OOD detection comparison with baseline

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library("ggpubr")
```

```
load_preproc <- function(dir)
{
    pwc_metrics <- read.csv(file.path(dir, "ens_pwc_metrics.csv"))
    cal_metrics <- read.csv(file.path(dir, "ens_cal_metrics.csv"))
    net_metrics <- read.csv(file.path(dir, "net_metrics.csv"))

    best_net_mets <- function(comb_id)
    {
        cur_net_mets <- net_metrics[as.logical(pwc_metrics[pwc_metrics$combination_id == comb_id, ][1, r
        comb_mets <- data.frame(as.list(comb_id)) %>% mutate(
            best_net_MSP_AUROC = max(cur_net_mets$MSP_AUROC),
            best_net_MLI_AUROC = max(cur_net_mets$MLI_AUROC),
            best_net_MSP_AUPRC = max(cur_net_mets$MSP_AUPRC),
            best_net_MLI_AUPRC = max(cur_net_mets$MLI_AUPRC))
    }

    best_net_aus <- dplyr::bind_rows(
        apply(unique(pwc_metrics[c("combination_id")]), 1, best_net_mets))
    pwc_metrics <- merge(pwc_metrics, best_net_aus)
    cal_metrics <- merge(cal_metrics, best_net_aus)

    cal_metrics <- cal_metrics %>%
        select(combination_id, combination_size, MSP_AUROC, MSP_AUPRC) %>%
        rename(BSL_MSP_AUPRC = MSP_AUPRC, BSL_MSP_AUROC = MSP_AUROC)

    pwc_metrics <- pwc_metrics %>% select(combination_id, combination_size, combining_method, coupling_r

    metrics <- left_join(pwc_metrics, cal_metrics)

    metrics <- metrics %>% mutate(DIFF_AUROC = MSP_AUROC - BSL_MSP_AUROC, DIFF_AUPRC = MSP_AUPRC - BSL_M
```

```
    for (co_m in unique(metrics$combining_method))
    {
        metrics_co <- metrics %>% filter(combining_method == co_m)
        for (cp_m in unique(metrics_co$coupling_method))
        {
            metrics_co_cp <- metrics_co %>% filter(coupling_method == cp_m)
            print(sprintf("Method %s+%s", co_m, cp_m))
            method <- paste0(co_m, "+", cp_m)
            print(ggdensity(data = metrics_co_cp, x = "DIFF_AUPRC", xlab = "AUPRC difference", title = m
            print(ggqqplot(data = metrics_co_cp, x = "DIFF_AUPRC", title = method))
            print(sprintf("Method %s+%s", co_m, cp_m))
            print(shapiro.test(metrics_co_cp$DIFF_AUPRC))
            print(ggdensity(data = metrics_co_cp, x = "DIFF_AUROC", xlab = "AUROC difference", title = m
            print(ggqqplot(data = metrics_co_cp, x = "DIFF_AUROC", title = method))
            print(sprintf("Method %s+%s", co_m, cp_m))
            print(shapiro.test(metrics_co_cp$DIFF_AUROC))
        }
    }
}
```

Shapiro-Wilk zero hypothesis is data normality. Accuracy differences seem to be normal, so we can proceed with the t-tests.

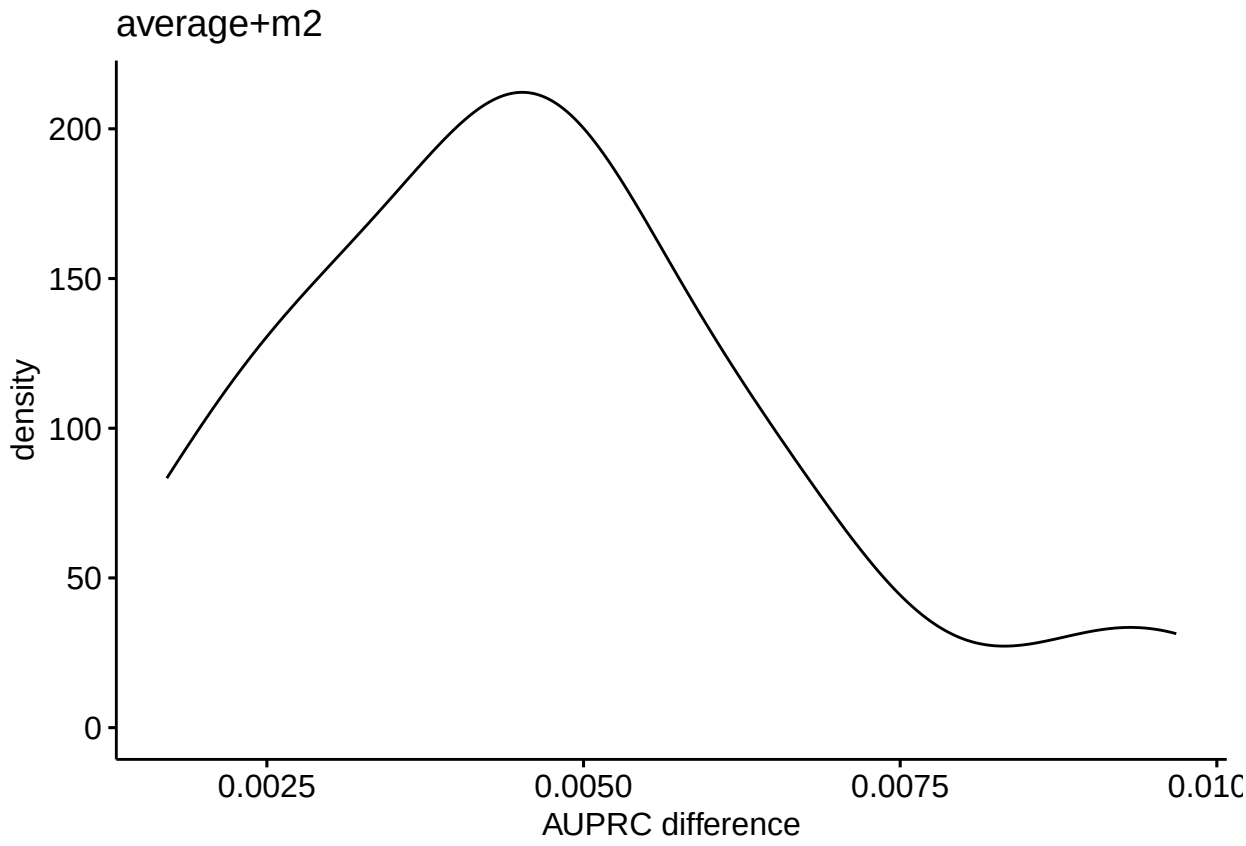
```
base_dir_C10 <- "/mnt/d/skola/1/weighted_ensembles/tests/test_cifar_ood_2022/C10vsC100_metrics"
base_dir_C100 <- "/mnt/d/skola/1/weighted_ensembles/tests/test_cifar_ood_2022/C100vsC10_metrics"

print("Processing C10 vs C100")
```

```
## [1] "Processing C10 vs C100"
```

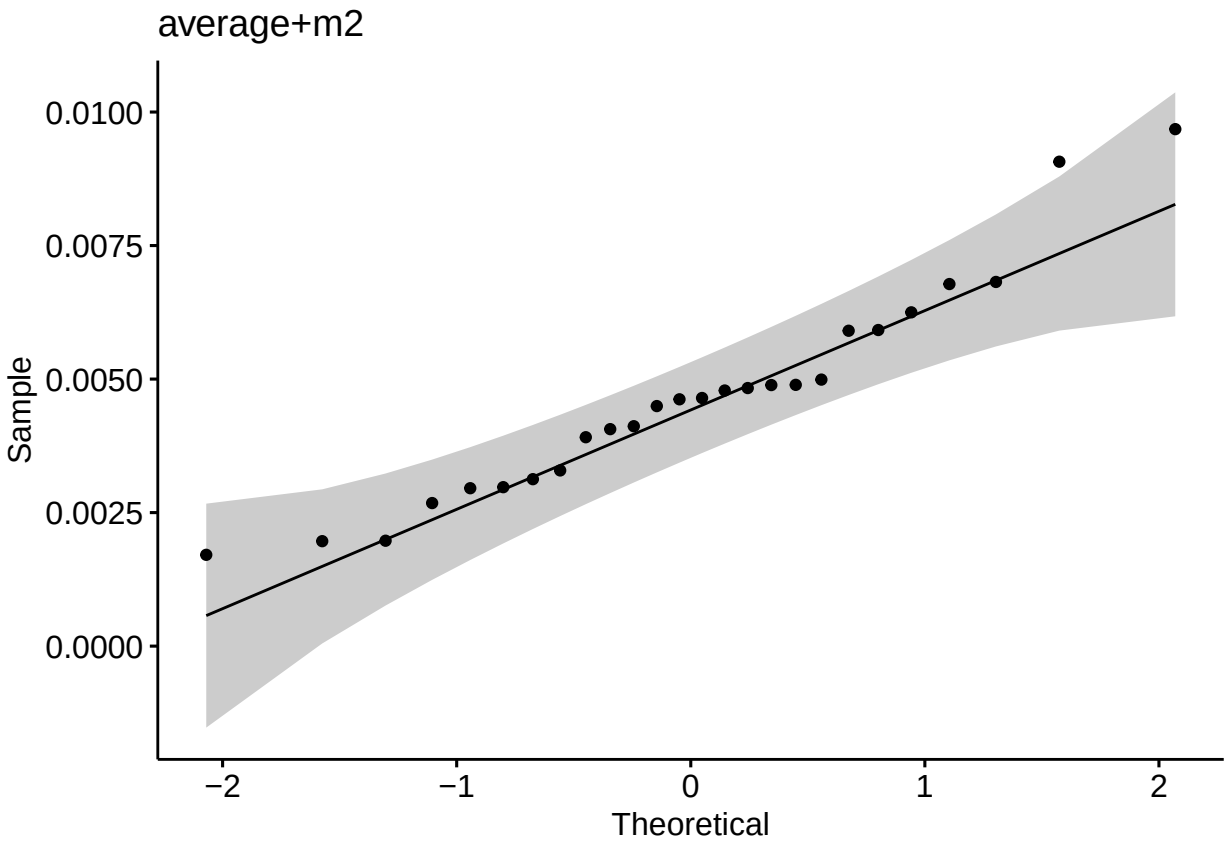
```
dfs <- load_preproc(base_dir_C10)
```

```
## Joining with `by = join_by(combination_id, combination_size)`
```

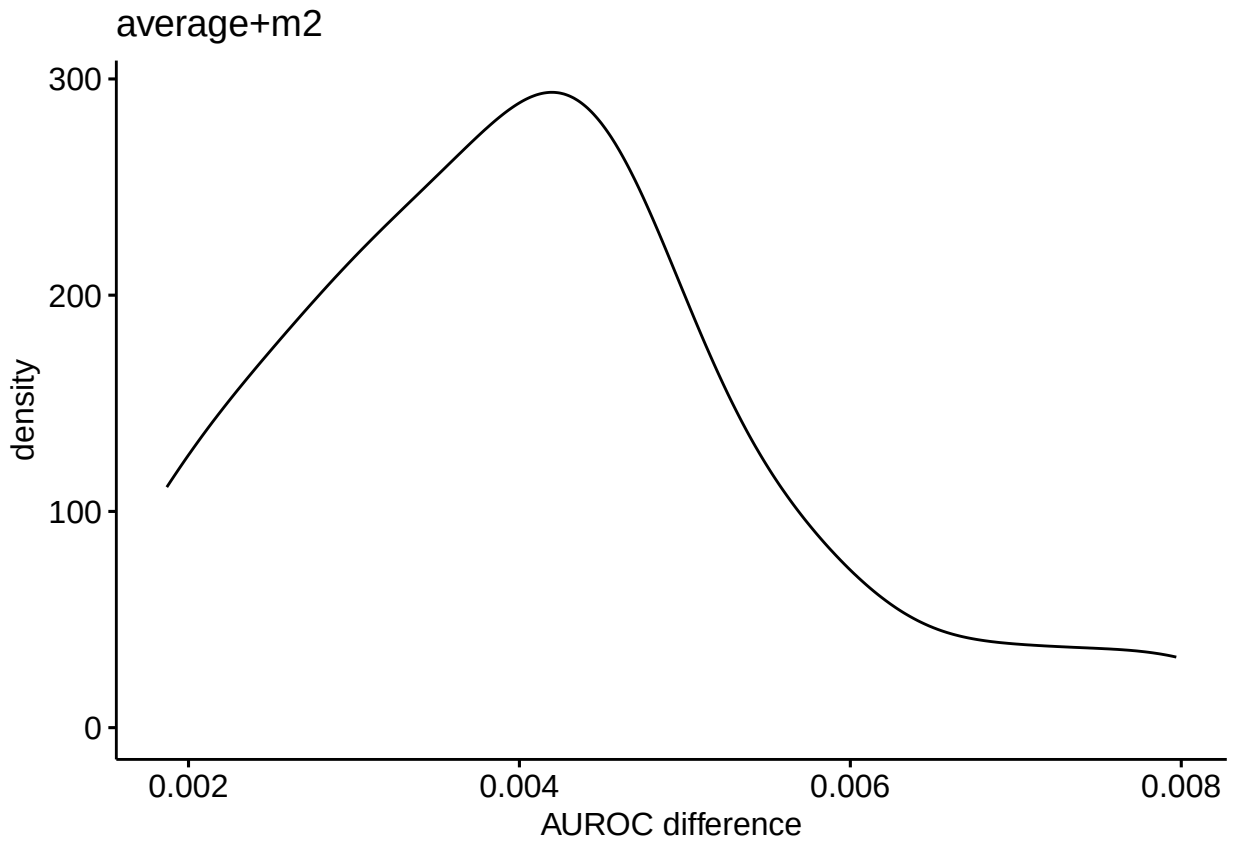
```
## [1] "Method average+m2"
```



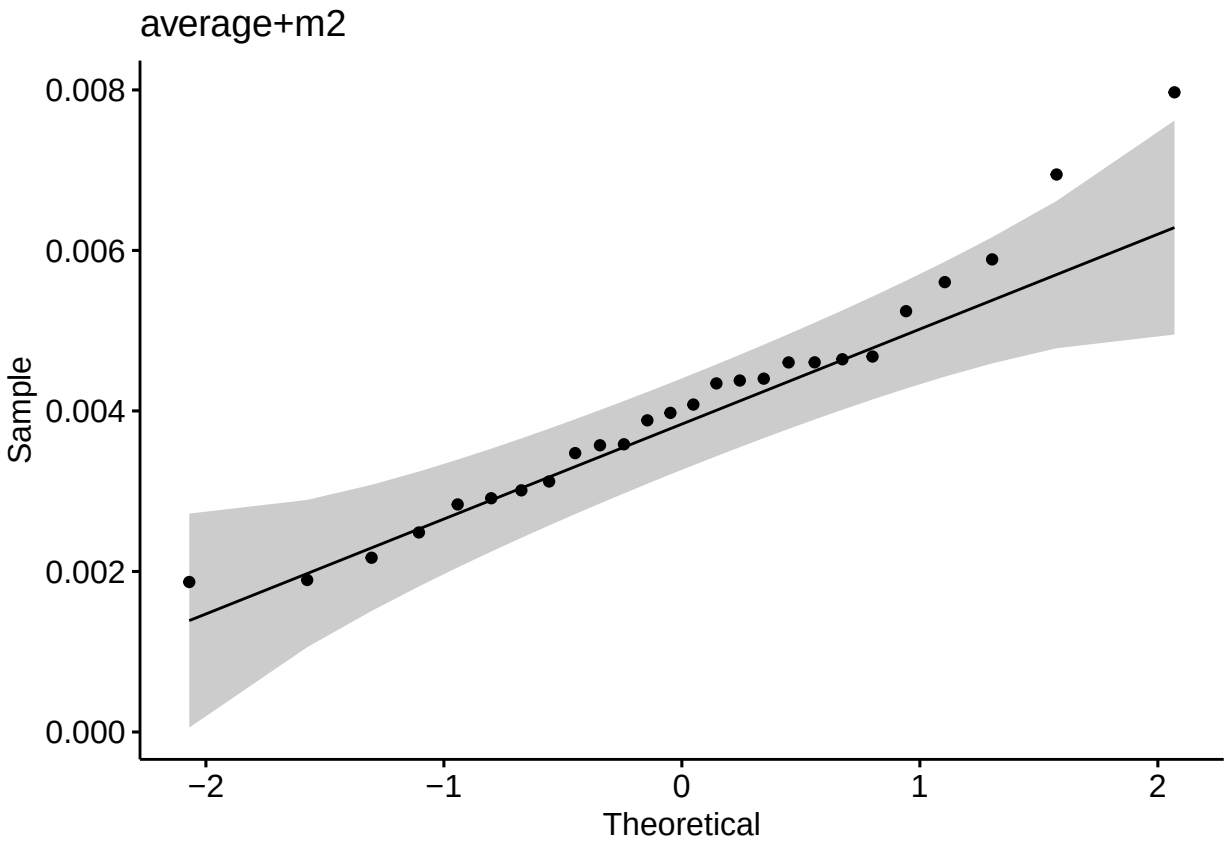

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



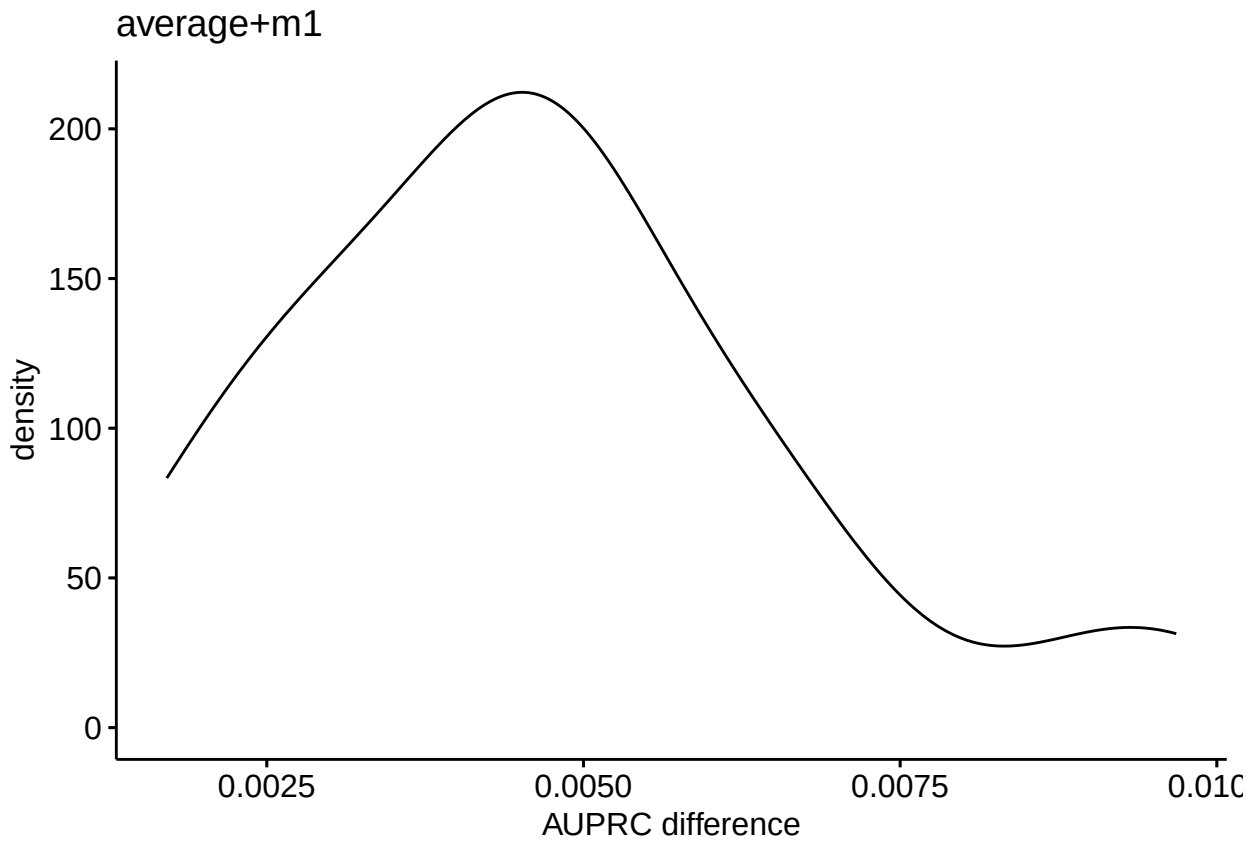

