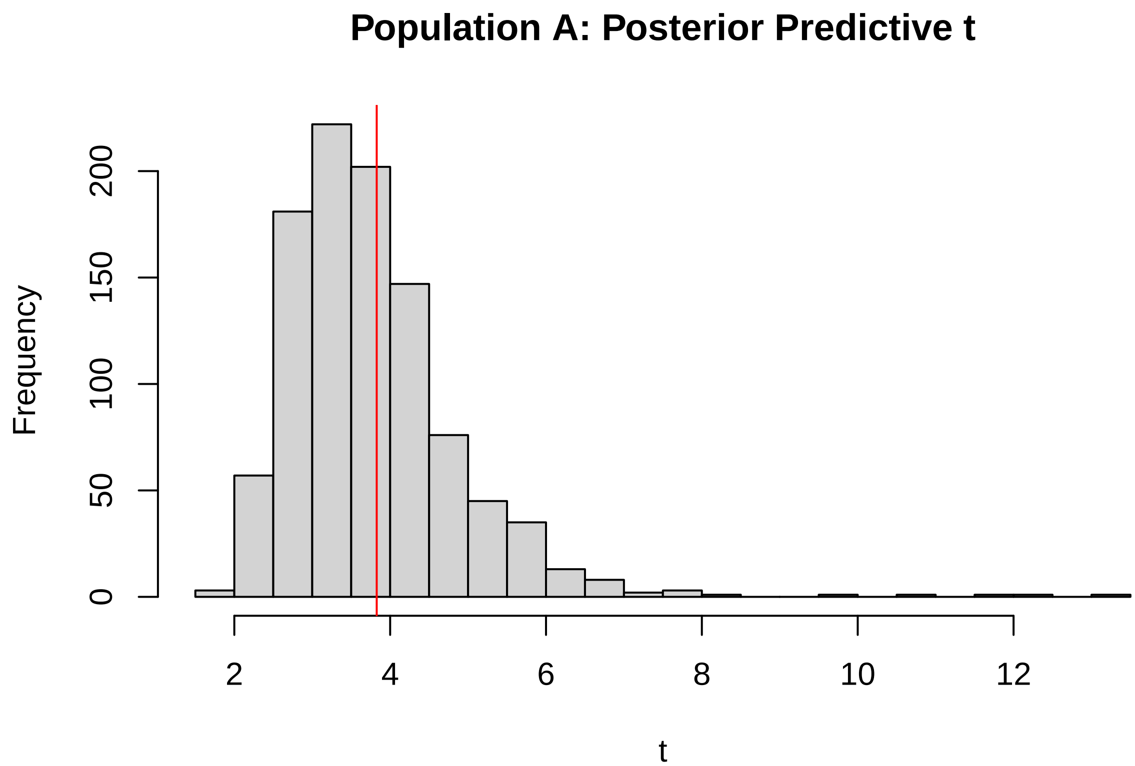
575_HW3

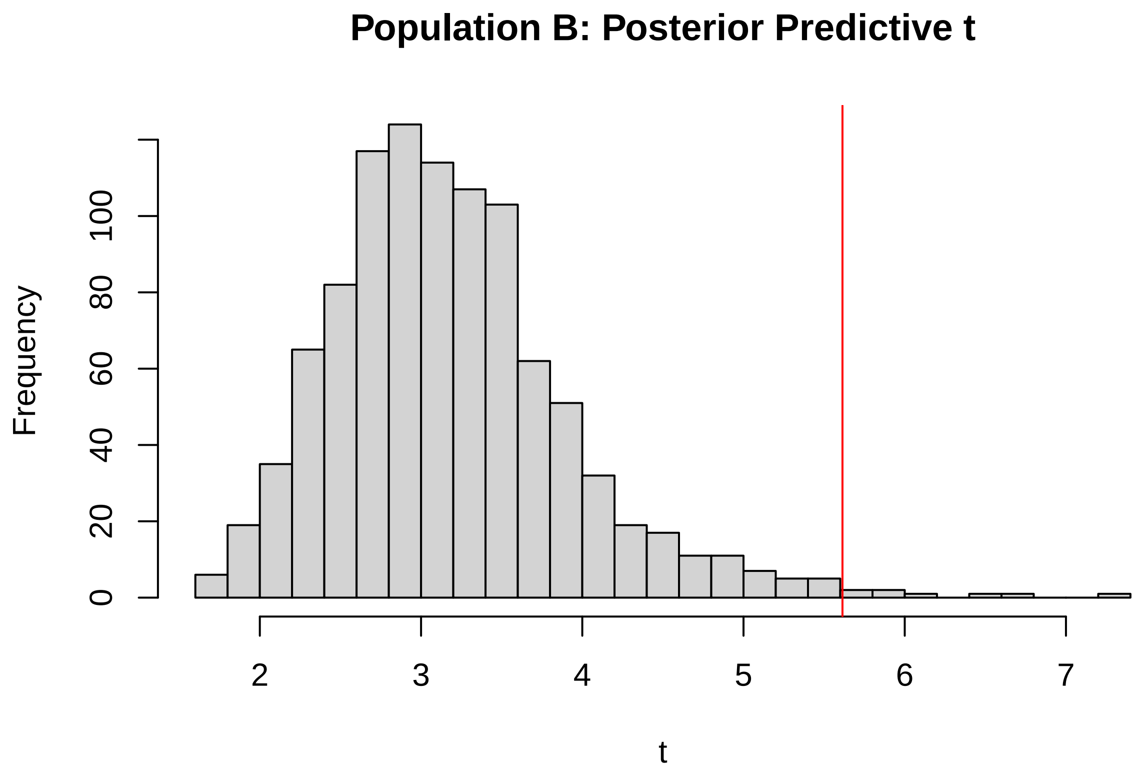
Matthew Stoebe

2025-02-14

```r
set.seed(123)
S <- 1000

y_A <- c(12,9,12,14,13,13,15,8,15,6)
sy_A <- sum(y_A); n_A <- length(y_A)

y_B <- c(11,11,10,9,9,8,7,10,6,8,8,9,7)
sy_B <- sum(y_B); n_B <- length(y_B)

#Prior A
a_A <- 120
b_A <- 10

#Prior B
a_B <- 12
b_B <- 1

post_a_A <- a_A + sy_A
post_b_A <- n_A + b_A

post_a_B <- a_B + sy_B
post_b_B <- n_B + b_B

# Population A
theta_A <- rgamma(S, post_a_A, post_b_A)
yA_rep <- sapply(theta_A, function(lambda) rpois(n_A, lambda))
t_rep_A <- apply(yA_rep, 2, function(y) mean(y)/sd(y))
t_obs_A <- mean(y_A)/sd(y_A)

hist(t_rep_A, main="Population A: Posterior Predictive t", breaks=30, xlab="t")
abline(v=t_obs_A, col="red")
```




```r
p_value_A <- mean(t_rep_A >= t_obs_A)

# Population B
theta_B <- rgamma(S, post_a_B, post_b_B)
yB_rep <- sapply(theta_B, function(lambda) rpois(n_B, lambda))
t_rep_B <- apply(yB_rep, 2, function(y) mean(y)/sd(y))
t_obs_B <- mean(y_B)/sd(y_B)

hist(t_rep_B, main="Population B: Posterior Predictive t", breaks=30, xlab="t")
abline(v=t_obs_B, col="red")
