HW #9

9., 9.4, 9.10, 9.12, 9.16, 9.20

Hayden Atchley

2022-11-27

9.1

a

flowers ~ light + time

b

flowers ~ light + time + light*time

9.4

Though the gestation differs, there are other explanatory variables in the model, so a difference in gestation does not necessarily imply the same type of difference in brain weight.

9.10

a

After accounting for all other effects, the estimated mean IQ is 8.3 points higher for those who received breast milk than those who didn't.

b

The 'Social class' and 'Mother's education' variables are categorical, so the specific numerical values of these variables are not especially meaningful. However, social class and education are somewhat ordered designations in reality, so provided the groups are ordered correctly the coefficient on these variables will give some information about the effect of these variables on the output. The actual coefficient value (the 'slope'), however, is meaningless (unless used as a comparison between models).

c

Though the actual p-value could be orders of magnitude smaller than 0.0001, any of these values provides overwhelming evidence of a relationship. For this study, a reported p-value of '<0.0001' is sufficient.

d

In this study, no babies were breast-fed (they were all fed from a tube), though some were fed with breast milk separately.

e

The intellectual development of a child is often correlated with the financial situation of the parents (and therefore the ability to provide for quality education), and so social class can have a huge effect on intellectual development. The education of the mother can also play an important role in intellectual development, as the mother can herself be a source of education. There could also be an 'inspiration' effect, where the child is inspired to further their education because their mother also did so.

