

## The Normal-Normal Model: An Introductory Homework on Bayesian Data Modeling

Bayesian statistical methods use *prior information* to help estimate unknown quantities of interest, also known as parameters. This homework introduces a fundamental statistical model, the normal-normal model. It is a model for data that is continuous and unimodal and that does not have outliers.

In this homework you will estimate your (average) blood pressure. Bayesian modeling at its most basic uses something called *prior information*, hopefully culled from an expert on the subject. You will be the expert on your own blood pressure. You will provide inputs to the calculations, then go measure your blood pressure. Formulas are given at the end for the Bayesian calculations. I'll illustrate the bigger picture in my talk.

Please follow these steps. If something is unclear, please email me ([robweiss@ucla.edu](mailto:robweiss@ucla.edu)) so that I can clarify.

This involves some modest calculations. An excel spreadsheet is provided (initially written by David Huang and modified by me) where you can enter the data and automatically see the results of the calculations. In the spreadsheet, you enter prior information and data in yellow cells; summaries of these appear in pinkish cells. Bayesian statistical calculations and inferences are in orange and classical statistical calculations and inferences appear in blue.

1. Elicitation. This is the novel part of Bayesian statistics. These numbers will go into your final estimates. You specify a plausible value for your blood pressure. Repeat these steps for systolic and diastolic blood pressures.
  - (a) Write down what you think your systolic blood pressure (SBP) is. Call this  $\mu_0$  (mu-naught or mu-zero). This is your *prior mean*. Best guess at SBP is:

$$\mu_0 = \underline{\hspace{2cm}}$$

- (b) Your best guess is allowed to be in error by some amount. You need to specify a *prior standard deviation* (SD) to assess how far off this best guess might be. We will refer to this SD as  $\tau$  (tau).

The idea is that 68% of the time your actual blood pressure should be within one standard deviation  $\tau$  of the best guess  $\mu_0$  and 95% of the time within two standard deviations  $2\tau$  of  $\mu_0$ . If your best guess is more than 2 standard deviations from the true SBP, you would be somewhat surprised. Guess at SBP SD is:

$$\tau = \underline{\hspace{2cm}}$$

- (c) Check. Calculate  $\mu_0 \pm \tau$  and calculate  $\mu_0 \pm 2\tau$ , these are 68% and 95% prior probability intervals respectively. Do the two calculations and enter your two intervals here. These are *prior probability intervals*, as they can be calculated before seeing the data.

SBP 68% interval is  $\left( \underline{\hspace{2cm}}, \underline{\hspace{2cm}} \right)$

SBP 95% interval is  $\left( \underline{\hspace{2cm}}, \underline{\hspace{2cm}} \right)$

- (d) Inspect your intervals from step 1c. Do they seem reasonable? Do you expect to find your blood pressure in the 95% interval? Is it too small? Is the 68% a reasonable range to find your true blood pressure in? At this point you are invited to change your estimated values  $\mu_0$  and  $\tau$  to make these intervals fit what you anticipate is your true blood pressure. The goal isn't to get the mean of your data in the interval, but the mean of your true BP in the interval.
- (e) Example 1. I think my pressure is high, but I never check it myself. If I set  $\mu = 150$ , and  $\tau = 30$ , my 95% interval is (90, 210) – seems unreasonable to include 90 and 210 in my interval. Perhaps 200 is too high as well. I should consider shrinking my sd, perhaps to 20 and perhaps increasing the point estimate slightly, say to 160, giving a 95% interval of (120, 200) – more reasonable, though still quite a large range. Perhaps  $\mu_0 = 160$  and  $\tau = 15$  for a final interval of (130, 190).
- (f) Example 2. I check my SBP often at the drugstore, and it usually measures in the low 120's. I set my prior mean  $\mu_0 = 122$ . I expect this to be within 10 of my true blood pressure, and set my prior sd  $\tau = 5$ .

- (g) Repeat the previous steps for your diastolic blood pressure. Best guess at DBP is:

$$\mu_0 = \underline{\hspace{2cm}}$$

Guess at DBP SD is:

$$\tau = \underline{\hspace{2cm}}$$

DBP 68% interval is  $\left( \underline{\hspace{2cm}}, \underline{\hspace{2cm}} \right)$

DBP 95% interval is  $\left( \underline{\hspace{2cm}}, \underline{\hspace{2cm}} \right)$

- (h) Check: Do DBP intervals seem reasonable? Fix if not
2. Now you will collect blood pressure data. This comes only after you have settled on values for  $\mu_0$  and  $\tau$  in the previous section.
- (a) Collect 5 repeated blood pressure observations. You can do this any place you have access to a blood pressure machine such as a drug store or a doctor's office. Ideally you would do this on 5 consecutive days at similar times, but for our purposes, it will be ok if you do 5 measures in a row.
- (b) Report the SBP data values.

$\underline{\hspace{2cm}}, \underline{\hspace{2cm}}, \underline{\hspace{2cm}}, \underline{\hspace{2cm}}, \underline{\hspace{2cm}}$

- (c) Calculate the SBP sample mean  $\bar{y}$  and the SBP sample standard deviation ( $n - 1$  denominator) which we will call  $\sigma$ .

$$\text{SBP Mean } \bar{y} = \underline{\hspace{2cm}}.$$

$$\text{SBP Standard Deviation } \sigma = \underline{\hspace{2cm}}.$$

- (d) Report the DBP data values.

$\underline{\hspace{2cm}}, \underline{\hspace{2cm}}, \underline{\hspace{2cm}}, \underline{\hspace{2cm}}, \underline{\hspace{2cm}}$

- (e) Calculate the DBP sample mean  $\bar{y}$  and the DBP sample standard deviation ( $n - 1$  denominator)  $\sigma$ .