```



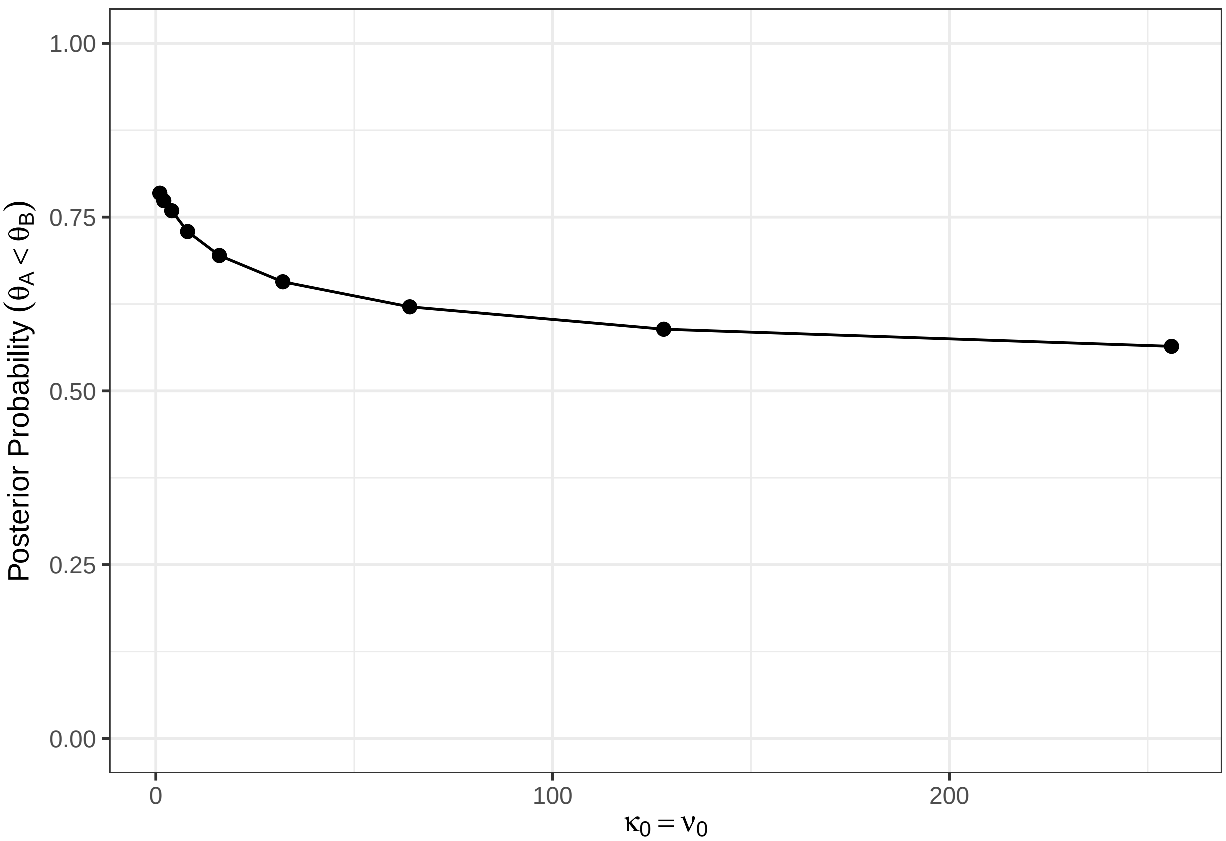

```
p_value_B <- mean(t_rep_B >= t_obs_B)

cat("Bayesian p-value (A):", p_value_A, "\nBayesian p-value (B):", p_value_B)
```

```
## Bayesian p-value (A): 0.394
## Bayesian p-value (B): 0.008
```

```
library(ggplot2)

sample_posterior <- function(ybar, s, n,
                             mu0, k0, alpha0, beta0,
                             M = 10000) {

  kn      <- k0 + n
  mun     <- (k0 * mu0 + n * ybar) / kn
  SS      <- (n - 1) * s^2

  alpha_n <- alpha0 + n/2
  beta_n  <- beta0 + 0.5 * (
    SS + (k0 * n / kn) * (ybar - mu0)^2
  )

  # Sample posterior for sigma^2
  sigma2_draws <- 1 / rgamma(M, shape = alpha_n, rate = beta_n)
```

```r
  # Sample posterior for theta
  theta_draws <- rnorm(M, mean = mun, sd = sqrt(sigma2_draws / kn))

  return(theta_draws)
}

nA     <- 16
ybarA  <- 75.2
sA     <- 7.3

nB     <- 16
ybarB  <- 77.5
sB     <- 8.1


mu0       <- 75
sigma0_sq <- 100

kappa_vals <- c(1, 2, 4, 8, 16, 32, 64, 128, 256)


M <- 100000

results <- data.frame(k0 = integer(), nu0 = integer(), prob = double())

for (k0 in kappa_vals) {

  # nu0 = k0
  nu0    <- k0
