INL2_analysis

Asli

2025-09-21

```
#Analysis Part of INL2

# 1) Data (Table 2)
# ----------------------------
x      = 1:10                          # programmers (index)
prog1  = c(104,102,159,168,150,151,111,105,137,124)       # Prog-1 minutes
prog2  = c(71.3,110,178,153,120,174,94.9,86.1,115,175)    # Prog-2 minutes

# ----------------------------
# 2) Descriptive statistics
# ----------------------------
summ = function(v) c(
  n      = length(v),
  mean   = mean(v),
  sd     = sd(v),
  median = median(v),
  IQR    = IQR(v),
  min    = min(v),
  max    = max(v)
)

rbind(
  `Prog-1` = summ(prog1),
  `Prog-2` = summ(prog2)
)
```

```
##         n   mean       sd median    IQR   min max
## Prog-1 10 131.10 25.04418  130.5 44.250 102.0 168
## Prog-2 10 127.73 39.54236  117.5 70.075  71.3 178
```

```
# ----------------------------
# 3) Map to IDE-A / IDE-B using Table 1
#    A->B for programmers {1,3,4,7,9}
# ----------------------------
A_then_B = c(1,3,4,7,9)
id       = 1:10

IDEA = ifelse(id %in% A_then_B, prog1, prog2)  # IDE-A times
IDEB = ifelse(id %in% A_then_B, prog2, prog1)  # IDE-B times

# Means by IDE (quick check)
c(mean_IDEA = mean(IDEA), mean_IDEB = mean(IDEB))
```

```
## mean_IDEA mean_IDEB
##    134.41    124.42
```

```