f

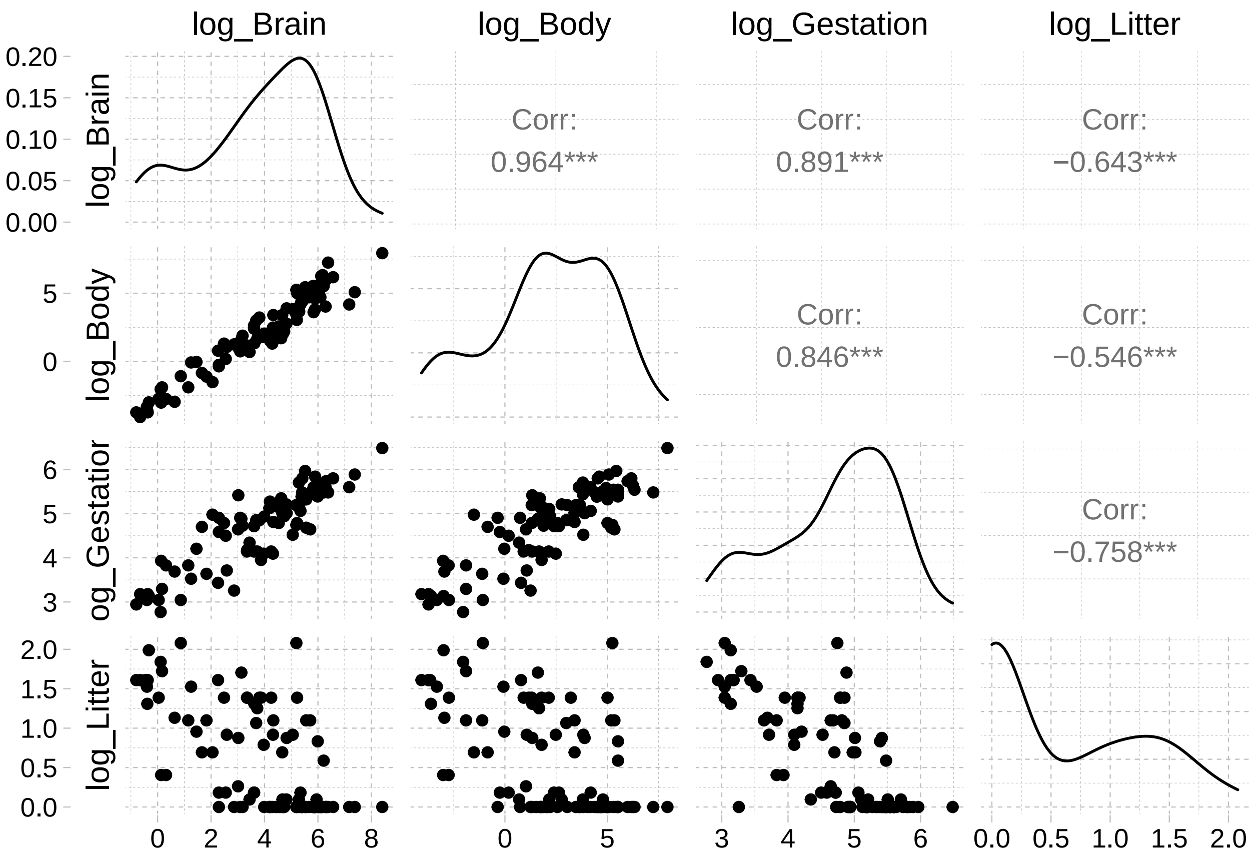
i

4.5

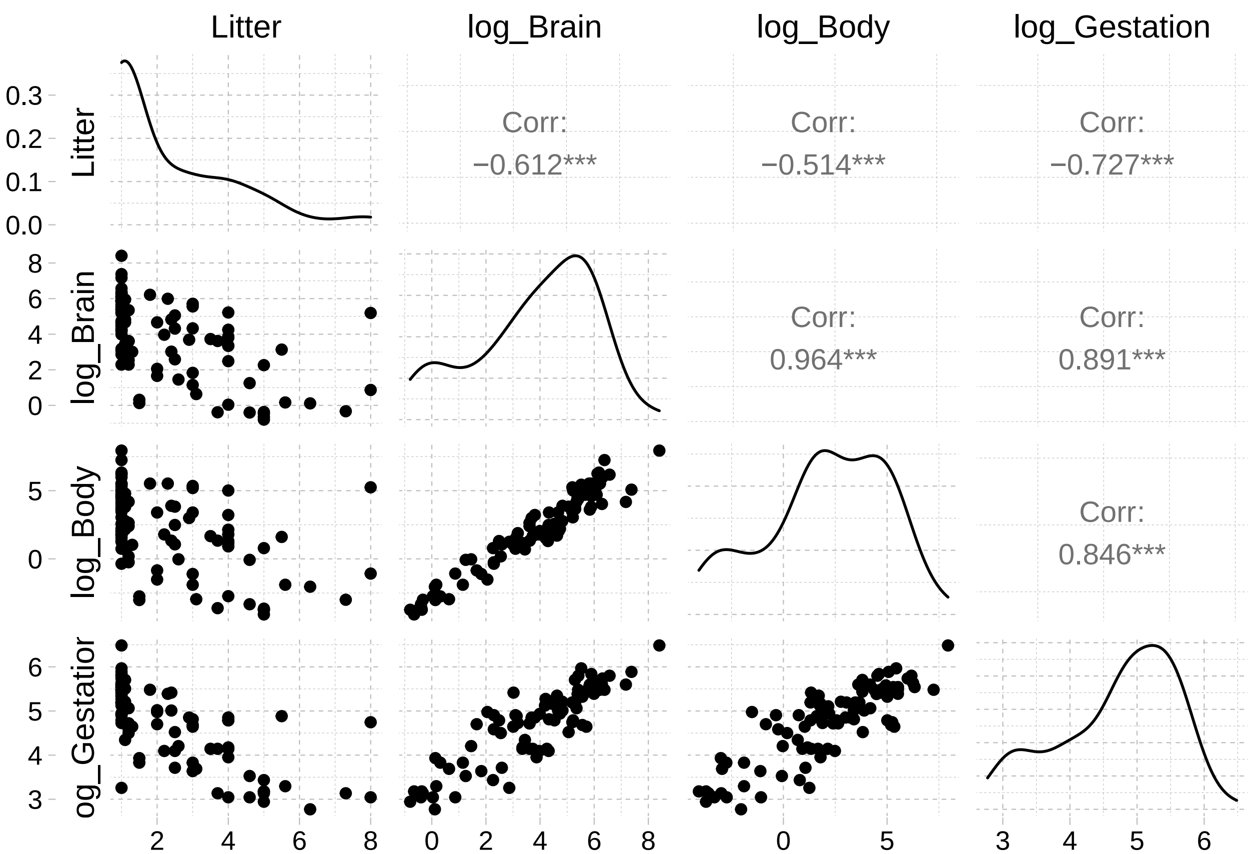
ii

The original model equated any amount of breast milk with exclusively breast milk, when in reality this was a continuous variable. The original coefficient could be said to represent an average increase for anyone who received any breast milk, whereas this continuous variable gives a more precise picture.