```
## [1] "Method average+m2"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.93822, p-value = 0.1218
```



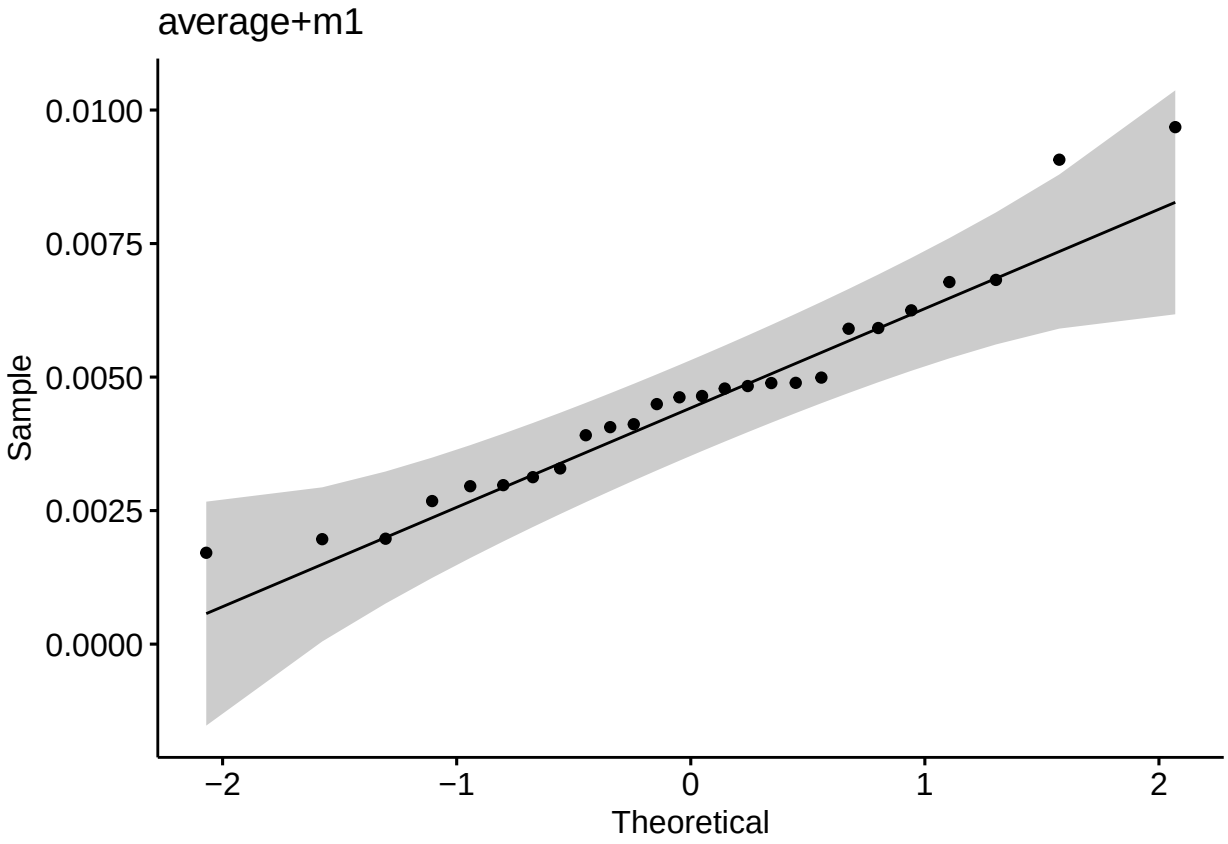

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



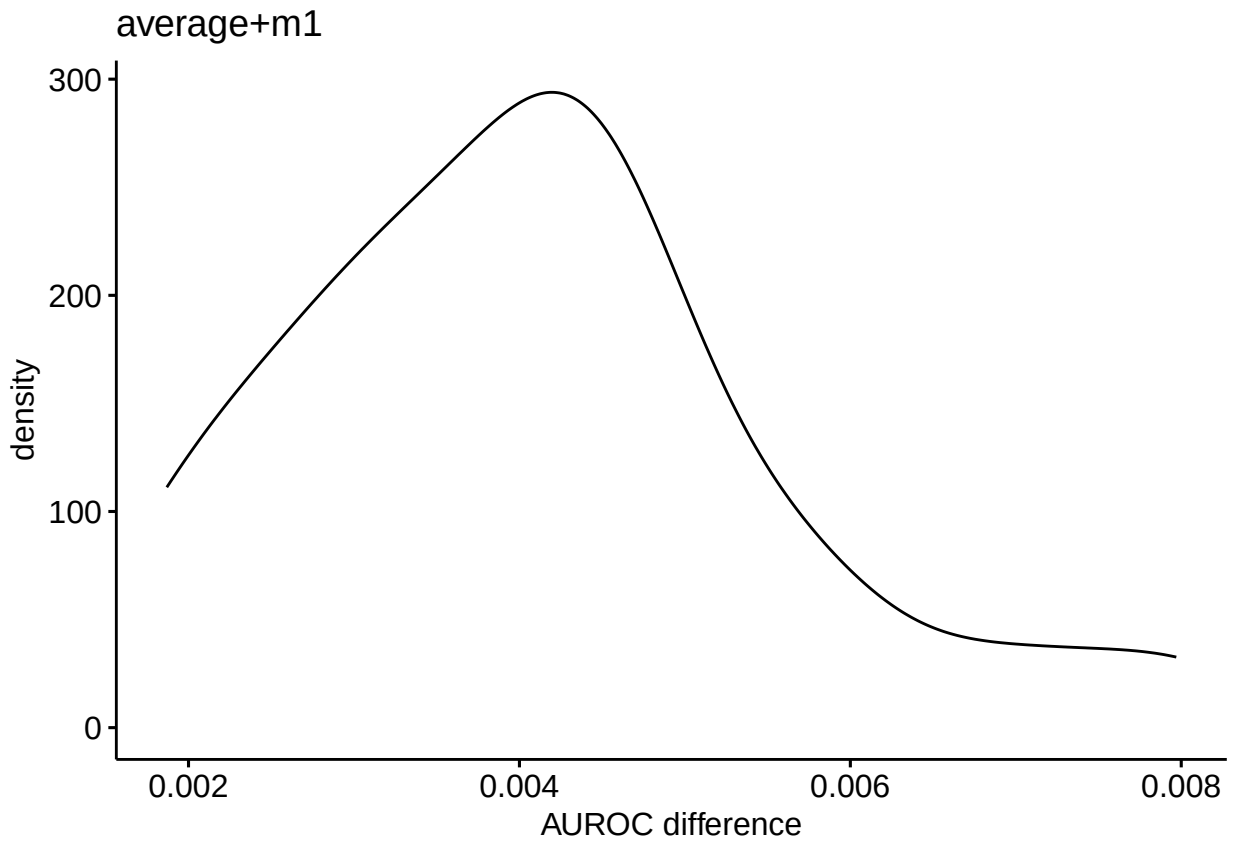

```
## [1] "Method average+m2"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.9533, p-value = 0.2767
##
## [1] "Method average+m1"
```



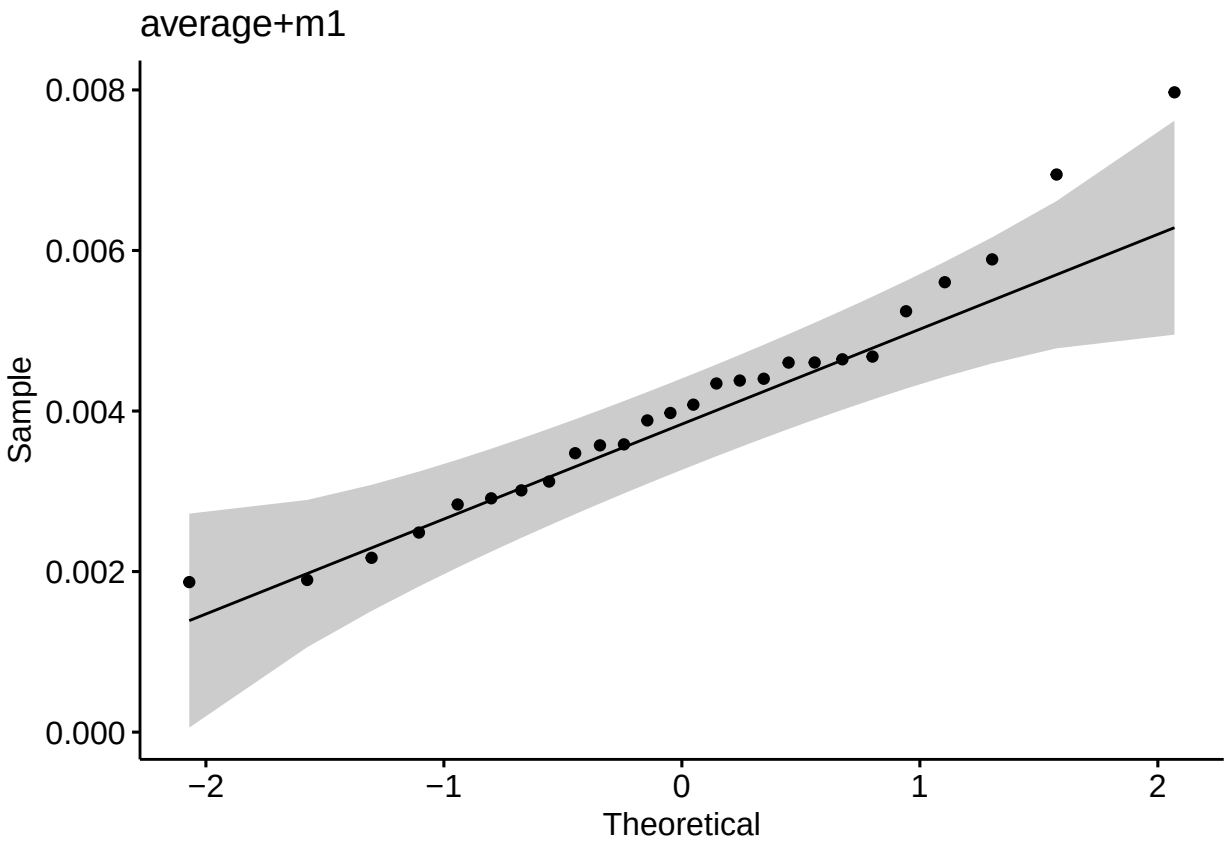

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



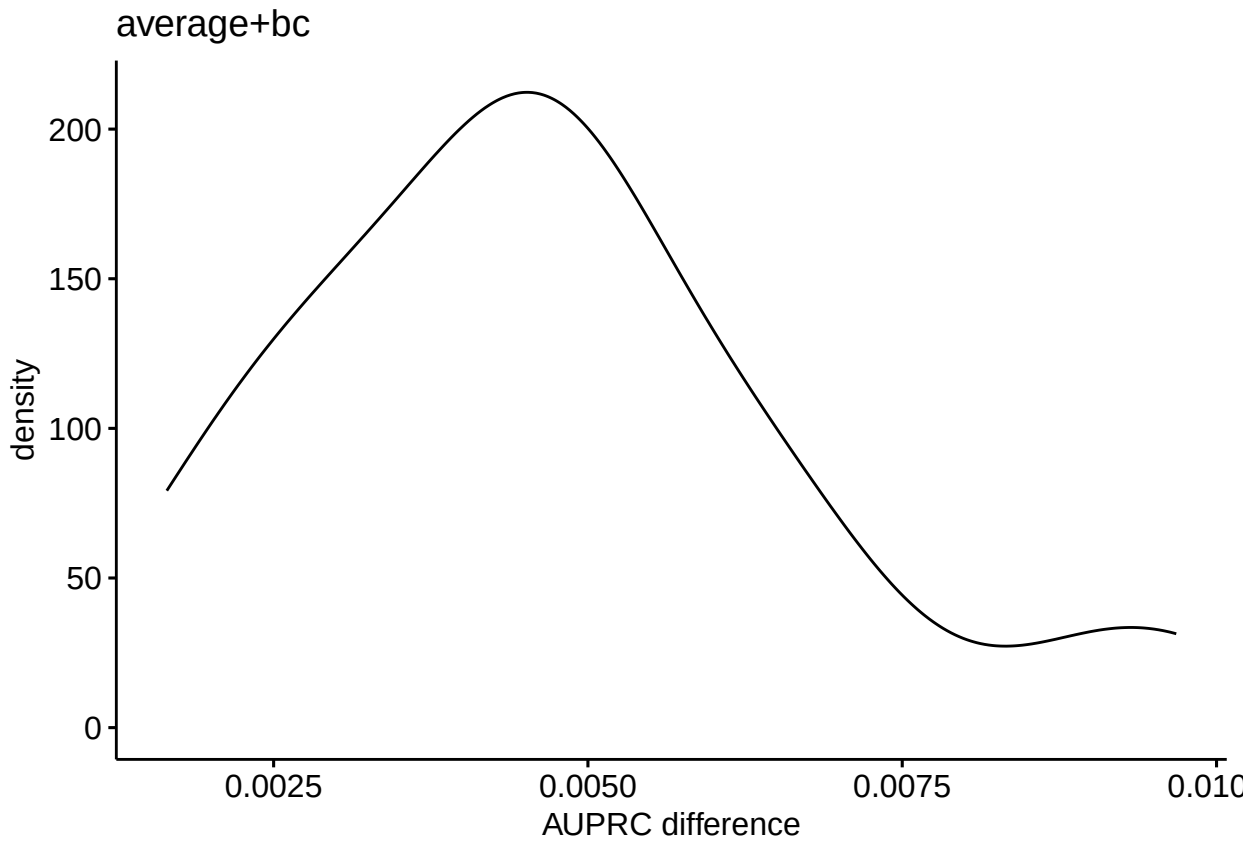

```
## [1] "Method average+m1"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.93821, p-value = 0.1217
```



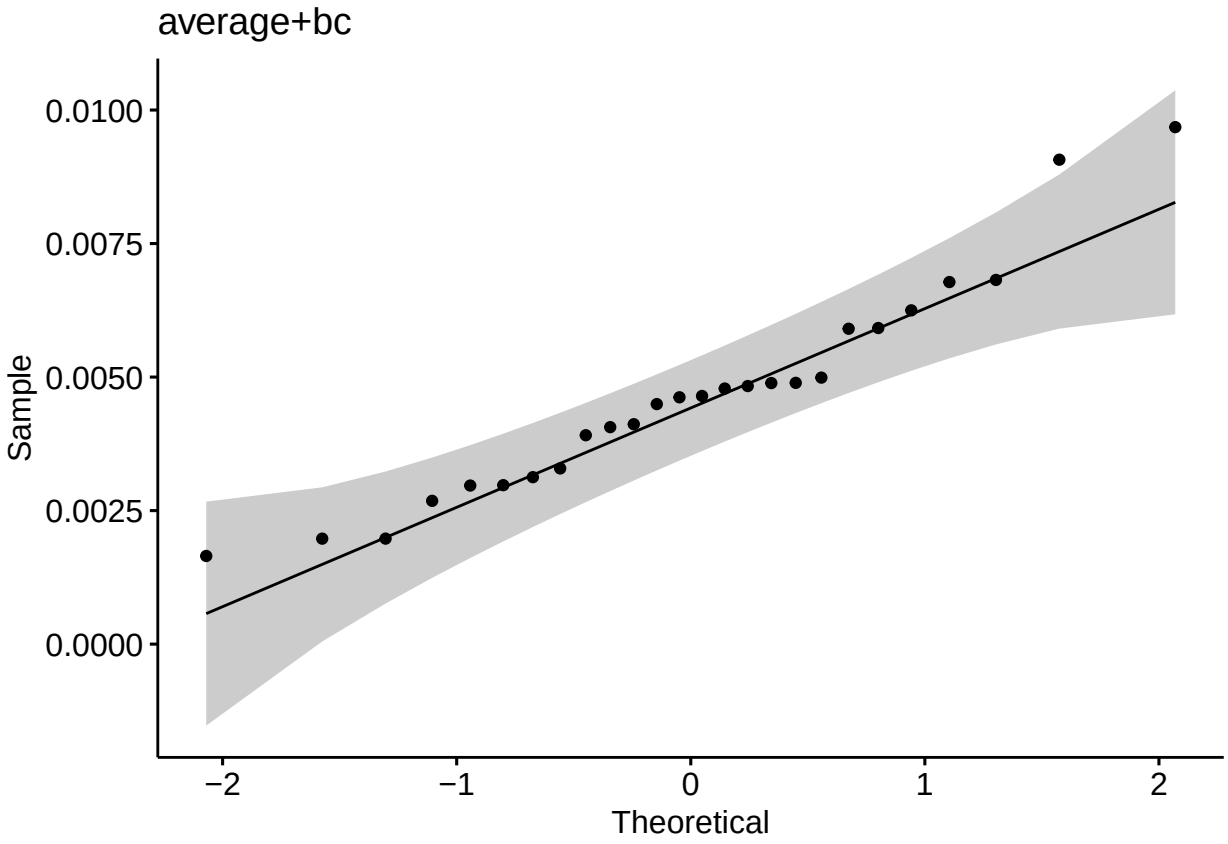

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



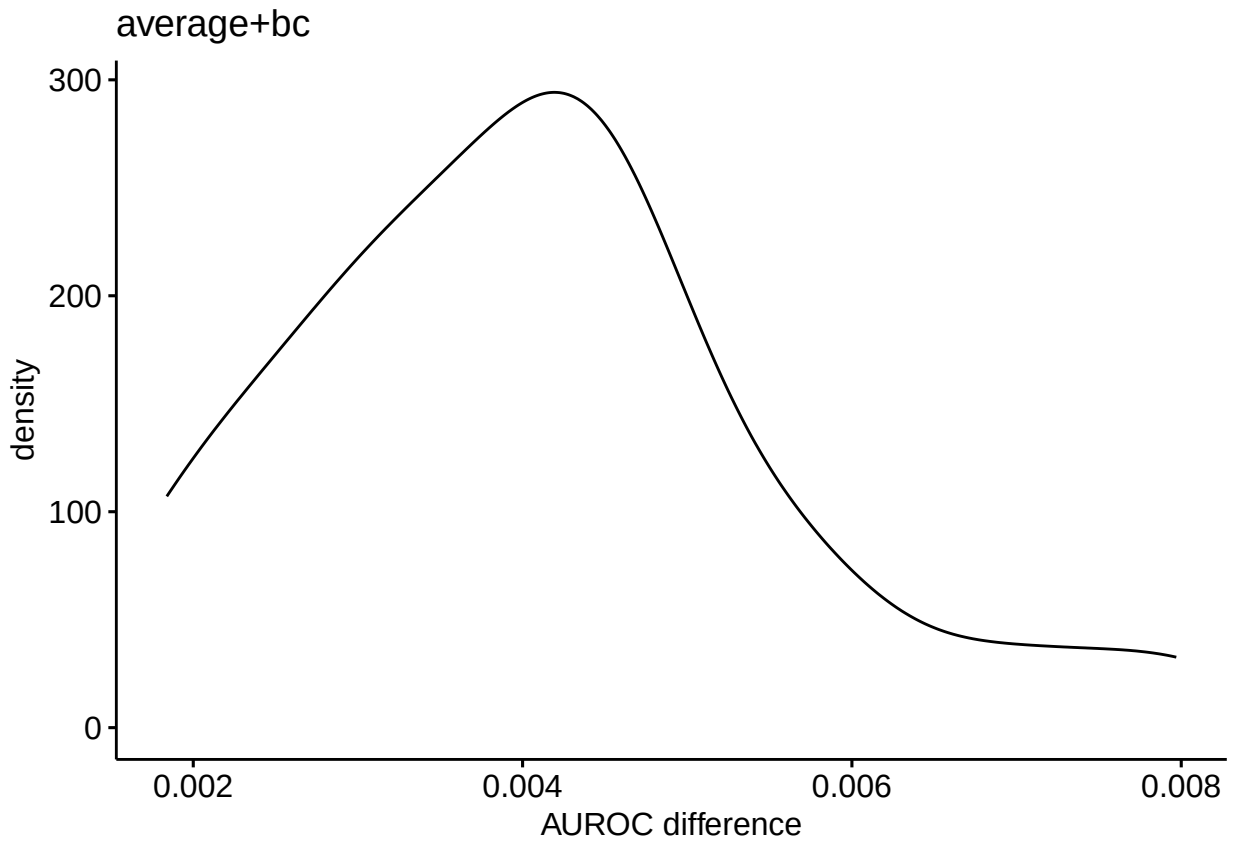