# Paired differences (choose one convention and stick to it)
# Here: diff = IDE-A – IDE-B (positive => A slower than B)
diff = IDEA - IDEB
c(mean_diff = mean(diff), median_diff = median(diff), sd_diff = sd(diff))
```

```
##   mean_diff median_diff     sd_diff
##     9.99000    15.55000    25.49882
```

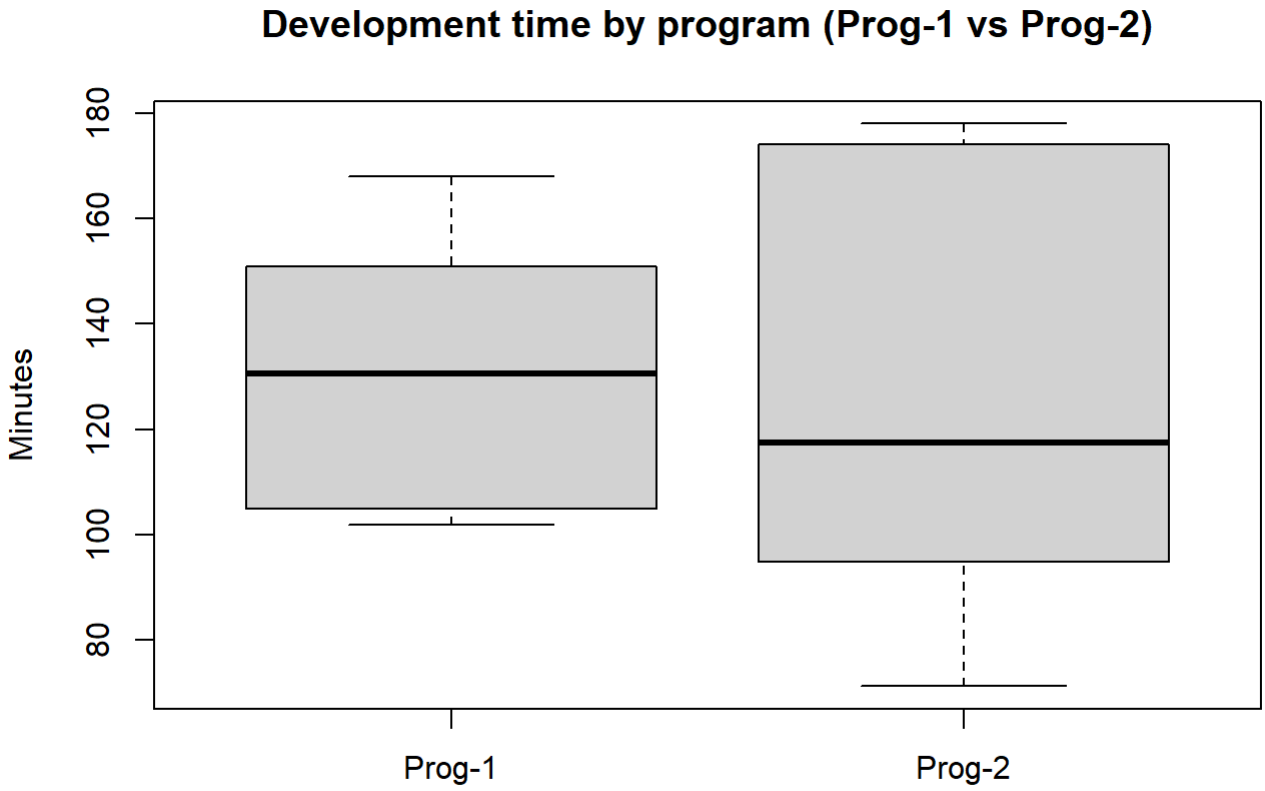
```
# ----------------------------
# 4) Hypothesis test (paired t-test)
#    H0: mu_A = mu_B  <=> mean(diff) = 0
#    H1: mu_A != mu_B
# ----------------------------
test  = t.test(IDEB, IDEA, paired = TRUE)  # equivalent to t.test(diff)
alpha = 0.05

cat("Paired t-test p-value =", test$p.value, "\n")
```

```
