**Normal Populations and Bayesian Statistics** **Math 422 - Workshop**

Suppose you are studying a population with a roughly normal distribution. You want to use a sample of $n$ individuals $x_1, \ldots, x_n$ to estimate the population mean $\mu$. To keep things simple, let's assume that you know the population standard deviation is $\sigma_0$. The following theorem tells you everything you need to know to do a Bayesian analysis.

**Theorem.** *If a population has a* $\text{Norm}(\mu, \sigma_0)$ *distribution with* $\sigma_0$ *known, and we use a normal distribution* $\text{Norm}(\mu_0, \tau_0)$ *as the prior distribution for* $\mu$*, then the posterior distribution for* $\mu$ *is also normally distributed. The mean of the posterior distribution is*

$$\frac{\alpha}{\alpha+\beta}\mu_0 + \frac{\beta}{\alpha+\beta}\bar{x}$$

*and the standard deviation is*

$$\sqrt{\frac{1}{\alpha+\beta}}$$

*where*

$$\alpha = \frac{1}{\tau_0^2} \qquad \textit{and} \qquad \beta = \frac{n}{\sigma_0^2}.$$

You can find a proof of this theorem in Example 7.1.2 of Evans and Rosenthal.

1. Last year, I collected data from 47 students in my introductory statistics class at Hampden-Sydney and their average height was $\bar{x} = 71.85$ inches. Heights for all men in the United States have approximately the $\text{Norm}(70, 3)$ distribution. I'd like to estimate the population mean height for all Hampden-Sydney students. I figure that the average Hampden-Sydney student is pretty close to the average for all men in the United States, so I might choose a prior distribution with mean 70 and standard deviation of just 1 inch (so $\mu_0 = 70$ and $\tau_0 = 1$). One way to think of this prior is that you are assuming Hampden-Sydney students are close to the average for men, but you wouldn't be surprised if the HSC average is 1 or 2 inches higher or lower. Use this information to find the posterior distribution for $\mu$.

2. According to the posterior distribution, what is the probability that the average height of Hampden-Sydney students is taller than the average for all men in the United States?

3. Find a 95% credible interval for the average height of all Hampden-Sydney students. Hint: To make a 95% credible interval using a normal posterior, just add and subtract 1.96 times the posterior standard deviation from the posterior mean.

4. What if we had started with a ridiculous prior that the average height of Hampden-Sydney students is around 7 ft tall ($\mu_0 = 84$)? If we also increased the standard deviation of the prior $\tau_0$ to 12 inches to reflect greater uncertainty, then what would the 95% credible interval be?

5. A general formula for a credible interval in this example is

$$\frac{\alpha}{\alpha+\beta}\mu_0 + \frac{\beta}{\alpha+\beta}\bar{x} \pm z^*\sqrt{\frac{1}{\alpha+\beta}}$$

where $\alpha = 1/\tau_0^2$ and $\beta = n/\sigma_0^2$ and $z^*$ is determined by the credibility level. What happens to this formula if $\tau_0$ is really large (i.e., what is the limit as $\tau_0 \to \infty$)?