

## Module 6 Lab Submission

Ben Tankus

Again we consider the Brain size data in the data set `case0902` from the `Sleuth3` library. You can read more about this data set by viewing the help file:

```
help(case0902)
```

```
## starting httpd help server ... done
```

```
head(case0902)
```

```
##           Species Brain   Body Gestation Litter
## 1           Aardvark   9.6   2.20         31    5.0
## 2           Acouchis   9.9   0.78         98    1.2
## 3 African elephant 4480.0 2800.00        655    1.0
## 4           Agoutis  20.3   2.80        104    1.3
## 5           Axis deer 219.0  89.00        218    1.0
## 6           Badger   53.0   6.00         60    2.2
```

In the previous Module, we considered a model using the log scale for all four variables, where the goal is to model log-brain-size as a function of log-gestation, log-littersize, and log-body-size. Here, we are going to arbitrarily set some of those predictor values to be missing:

```
set.seed(123)
brainMiss <- case0902
n <- nrow(brainMiss)
brainMiss$Body[sample(1:n, 10, replace=FALSE)] <- NA
brainMiss$Litter[sample(1:n, 7, replace=FALSE)] <- NA
brainMiss$Gestation[sample(1:n, 15, replace=FALSE)] <- NA
```

1. Create a visual representation of the missingness pattern in this data, using one of the methods demonstrated in the lab example.

```
brainMiss$zip <- rownames(brainMiss)
brainMiss_long <- gather(brainMiss, variable, value, -zip)
```

```
## Warning: attributes are not identical across measure variables;
## they will be dropped
```

```
(brainMiss_long)
```

##	zip	variable	value
## 1	1	Species	Aardvark
## 2	2	Species	Acouchis
## 3	3	Species	African elephant
## 4	4	Species	Agoutis
## 5	5	Species	Axis deer
## 6	6	Species	Badger
## 7	7	Species	Barbary sheep
## 8	8	Species	Barking deer
## 9	9	Species	Bat-eared fox
## 10	10	Species	Beaked whale
## 11	11	Species	Beaver
## 12	12	Species	Black buck antelope
## 13	13	Species	Bush baby
## 14	14	Species	Canadian beaver
## 15	15	Species	Capybara
## 16	16	Species	Caribou
## 17	17	Species	Cattle
## 18	18	Species	Chimpanzee
## 19	19	Species	Chinchilla
## 20	20	Species	Deer mouse I
## 21	21	Species	Deer mouse II
## 22	22	Species	Deer mouse III
## 23	23	Species	Deer mouse IV
## 24	24	Species	Dog
## 25	25	Species	Dolphin
## 26	26	Species	Domestic cat
## 27	27	Species	Domestic goat
## 28	28	Species	Domestic pig
## 29	29	Species	Domestic sheep
## 30	30	Species	Duikers
## 31	31	Species	Eland
## 32	32	Species	Elephant shrew I
## 33	33	Species	Elephant shrew II
## 34	34	Species	Elk
## 35	35	Species	Fallow deer
## 36	36	Species	Flying squirrel
## 37	37	Species	Fur seal
## 38	38	Species	Gentle lemur
## 39	39	Species	Gorilla
## 40	40	Species	Gray fox
## 41	41	Species	Grizzly bear
## 42	42	Species	Guinea pig
## 43	43	Species	Hamadryas baboon
## 44	44	Species	Hamster I
## 45	45	Species	Hamster II
## 46	46	Species	Harp seal
## 47	47	Species	Hedgehog
## 48	48	Species	Hippopotamus
## 49	49	Species	Hopping mouse
## 50	50	Species	Horse
## 51	51	Species	House mouse
## 52	52	Species	Howler monkey
## 53	53	Species	Human being

