

2025-06-27

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
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

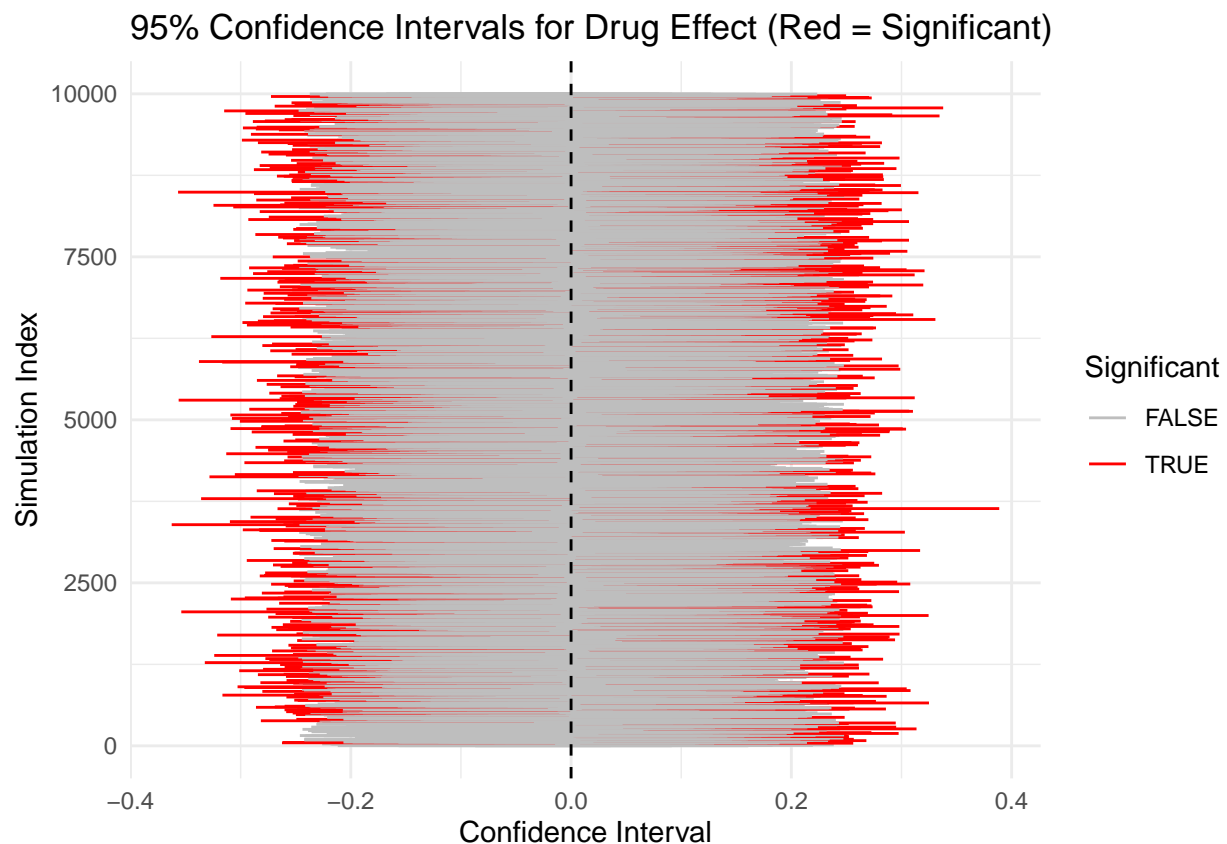
set.seed(123)
n_sim <- 10000
n <- 500
alpha <- 0.05

lower_bounds <- numeric(n_sim)
upper_bounds <- numeric(n_sim)
significant <- logical(n_sim)

for (i in 1:n_sim) {
  treatment <- rnorm(n, mean = 0, sd = 1)
  control <- rnorm(n, mean = 0, sd = 1)
  diff <- mean(treatment) - mean(control)
  se <- sqrt(1/n + 1/n)
  margin <- qnorm(1 - alpha/2) * se
  lower_bounds[i] <- diff - margin
  upper_bounds[i] <- diff + margin
  significant[i] <- (lower_bounds[i] > 0 | upper_bounds[i] < 0)
}
```

```
df <- data.frame(
  index = 1:n_sim,
  lower = lower_bounds,
  upper = upper_bounds,
  significant = significant
)

ggplot(df, aes(y = index)) +
  geom_segment(aes(x = lower, xend = upper, yend = index, color = significant)) +
  geom_vline(xintercept = 0, linetype = "dashed") +
  scale_color_manual(values = c("FALSE" = "gray", "TRUE" = "red")) +
  labs(
    title = "95% Confidence Intervals for Drug Effect (Red = Significant)",
    x = "Confidence Interval",
    y = "Simulation Index",
    color = "Significant"
  ) +
  theme_minimal()
```