9.12

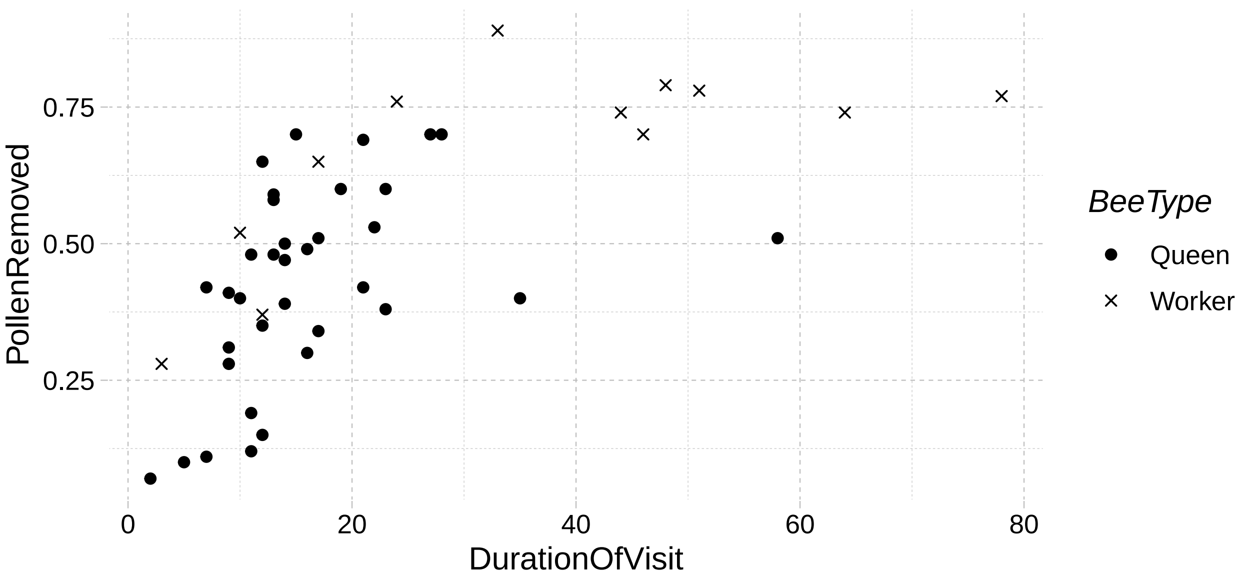


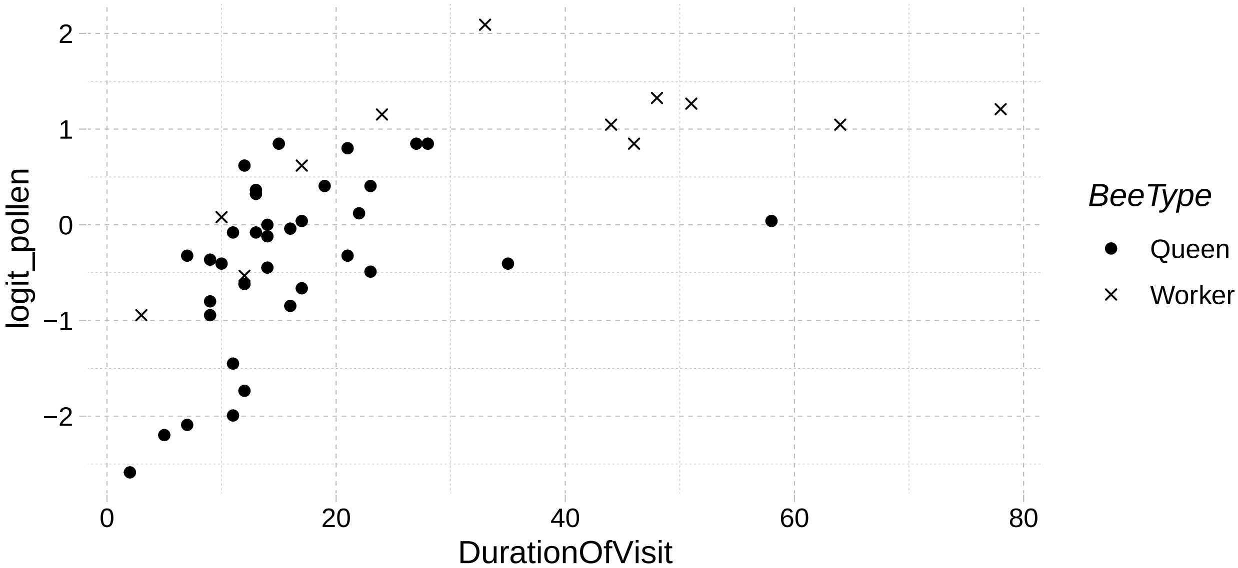
term	estimate	std.error	statistic	p.value
(Intercept)	0.8548	0.6617	1.29	0.1996
log_Body	0.5751	0.0326	17.65	0.0000
log_Gestation	0.4179	0.1408	2.97	0.0038
log_Litter	-0.3101	0.1159	-2.67	0.0089