##	54	54	Species	Hyrax
##	55	55	Species	Jack rabbit
##	56	56	Species	Kinkajou
##	57	57	Species	Leaf monkey
##	58	58	Species	Lemur
##	59	59	Species	Leopard
##	60	60	Species	Lion
##	61	61	Species	Llama
##	62	62	Species	Long-nose armadillo
##	63	63	Species	Lynx
##	64	64	Species	Nutria
##	65	65	Species	Orangutan
##	66	66	Species	Porcupine I
##	67	67	Species	Porcupine II
##	68	68	Species	Porcupine III
##	69	69	Species	Porpoise
##	70	70	Species	Pygmy gerbil
##	71	71	Species	Pygmy hippopotamus
##	72	72	Species	Quokka
##	73	73	Species	Raccoon
##	74	74	Species	Rat I
##	75	75	Species	Rat II
##	76	76	Species	Red deer
##	77	77	Species	Red fox
##	78	78	Species	Rhesus monkey I
##	79	79	Species	Rhesus monkey II
##	80	80	Species	Ring-tail monkey
##	81	81	Species	Sambar
##	82	82	Species	Sea lion
##	83	83	Species	Slow loris
##	84	84	Species	Spider monkey I
##	85	85	Species	Spider monkey II
##	86	86	Species	Tapir
##	87	87	Species	Tiger
##	88	88	Species	Tree shrew
##	89	89	Species	Tree squirrel
##	90	90	Species	Vervet guenon
##	91	91	Species	Vicuna
##	92	92	Species	Weddell seal
##	93	93	Species	Western baboon
##	94	94	Species	White-handed gibbon
##	95	95	Species	Wild boar
##	96	96	Species	Yak
##	97	1	Brain	9.6
##	98	2	Brain	9.9
##	99	3	Brain	4480
##	100	4	Brain	20.3
##	101	5	Brain	219
##	102	6	Brain	53
##	103	7	Brain	210
##	104	8	Brain	124
##	105	9	Brain	28.5
##	106	10	Brain	500
##	107	11	Brain	45

##	108	12	Brain	200
##	109	13	Brain	9.9
##	110	14	Brain	40
##	111	15	Brain	76
##	112	16	Brain	288
##	113	17	Brain	456
##	114	18	Brain	360
##	115	19	Brain	5.25
##	116	20	Brain	0.68
##	117	21	Brain	0.63
##	118	22	Brain	0.52
##	119	23	Brain	0.69
##	120	24	Brain	70.2
##	121	25	Brain	1600
##	122	26	Brain	28.4
##	123	27	Brain	106
##	124	28	Brain	180
##	125	29	Brain	125
##	126	30	Brain	93
##	127	31	Brain	480
##	128	32	Brain	1.14
##	129	33	Brain	1.37
##	130	34	Brain	365
##	131	35	Brain	223
##	132	36	Brain	1.89
##	133	37	Brain	355
##	134	38	Brain	7.8
##	135	39	Brain	406
##	136	40	Brain	37.3
##	137	41	Brain	400
##	138	42	Brain	4.28
##	139	43	Brain	183
##	140	44	Brain	0.67
##	141	45	Brain	1.12
##	142	46	Brain	442
##	143	47	Brain	3.5
##	144	48	Brain	590
##	145	49	Brain	1.18
##	146	50	Brain	712
##	147	51	Brain	0.45
##	148	52	Brain	54
##	149	53	Brain	1300
##	150	54	Brain	20.5
##	151	55	Brain	13.3
##	152	56	Brain	31.2
##	153	57	Brain	65.5
##	154	58	Brain	22
##	155	59	Brain	157
##	156	60	Brain	260
##	157	61	Brain	225
##	158	62	Brain	12
##	159	63	Brain	75
##	160	64	Brain	23
##	161	65	Brain	343