```
signif_count <- sum(df$significant)
prop_signif <- mean(df$significant)
cat("Number of significant results:", signif_count, "\n")
```

```
## Number of significant results: 512
```

```
cat("Proportion of significant results:", prop_signif, "\n")
```

```
## Proportion of significant results: 0.0512
```

```
set.seed(1234)
treatment_sample <- rnorm(n, mean = 0, sd = 1)
control_sample <- rnorm(n, mean = 0, sd = 1)

mean_treatment <- mean(treatment_sample)
mean_control <- mean(control_sample)
var_treatment <- var(treatment_sample)
var_control <- var(control_sample)

mean_diff <- mean_treatment - mean_control

se_diff <- sqrt(var_treatment/n + var_control/n)

z_critical <- qnorm(1 - alpha/2)
margin_of_error <- z_critical * se_diff
lower_bound <- mean_diff - margin_of_error
upper_bound <- mean_diff + margin_of_error

cat("Treatment Mean:", mean_treatment, "\n")
```

```
## Treatment Mean: 0.001838821
```

```
cat("Treatment Variance:", var_treatment, "\n")
```

```
## Treatment Variance: 1.07084
```

```
cat("Control Mean:", mean_control, "\n")
```

```
## Control Mean: -0.05503322
```

```
cat("Control Variance:", var_control, "\n")
```

```
## Control Variance: 0.9188981
```

```
cat("Difference in Means:", mean_diff, "\n")
```

```
## Difference in Means: 0.05687204
```

```
cat("Standard Error:", se_diff, "\n")
```

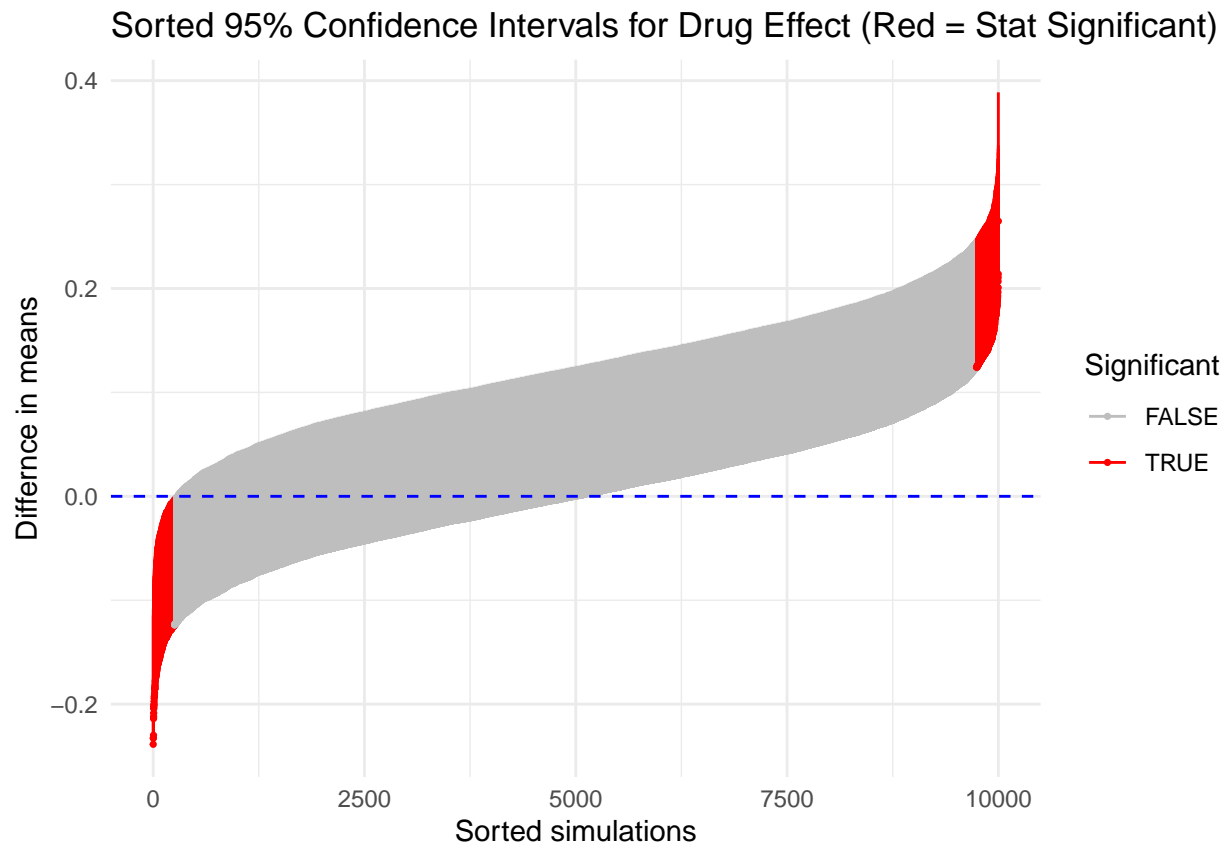
```
## Standard Error: 0.06308309
```