## Paired t-test p-value = 0.2467096
```

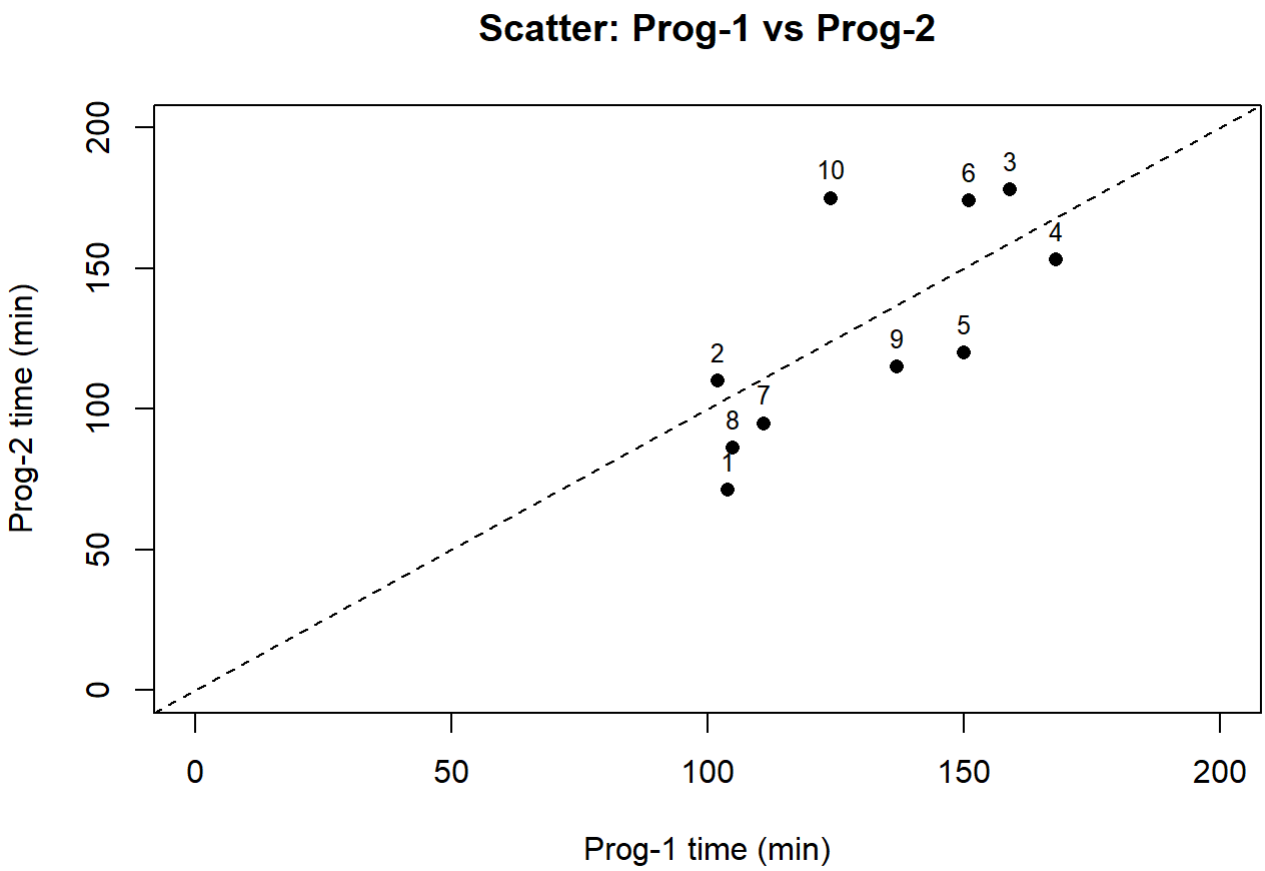
```
if (test$p.value < alpha) {
  cat("Decision: Reject H0 (IDEs differ).\n")
} else {
  cat("Decision: Fail to reject H0 (IDEs considered the same in this sample).\n")
}
```

```
## Decision: Fail to reject H0 (IDEs considered the same in this sample).
```

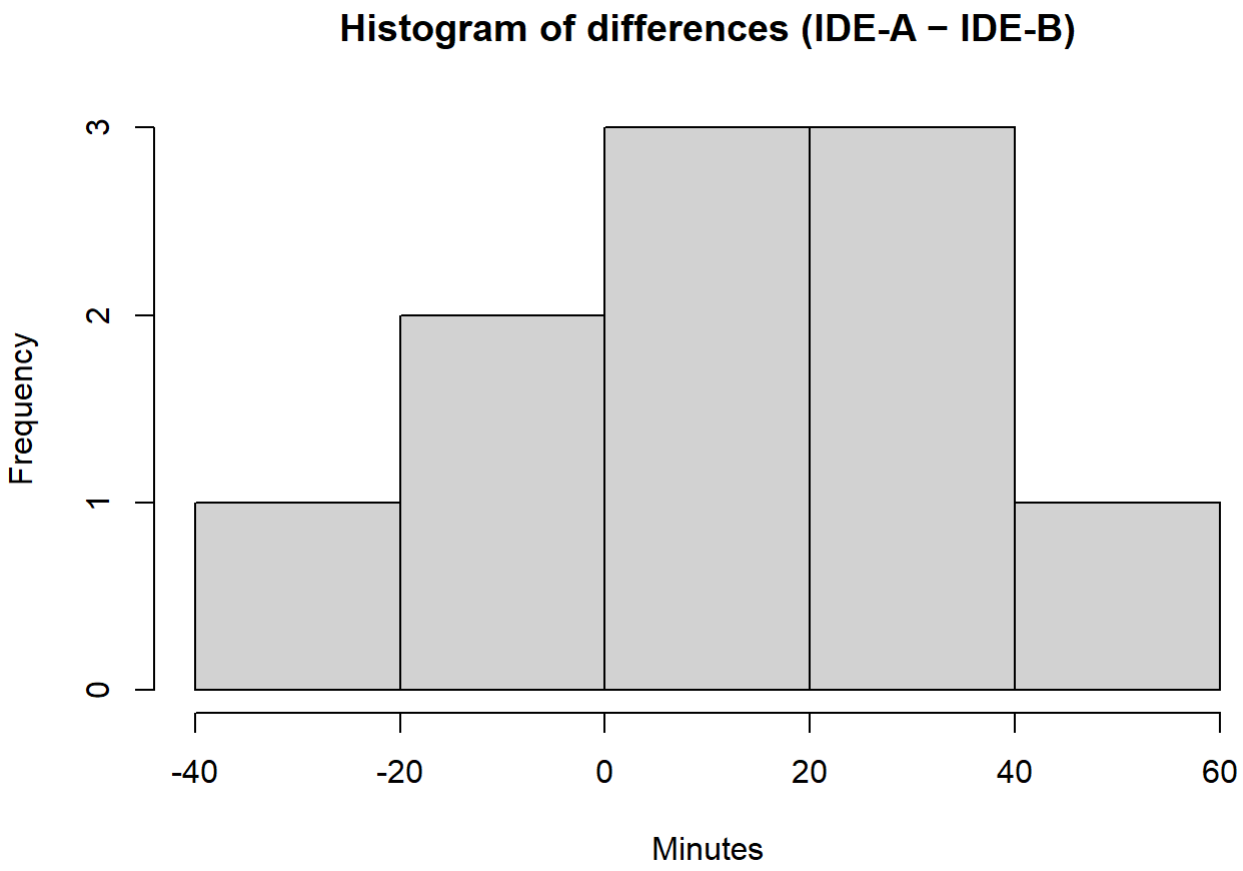
```
# ----------------------------
# 5) Visualizations
# ----------------------------

# 5a) Box plot by Program (Prog-1 vs Prog-2)
boxplot(
  list(`Prog-1` = prog1, `Prog-2` = prog2),
  ylab = "Minutes",
  main = "Development time by program (Prog-1 vs Prog-2)"