##	162	66	Brain	37
##	163	67	Brain	37
##	164	68	Brain	24
##	165	69	Brain	537
##	166	70	Brain	1.04
##	167	71	Brain	260
##	168	72	Brain	17.5
##	169	73	Brain	41.6
##	170	74	Brain	0.72
##	171	75	Brain	2.38
##	172	76	Brain	435
##	173	77	Brain	48
##	174	78	Brain	84.6
##	175	79	Brain	107
##	176	80	Brain	73
##	177	81	Brain	383
##	178	82	Brain	363
##	179	83	Brain	12.8
##	180	84	Brain	114
##	181	85	Brain	109
##	182	86	Brain	250
##	183	87	Brain	302
##	184	88	Brain	3.15
##	185	89	Brain	6.23
##	186	90	Brain	67
##	187	91	Brain	198
##	188	92	Brain	550
##	189	93	Brain	179
##	190	94	Brain	102
##	191	95	Brain	185
##	192	96	Brain	334
##	193	1	Body	2.2
##	194	2	Body	0.78
##	195	3	Body	2800
##	196	4	Body	2.8
##	197	5	Body	89
##	198	6	Body	6
##	199	7	Body	66
##	200	8	Body	16
##	201	9	Body	3.2
##	202	10	Body	250
##	203	11	Body	25
##	204	12	Body	39
##	205	13	Body	0.7
##	206	14	Body	<NA>
##	207	15	Body	30
##	208	16	Body	110
##	209	17	Body	520
##	210	18	Body	45
##	211	19	Body	0.43
##	212	20	Body	0.027
##	213	21	Body	0.026
##	214	22	Body	0.017
##	215	23	Body	0.024

##	216	24	Body	8.5
##	217	25	Body	<NA>
##	218	26	Body	2.5
##	219	27	Body	30
##	220	28	Body	190
##	221	29	Body	49
##	222	30	Body	13
##	223	31	Body	<NA>
##	224	32	Body	0.049
##	225	33	Body	0.064
##	226	34	Body	120
##	227	35	Body	80
##	228	36	Body	0.052
##	229	37	Body	250
##	230	38	Body	0.22
##	231	39	Body	140
##	232	40	Body	3.8
##	233	41	Body	250
##	234	42	Body	<NA>
##	235	43	Body	<NA>
##	236	44	Body	0.036
##	237	45	Body	0.13
##	238	46	Body	110
##	239	47	Body	0.93
##	240	48	Body	1400
##	241	49	Body	0.15
##	242	50	Body	<NA>
##	243	51	Body	<NA>
##	244	52	Body	7.7
##	245	53	Body	65
##	246	54	Body	3.8
##	247	55	Body	2.9
##	248	56	Body	2
##	249	57	Body	5.8
##	250	58	Body	2.1
##	251	59	Body	46
##	252	60	Body	180
##	253	61	Body	93
##	254	62	Body	3.7
##	255	63	Body	12
##	256	64	Body	5
##	257	65	Body	37
##	258	66	Body	11
##	259	67	Body	<NA>
##	260	68	Body	6.6
##	261	69	Body	56
##	262	70	Body	0.065
##	263	71	Body	150
##	264	72	Body	3.5
##	265	73	Body	5.3
##	266	74	Body	0.05
##	267	75	Body	0.34
##	268	76	Body	200
##	269	77	Body	6

##	270	78	Body	6
##	271	79	Body	<NA>
##	272	80	Body	3.7
##	273	81	Body	120
##	274	82	Body	100
##	275	83	Body	1.2
##	276	84	Body	9.1
##	277	85	Body	7.7
##	278	86	Body	230
##	279	87	Body	210
##	280	88	Body	0.15
##	281	89	Body	0.33
##	282	90	Body	4.6
##	283	91	Body	45
##	284	92	Body	400
##	285	93	Body	<NA>
##	286	94	Body	5.5
##	287	95	Body	150
##	288	96	Body	250
##	289	1	Gestation	31
##	290	2	Gestation	98
##	291	3	Gestation	655
##	292	4	Gestation	104
##	293	5	Gestation	218
##	294	6	Gestation	60
##	295	7	Gestation	<NA>
##	296	8	Gestation	183
##	297	9	Gestation	<NA>
##	298	10	Gestation	240
##	299	11	Gestation	128
##	300	12	Gestation	180
##	301	13	Gestation	135
##	302	14	Gestation	128
##	303	15	Gestation	<NA>
##	304	16	Gestation	225
##	305	17	Gestation	280
##	306	18	Gestation	230
##	307	19	Gestation	110
##	308	20	Gestation	23
##	309	21	Gestation	23
##	310	22	Gestation	24
##	311	23	Gestation	24
##	312	24	Gestation	63
##	313	25	Gestation	360
##	314	26	Gestation	<NA>
##	315	27	Gestation	151
##	316	28	Gestation	115
##	317	29	Gestation	150
##	318	30	Gestation	120
##	319	31	Gestation	255
##	320	32	Gestation	<NA>
##	321	33	Gestation	46
##	322	34	Gestation	235
##	323	35	Gestation	240