DBP Mean  $\bar{y} =$  \_\_\_\_\_

DBP Standard Deviation  $\sigma =$  \_\_\_\_\_.

- (f) We will treat the sample standard deviation  $\sigma$  that you calculated as if it is known. This is an approximation to keep the example simple.
3. Inference. We will calculate two example Bayesian inferences for this example (i) with all five data points and (ii) as if you had collected 1 data point. To distinguish the calculations, we will subscript the inferences with a 5 or a 1 to denote that it is based on all 5 data points or on just 1 data point. We will repeat for SBP and DBP. You will use your SBP  $\sigma$  that you calculated for both SBP examples.

- (a) Calculate the SBP posterior mean  $\bar{\mu}$  and SBP posterior standard deviation  $S$ . *Formulas for  $S$  and  $\bar{\mu}$  are given at the end of this work sheet.*

$n = 5$  SBP posterior mean  $\bar{\mu}_5 =$  \_\_\_\_\_

$n = 5$  SBP posterior sd  $S_5 =$  \_\_\_\_\_.

- (b) Calculate the SBP posterior 95% interval as  $\bar{\mu}_5 \pm 2 * S_5$ .

SBP 95% posterior interval is  $\left( \text{_____}, \text{_____} \right)$

- (c) Suppose you had only the first data point. Your sample size is  $n = 1$ . Now calculate the posterior mean and standard deviation. (Note:  $\sigma$  remains the same as previous step, only  $n$  changes.)

SBP posterior mean  $\bar{\mu}_1 =$  \_\_\_\_\_

$n = 1$  SBP posterior sd  $S_1 =$  \_\_\_\_\_.

- (d) Calculate the  $n = 1$  SBP posterior 95% interval as  $\bar{\mu}_1 \pm 2 * S_1$ .

SBP 95% posterior interval is  $\left( \text{—————}, \text{—————} \right)$

- (e) Repeat for DBP. Calculate the DBP posterior mean  $\bar{\mu}$  and DBP posterior standard deviation  $S$ .

$n = 5$  DBP posterior mean  $\bar{\mu}_5 = \text{—————}$

$n = 5$  DBP posterior sd  $S_5 = \text{—————}$ .

DBP 95% posterior interval is  $\left( \text{—————}, \text{—————} \right)$

- (f) DBP. Suppose you had only the first data point. Your sample size is  $n = 1$ . Now calculate the posterior mean and standard deviation. (Note:  $\sigma$  remains the same as previous step, only  $n$  changes.)

$n = 1$  DBP posterior mean  $\bar{\mu}_1 = \text{—————}$

$n = 1$  DBP posterior sd  $S_1 = \text{—————}$ .

DBP 95% posterior interval is  $\left( \text{—————}, \text{—————} \right)$

- (g) Classical inferences. The classical statistical inference is  $\hat{\mu} = \bar{y}$ , with standard error  $SE = \sigma/\sqrt{n}$  and the classical confidence interval (CI)  $\bar{y} \pm 2 * SE$ . Since we are pretending to know  $\sigma$ , we can calculate the CIs for both  $n = 1$  and  $n = 5$ .

$n = 1$  SBP estimate  $\bar{y}_1 = \text{—————}$

$n = 1$  SBP SE =  $\text{—————}$ .

SBP 95% confidence interval is  $\left( \text{—————}, \text{—————} \right)$

$n = 5$  SBP estimate  $\bar{y}_5 =$  \_\_\_\_\_

$n = 5$  SBP SE = \_\_\_\_\_.

SBP 95% confidence interval is  $\left( \text{_____}, \text{_____} \right)$

4. Draw conclusions. The quality of the conclusions depends on the data behaving well and on the quality of the prior specification of  $\mu_0$  and  $\tau$ .
  - (a) Notice that the classical interval is wider than the Bayesian intervals for the same sample size, particularly for  $n = 1$ . That is, the posterior standard deviation  $S$  is smaller than the classical SE.
  - (b) Sample of size 1. For your inference, do you prefer your classical interval or your Bayesian interval? Which contains more sensible values? Answer the same questions for  $n = 5$ .
  - (c) Compare your point estimates  $(\bar{\mu}, \hat{\mu})$ , based on  $n = 1$  and  $n = 5$ . For each sample size, which seems more sensible to you?
5. Comparison to values from the Nurse's Blood Pressure study which measured ambulatory blood pressure over several days of waking periods for 203 nurses. From this study, we can estimate  $\mu_0 = 118$ ,  $\tau = 8$ ,  $\sigma = 13$ . For a single observation, the SE is 13, the Bayes  $S = 6.8$ . The Bayes interval will be roughly half as long as the classical interval. For  $n = 5$ , the SE = 5.8, and the Bayes  $S = 4.7$ . Bayes intervals are more than 20% shorter than the classical intervals. At  $n = 10$ , the Bayes intervals are still 10% shorter at the same cost.

## Formulas

The posterior mean  $\bar{\mu}$  and posterior standard deviation  $S$  are common Bayesian inferences in this model. You can construct a 95% posterior interval

for your blood pressure as  $\bar{\mu} \pm 2 * S$ .

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (1)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^5 (y_i - \bar{y})^2}{n - 1}} \quad (2)$$

$$S^2 = \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \quad (3)$$

$$\bar{\mu} = S^2 \left( \frac{\mu_0}{\tau^2} + \frac{\bar{y}n}{\sigma^2} \right) \quad (4)$$

Take the square root of  $S^2$  to get the posterior standard deviation  $S = \sqrt{S^2}$ .