

# Population Genetics Homework - Week 14

Christian Polania

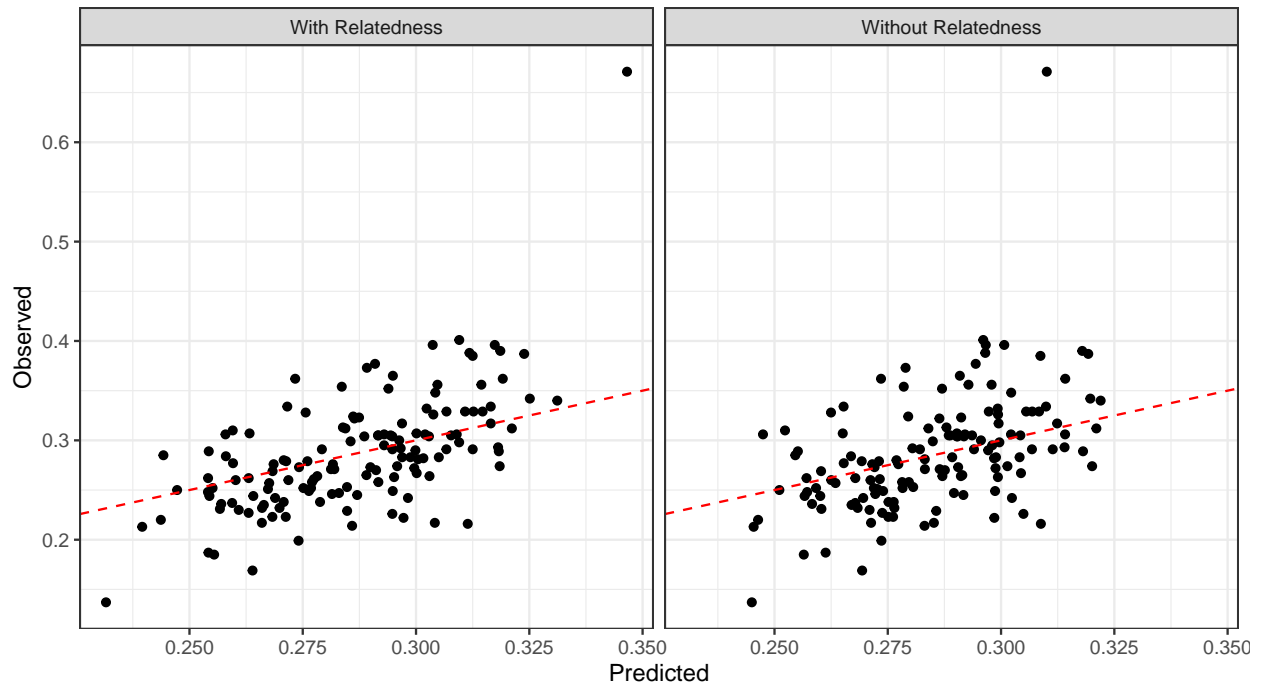
Just to remind both you and myself: I'm using the pre-analyzed data provided with Lab 13, and you can let me know if you'd prefer that I run the rstan code myself once I have more processing power.

1. All three analyses (for Mass, PD, and TDT) are below. Shown are parameters, plots, and  $R^2$  estimates.

```
fit_Mass <- brm(Mass ~ X34862 + X29507 + X44522 + X40856 + X89362 + X65887 + X36978 +  
  X89363 + X89364 + X17757 + (1|gr(sample, cov = rel)),  
  data = dat,  
  family = gaussian(),  
  set_prior(horseshoe(df = 3, par_ratio = 0.5)),  
  data2 = list(rel = rel),  
  iter = 5000,  
  refresh = 0)
```

```
fixef(fit_Mass, summary = TRUE)
```

	Estimate	Est.Error	Q2.5	Q97.5
Intercept	0.285993848	0.016742064	0.253073353	0.3186890982
X34862	0.008162337	0.006829009	-0.001841795	0.0225749682
X29507	-0.004145400	0.005742303	-0.018268967	0.0040484745
X44522	-0.007589986	0.006498457	-0.021634709	0.0016828795
X40856	0.006162575	0.006912101	-0.003279557	0.0221739064
X89362	0.002922440	0.008446758	-0.011541495	0.0226350469
X65887	-0.011186976	0.007072699	-0.024974730	0.0005560791
X36978	0.005982348	0.006067908	-0.002512748	0.0198691084
X89363	0.003432673	0.008345404	-0.010651947	0.0243085867
X89364	0.003040303	0.008495817	-0.011527210	0.0229228271
X17757	-0.004508206	0.005901391	-0.018672216	0.0037772402