## 324	36 Gestation	<NA>
## 325	37 Gestation	254
## 326	38 Gestation	145
## 327	39 Gestation	265
## 328	40 Gestation	63
## 329	41 Gestation	<NA>
## 330	42 Gestation	<NA>
## 331	43 Gestation	<NA>
## 332	44 Gestation	21
## 333	45 Gestation	16
## 334	46 Gestation	240
## 335	47 Gestation	34
## 336	48 Gestation	240
## 337	49 Gestation	27
## 338	50 Gestation	330
## 339	51 Gestation	19
## 340	52 Gestation	139
## 341	53 Gestation	270
## 342	54 Gestation	225
## 343	55 Gestation	41
## 344	56 Gestation	77
## 345	57 Gestation	168
## 346	58 Gestation	135
## 347	59 Gestation	92
## 348	60 Gestation	108
## 349	61 Gestation	330
## 350	62 Gestation	120
## 351	63 Gestation	60
## 352	64 Gestation	132
## 353	65 Gestation	270
## 354	66 Gestation	112
## 355	67 Gestation	112
## 356	68 Gestation	113
## 357	69 Gestation	270
## 358	70 Gestation	21
## 359	71 Gestation	205
## 360	72 Gestation	26
## 361	73 Gestation	63
## 362	74 Gestation	23
## 363	75 Gestation	21
## 364	76 Gestation	<NA>
## 365	77 Gestation	52
## 366	78 Gestation	<NA>
## 367	79 Gestation	165
## 368	80 Gestation	180
## 369	81 Gestation	<NA>
## 370	82 Gestation	343
## 371	83 Gestation	<NA>
## 372	84 Gestation	140
## 373	85 Gestation	140
## 374	86 Gestation	390
## 375	87 Gestation	104
## 376	88 Gestation	46
## 377	89 Gestation	38

