

Inference from grouped normal data

We have the following data for the heights of men from a local university. (This a fake story and fake data. I have actually simulated the data myself.)

Height interval (cm)	Frequency
less than 165	17
165–170	41
170–175	56
175–180	45
180–185	25
over 185	16

Our model for the data is that we have grouped observations from a normal population $N(\mu, \sigma^2)$. Our observations are multinomial data arising from the grouping. Notice that the probabilities of the categories are easy to express in terms of the distribution function $\Phi(x | \mu, \sigma^2)$ of the normal distribution. Now, write a formula for the likelihood. We take as our prior the standard noninformative improper prior, which is proportional to $1/\sigma$.

We next form initial guesses for the parameters. To do this, create an artificial dataset by replacing each grouped observation by its bin midpoint. So we get, e.g., 41 artificial observation at the value 167.5. For the lowest and highest category, we can use, e.g., the values 162.5 and 187.5 for the bin midpoints. Next calculate the sample mean and sample variance of these artificial observations. These are the first guesses for the values of μ and σ^2 .

Finish by doing posterior simulations from the model using your method of choice and summarize the posterior, e.g., by drawing scatter plots and by calculating other posterior summaries.