  alpha0 <- nu0 / 2
  beta0  <- (nu0 * sigma0_sq) / 2

  thetaA_draws <- sample_posterior(
    ybar = ybarA, s = sA, n = nA,
    mu0  = mu0,   k0 = k0,
    alpha0 = alpha0, beta0 = beta0,
    M = M
  )

  # Posterior draws for group B
  thetaB_draws <- sample_posterior(
    ybar = ybarB, s = sB, n = nB,
    mu0  = mu0,   k0 = k0,
    alpha0 = alpha0, beta0 = beta0,
    M = M
  )

  probA_lt_B <- mean(thetaA_draws < thetaB_draws)

  results <- rbind(results,
                   data.frame(k0 = k0,
                              nu0 = nu0,
                              prob = probA_lt_B))
```

```
}



ggplot(results, aes(x = k0, y = prob)) +
  geom_point(size = 2) +
  geom_line() +
  labs(
    x = expression(kappa[0] == nu[0]),
    y = expression("Posterior Probability"~(theta[A]<theta[B]))
  ) +
  theme_bw() +
  ylim(0,1)
```



The probability remains above .5 consistently and indicates robust evidence that thetaB is greater than thetaA. as our k0 grows the probability dips and asymptopes towards .5 which indicates sensitivitity in how strongly we shring the mean towards 75 ie how strong our prior knowledge is.