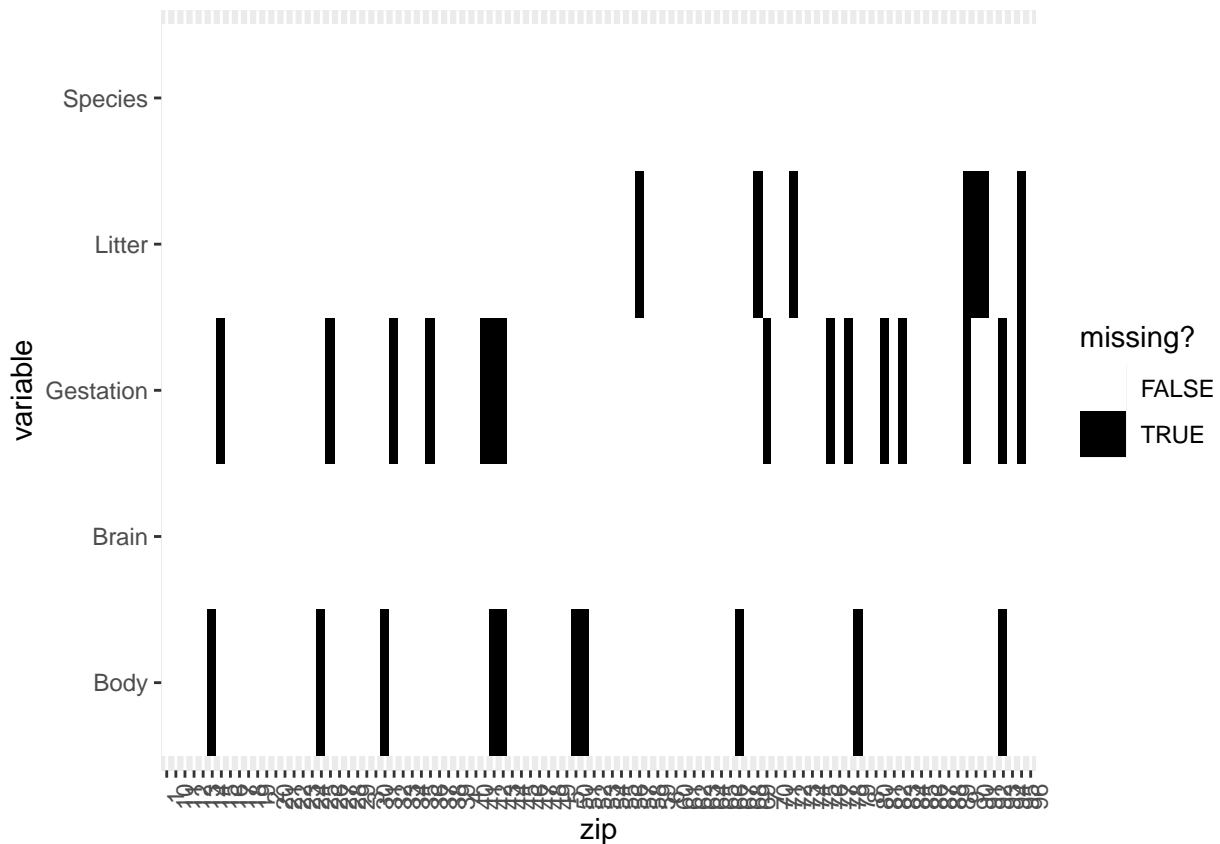


##	378	90	Gestation	195
##	379	91	Gestation	300
##	380	92	Gestation	310
##	381	93	Gestation	<NA>
##	382	94	Gestation	210
##	383	95	Gestation	<NA>
##	384	96	Gestation	255
##	385	1	Litter	5
##	386	2	Litter	1.2
##	387	3	Litter	1
##	388	4	Litter	1.3
##	389	5	Litter	1
##	390	6	Litter	2.2
##	391	7	Litter	1.2
##	392	8	Litter	1.1
##	393	9	Litter	<NA>
##	394	10	Litter	1.8
##	395	11	Litter	4
##	396	12	Litter	1
##	397	13	Litter	1
##	398	14	Litter	2.9
##	399	15	Litter	3
##	400	16	Litter	1
##	401	17	Litter	1
##	402	18	Litter	1
##	403	19	Litter	2
##	404	20	Litter	3.7
##	405	21	Litter	5
##	406	22	Litter	5
##	407	23	Litter	5
##	408	24	Litter	4
##	409	25	Litter	1
##	410	26	Litter	4
##	411	27	Litter	2
##	412	28	Litter	8
##	413	29	Litter	2.4
##	414	30	Litter	1
##	415	31	Litter	1
##	416	32	Litter	1.5
##	417	33	Litter	1.5
##	418	34	Litter	1
##	419	35	Litter	1
##	420	36	Litter	3.1
##	421	37	Litter	1
##	422	38	Litter	2
##	423	39	Litter	1
##	424	40	Litter	3.7
##	425	41	Litter	2.3
##	426	42	Litter	2.6
##	427	43	Litter	1
##	428	44	Litter	4.6
##	429	45	Litter	6.3
##	430	46	Litter	1
##	431	47	Litter	4.6