)
```



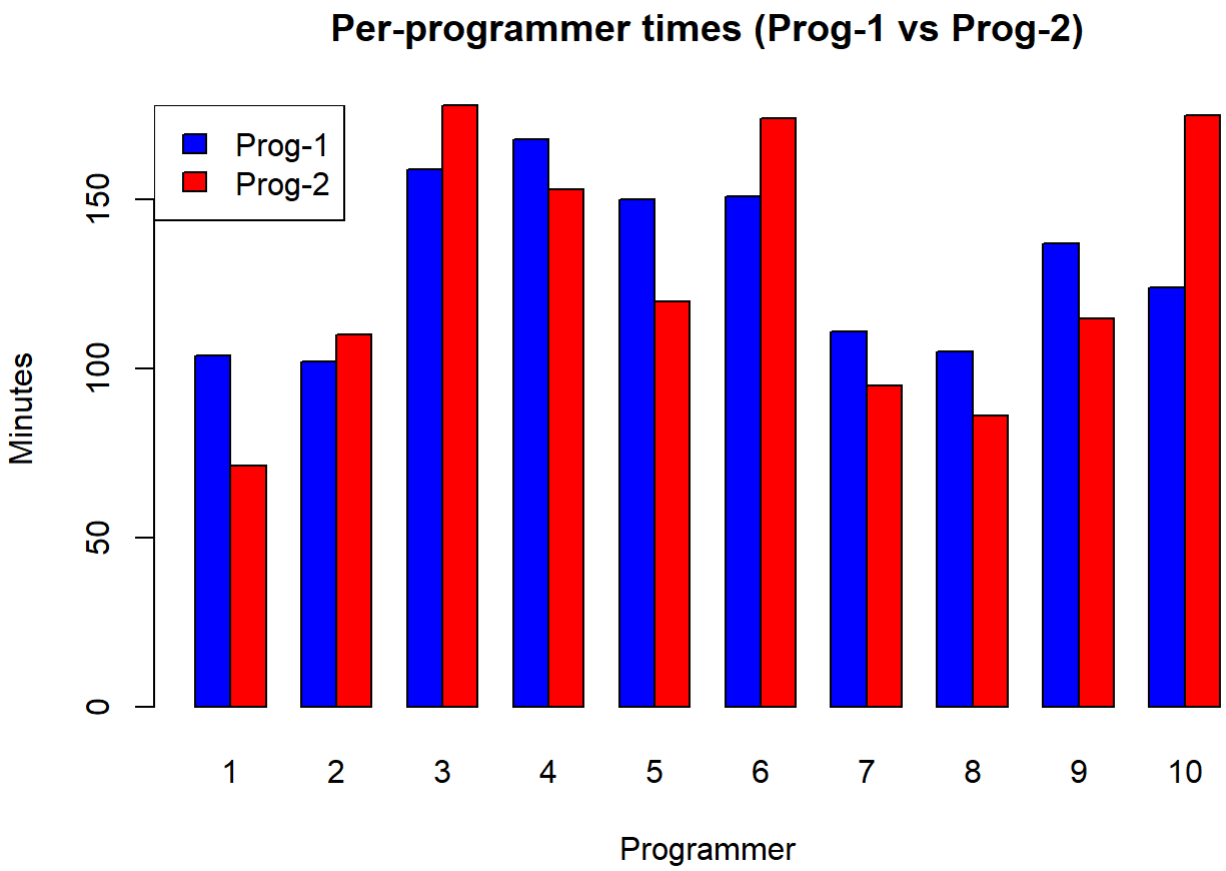

```
# 5b) Scatter: Prog-1 (x) vs Prog-2 (y)
plot(
  prog1, prog2,
  xlab = "Prog-1 time (min)",
  ylab = "Prog-2 time (min)",
  main = "Scatter: Prog-1 vs Prog-2",
  pch  = 16,                    # filled circle
  xlim = c(0, 200), ylim = c(0, 200)
)
abline(0, 1, lty = 2)           # line of equality y = x
text(prog1, prog2, labels = 1:10, pos = 3, cex = 0.8)
```



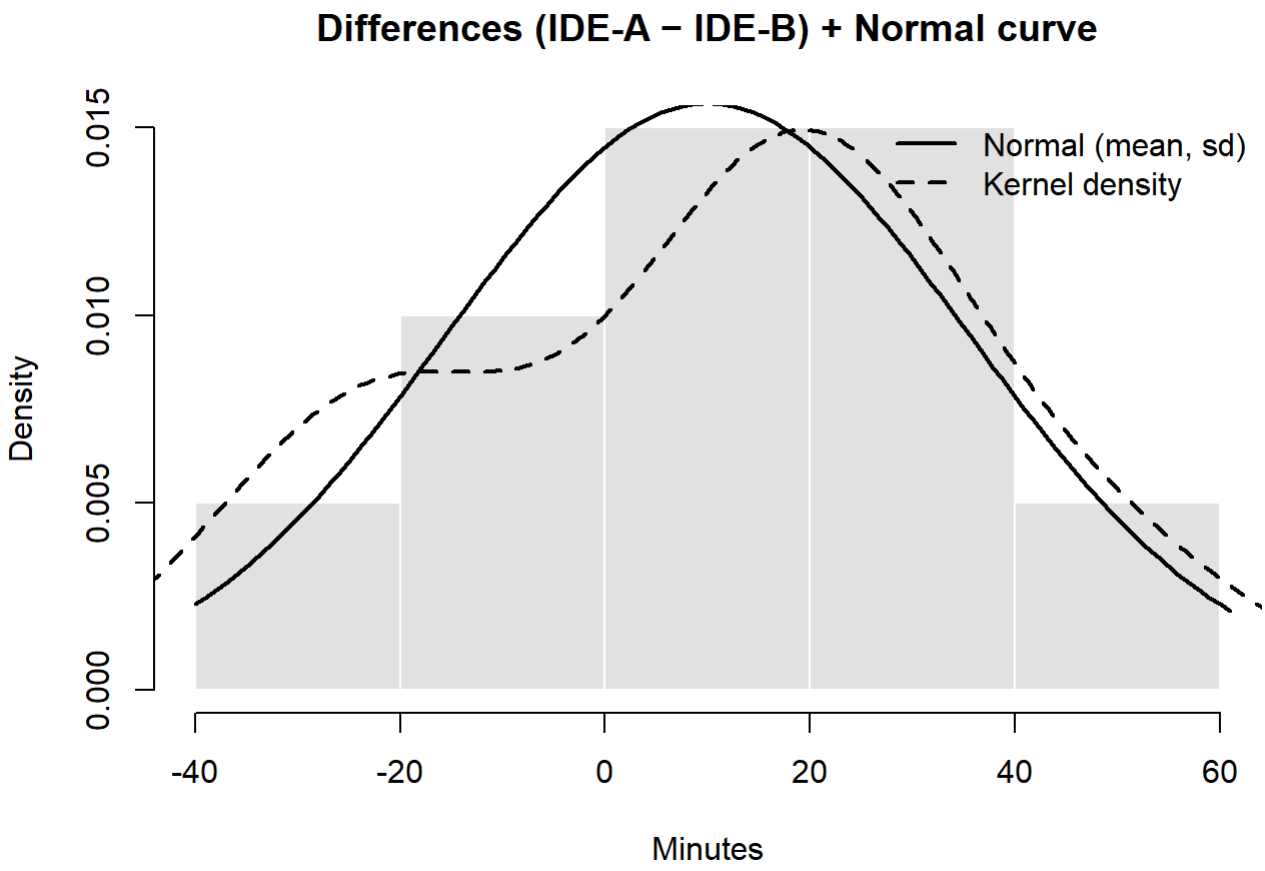

```
# 5c) Histogram of paired differences (IDE-A – IDE-B)
hist(
  diff,
  main = "Histogram of differences (IDE-A – IDE-B)",
  xlab = "Minutes"
)
```



