Statistical Rethinking 2024

Week 2 lecture notes

2024-01-16

Table of contents

Linear regression as an Owl

```
data(Howell1)
d <- Howell1 %>%
  filter(age>17)
```

Bayes' Rule as a regression

Bayesian regression through grid approximation

```
mu.list <- seq( from=150, to=160 , length.out=100 )

frq(mu.list)
```

```
x <numeric>
# total N=100 valid N=100 mean=155.00 sd=2.93
```

Value	N	Raw %	Valid %	Cum. %
150.00	1	1	1	1.00
150.10	1	1	1	2.00
150.20	1	1	1	3.00
150.30	1	1	1	4.00
150.40	1	1	1	5.00
150.51	1	1	1	6.00
150.61	1	1	1	7.00
150.71	1	1	1	8.00
150.81	1	1	1	9.00
150.91	1	1	1	10.00
151.01	1	1	1	11.00
151.11	1	1	1	12.00
151.21	1	1	1	13.00
151.31	1	1	1	14.00
151.41	1	1	1	15.00
151.52	1	1	1	16.00
151.62	1	1	1	17.00
151.72	1	1	1	18.00
151.82	1	1	1	19.00
151.92	1	1	1	20.00
152.02	1	1	1	21.00
152.12	1	1	1	22.00
152.22	1	1	1	23.00
152.32	1	1	1	24.00
152.42	1	1	1	25.00
152.53	1	1	1	26.00
152.63	1	1	1	27.00
152.73	1	1	1	28.00
152.83	1	1	1	29.00
152.93	1	1	1	30.00
153.03	1	1	1	31.00
153.13	1	1	1	32.00
153.23	1	1	1	33.00
153.33	1	1	1	34.00
153.43	1	1	1	35.00
153.54	1	1	1	36.00
153.64	1	1	1	37.00
153.74	1	1	1	38.00
153.84	1	1	1	39.00

153.94	1	1	1	40.00
154.04	1	1	1	41.00
154.14	1	1	1	42.00
154.24	1	1	1	43.00
154.34	1	1	1	44.00
154.44	1	1	1	45.00
154.55	1	1	1	46.00
154.65	1	1	1	47.00
154.75	1	1	1	48.00
154.85	1	1	1	49.00
154.95	1	1	1	50.00
155.05	1	1	1	51.00
155.15	1	1	1	52.00
155.25	1	1	1	53.00
155.35	1	1	1	54.00
155.45	1	1	1	55.00
155.56	1	1	1	56.00
155.66	1	1	1	57.00
155.76	1	1	1	58.00
155.86	1	1	1	59.00
155.96	1	1	1	60.00
156.06	1	1	1	61.00
156.16	1	1	1	62.00
156.26	1	1	1	63.00
156.36	1	1	1	64.00
156.46	1	1	1	65.00
156.57	1	1	1	66.00
156.67	1	1	1	67.00
156.77	1	1	1	68.00
156.87	1	1	1	69.00
156.97	1	1	1	70.00
157.07	1	1	1	71.00
157.17	1	1	1	72.00
157.27	1	1	1	73.00
157.37	1	1	1	74.00
157.47	1	1	1	75.00
157.58	1	1	1	76.00
157.68	1	1	1	77.00
157.78	1	1	1	78.00
157.88	1	1	1	79.00
157.98	1	1	1	80.00

```
158.08 | 1 |     1 |       1 |  81.00
158.18 | 1 |     1 |       1 |  82.00
158.28 | 1 |     1 |       1 |  83.00
158.38 | 1 |     1 |       1 |  84.00
158.48 | 1 |     1 |       1 |  85.00
158.59 | 1 |     1 |       1 |  86.00
158.69 | 1 |     1 |       1 |  87.00
158.79 | 1 |     1 |       1 |  88.00
158.89 | 1 |     1 |       1 |  89.00
158.99 | 1 |     1 |       1 |  90.00
159.09 | 1 |     1 |       1 |  91.00
159.19 | 1 |     1 |       1 |  92.00
159.29 | 1 |     1 |       1 |  93.00
159.39 | 1 |     1 |       1 |  94.00
159.49 | 1 |     1 |       1 |  95.00
159.60 | 1 |     1 |       1 |  96.00
159.70 | 1 |     1 |       1 |  97.00
159.80 | 1 |     1 |       1 |  98.00
159.90 | 1 |     1 |       1 |  99.00
160.00 | 1 |     1 |       1 | 100.00
  <NA> | 0 |     0 |    <NA> |   <NA>
```

```r
sigma.list <- seq( from=7 , to=9 , length.out=100 )

frq(sigma.list)
```

```
x <numeric>
# total N=100 valid N=100 mean=8.00 sd=0.59

Value | N | Raw % | Valid % | Cum. %
-------------------------------------
 7.00 | 1 |     1 |       1 |   1.00
 7.02 | 1 |     1 |       1 |   2.00
 7.04 | 1 |     1 |       1 |   3.00
 7.06 | 1 |     1 |       1 |   4.00
 7.08 | 1 |     1 |       1 |   5.00
 7.10 | 1 |     1 |       1 |   6.00
 7.12 | 1 |     1 |       1 |   7.00
 7.14 | 1 |     1 |       1 |   8.00
```