##	432	48	Litter	1
##	433	49	Litter	5.6
##	434	50	Litter	1
##	435	51	Litter	5
##	436	52	Litter	1
##	437	53	Litter	1
##	438	54	Litter	2.4
##	439	55	Litter	2.5
##	440	56	Litter	1.1
##	441	57	Litter	<NA>
##	442	58	Litter	1
##	443	59	Litter	2.5
##	444	60	Litter	3
##	445	61	Litter	1
##	446	62	Litter	4
##	447	63	Litter	2.5
##	448	64	Litter	5.5
##	449	65	Litter	1
##	450	66	Litter	1.2
##	451	67	Litter	1.2
##	452	68	Litter	1
##	453	69	Litter	<NA>
##	454	70	Litter	4
##	455	71	Litter	1
##	456	72	Litter	<NA>
##	457	73	Litter	3.5
##	458	74	Litter	7.3
##	459	75	Litter	8
##	460	76	Litter	1
##	461	77	Litter	4
##	462	78	Litter	1
##	463	79	Litter	1.1
##	464	80	Litter	1
##	465	81	Litter	1.1
##	466	82	Litter	1
##	467	83	Litter	1.2
##	468	84	Litter	1
##	469	85	Litter	1
##	470	86	Litter	1
##	471	87	Litter	3
##	472	88	Litter	3
##	473	89	Litter	3
##	474	90	Litter	<NA>
##	475	91	Litter	<NA>
##	476	92	Litter	1
##	477	93	Litter	1
##	478	94	Litter	1
##	479	95	Litter	<NA>
##	480	96	Litter	1

```
qplot(zip, variable, data = brainMiss_long, geom = "tile",
fill = is.na(value)) +
scale_fill_manual("missing?",
values = c('TRUE' = "black", 'FALSE' = "white")) +
```

```
theme(axis.text.x = element_text(angle = 90))
```



2. Use the `amelia` function to perform multiple imputation for these missing values, based on `m=50` imputed datasets.

```
brainMiss <- case0902
n <- nrow(brainMiss)
brainMiss$Body[sample(1:n, 10, replace=FALSE)] <- NA
brainMiss$Litter[sample(1:n, 7, replace=FALSE)] <- NA
brainMiss$Gestation[sample(1:n, 15, replace=FALSE)] <- NA

brainMissQuant <- subset(brainMiss, select = -Species)

n.imp <- 50
ame.imp <- amelia(brainMissQuant, m=n.imp, p2s = 0, log = 2:4)
```

3. Fit models using each of the imputed datasets

```
betas <- matrix(0, nrow=n.imp, ncol=4)
ses <- matrix(0, nrow=n.imp, ncol=4)

for(i in 1:n.imp){
  newMod <- lm(log(Brain) ~ log(Body) + log(Gestation) + log(Litter), data=ame.imp$imputations[[i]])
```

```

betas[i,] <- coef(newMod)
14
ses[i,] <- coef(summary(newMod))[,2]
}

```

4. Use the `mi.meld` function to find the multiple imputation estimates of the coefficients.

```

mi.meld(q = betas, se = ses)

## $q.mi
##           [,1]      [,2]      [,3]      [,4]
## [1,] 0.2368904 0.5377797 0.5587142 -0.3184645
##
## $se.mi
##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.288768 0.06001072 0.273028 0.2040684

```

Recall the coefficients from the model constructed on the full data (with no missing values):

```

origMod <- lm(log(Brain) ~ log(Gestation) + log(Litter) + log(Body), data=case0902)
summary(origMod)

##
## Call:
## lm(formula = log(Brain) ~ log(Gestation) + log(Litter) + log(Body),
##     data = case0902)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95415 -0.29639 -0.03105  0.28111  1.57491
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.85482    0.66167   1.292  0.19962
## log(Gestation) 0.41794    0.14078   2.969  0.00381 **
## log(Litter)   -0.31007    0.11593  -2.675  0.00885 **
## log(Body)      0.57507    0.03259  17.647 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4748 on 92 degrees of freedom
## Multiple R-squared:  0.9537, Adjusted R-squared:  0.9522
## F-statistic: 631.6 on 3 and 92 DF, p-value: < 2.2e-16

```

5. How do the coefficient estimates from the multiple imputation compare to the coefficient estimates on the original full data?

The coefficient estimates from the imputation are in the same ballpark as the original full data. Imputation via linear modeling in this case seems to be an over-estimate by roughly one standard error.