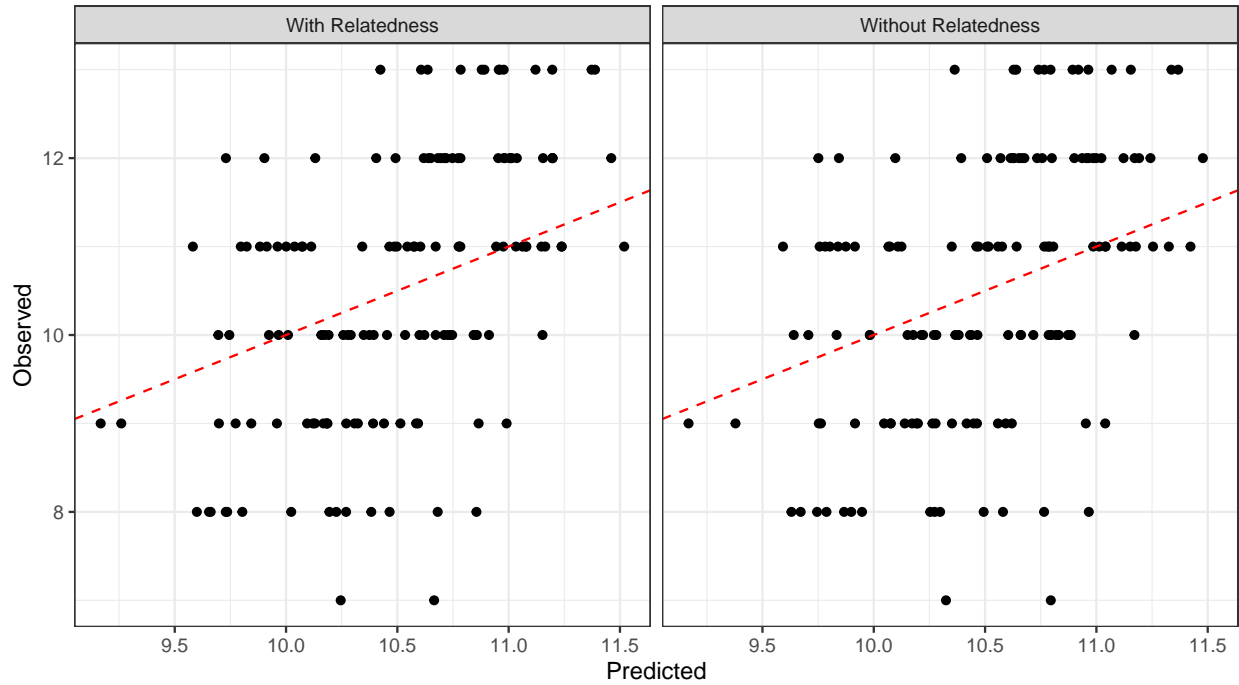


	R2	error	Q2.5	Q97.5
with relatedness	0.2224987	0.07736938	0.07705219	0.3760999
without relatedness	0.1457340	0.06627432	0.01821219	0.2744413
ratio	0.6549879	0.85659620	NA	NA

```
fit_PD <- brm(PD ~ X10472 + X66820 + X13105 + X79056 + X86539 + X86541 + X86543 +
  X86542 + X82900 + X21382 + (1|gr(sample, cov = rel)),
  data = dat,
  family = gaussian(),
  set_prior(horseshoe(df = 3, par_ratio = 0.5)),
  data2 = list(rel = rel),
  iter = 5000,
  refresh = 0)
```

```
fixef(fit_PD, summary = TRUE)
```

	Estimate	Est.Error	Q2.5	Q97.5
Intercept	10.16860852	0.3922227	9.373987261	10.93404245
X10472	0.34934929	0.2004970	-0.005732462	0.73989953
X66820	-0.20748275	0.1716010	-0.575332639	0.04179527
X13105	-0.26570189	0.1866982	-0.644572861	0.02142044
X79056	-0.09228118	0.1429711	-0.443913353	0.12553834
X86539	0.05553486	0.2228142	-0.351407176	0.58070713
X86541	0.06046770	0.2206207	-0.367937698	0.59080390
X86543	0.05769050	0.2203297	-0.356743158	0.57726105
X86542	0.05907905	0.2203531	-0.346266479	0.56235522
X82900	0.09835803	0.1333603	-0.104908084	0.41437781
X21382	0.13936613	0.1572623	-0.076562284	0.50712271