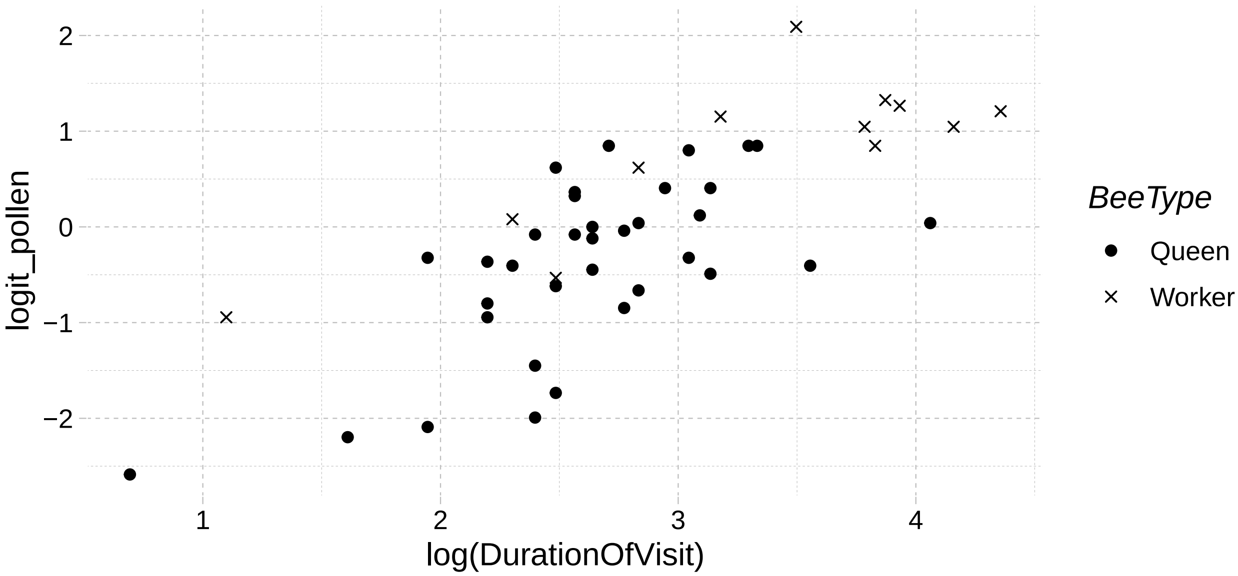


The fit of `log_Brain ~ Litter` appears *slightly* better than the fit of `log_Brain ~ log_Litter`, but neither appears to be a very tight fit.

9.16







Of these three plots, `logit_pollen ~ log(DurationOfVisit)` seems the most promising for linear regression. We fit this model:

