```
# Install needed packages (run once)
install.packages("survival")
install.packages("flexsurv")

# Load libraries
library(survival)
library(flexsurv)

# 1. Read your CSV
df <- read.csv("hypothetical_os_data.csv")   # adjust path/filename

# Check structure
head(df)

# 2. Create survival object
surv_obj <- Surv(time = df$time, event = df$event)

# 3. Fit models
fit_exp     <- flexsurvreg(surv_obj ~ 1, dist = "exp")
fit_gamma   <- flexsurvreg(surv_obj ~ 1, dist = "gamma")
fit_gengamma <- flexsurvreg(surv_obj ~ 1, dist = "gengamma")
fit_gompertz <- flexsurvreg(surv_obj ~ 1, dist = "gompertz")
fit_llogis  <- flexsurvreg(surv_obj ~ 1, dist = "llogis")
fit_lnorm   <- flexsurvreg(surv_obj ~ 1, dist = "lnorm")
fit_weibull <- flexsurvreg(surv_obj ~ 1, dist = "weibull")

# 4. Extract parameter estimates
coef(fit_exp)     # log(rate)
coef(fit_gamma)   # log(shape), log(rate)
coef(fit_gengamma) # mu, log(sigma), Q
coef(fit_gompertz) # shape, rate
```

```
coef(fit_llogis)   # log(scale), log(shape)
coef(fit_lnorm)    # mu (meanlog), log sigma = log(sdlog)
coef(fit_weibull)  # log(shape), log(scale)
```

```
# 1. Read CSV
df <- read.csv("hypothetical_os_data.csv")  # adjust path/filename

# Make sure event is numeric
df$event <- as.numeric(df$event)

# 2. Survival object
surv_obj <- Surv(time = df$time, event = df$event)

# 3. Fit some models
fit_exp     <- flexsurvreg(surv_obj ~ 1, dist = "exp")
fit_weibull  <- flexsurvreg(surv_obj ~ 1, dist = "weibull")
fit_lnorm    <- flexsurvreg(surv_obj ~ 1, dist = "lnorm")
fit_llogis   <- flexsurvreg(surv_obj ~ 1, dist = "llogis")

# 4. Plot KM with model fits
plot(survfit(surv_obj ~ 1), xlab = "Time (months)", ylab = "Survival probability",
    main = "Kaplan–Meier with parametric fits", conf.int = FALSE,
    lwd = 3)
```

```
lines(fit_exp, col = "blue", lwd = 2)
lines(fit_weibull, col = "red", lwd = 2)
lines(fit_lnorm, col = "darkgreen", lwd = 2)
lines(fit_llogis, col = "purple", lwd = 2)

legend("topright",
       legend = c("Exponential", "Weibull", "Lognormal", "Log-logistic"),
       col = c("blue", "red", "darkgreen", "purple"),
       lwd = 2, bty = "n")
```

```
# --- 1. Read CSV ---
df <- read.csv("hypothetical_os_data.csv")  # adjust path/filename

# Make sure event is numeric
df$event <- as.numeric(df$event)

# --- 2. Survival object ---
surv_obj <- Surv(time = df$time, event = df$event)

# --- 3. Helper function for safe fitting ---
safe_fit <- function(dist) {
  fit <- try(flexsurvreg(surv_obj ~ 1, dist = dist), silent = TRUE)
```

```
 if (inherits(fit, "try-error")) return(NULL)
 return(fit)
}

# --- 4. Distributions to test ---
dists <- c("exp", "weibull", "lnorm", "llogis", "gamma", "gengamma", "gompertz")
dist_names <- c("Exponential", "Weibull", "Lognormal",
                "Log-logistic", "Gamma", "Generalized Gamma", "Gompertz")

# --- 5. Fit models ---
fits <- setNames(lapply(dists, safe_fit), dist_names)

# --- 6. Keep only successful fits ---
fits_ok <- fits[!vapply(fits, is.null, logical(1))]

# --- 7. Create comparison table ---
model_comp <- data.frame(
 Distribution = names(fits_ok),
 AIC = sapply(fits_ok, AIC),
 BIC = sapply(fits_ok, BIC),
 row.names = NULL,
 check.names = FALSE
)

# Sort by AIC (best fit first)
model_comp <- model_comp[order(model_comp$AIC), ]

print(model_comp)

# --- 8. Plot KM with model fits ---
plot(survfit(surv_obj ~ 1),
```

```
    xlab = "Time (months)", ylab = "Survival probability",
    main = "Kaplan–Meier with parametric fits", conf.int = FALSE)

cols <- c("blue","red","darkgreen","purple","orange","brown","pink")
i <- 1
for (f in fits_ok) {
  lines(f, col = cols[i], lwd = 2)
  i <- i + 1
}

legend("topright",
      legend = names(fits_ok),
      col = cols[seq_along(fits_ok)],
      lwd = 2, bty = "n")

# --- 9. Export results ---
write.csv(model_comp, "AIC_BIC_comparison_os.csv", row.names = FALSE)  # adjust path/filename

# Read data
df <- read.csv("hypothetical_os_data.csv")  # adjust path/filename

# KM fit
```

```
km_fit <- survfit(Surv(time, event) ~ 1, data = df)

# Extract time & survival
km_data <- data.frame(
  time = km_fit$time,
  survival = km_fit$surv
)

# Export to Excel/CSV
write.csv(km_data, "KM_curve_os.csv", row.names = FALSE)  # adjust path/filename
```

```
pfs <- read.csv("KM_curve_pfs.csv")  # adjust path/filename
os  <- read.csv("KM_curve_os.csv")  # adjust path/filename

# Base plot with OS
plot(os$time, os$survival, type = "s", lwd = 2, col = "red",
     xlab = "Time (months)", ylab = "Survival probability",
     ylim = c(0, 1), main = "PFS vs OS")

# Add PFS curve
lines(pfs$time, pfs$survival, type = "s", lwd = 2, col = "blue")

# Add legend
```

```
legend("topright",
       legend = c("OS", "PFS"),
       col = c("red", "blue"),
       lwd = 2, bty = "n")
```