7.16	1	1	1	9.00
7.18	1	1	1	10.00
7.20	1	1	1	11.00
7.22	1	1	1	12.00
7.24	1	1	1	13.00
7.26	1	1	1	14.00
7.28	1	1	1	15.00
7.30	1	1	1	16.00
7.32	1	1	1	17.00
7.34	1	1	1	18.00
7.36	1	1	1	19.00
7.38	1	1	1	20.00
7.40	1	1	1	21.00
7.42	1	1	1	22.00
7.44	1	1	1	23.00
7.46	1	1	1	24.00
7.48	1	1	1	25.00
7.51	1	1	1	26.00
7.53	1	1	1	27.00
7.55	1	1	1	28.00
7.57	1	1	1	29.00
7.59	1	1	1	30.00
7.61	1	1	1	31.00
7.63	1	1	1	32.00
7.65	1	1	1	33.00
7.67	1	1	1	34.00
7.69	1	1	1	35.00
7.71	1	1	1	36.00
7.73	1	1	1	37.00
7.75	1	1	1	38.00
7.77	1	1	1	39.00
7.79	1	1	1	40.00
7.81	1	1	1	41.00
7.83	1	1	1	42.00
7.85	1	1	1	43.00
7.87	1	1	1	44.00
7.89	1	1	1	45.00
7.91	1	1	1	46.00
7.93	1	1	1	47.00
7.95	1	1	1	48.00
7.97	1	1	1	49.00

7.99	1	1	1	50.00
8.01	1	1	1	51.00
8.03	1	1	1	52.00
8.05	1	1	1	53.00
8.07	1	1	1	54.00
8.09	1	1	1	55.00
8.11	1	1	1	56.00
8.13	1	1	1	57.00
8.15	1	1	1	58.00
8.17	1	1	1	59.00
8.19	1	1	1	60.00
8.21	1	1	1	61.00
8.23	1	1	1	62.00
8.25	1	1	1	63.00
8.27	1	1	1	64.00
8.29	1	1	1	65.00
8.31	1	1	1	66.00
8.33	1	1	1	67.00
8.35	1	1	1	68.00
8.37	1	1	1	69.00
8.39	1	1	1	70.00
8.41	1	1	1	71.00
8.43	1	1	1	72.00
8.45	1	1	1	73.00
8.47	1	1	1	74.00
8.49	1	1	1	75.00
8.52	1	1	1	76.00
8.54	1	1	1	77.00
8.56	1	1	1	78.00
8.58	1	1	1	79.00
8.60	1	1	1	80.00
8.62	1	1	1	81.00
8.64	1	1	1	82.00
8.66	1	1	1	83.00
8.68	1	1	1	84.00
8.70	1	1	1	85.00
8.72	1	1	1	86.00
8.74	1	1	1	87.00
8.76	1	1	1	88.00
8.78	1	1	1	89.00
8.80	1	1	1	90.00

```
8.82 | 1 |     1 |       1 |  91.00
8.84 | 1 |     1 |       1 |  92.00
8.86 | 1 |     1 |       1 |  93.00
8.88 | 1 |     1 |       1 |  94.00
8.90 | 1 |     1 |       1 |  95.00
8.92 | 1 |     1 |       1 |  96.00
8.94 | 1 |     1 |       1 |  97.00
8.96 | 1 |     1 |       1 |  98.00
8.98 | 1 |     1 |       1 |  99.00
9.00 | 1 |     1 |       1 | 100.00
<NA> | 0 |     0 |    <NA> |   <NA>
```

```
post <- expand.grid( mu=mu.list , sigma=sigma.list )

post %>%
  head() %>%
  flextable()
```

