Sampling and Inference

Soumya Ray

Oct 11, 2014

Sampling Process and Inferential Error

We can create a population and then create a noisy sampling frame. This increases the 'dispersion' of the sample by adding noise to it.

This code block creates a population and noise vectors of size N, creates a sampling frame with noise, and samples from it.

```
N = 10000
n = 150
pop <- rnorm(N, mean = 22, sd = 5)
noise <- rnorm(length(pop), sd = 4)
noisy.frame <- pop + noise
noisy.sample <- sample(noisy.frame, size = n)
```

We can now use the `plot` command to visualize the dispersion of the original population (solid curve), compared to that of the sample (dashed):

```
# Plot the population and noisy sampling frame
plot(density(pop), main = NULL)
lines(density(noisy.sample), lty = "dashed")
```

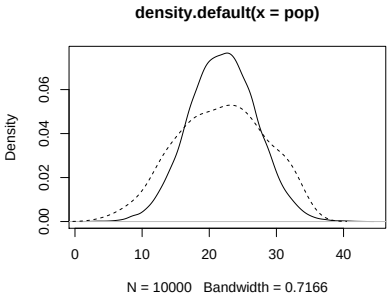

Figure 1: plot figure

And now the same plot using `ggplot`. This time we have to create data.frame objects of the population and sample data:

```
pop.df = as.data.frame(pop)
sample.df = as.data.frame(noisy.sample)
ggplot() + geom_density(data = pop.df, aes(pop)) +
    geom_density(data = sample.df, aes(noisy.sample),
        linetype = "dashed")
```

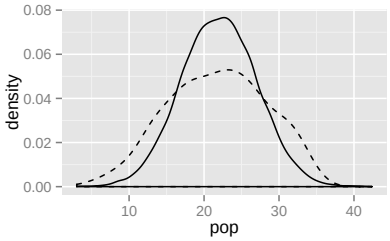

Figure 2: ggplot figure