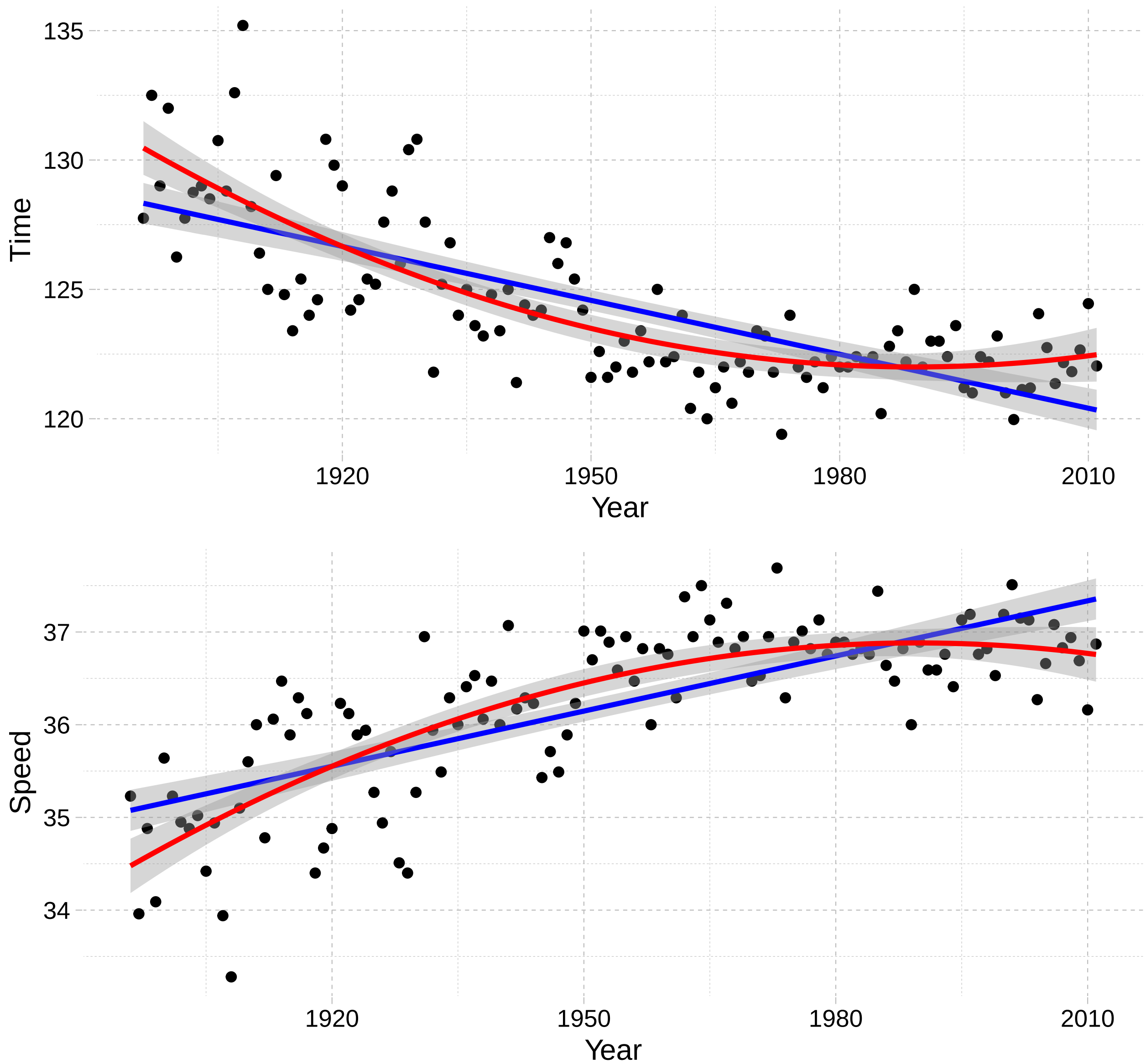
term	estimate	std.error	statistic	p.value
(Intercept)	-3.04	0.511	-5.94	0.0000
log(DurationOfVisit)	1.01	0.190	5.32	0.0000
BeeTypeWorker	1.38	0.872	1.58	0.1217
log(DurationOfVisit):BeeTypeWorker	-0.27	0.282	-0.96	0.3416

This shows no evidence that the amount of pollen removed depends on the type of bee after accounting for all other variables. We fit without the interaction term:

term	estimate	std.error	statistic	p.value
(Intercept)	-2.71	0.384	-7.07	0.0000
log(DurationOfVisit)	0.89	0.140	6.34	0.0000
BeeTypeWorker	0.57	0.236	2.41	0.0202

This gives strong evidence that worker bees remove more pollen than queen bees, accounting for duration of visit.