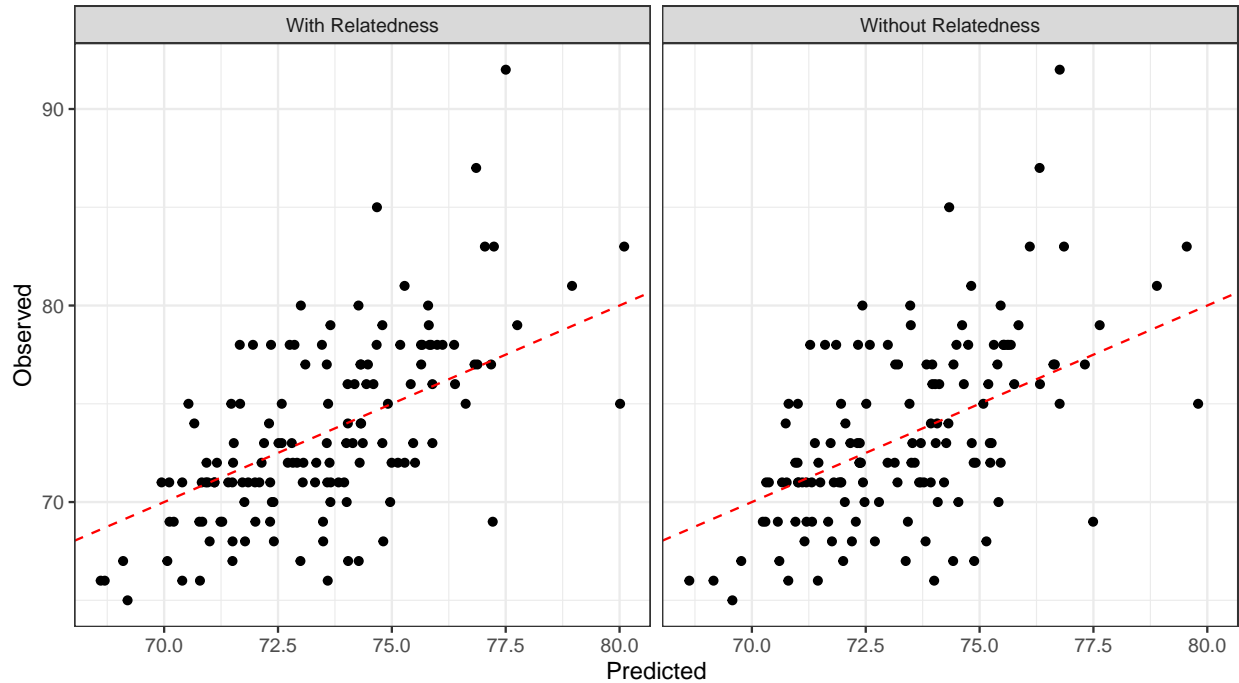


	R2	error	Q2.5	Q97.5
with relatedness	0.1635117	0.06547833	0.04299609	0.2990859
without relatedness	0.1395698	0.06243714	0.02299428	0.2662653
ratio	0.8535769	0.95355416	NA	NA

```
fit_TDT <- brm(TDT ~ X4612 + X36978 + X16239 + X71727 + X16238 + X54640 + X54640 +
  X44522 + X65887 + X86539 + (1|gr(sample, cov = rel)),
  data = dat,
  family = gaussian(),
  set_prior(horseshoe(df = 3, par_ratio = 0.5)),
  data2 = list(rel = rel),
  iter = 5000,
  refresh = 0)
```

```
fixef(fit_TDT, summary = TRUE)
```

	Estimate	Est.Error	Q2.5	Q97.5
Intercept	76.4772063	1.4653530	73.59547434	79.41005660
X4612	0.8992803	0.6073762	-0.13493409	2.13581694
X36978	1.0066151	0.4965217	0.01774889	1.94943427
X16239	-0.6711920	0.8467994	-2.38904414	0.93093932
X71727	-1.2319780	0.6069259	-2.42426907	-0.03978019
X16238	-0.5959653	0.8330811	-2.30920731	0.96019494
X54640	0.3818332	0.4483295	-0.41860141	1.32565704
X44522	-1.1129431	0.4879845	-2.05554324	-0.11302585
X65887	-1.2933417	0.4817098	-2.19305459	-0.27222447
X86539	-0.7982282	0.5348823	-1.88176639	0.12608791



	R2	error	Q2.5	Q97.5
with relatedness	0.3089830	0.07283663	0.1669808	0.4567522
without relatedness	0.2563782	0.06692869	0.1133732	0.3772701
ratio	0.8297486	0.91888769	NA	NA

2. Including the relatedness data, the first ten loci account for a sizable proportion of the variation in each trait.

- Mass: 22.4%
- PD: 16.2%
- TDT: 31.0%

3. Excluding relatedness, the explanatory power of ten loci from TDT seems higher than that of the other two traits. Adding relatedness helps each trait about equally, except for PD.

4. By including a relatedness matrix, one is solving the problem of the similarity of two related genomes; that is, two related moths will be similar in trait and similar at most loci, but that doesn't mean that every locus that's similar between the two is contributing to the trait.

My guess for why that can go wrong: considering will trade one problem for another: the way that the sampled population is unique in comparison to other populations will be confounded with genotype-phenotype association. This should be there anyways, but maybe accounting for individual relatedness amplifies that larger-scale problem? All I know for sure is that sampling across as many populations as possible would solve it either way.