```
## [1] "Method average+m1"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.95329, p-value = 0.2765
##
## [1] "Method average+bc"
```



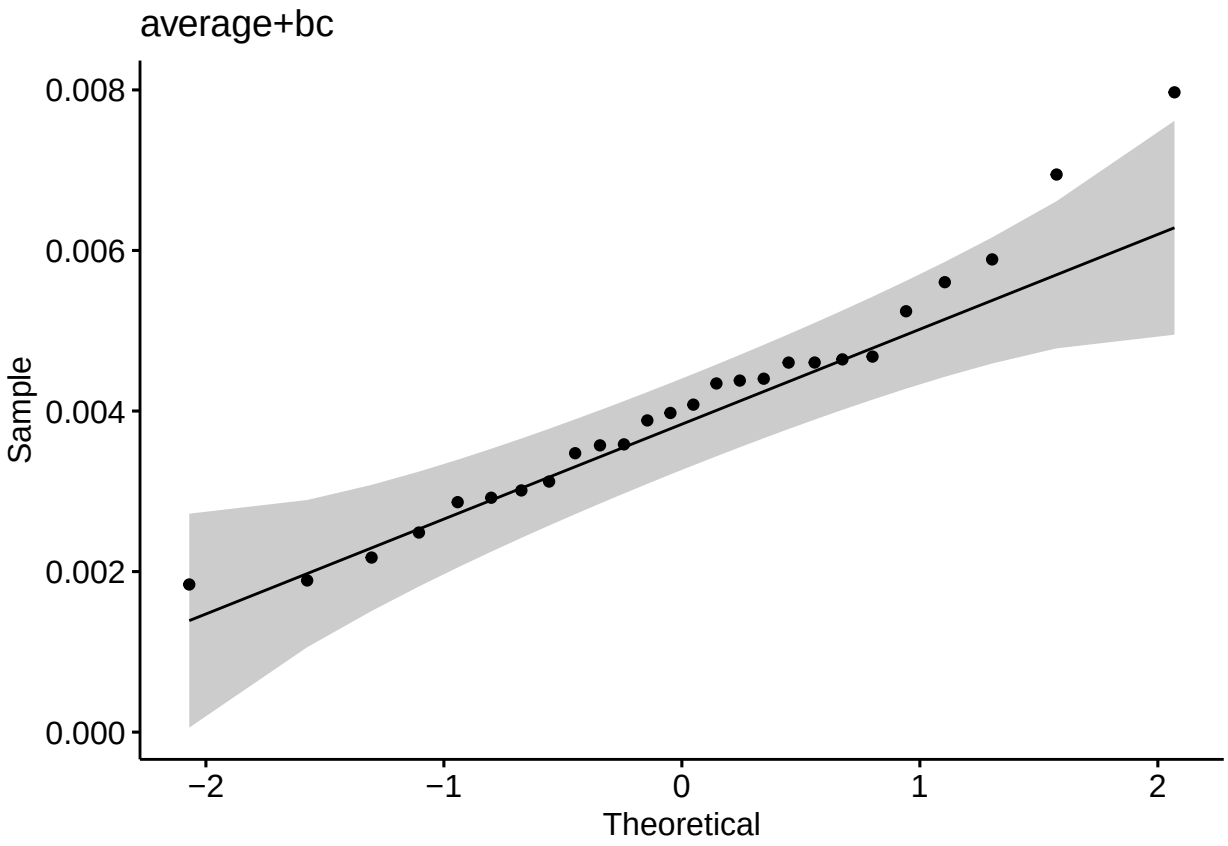

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



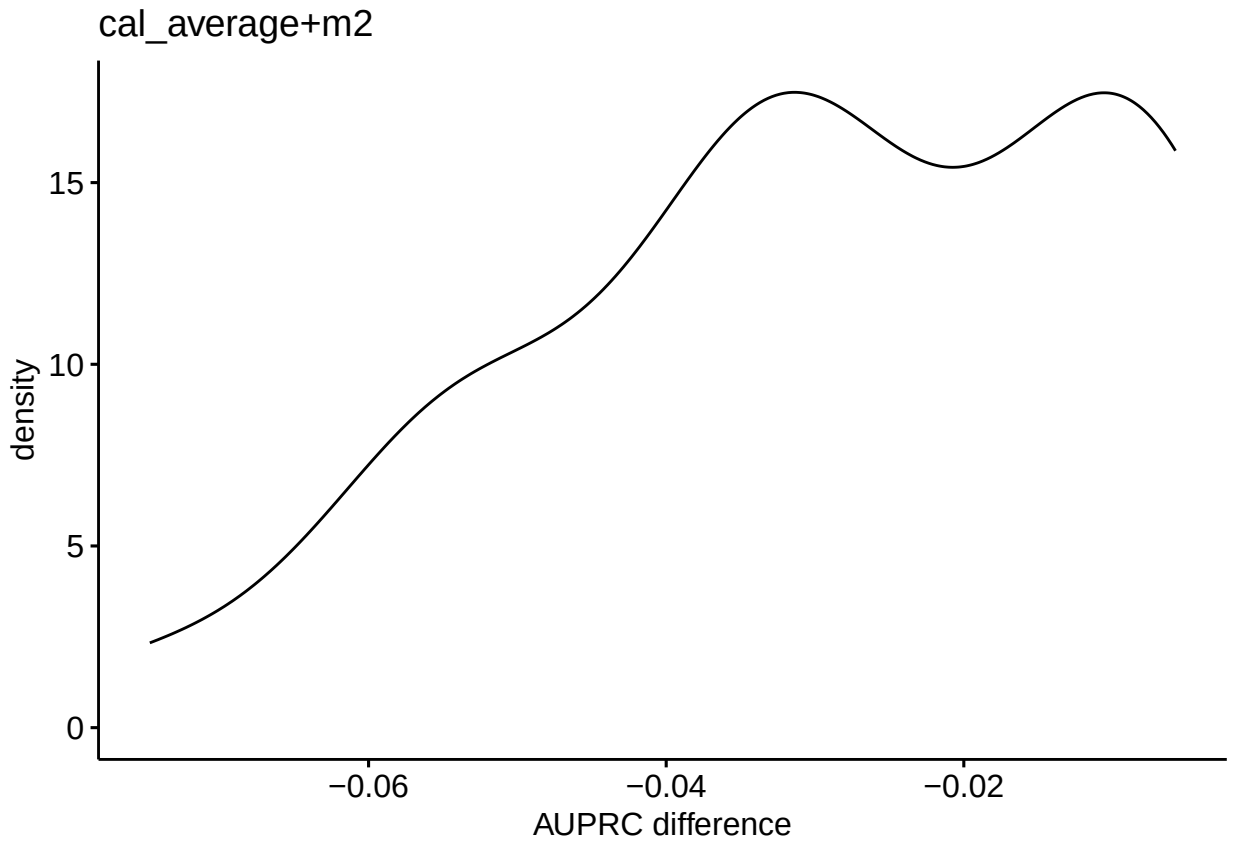

```
## [1] "Method average+bc"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.93975, p-value = 0.1325
```



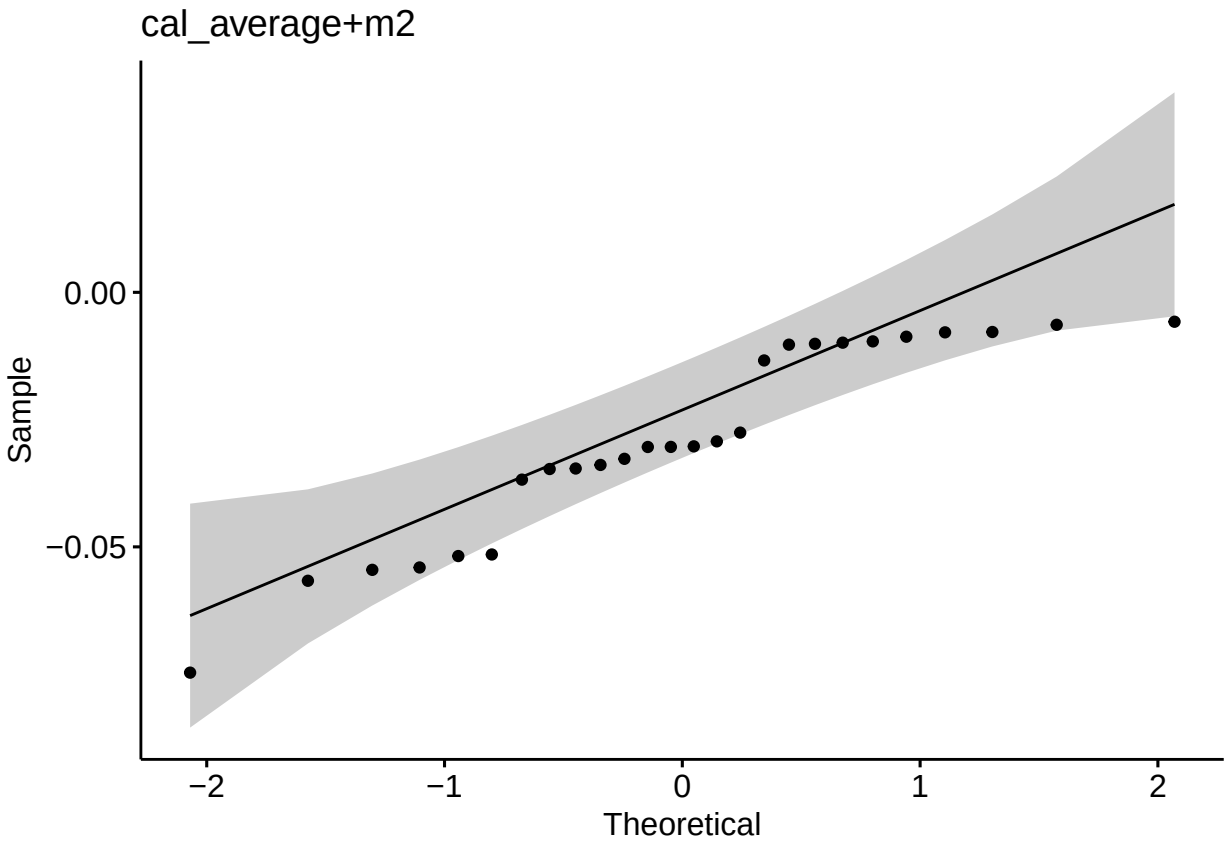

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



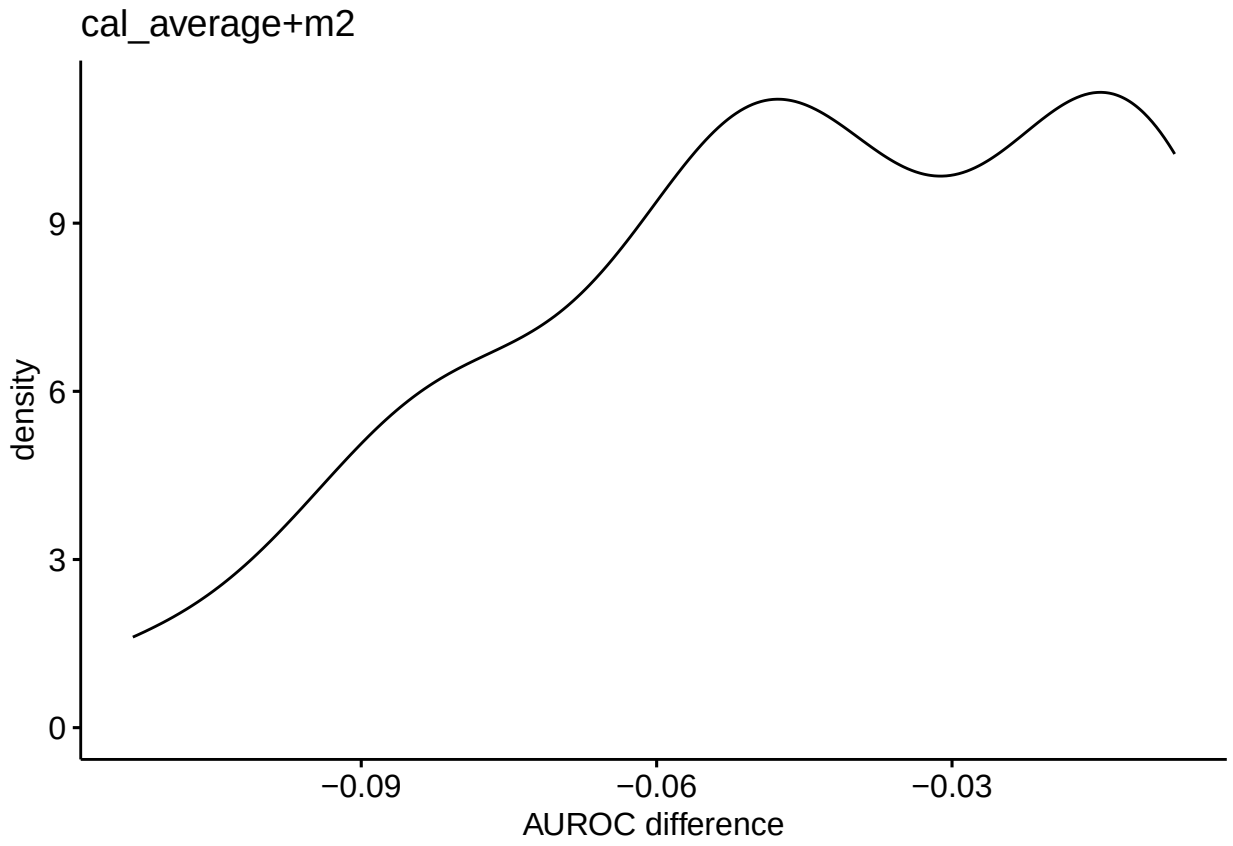

```
## [1] "Method average+bc"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.95423, p-value = 0.2907
##
## [1] "Method cal_average+m2"
```



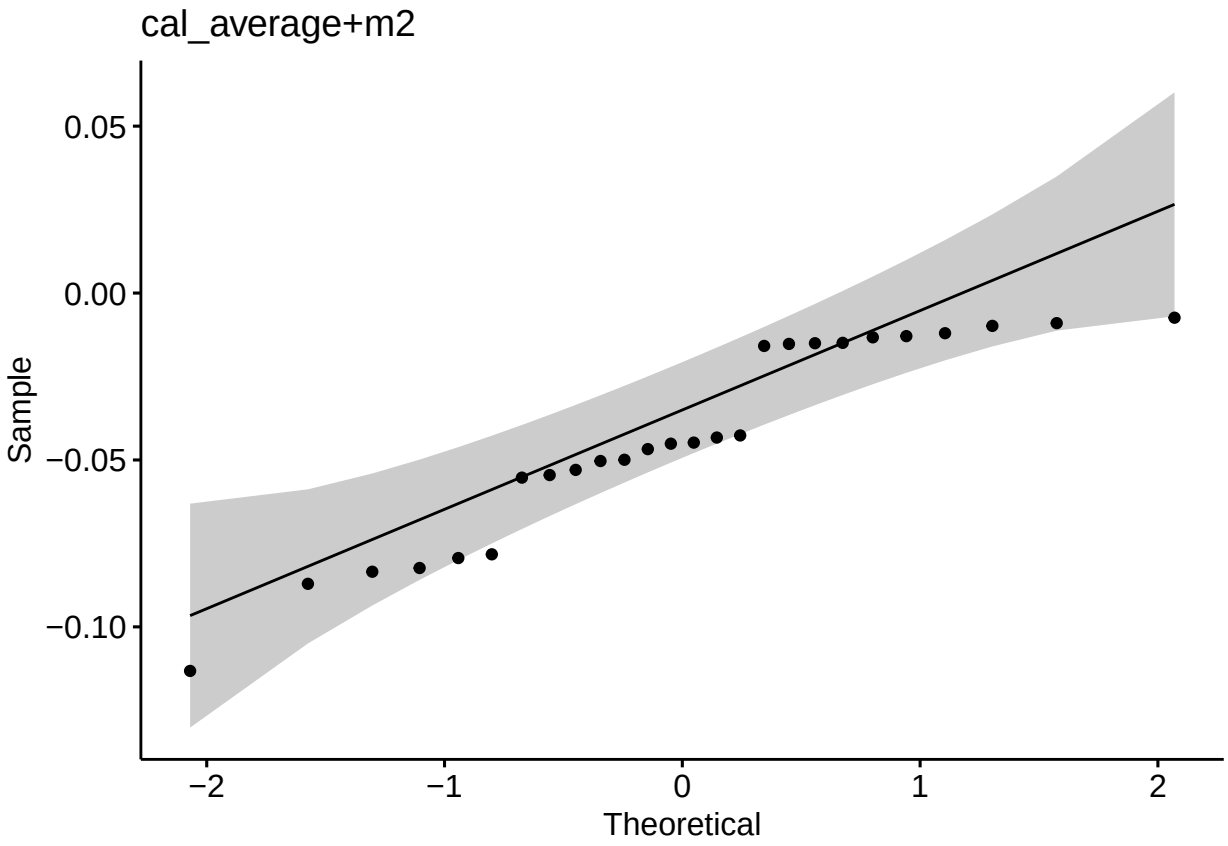

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



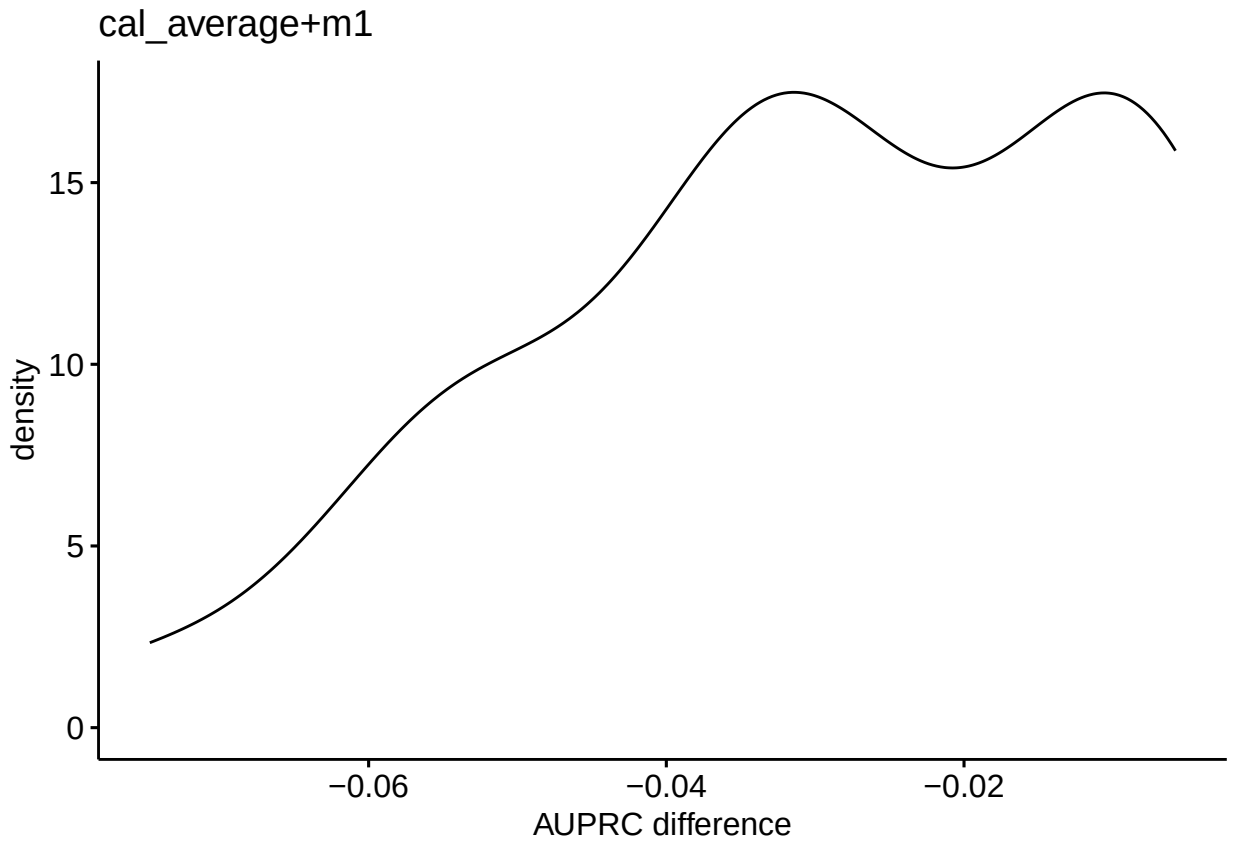

