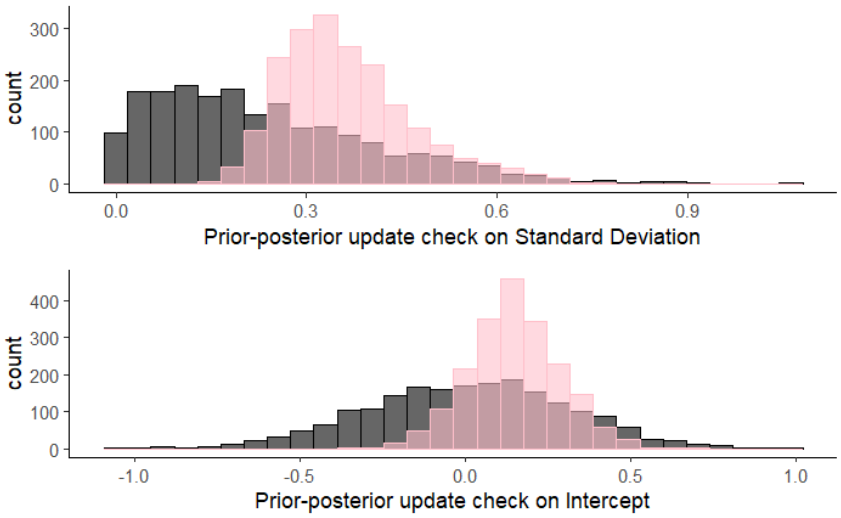
The model from part 1's model was applied, using identical priors.

Prior and posterior predictive checks on Empirical data



Prior values appear to be within the appropriate range of scale. Although there is some uncertainty in the model's predictions, it appears that it closely represents the general distribution of the relevant variable. To determine how successfully the model upgrades from the priors after being subjected to the real data, a prior-posterior update check may be helpful.

Ioana Luisa Forcas
au693527



Plots show that both posterior distributions grow significantly more certain and learn from the prior. The priors may not be changed since it is not being pushed at the tails of the prior distribution.

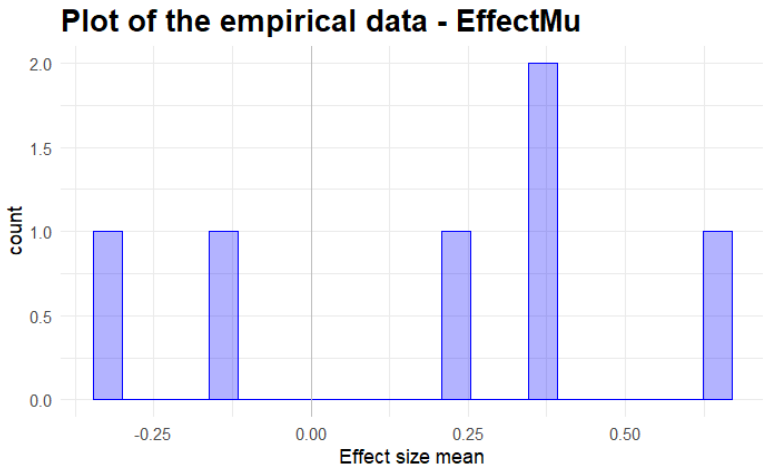
The distribution from the model with the real data resembles the distribution of the simulated data with asymmetric publication bias when comparing the distribution of effect sizes across studies (we can see that by looking at the posterior-predictive checks plots of both simulated and real data).

Even though the number of potential values in the model using simulated data is bigger than that in the model involving empirical data, it may be plausible to assume that the empirical data also contains asymmetric publication bias.

We should exercise caution when analyzing the results of published research since they may be positively biased if the true data is genuinely subject to publication bias because many studies are not accepted for publication.

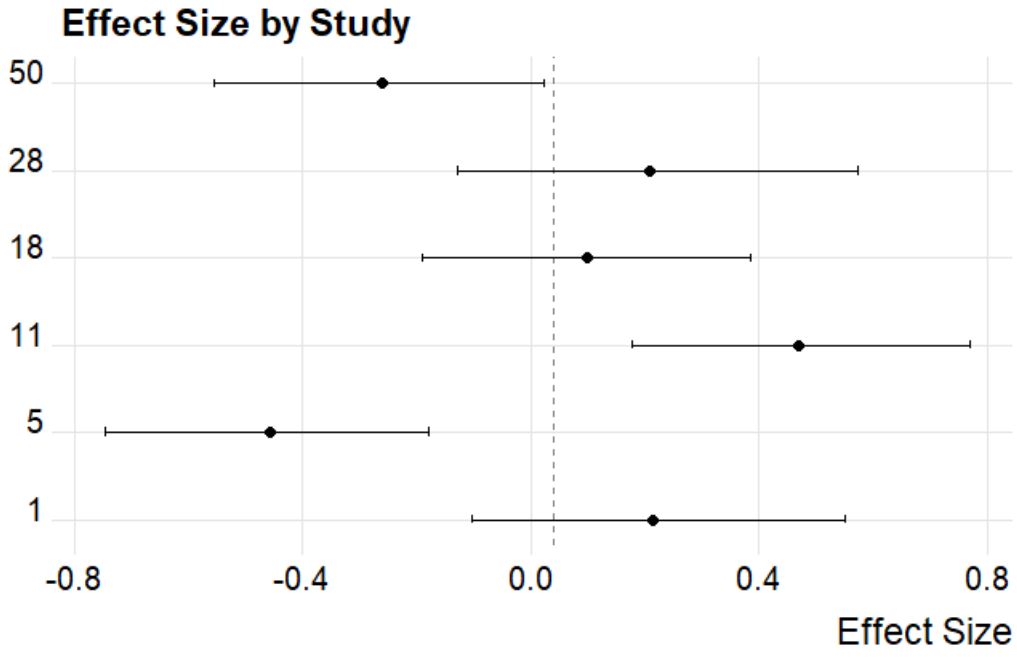
In this particular instance, the anticipated pitch results are made public.

Ioana Luisa Forcas
au693527



The graph displays the studies' impact sizes from the empirical data set. As the data set contains several NAs, only 6 mean impact sizes are calculated in this instance.

However, the actual data distribution resembles the simulated studies on the left the most, which may once again point to this kind of publishing bias (of course, bearing in mind that the data on the left is purely simulated).



The model's estimated impacts are shown in this plot. Each estimated study impact is shown by the mean and upper/lower quantiles, while the dashed line shows the mean effect size.

The fifth, eleventh, and fifty-first investigations appear to miss the average mean. The fifth and eleventh studies stand out from the others the most.

Ioana Luisa Forcas
au693527