```
cat("95% Confidence Interval: [", lower_bound, ",", upper_bound, "]\n")
```

```
## 95% Confidence Interval: [ -0.06676854 , 0.1805126 ]
```

```
df <- df %>%
  mutate(
    mean_diff = (lower + upper) / 2,
    margin = (upper - lower) / 2
  ) %>%
  arrange(mean_diff) %>%
  mutate(sorted_index = row_number())

ggplot(df, aes(x = sorted_index, y = mean_diff, color = significant)) +
  geom_errorbar(aes(ymin = mean_diff, ymax = mean_diff + margin), width = 0.4) +
  geom_point(size = 0.6) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "blue") +
  scale_color_manual(values = c("FALSE" = "gray", "TRUE" = "red")) +
  labs(
    title = "Sorted 95% Confidence Intervals for Drug Effect (Red = Stat Significant)",
    x = "Sorted simulations",
    y = "Difference in means",
    color = "Significant"
  ) +
  theme_minimal()
```



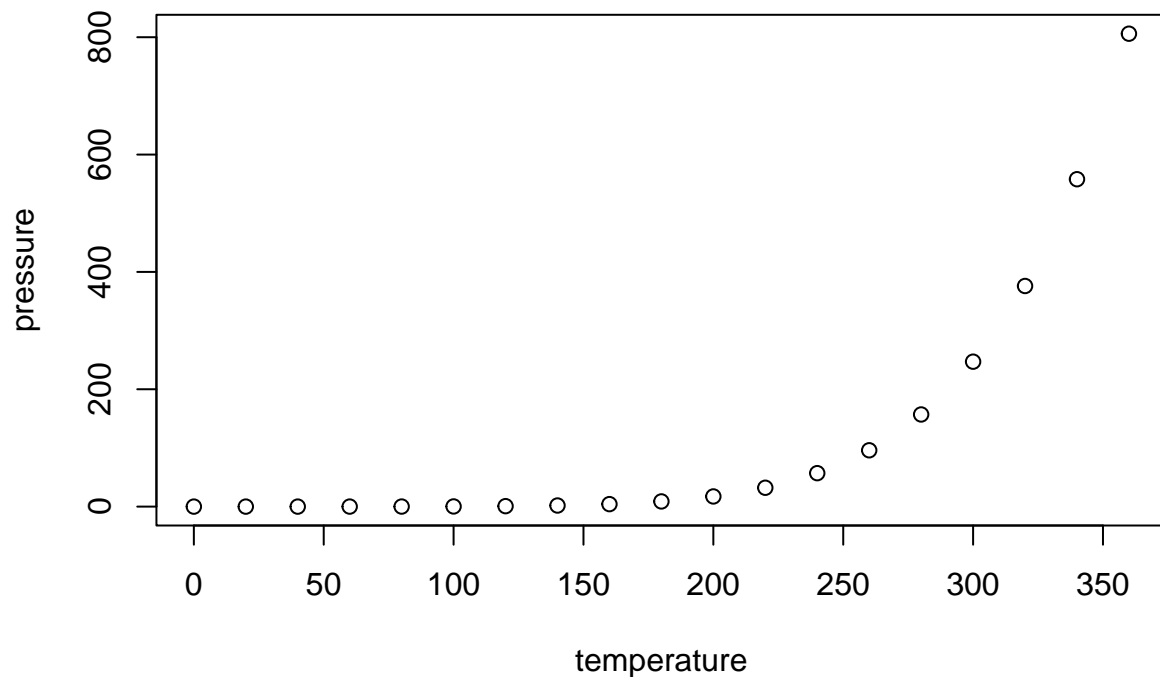
```
#Number of significant results: 512  
#the company's policy is problematic because while the "Proportion of significant results" is around  
#five percent, it means that this porportion of the intervals will not contain the true effect  
#
```

```
summary(cars)
```

```
##      speed      dist  
##  Min.   : 4.0    Min.   :  2.00  
## 1st Qu.:12.0    1st Qu.: 26.00  
##  Median :15.0    Median : 36.00  
##   Mean  :15.4    Mean   : 42.98  
## 3rd Qu.:19.0    3rd Qu.: 56.00  
##   Max.  :25.0    Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.