9.20

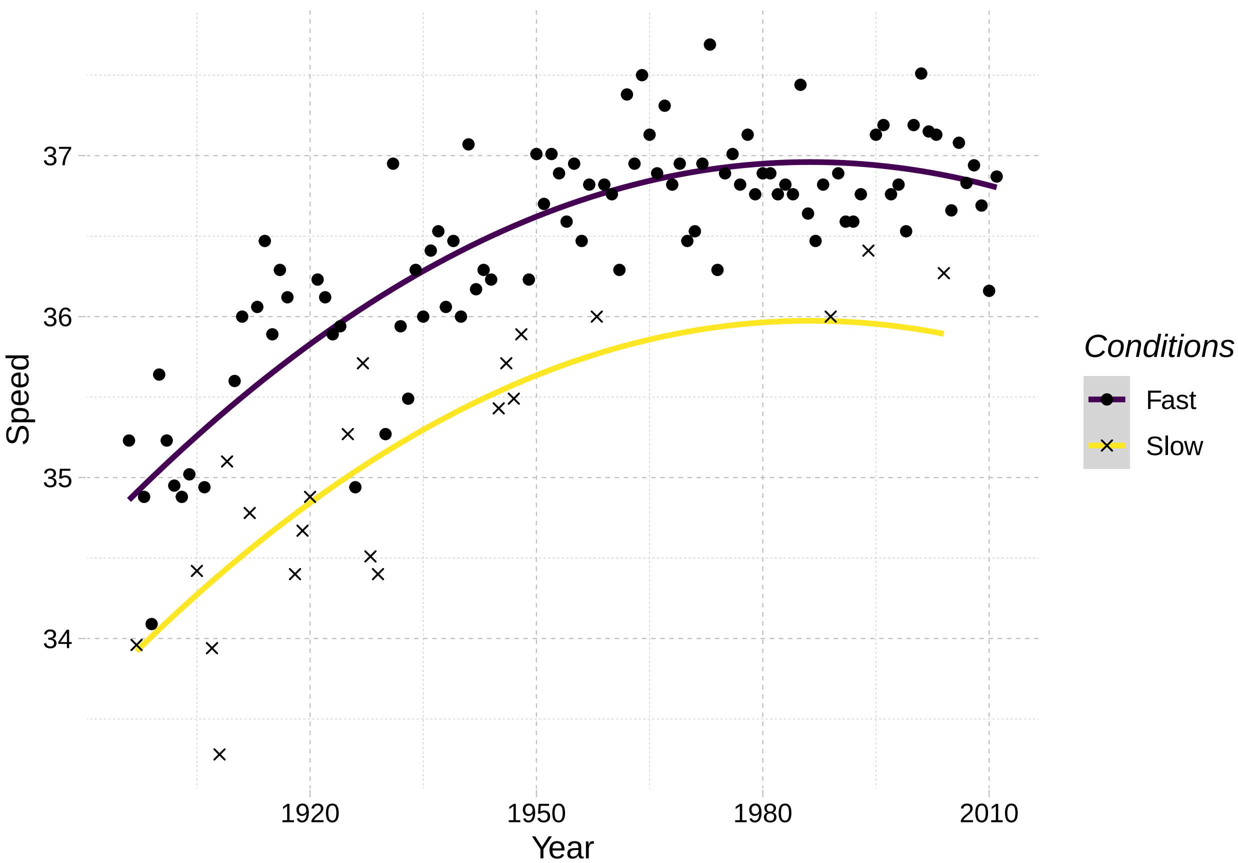


Using either speed or time, a quadratic regression is significantly better than a linear one. However, both models at a glance appear to have about equal fit. A summary of both models:

	Speed Model				Time Model			
term	estimate	std.error	statistic	p.value	estimate	std.error	statistic	p.value
(Intercept)	36.22	0.050	720.9	0e+00	124.3	0.18	699.9	0e+00
poly(Year, 2)1	7.15	0.541	13.2	0e+00	-25.1	1.91	-13.1	0e+00
poly(Year, 2)2	-2.96	0.541	-5.5	3e-07	10.6	1.91	5.5	2e-07

Both models fit exactly as well as each other (even the standard errors of the `Year` terms are reciprocals). I will use the `Speed` model going forward.

Adding the track conditions to the model:



term	estimate	std.error	statistic	p.value
(Intercept)	36.39	0.041	890.92	0.00e+00
poly(Year, 2)1	6.09	0.410	14.86	0.00e+00
poly(Year, 2)2	-2.78	0.396	-7.03	1.76e-10
ConditionsSlow	-0.99	0.099	-9.97	0.00e+00

Based on this model, winning track speeds are on average about 1 mph faster in fast conditions than slow conditions.

Adding the number of horses in the race to the model:

term	estimate	std.error	statistic	p.value
(Intercept)	36.68	0.138	265.82	0.000
poly(Year, 2)1	6.79	0.520	13.06	0.000
poly(Year, 2)2	-2.95	0.398	-7.42	0.000
ConditionsSlow	-0.97	0.098	-9.90	0.000
Starters	-0.02	0.010	-2.13	0.035

This shows pretty good evidence of the number of horses having a slight effect on the winning speed. With an interaction term:

term	estimate	std.error	statistic	p.value
(Intercept)	36.74	0.154	238.60	0.000
poly(Year, 2)1	6.86	0.527	13.02	0.000
poly(Year, 2)2	-2.91	0.401	-7.26	0.000
ConditionsSlow	-1.18	0.254	-4.62	0.000
Starters	-0.02	0.011	-2.30	0.023
ConditionsSlow:Starters	0.02	0.018	0.89	0.377

This shows no evidence of an interaction between number of horses and track conditions, i.e. that the affect of the number of horses is different depending on the track conditions.