```
## [1] "Method cal_average+m2"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.90174, p-value = 0.01713
```



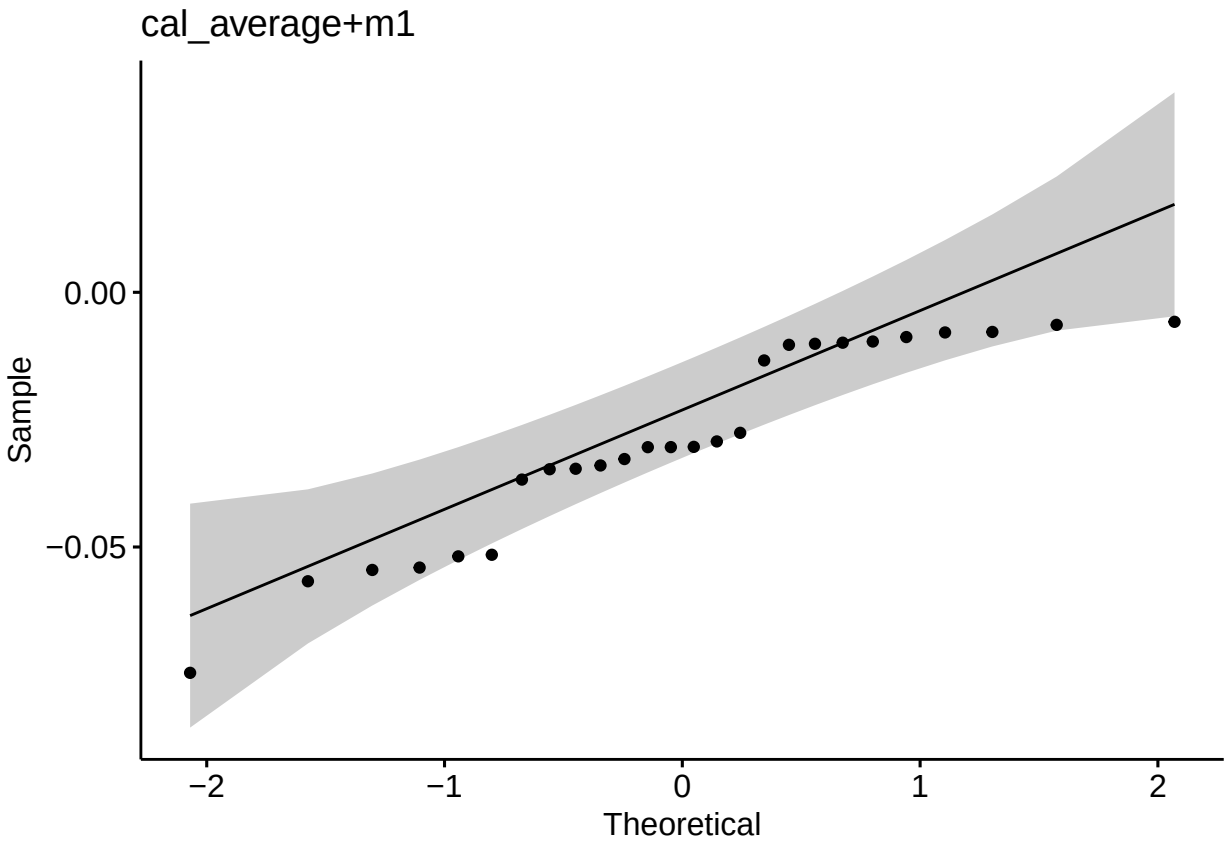

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



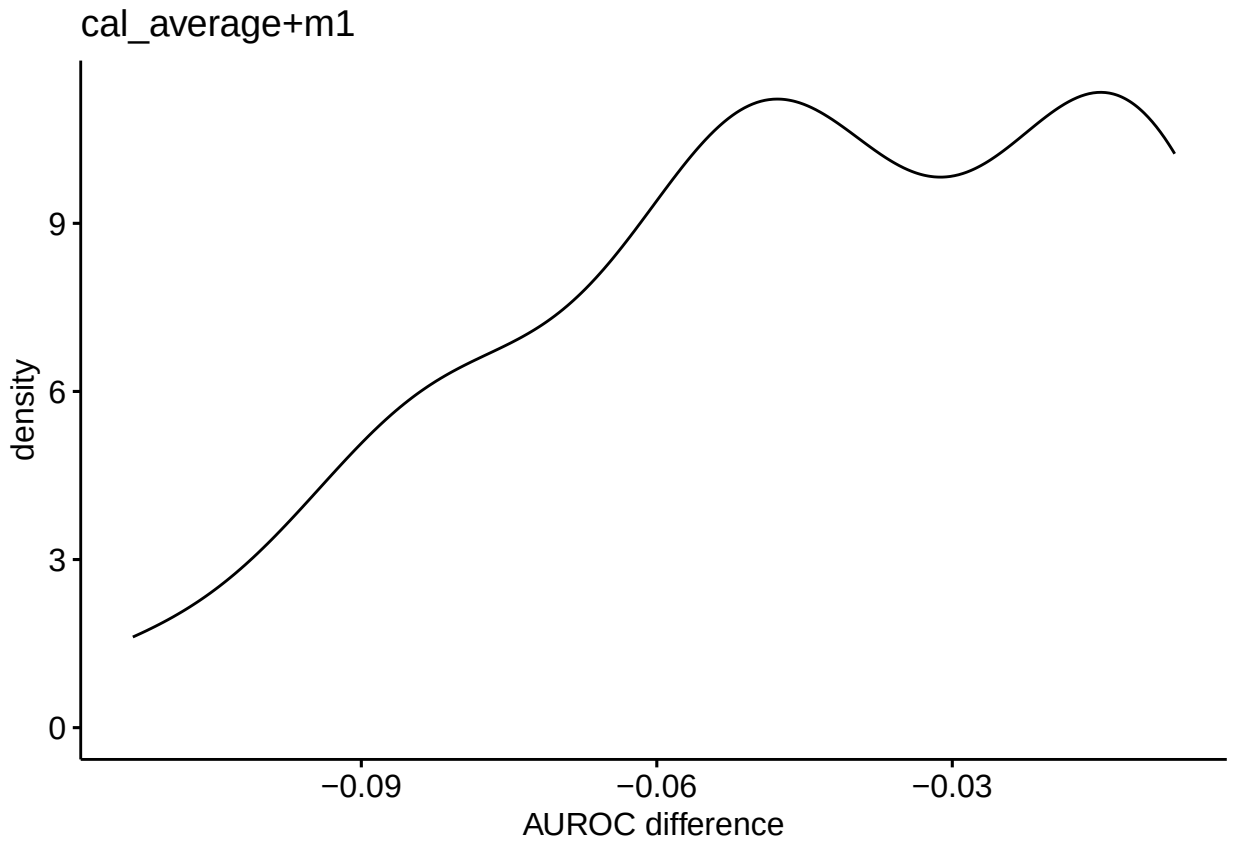

```
## [1] "Method cal_average+m2"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.90118, p-value = 0.01664
##
## [1] "Method cal_average+m1"
```



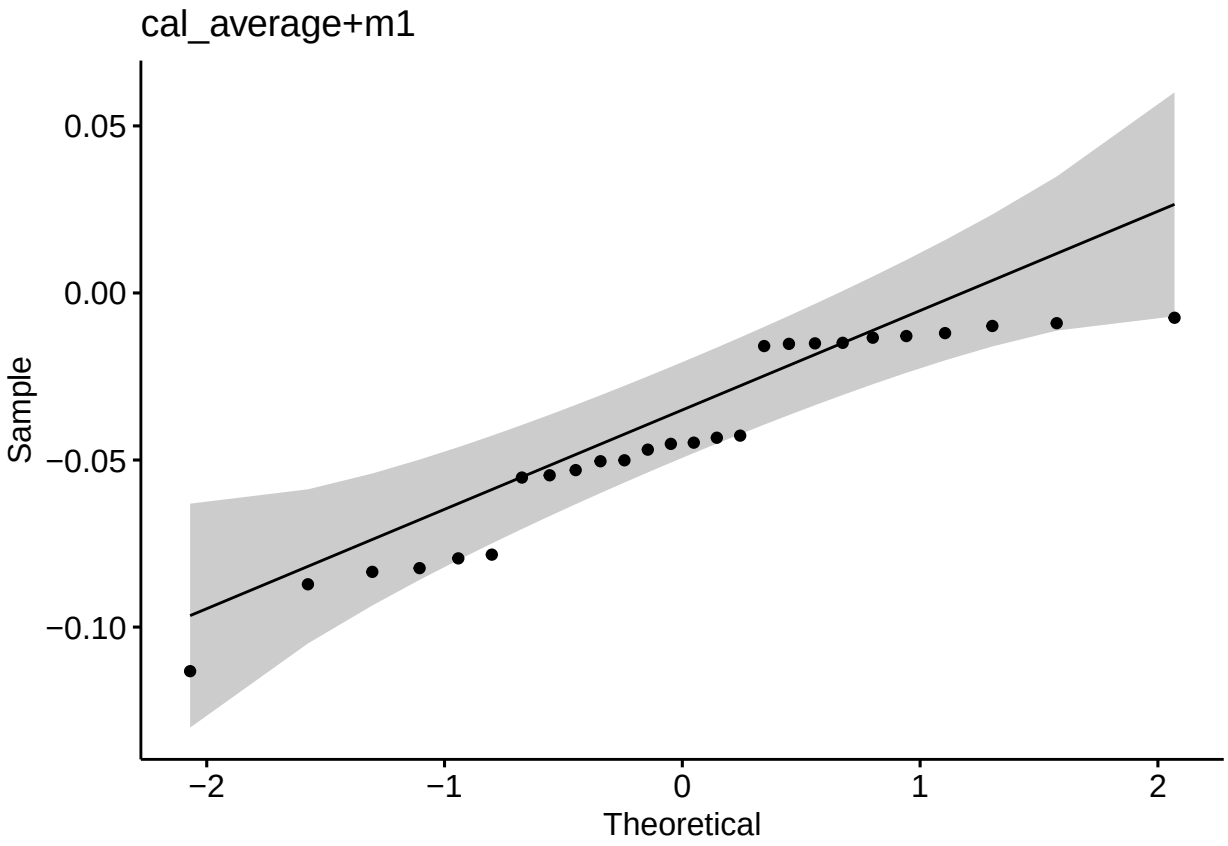

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



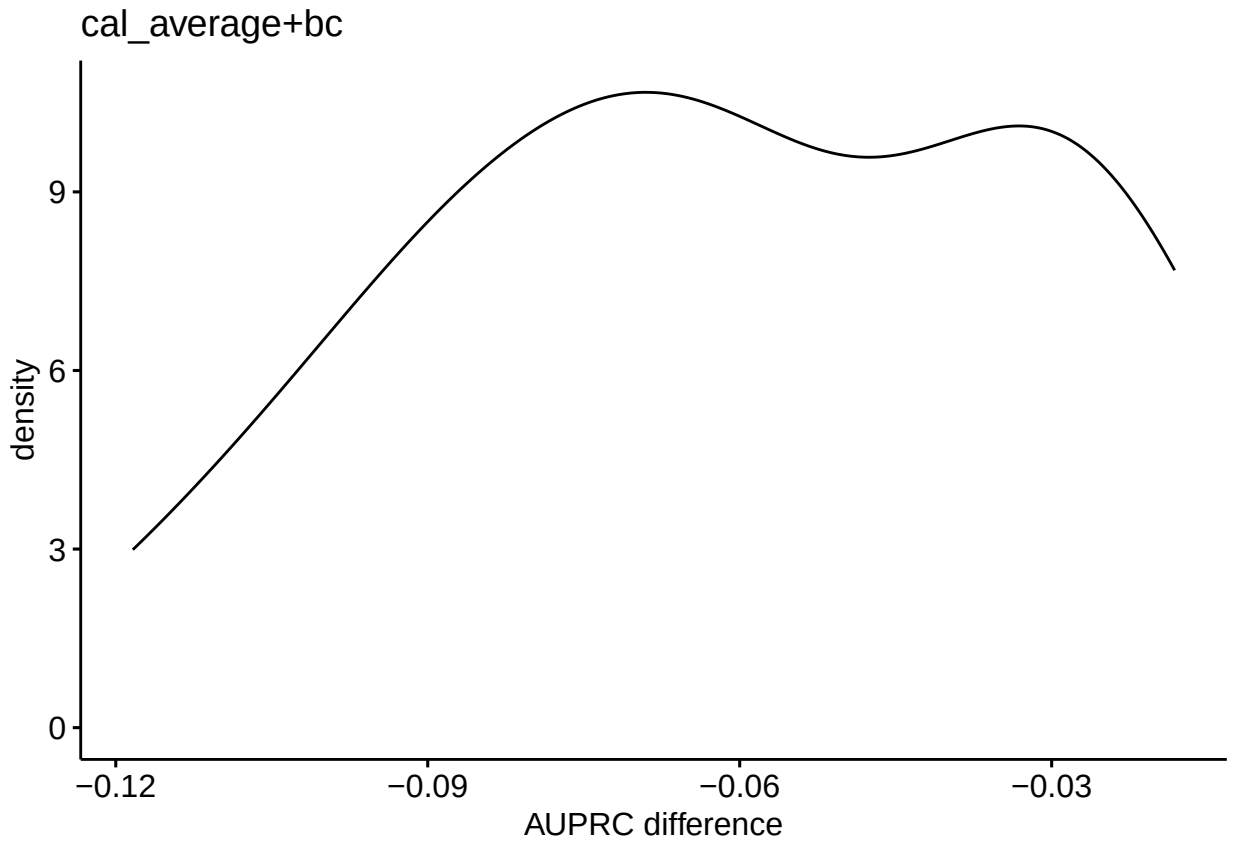

```
## [1] "Method cal_average+m1"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.90169, p-value = 0.01708
```



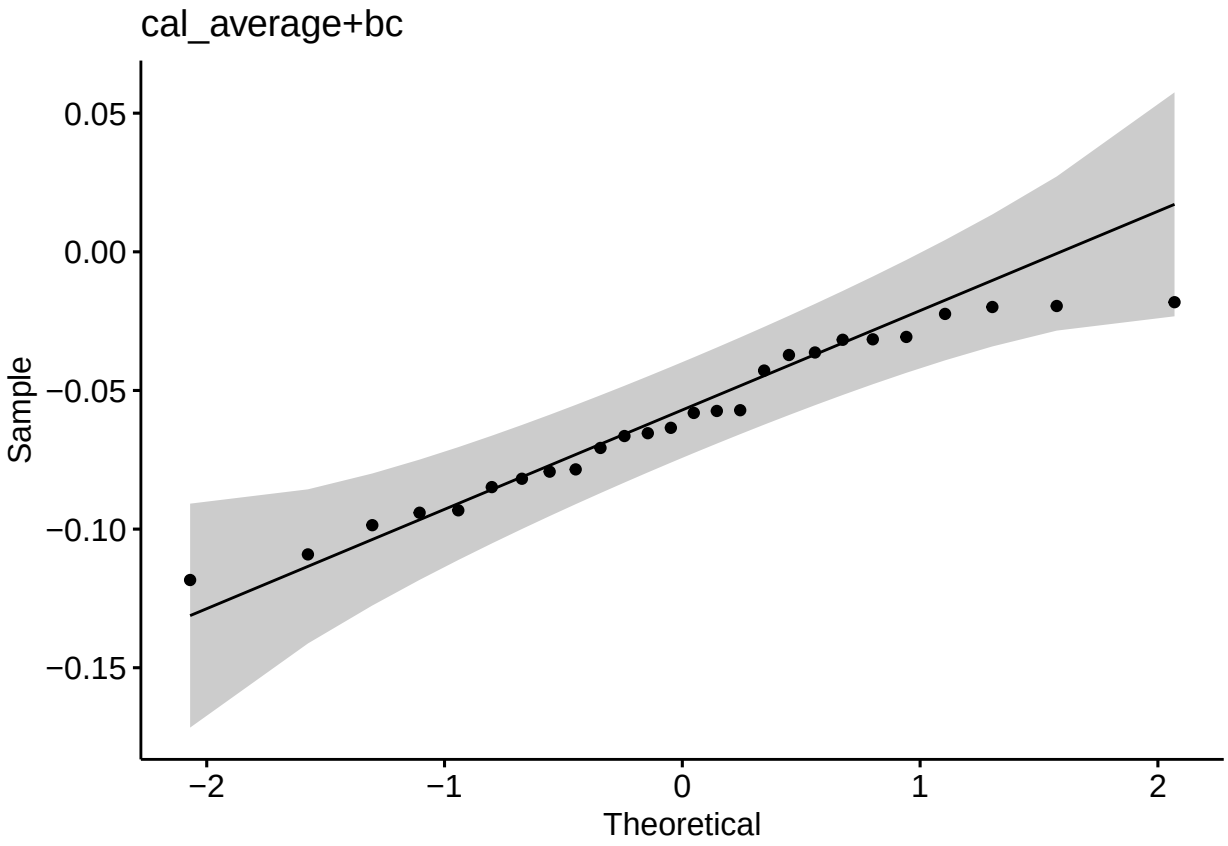

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



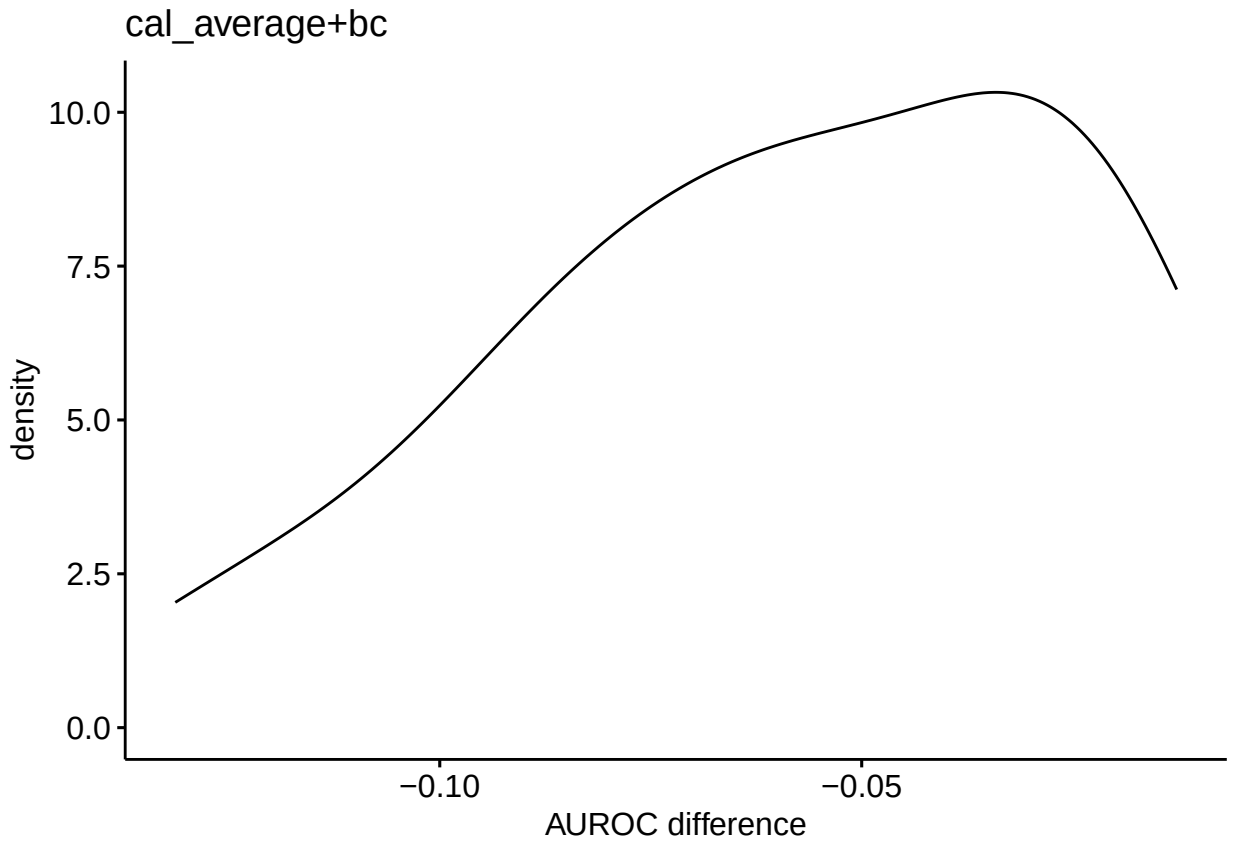