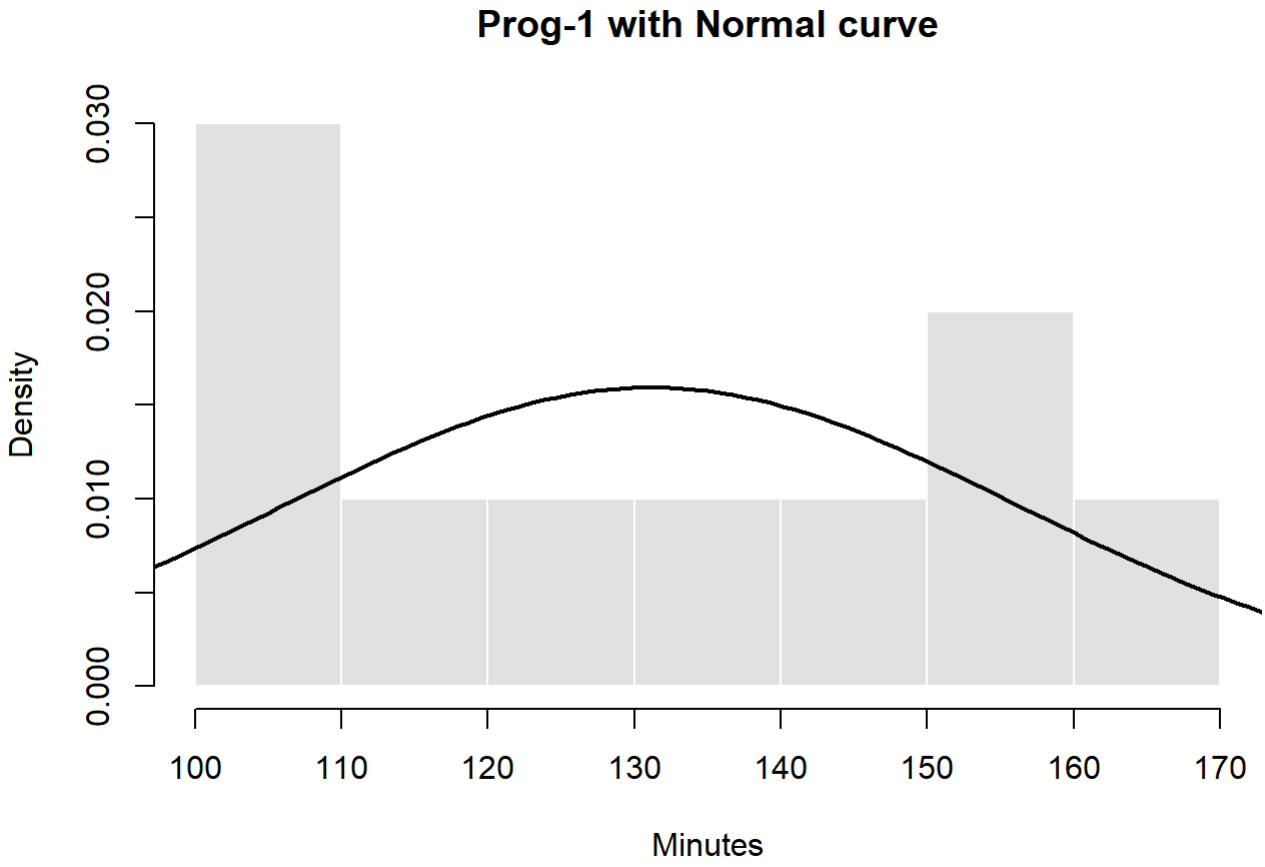

```
# 5d) Per-programmer barplot (side-by-side)
barplot(
  height    = rbind(prog1, prog2),  # 2 x 10 matrix (rows = programs, cols = programmers)
  beside    = TRUE,                 # side-by-side bars
  col       = c("blue", "red"),
  names.arg = 1:10,                 # programmer index on x-axis
  xlab      = "Programmer",
  ylab      = "Minutes",
  main      = "Per-programmer times (Prog-1 vs Prog-2)"
)
legend("topleft", fill = c("blue", "red"), legend = c("Prog-1", "Prog-2"))
```



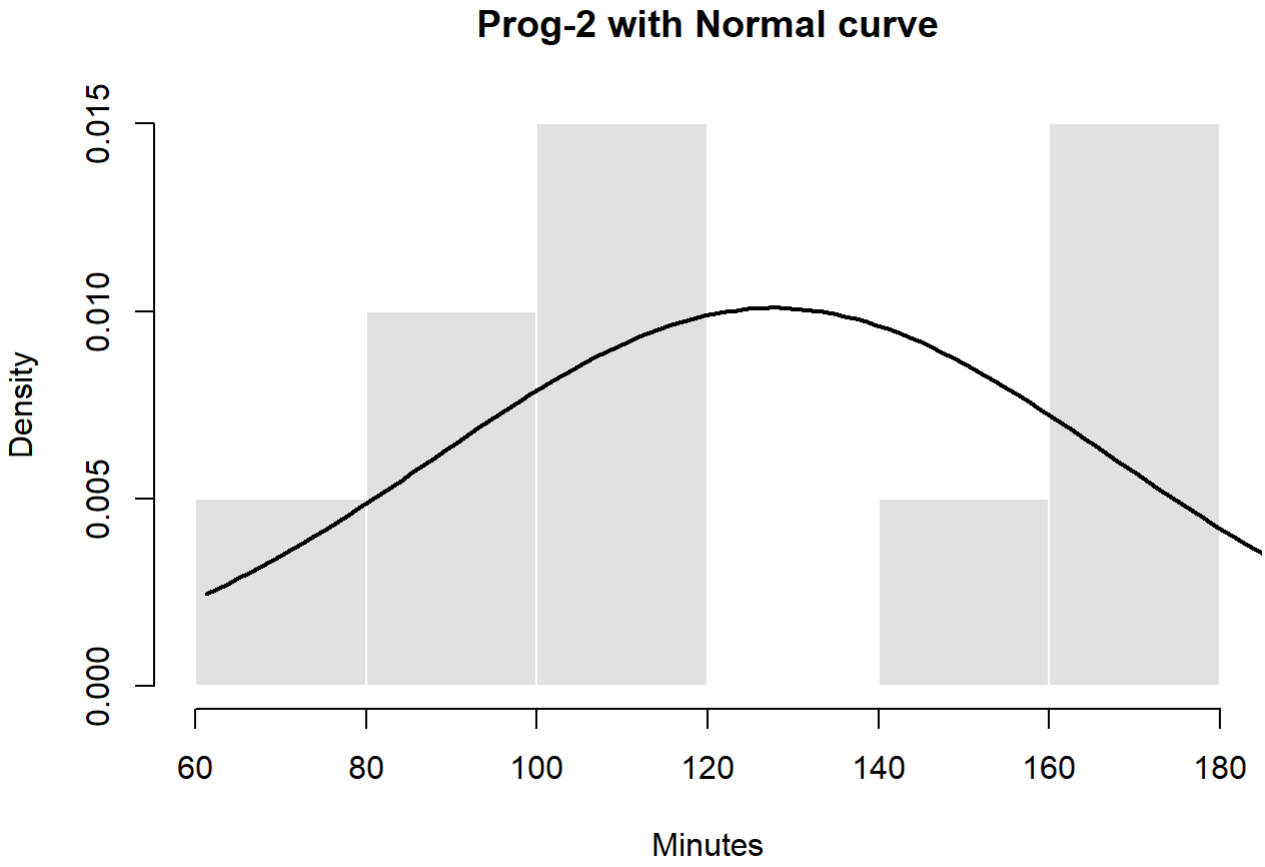

```