mu	sigma
150	7
150	7
150	7
150	7
150	7
151	7

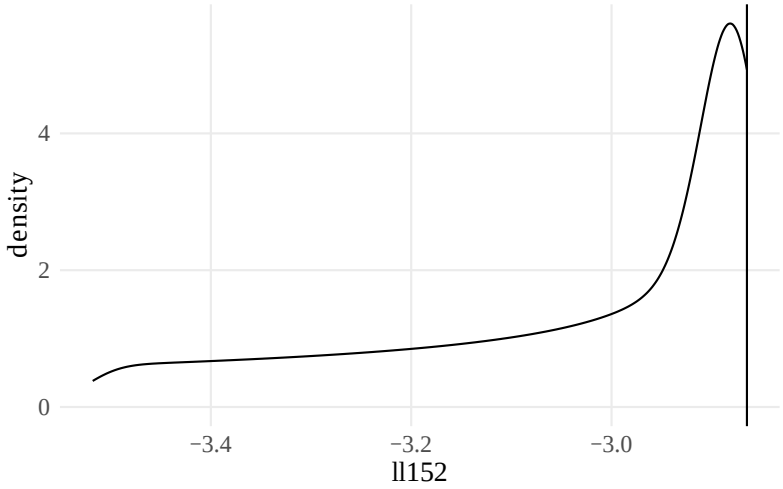
```
ht <- 152
mu <- 150
sig <- 7

post <- post %>%
  mutate(ll152 = dnorm(ht, mu, sig, log=T))

psych::describe(post)
```

n	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
	2.92	155	155	3.74	150	160	10	-6.94e-16	-1.2	0.0292
	0.583	8	8	0.749	7	9	2	3.53e-16	-1.2	0.00583
94	0.195	-2.96	-3.02	0.131	-3.52	-2.86	0.653	-0.94	-0.414	0.00195

```
ggplot(post, aes(ll152)) +
  geom_vline(xintercept=max(post$ll152)) +
  geom_density()
```




```
tll <- sum(post$ll152)
```

The log likelihood for the particular parameter value of mu=150 and sigma=7, for the datum of height=152, is -30,434.

```
d <- d %>%
  mutate(ll150_7=dnorm(height, mu, sig, log=T))

d %>%
  head() %>%
  flextable()
```

height	weight	age	male	ll150_7
152	47.8	63	1	-2.90
140	36.5	63	0	-3.95
137	31.9	65	0	-4.72
157	53.0	41	1	-3.34
145	41.3	51	0	-3.08
164	63.0	35	1	-4.82

```
sum(d$ll150_7)
```

```
[1] -1299
```

```
# textbook code

#post$LL <- sapply( 1:nrow(post) , function(i) sum(
#  dnorm( d$height , post$mu[i] , post$sigma[i] , log=TRUE ) ) )
```

```
post <- post %>%
  mutate(LL = sapply(1:nrow(post), function(i)
    dnorm(d$height,
          post$mu[i],
          post$sigma[i],
          log=T) %>%
      sum()
  ))
```

```
post <- post %>%
  mutate(LL = sapply(1:nrow(post), function(i)
    sum(
    dnorm(d$height,
          post$mu[i],
          post$sigma[i],
          log=T)
  )))
```

```
post %>%
  head() %>%
  flextable()
```

mu	sigma	ll152	LL
150	7	-2.91	-1,299
150	7	-2.90	-1,296
150	7	-2.90	-1,293
150	7	-2.89	-1,289
150	7	-2.89	-1,286
151	7	-2.89	-1,283

```
post$prod <- post$LL + dnorm( post$mu , 178 , 20 , TRUE ) +
  dunif( post$sigma , 0 , 50 , TRUE )
```

```
post$prob <- exp( post$prod - max(post$prod) )
```