```
## [1] "Method cal_average+m1"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.90117, p-value = 0.01663
##
## [1] "Method cal_average+bc"
```



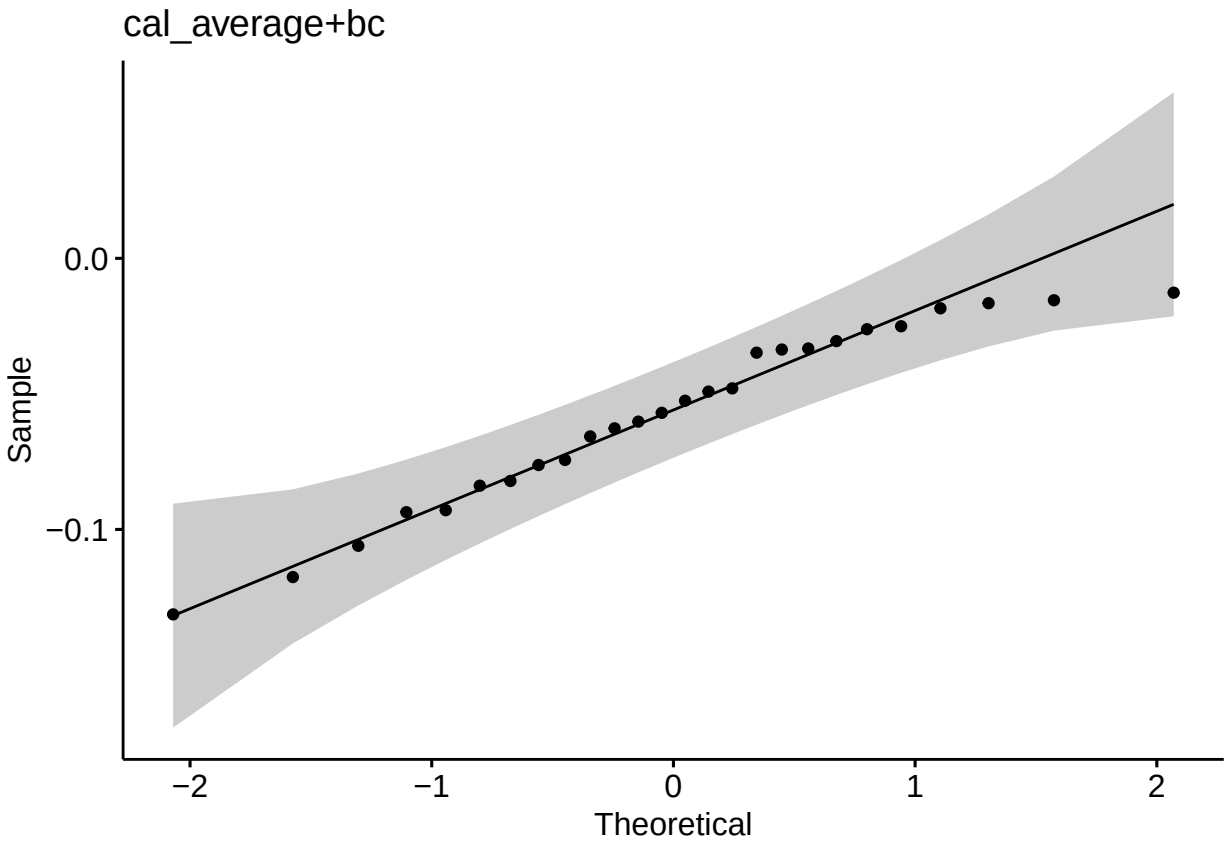

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



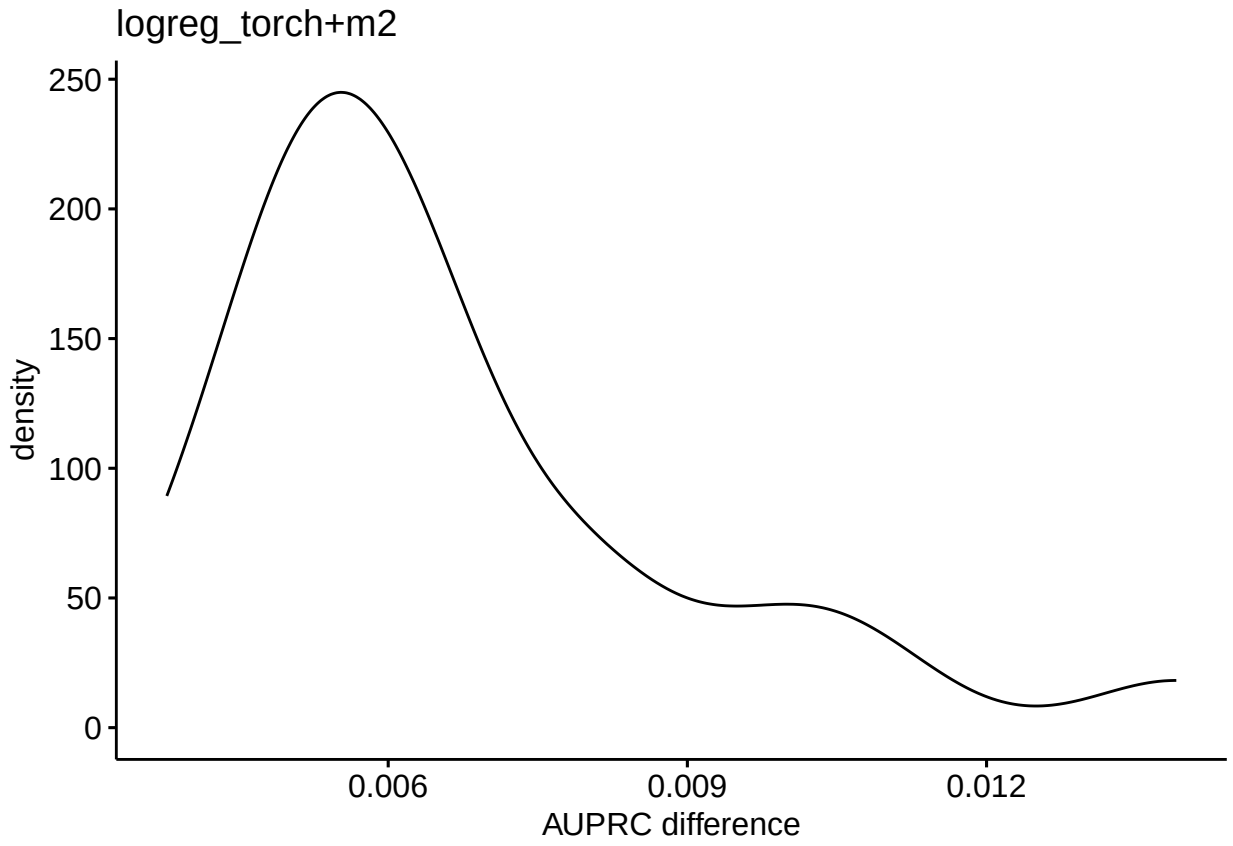

```
## [1] "Method cal_average+bc"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.95216, p-value = 0.2604
```



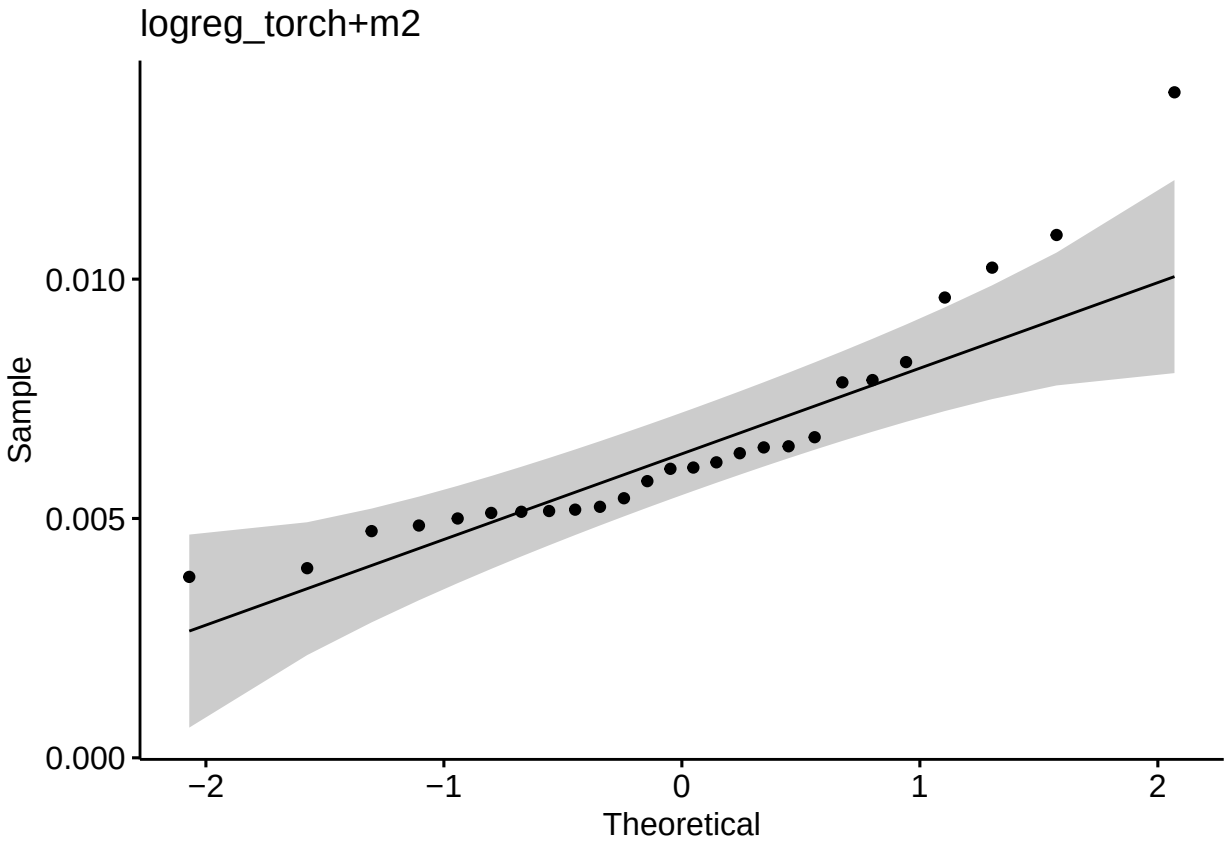

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



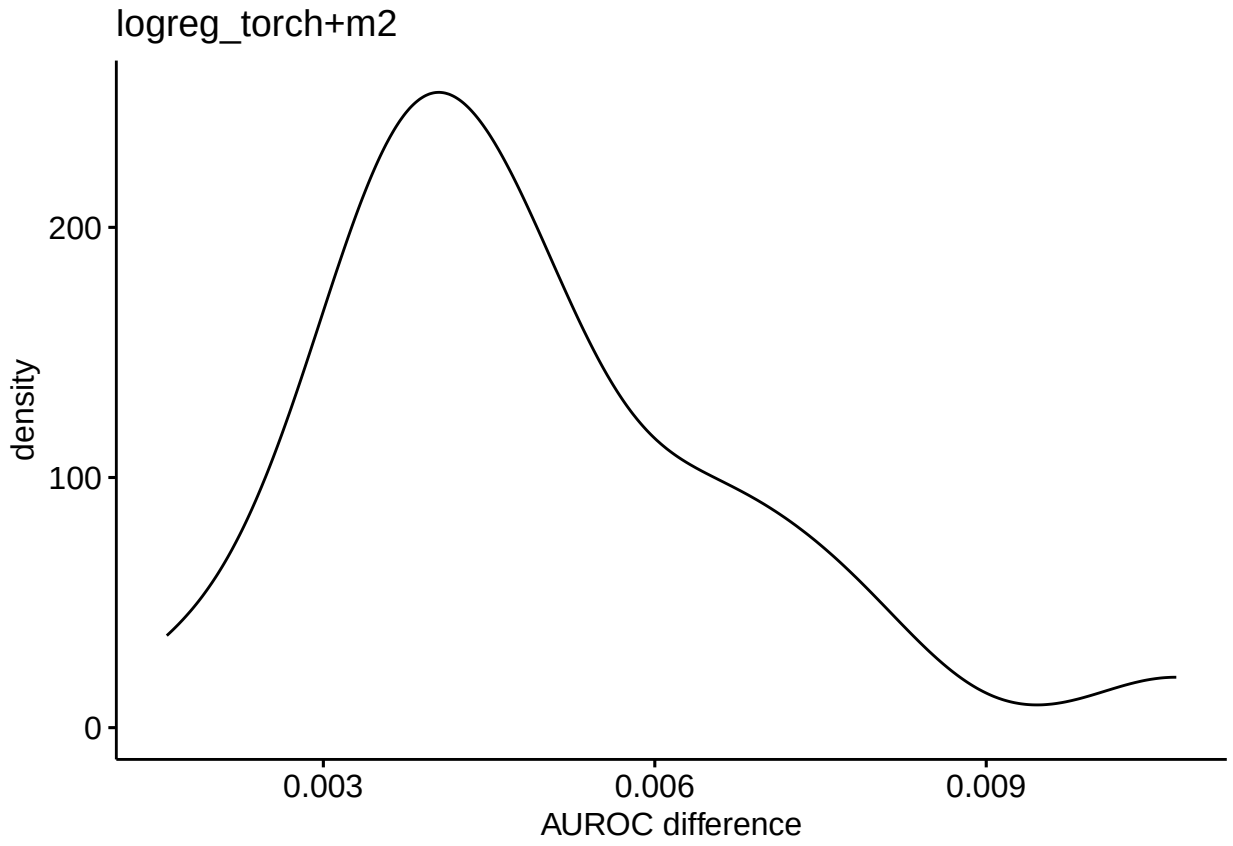

```
## [1] "Method cal_average+bc"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.95062, p-value = 0.2398
##
## [1] "Method logreg_torch+m2"
```



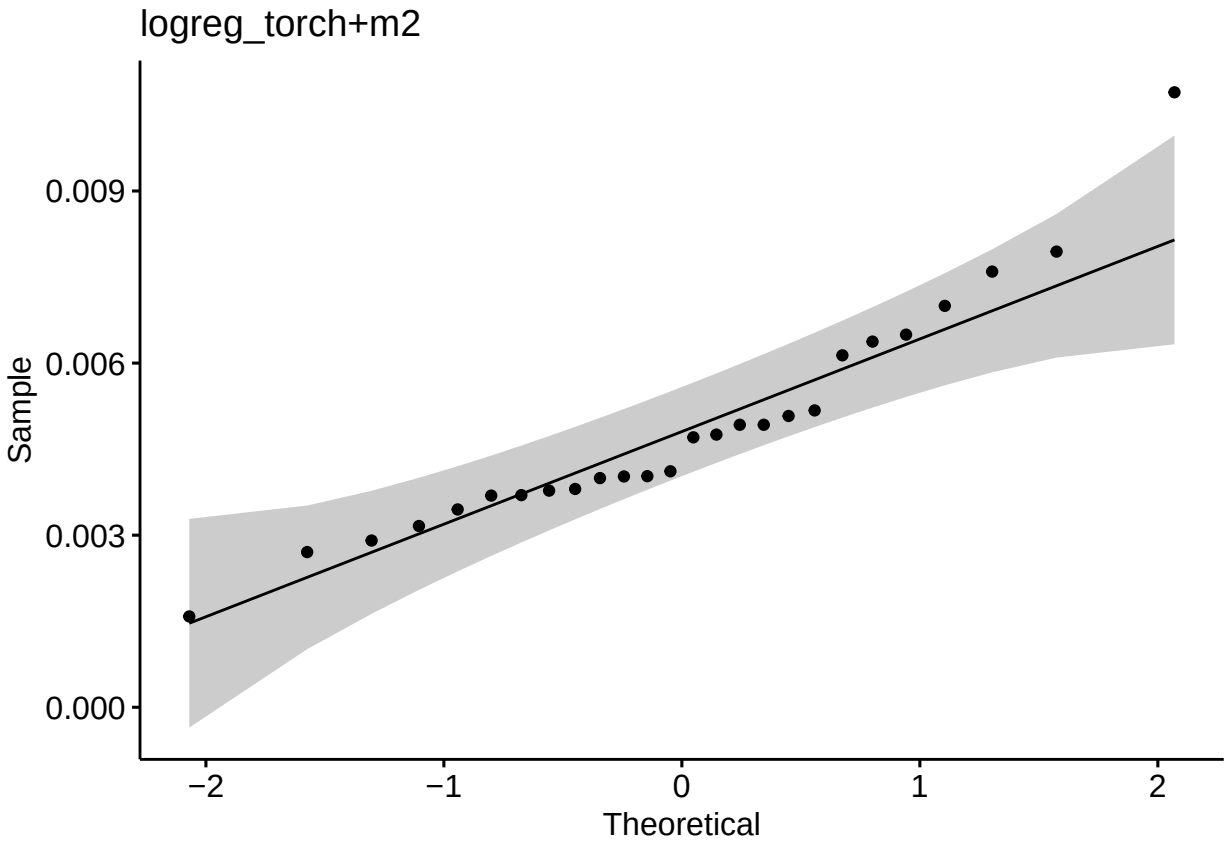

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



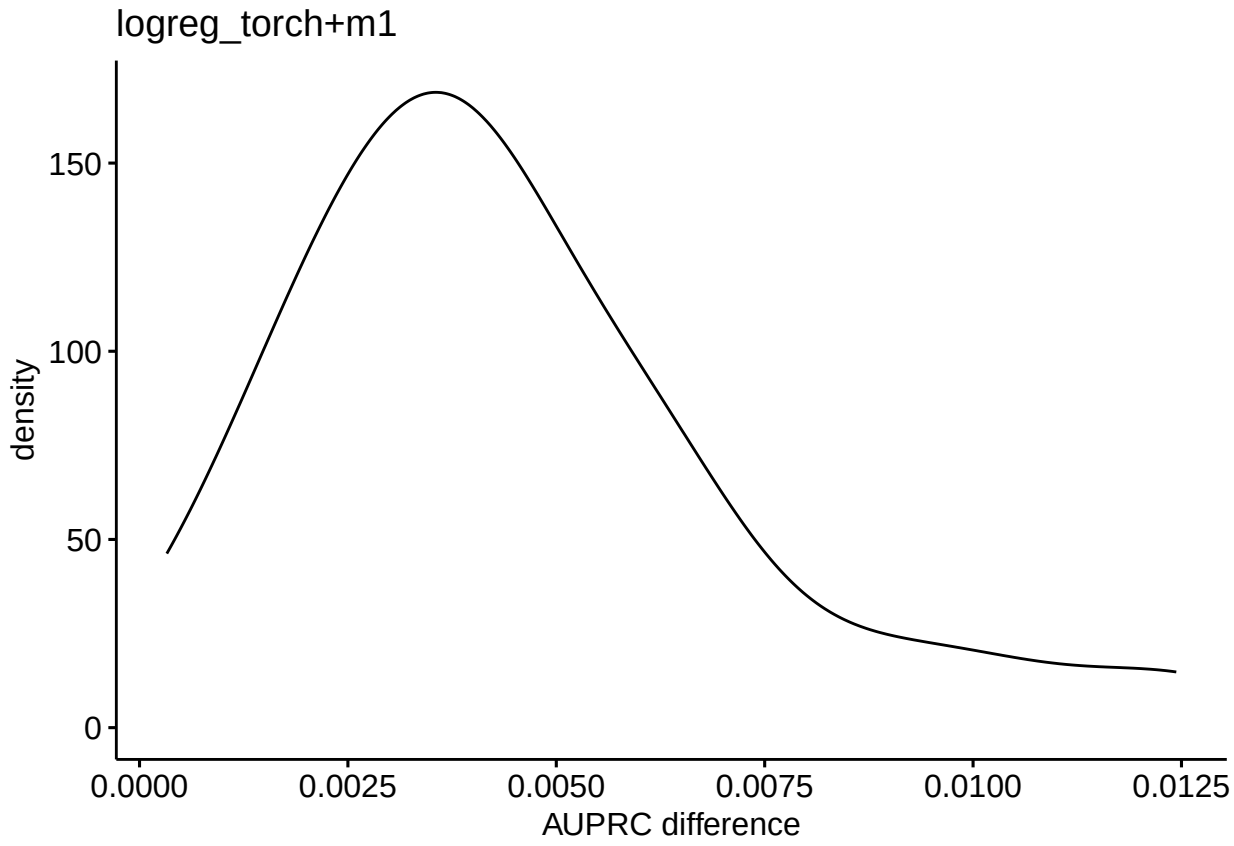