# ----------------------------
# 6) Bell-shaped (normal) curves
# ----------------------------

# 6a) Differences with fitted normal curve
m = mean(diff); s = sd(diff)
hist(
  diff,
  freq  = FALSE,                   # density scale
  main  = "Differences (IDE-A – IDE-B) + Normal curve",
  xlab  = "Minutes",
  col   = "grey90",
  border = "white"
)
curve(
  dnorm(x, mean = m, sd = s),
  from = min(diff) - 10, to = max(diff) + 10,
  add  = TRUE, lwd = 2
)
# Optional: kernel density overlay
lines(density(diff), lwd = 2, lty = 2)
legend("topright", lwd = 2, lty = c(1,2),
       legend = c("Normal (mean, sd)", "Kernel density"), bty = "n")
```




```
# 6b) Prog-1 with normal curve
m1 = mean(prog1); s1 = sd(prog1)
hist(
  prog1, freq = FALSE, col = "grey90", border = "white",
  main = "Prog-1 with Normal curve", xlab = "Minutes"
)
curve(dnorm(x, mean = m1, sd = s1),
      from = min(prog1) - 10, to = max(prog1) + 10,
      add = TRUE, lwd = 2)
```




```
# 6c) Prog-2 with normal curve
m2 = mean(prog2); s2 = sd(prog2)
hist(
  prog2, freq = FALSE, col = "grey90", border = "white",
  main = "Prog-2 with Normal curve", xlab = "Minutes"
)
curve(dnorm(x, mean = m2, sd = s2),
      from = min(prog2) - 10, to = max(prog2) + 10,
      add = TRUE, lwd = 2)
```


Prog-2 with Normal curve
Density
0.000
0.005
0.010
0.015
60
80
100
120
140
160
180
Minutes