STAT 638: Solution for Homework #4

5.1 The results in this exercise are based on 100,000 draws from the posterior.

a)

	Parameter	Mean	95% credible interval
School 1	μ	9.294	$(7.775, 10.828)$
	σ	3.905	$(2.996, 5.169)$
School 2	μ	6.948	$(5.140, 8.750)$
	σ	4.394	$(3.346, 5.874)$
School 3	μ	7.812	$(6.181, 9.455)$
	σ	3.750	$(2.801, 5.125)$

b)

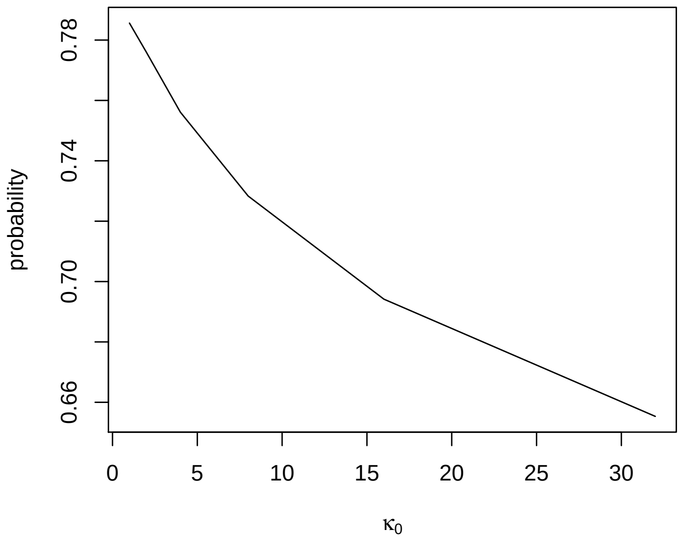
ijk	$P(\theta_i < \theta_j < \theta_k)$
123	.00600
132	.00364
213	.08612
231	.67025
312	.01554
321	.21845

c)

ijk	$P(\tilde{Y}_i < \tilde{Y}_j < \tilde{Y}_k)$
123	.10462
132	.10415
213	.18308
231	.26882
312	.13838
321	.20095

d) $P(\theta_1 > \theta_2 \cap \theta_1 > \theta_3) = .67025 + .21845 = .8887$, $P(\tilde{Y}_1 > \tilde{Y}_2 \cap \tilde{Y}_1 > \tilde{Y}_3) = .26882 + .20095 = .46977$

5.2 The probabilities in the plot were obtained by generating 100,000 draws from the posterior for each choice of (κ_0, ν_0).



When ν_0 and κ_0 are small, there is less prior information. So, when the data are allowed to speak for themselves, there is about a 78% chance that method B yields higher scores on average than does method A. As ν_0 and κ_0 increase, there is more and more prior information indicating that the two methods produce similar scores, and hence the posterior probability that method B yields higher scores on average than does method A becomes smaller.

5.5 a) The log-likelihood is

$$\ell(\theta, \psi) = -(n/2)\log(2\pi) + (n/2)\log\psi - (\psi/2)\sum_{i=1}^{n}(y_i - \theta)^2.$$

b) Let $s^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/n$. The density p_U is proportional to

$$\sqrt{\psi}\exp\left(-\frac{\psi}{2n}\sum_{i=1}^{n}(y_i - \theta)^2\right) = \sqrt{\psi}\exp\left[-\frac{\psi}{2}\left(s^2 + (\theta - \bar{y})^2\right)\right].$$

It follows that p_U is normal-gamma with ψ distributed gamma$(1, s^2/2)$ and θ given ψ $N(\bar{y}, 1/\psi)$.

c) Using the results from p. 97 of the notes, the posterior is normal-gamma with ψ distributed gamma$((n{+}2)/2, (n{+}1)s^2/2)$ and θ given ψ $N(\bar{y}, (1/\psi)/(n{+}1))$. Technically, this wouldn't be a posterior because the data were used to construct the prior. However, practically speaking it seems ok to use as a posterior because the "prior" contains only as much information as there would be in a single observation.