```
## [1] "Method logreg_torch+m2"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.84958, p-value = 0.001396
```



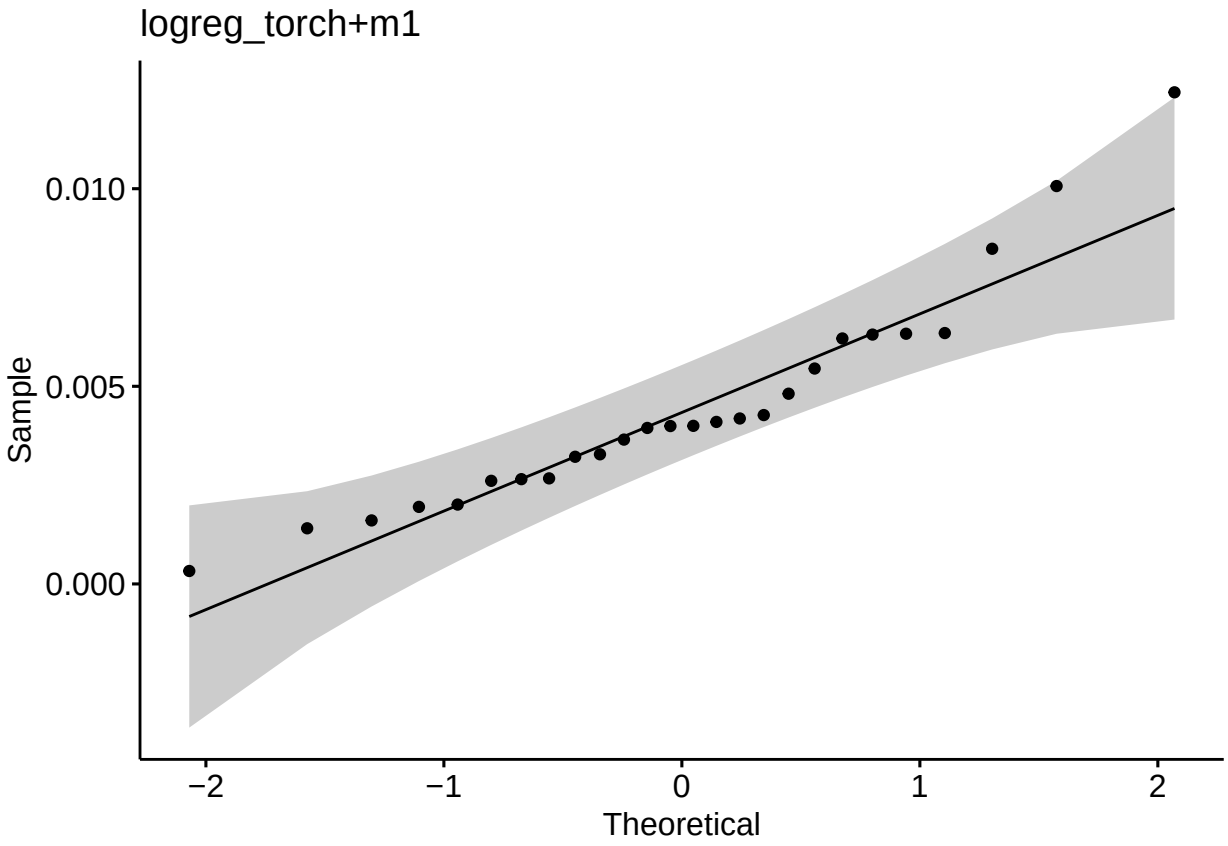

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



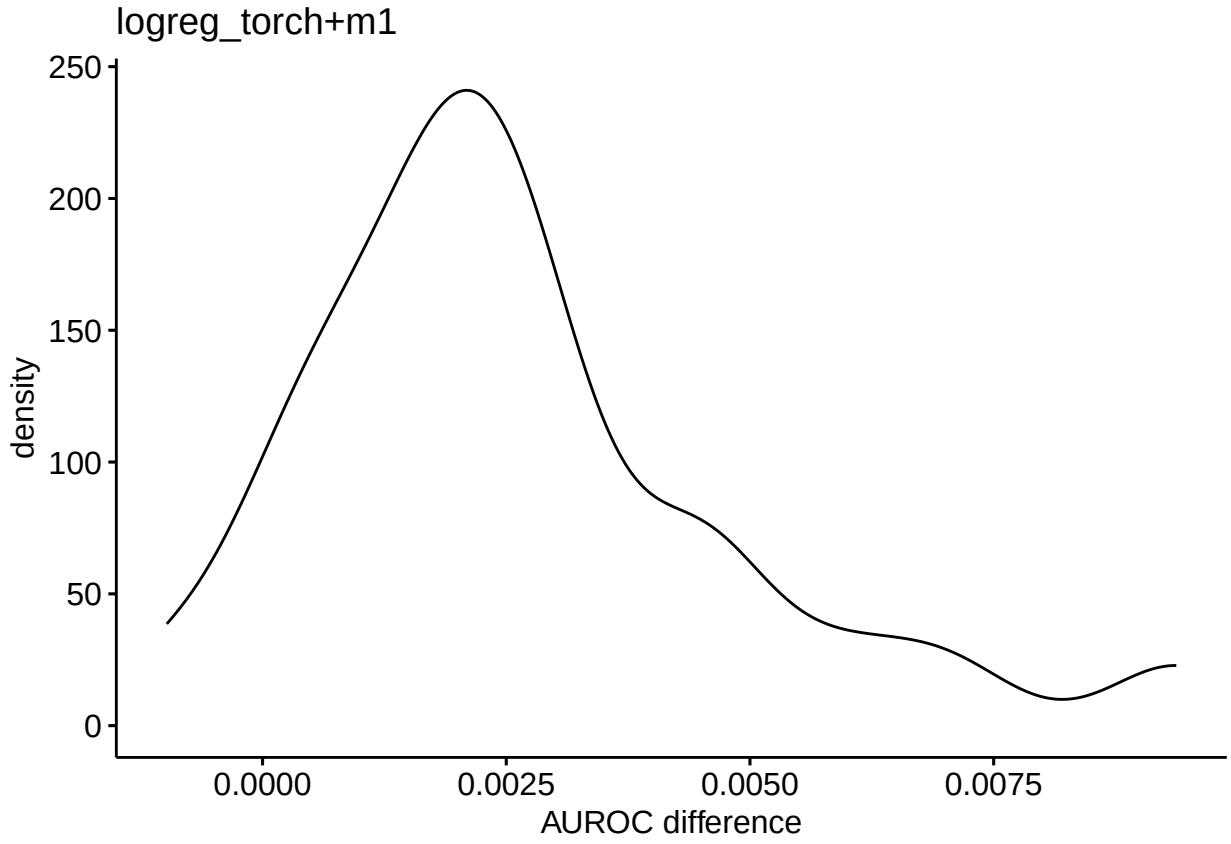

```
## [1] "Method logreg_torch+m2"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.92006, p-value = 0.04511
##
## [1] "Method logreg_torch+m1"
```



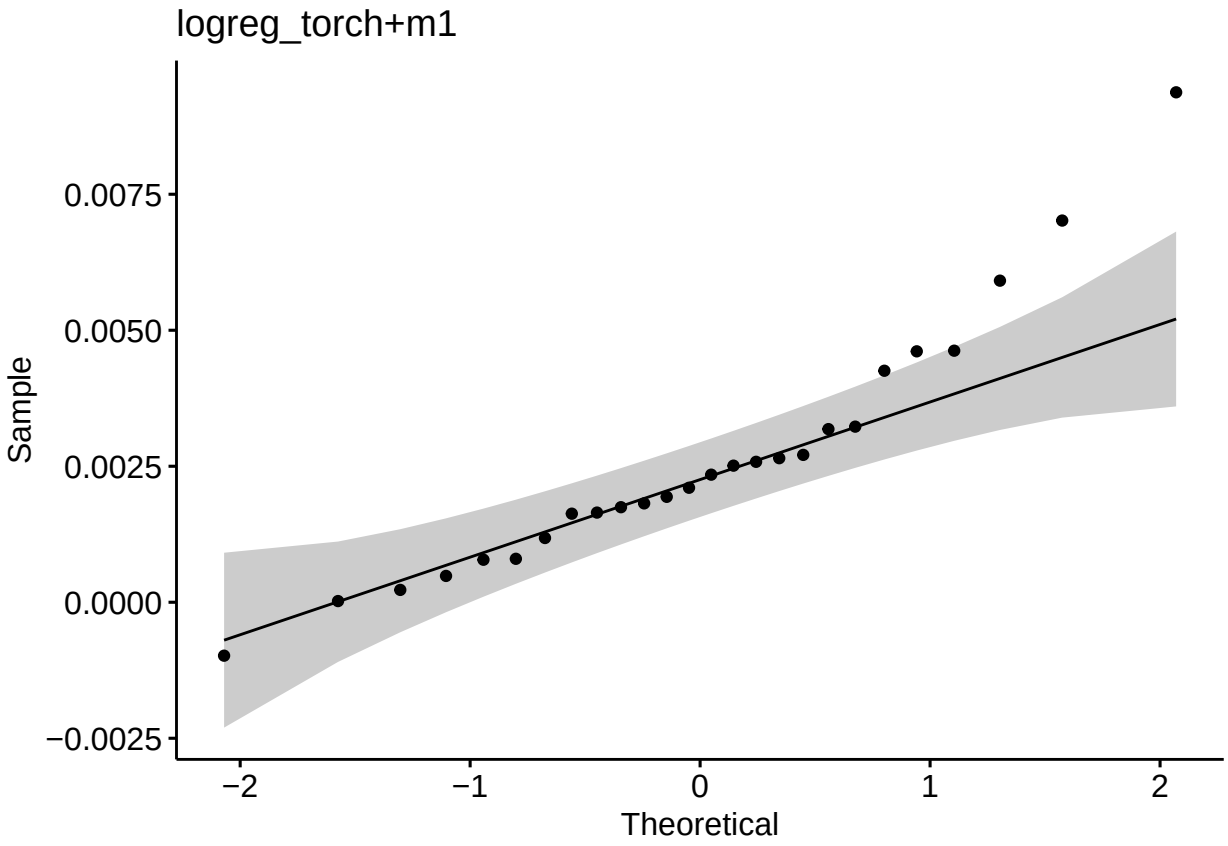

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



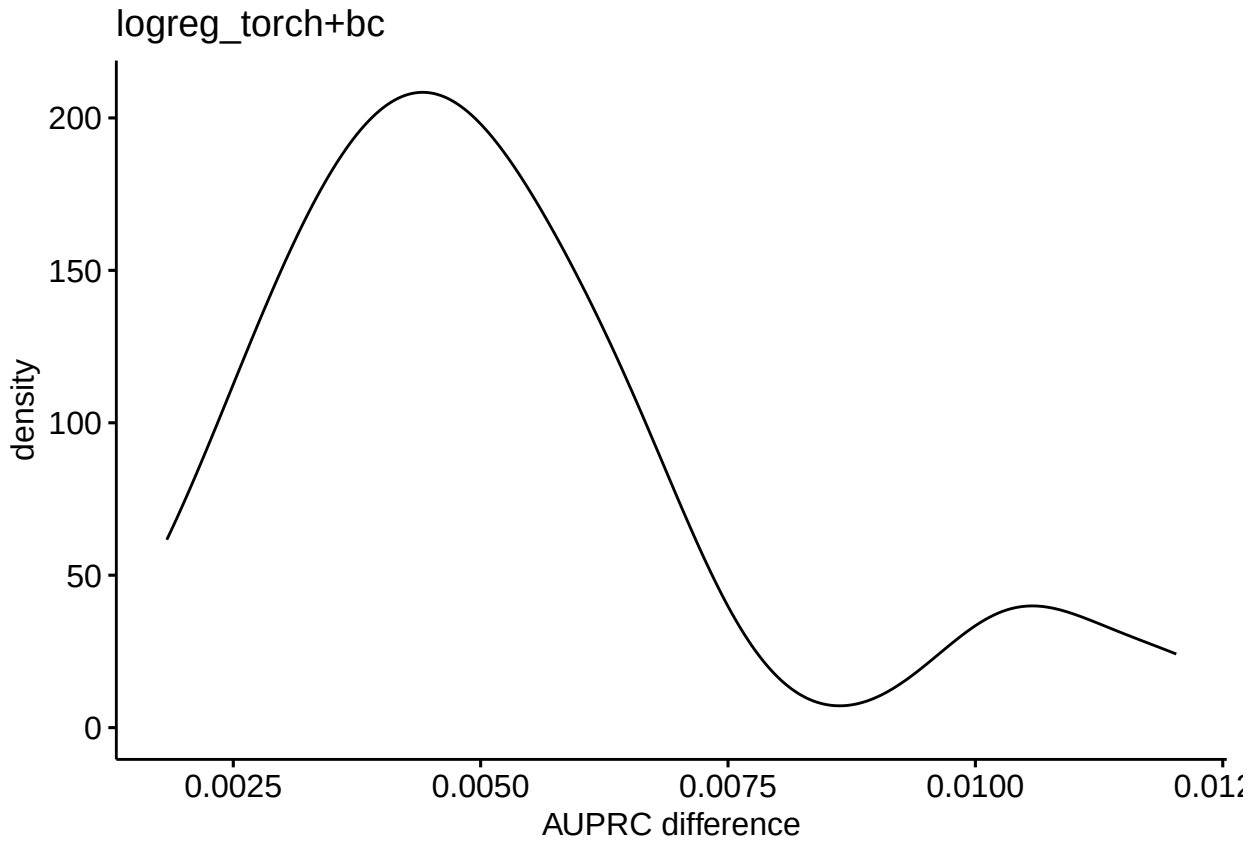

```
## [1] "Method logreg_torch+m1"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.90823, p-value = 0.02402
```



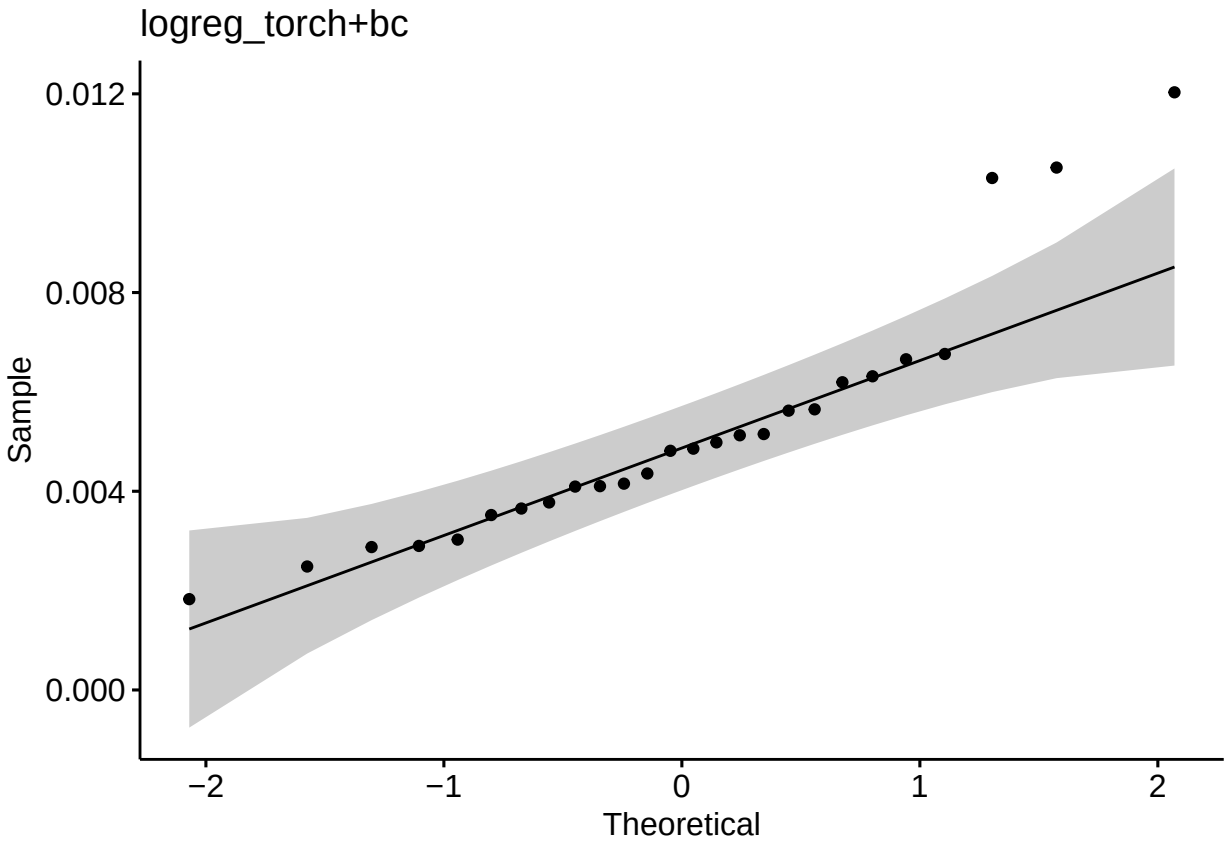

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



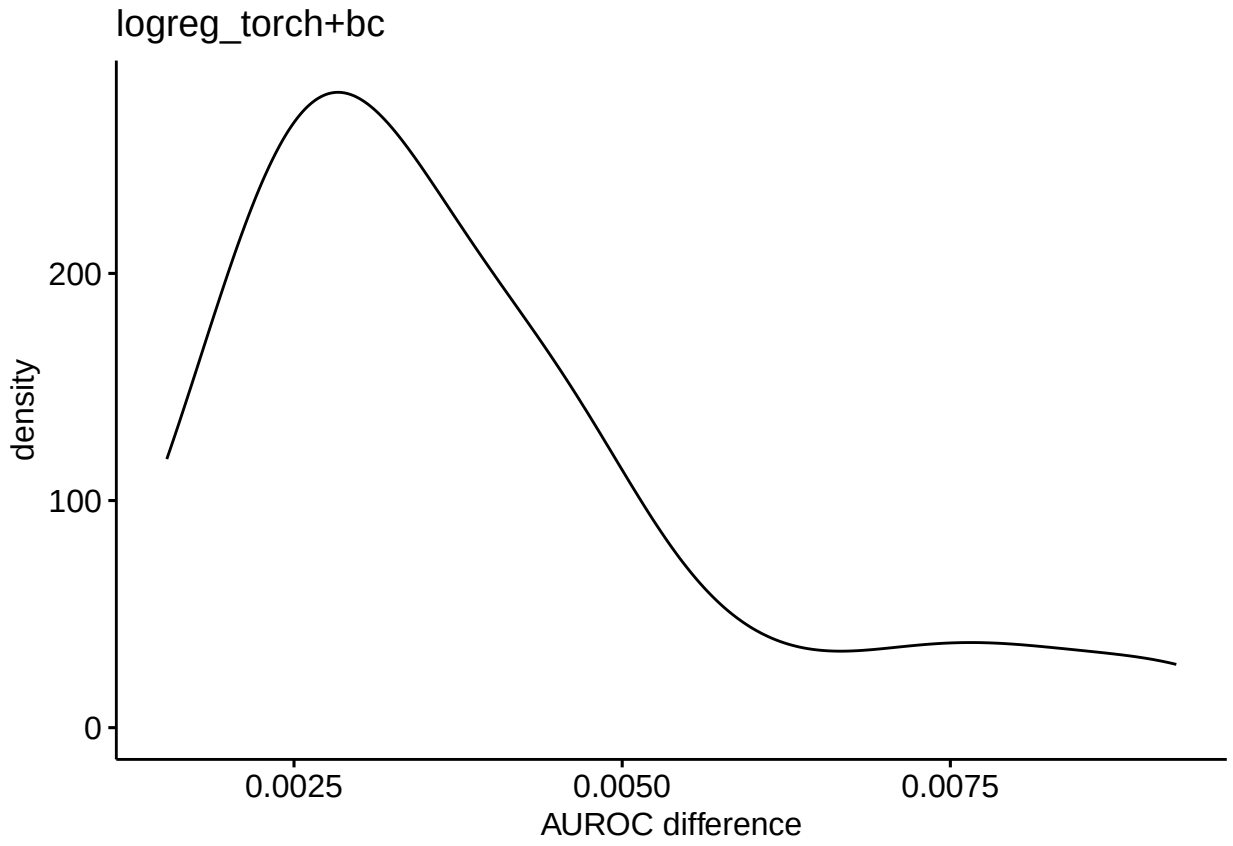

```
## [1] "Method logreg_torch+m1"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.91527, p-value = 0.03488
##
## [1] "Method logreg_torch+bc"
```



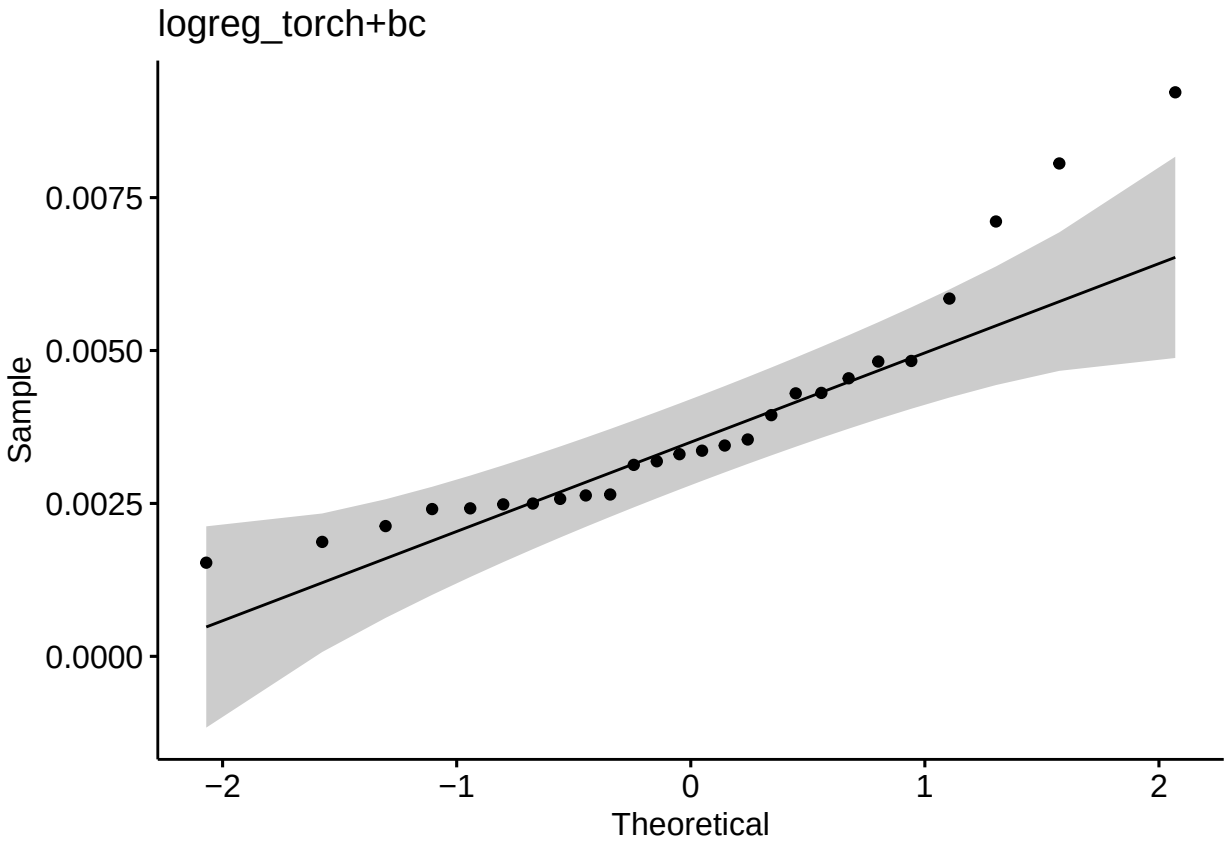

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



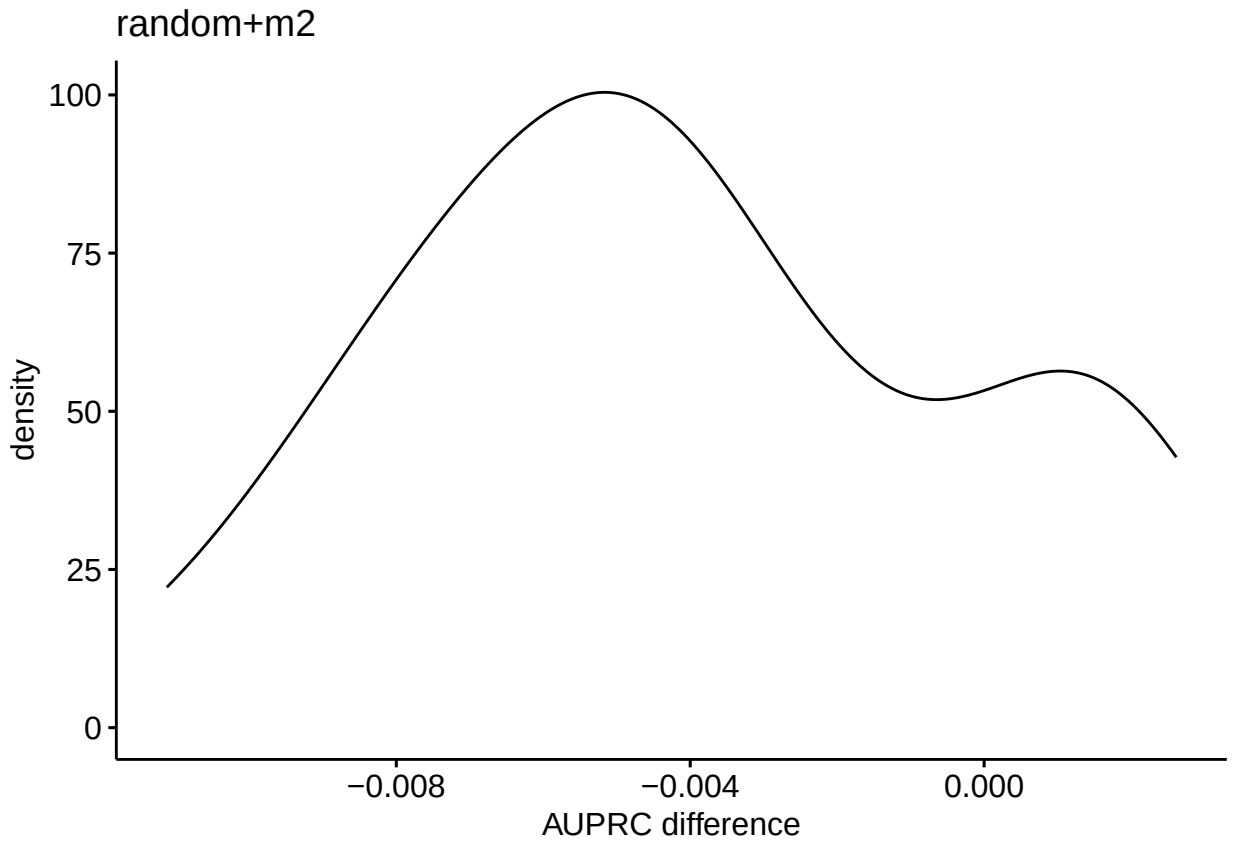

```
## [1] "Method logreg_torch+bc"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.86821, p-value = 0.003277
```



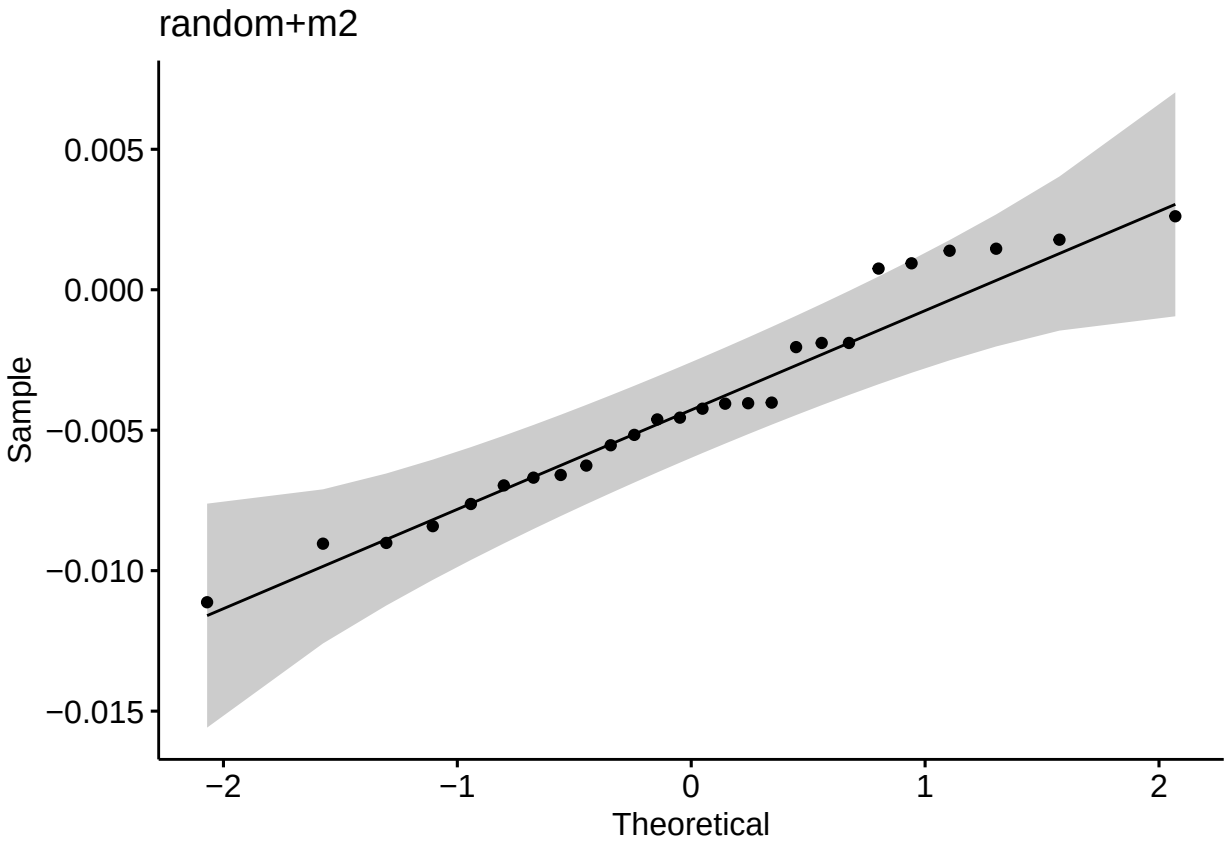

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



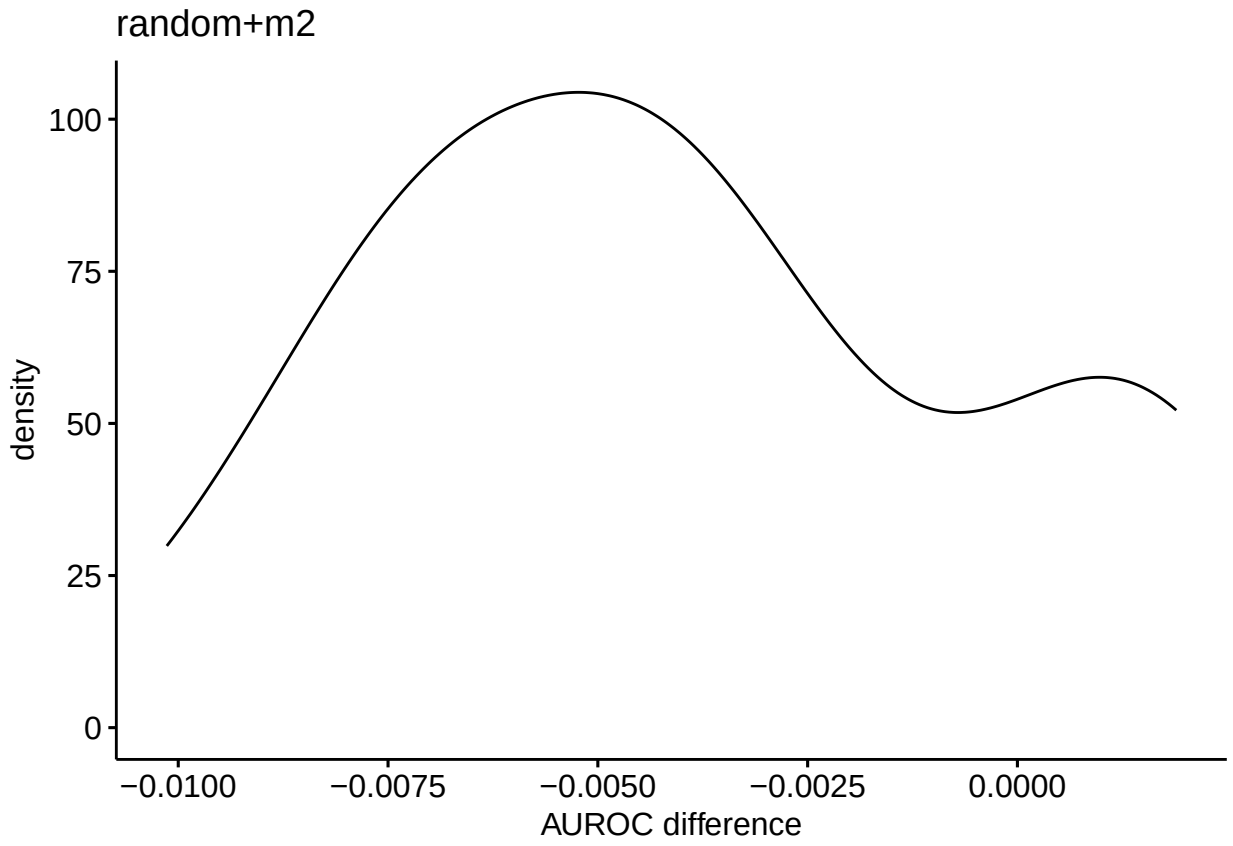

```
## [1] "Method logreg_torch+bc"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.85766, p-value = 0.00201
##
## [1] "Method random+m2"
```



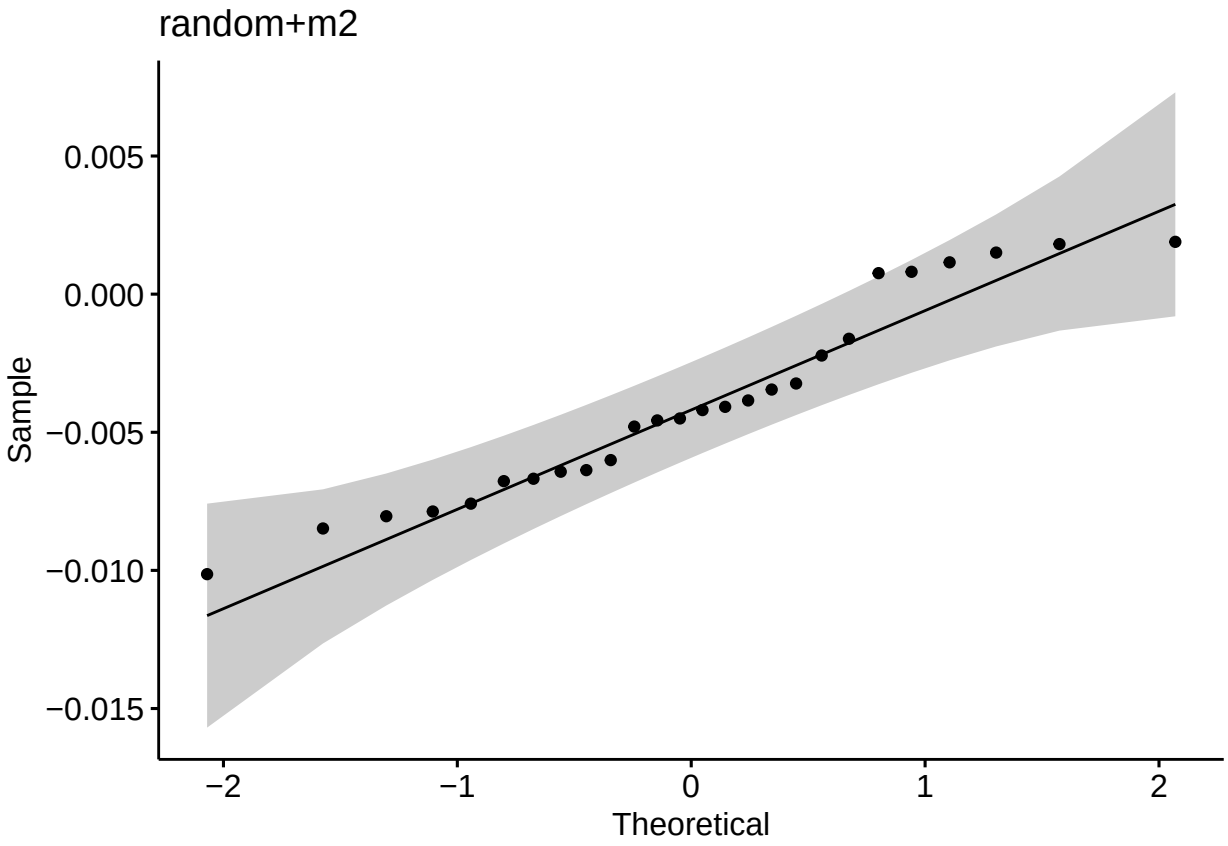

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



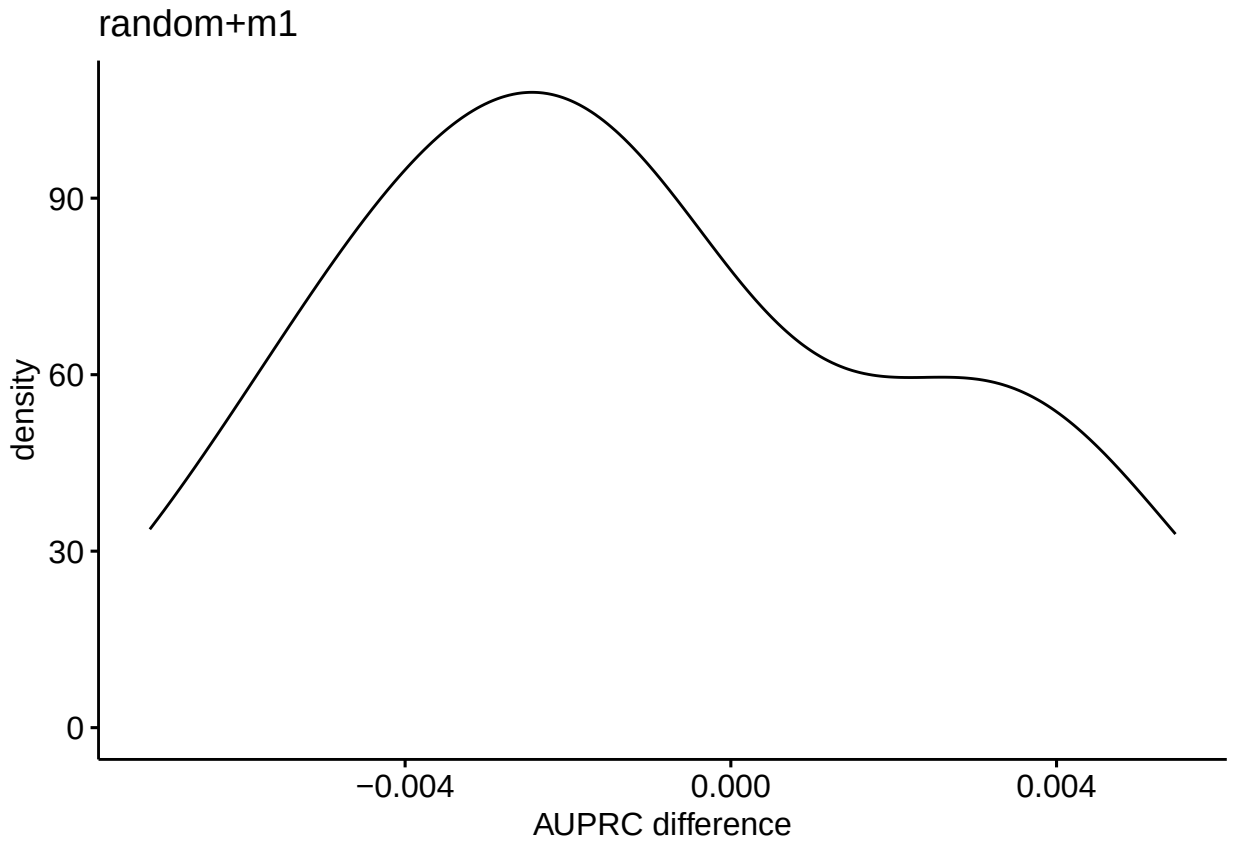

```
## [1] "Method random+m2"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.95471, p-value = 0.2982
```



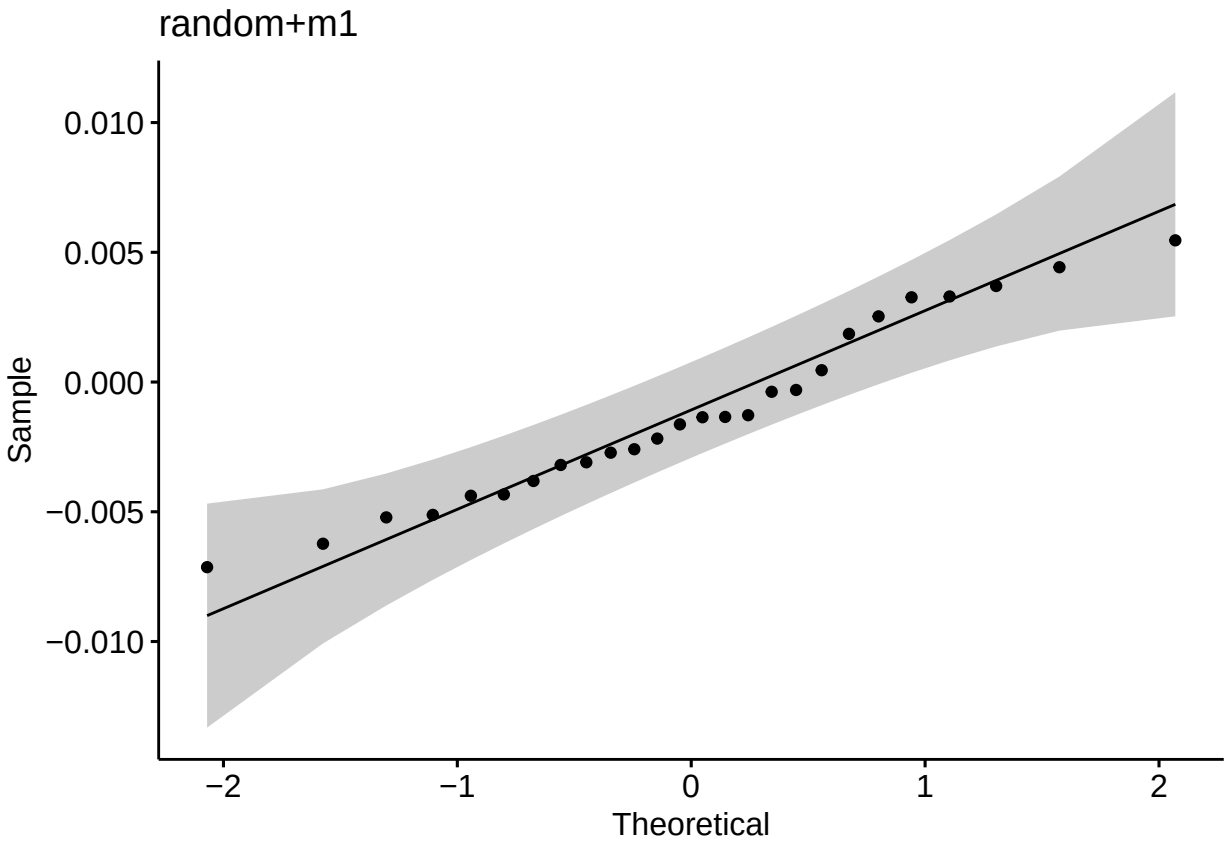

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



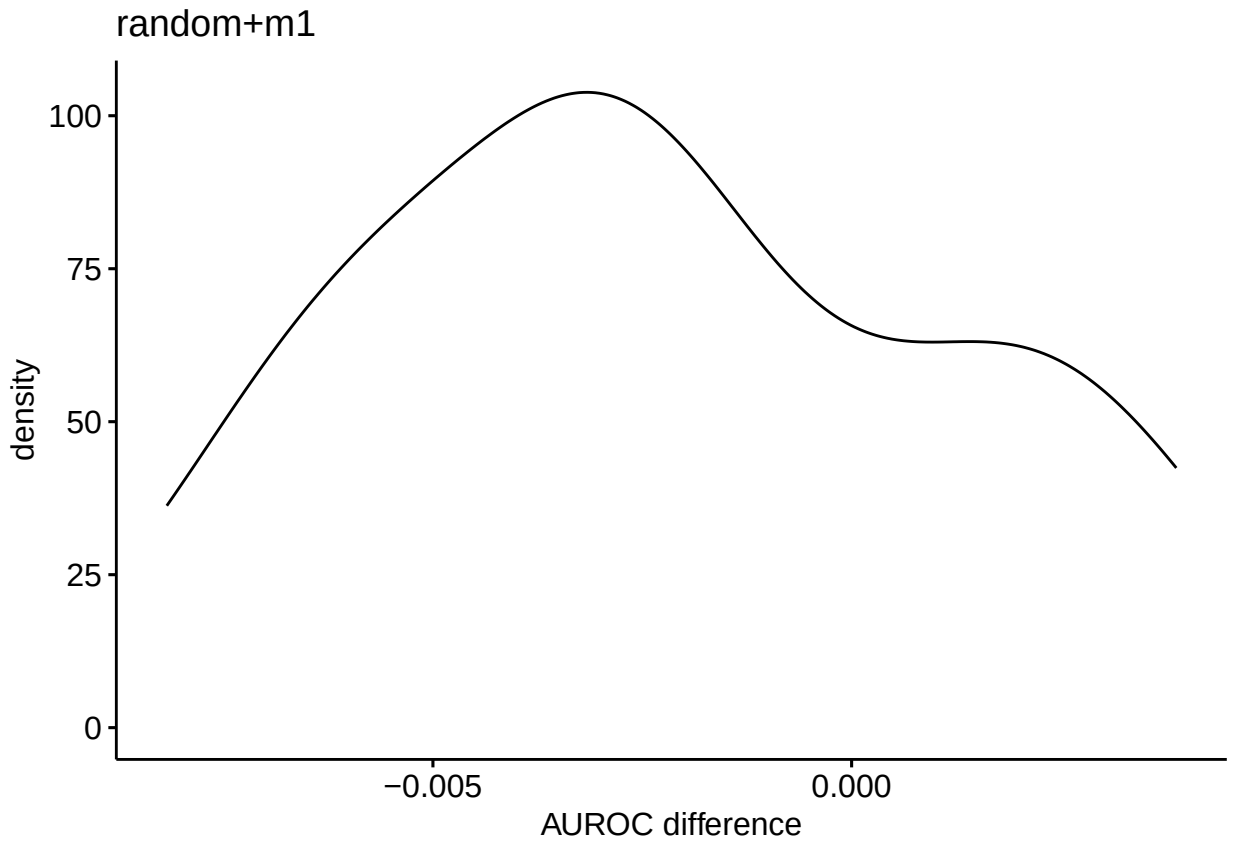

```
## [1] "Method random+m2"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.93671, p-value = 0.1121
##
## [1] "Method random+m1"
```



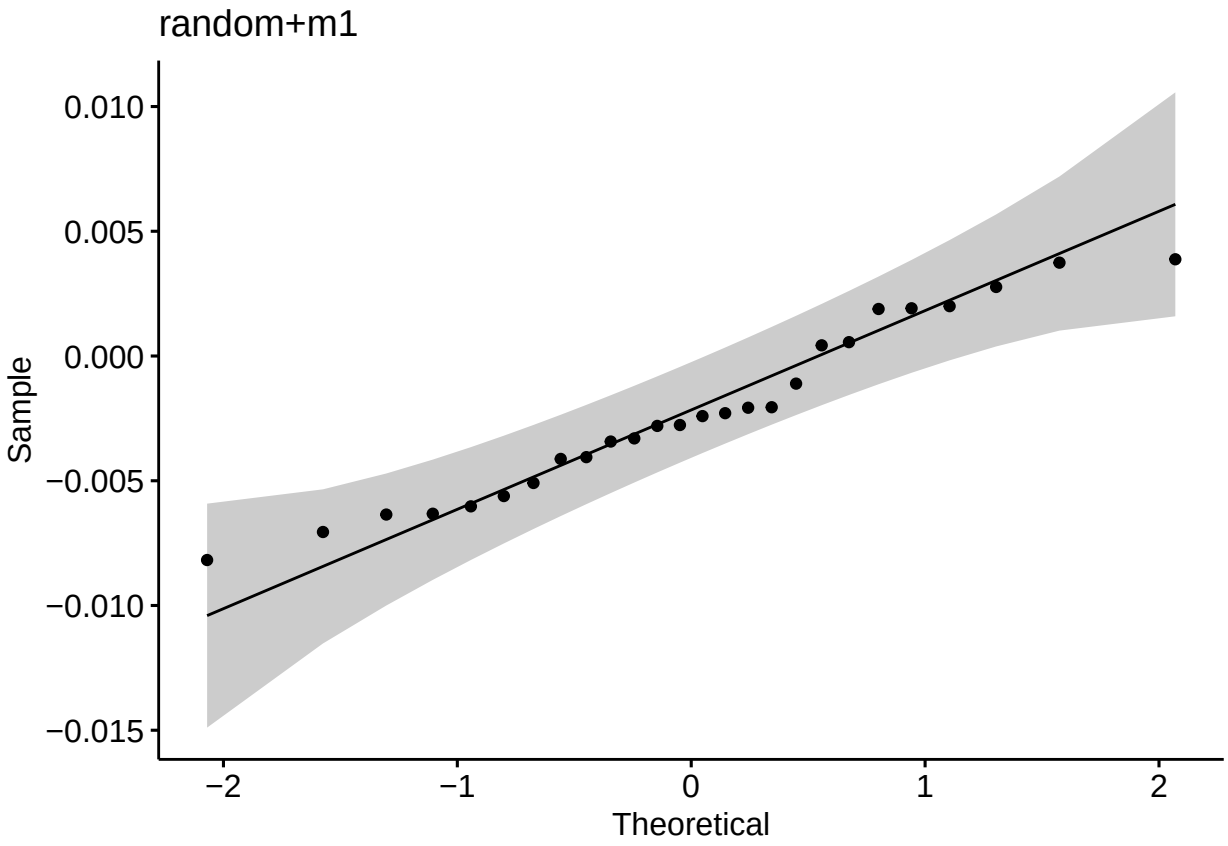

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



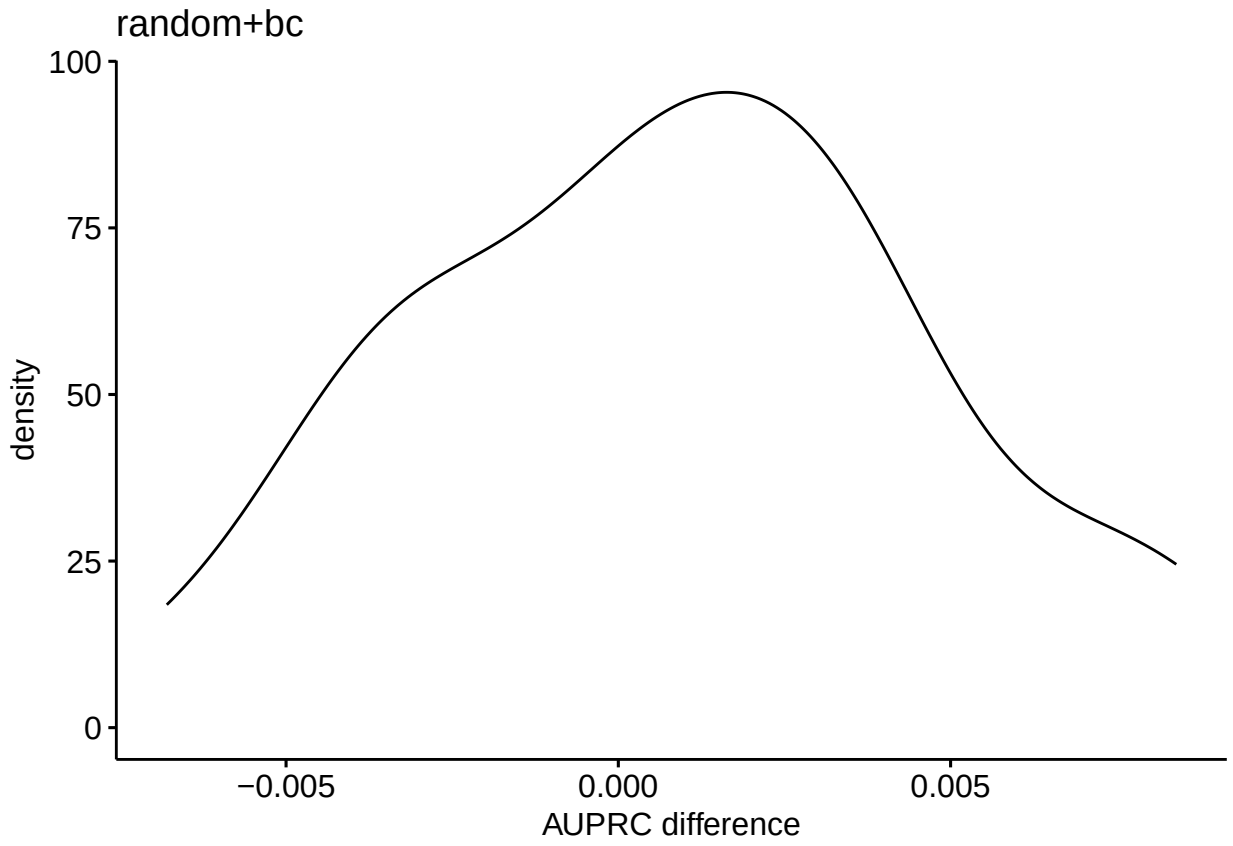

```
## [1] "Method random+m1"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.96382, p-value = 0.4722
```



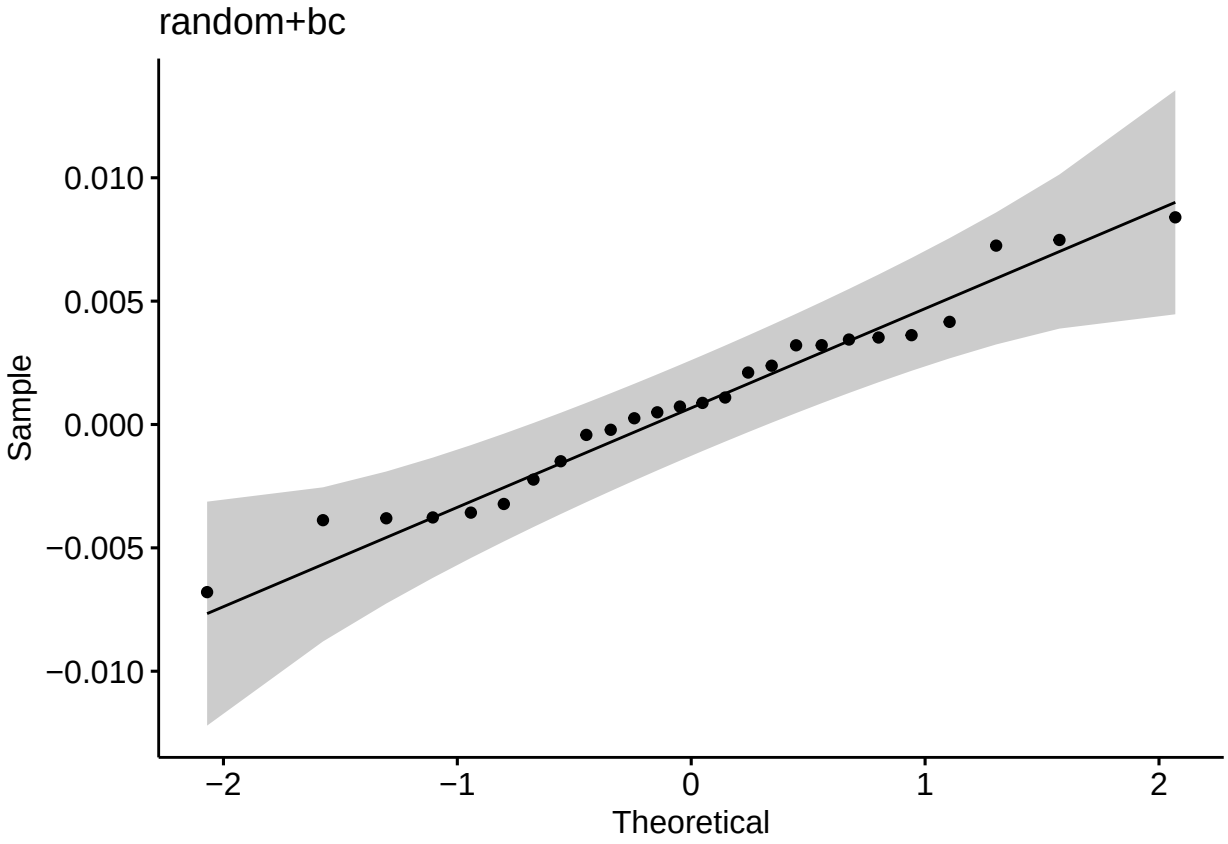

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```



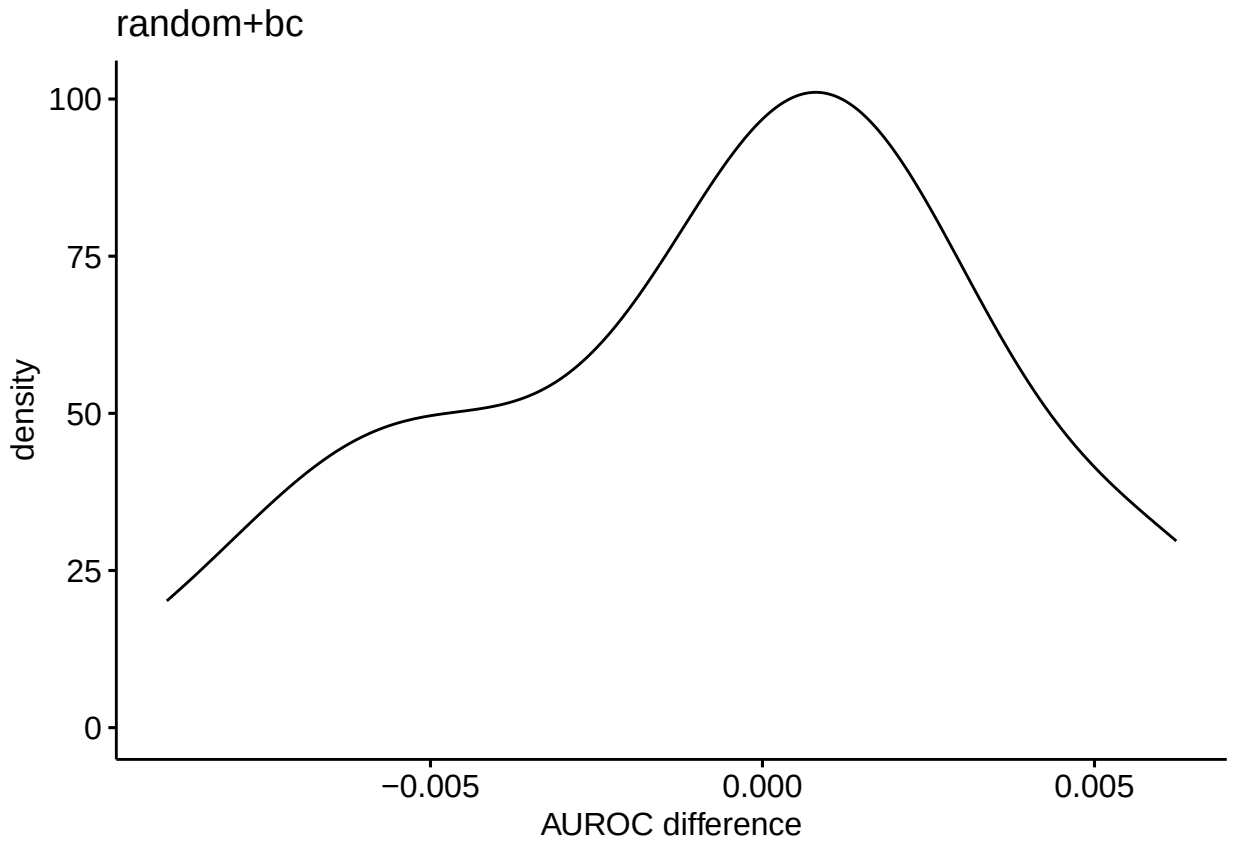

```
## [1] "Method random+m1"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.95634, p-value = 0.3246
##
## [1] "Method random+bc"
```



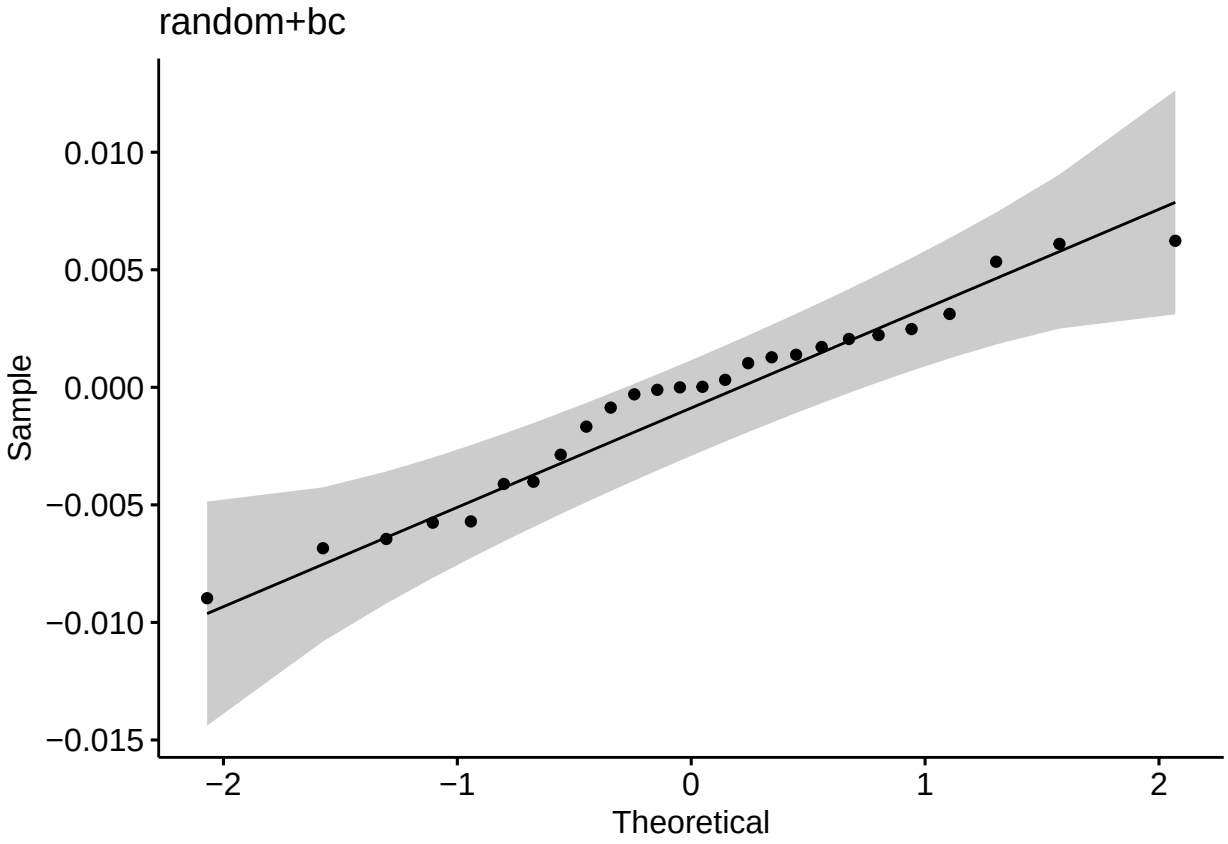

```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```




```
## [1] "Method random+bc"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUPRC
## W = 0.97148, p-value = 0.6619
```




```
## Warning: The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
## The following aesthetics were dropped during statistical transformation: sample
## i This can happen when ggplot fails to infer the correct grouping structure in
##   the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
##   variable into a factor?
```




```
## [1] "Method random+bc"
##
##  Shapiro-Wilk normality test
##
## data:  metrics_co_cp$DIFF_AUROC
## W = 0.96176, p-value = 0.4273
```