Exploring Factors of Tulip Germination

Stephen Merrill

Brigham Young University

February 11, 2017

Background

- Tulips require a period of cool dormancy "Chilling Time" in order to germinate
- Over 75 species exist - 12 considered, with one removed
- Netherlands - 9 million bulbs produced, 25% of agricultural exports, cultural importance
- Temperature trends carry increased risk of flooding, devastating to tulips

Goals

- Understand the effect of chilling time on germination
- Is there a population effect? Does it vary from population to population?
- What is the ideal chilling time? Does it vary in population?
- How would a decrease in chilling time from 10 to 8 weeks affect germination?

Data

- Germination response is a 0-1 categorical variable
- 210 observations of 11 populations, 30 assigned to each Chilling Time (continuous, in weeks)

	0	2	4	6	8	10	12
1	0.40	0.97	0.83	0.87	0.87	0.97	0.90
2	0.13	0.53	0.73	0.73	0.83	0.90	0.83
3	0.00	0.53	0.80	0.83	0.97	0.90	0.87
4	0.00	0.17	0.53	0.60	0.73	0.90	0.73
5	0.33	0.87	0.67	0.73	0.70	0.57	0.50
6	0.00	0.03	0.07	0.40	0.43	0.80	0.67
7	0.00	0.00	0.10	0.33	0.47	0.83	0.67
8	0.00	0.03	0.27	0.33	0.33	0.30	0.30
9	0.00	0.00	0.00	0.00	0.07	0.60	0.60
10	0.00	0.17	0.10	0.53	0.87	0.87	0.83
11	0.00	0.00	0.20	0.23	0.67	0.83	0.47

Data

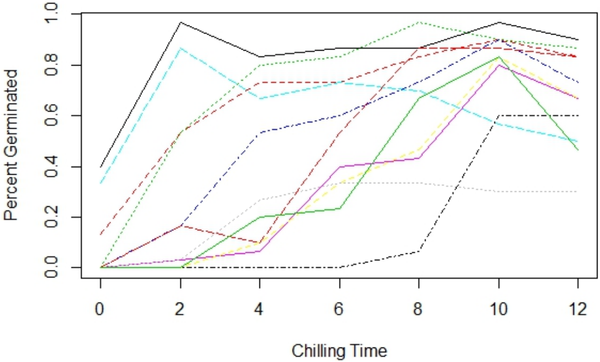

Figure 1: Clear nonlinear relationship, but most populations seem to be maximized at 10 weeks

Logistic Regression Model

$$p = Prob(Y = 1)$$

$$Y_i \overset{ind}{\sim} Bernoulli(p_i)$$

$$log(\tfrac{p_i}{1-p_i}) = x_i'\beta \implies p_i = \tfrac{exp(x_i'\beta)}{1+exp(x_i'\beta)}$$

- Logistic function gives output between 0 & 1, allowing for probabilities
- β_0: Intercept
- $\beta_1 - \beta_6$: Effect of the Chilling Time natural splines
- $\beta_7 - \beta_{16}$: Effect of each population
- $\beta_{17} - \beta_{76}$: Effect of interactions between each spline and each population
- β's are not interpretable
- Used to answer questions 2 and 3 but β's can't be used to calculate test statistics for population differences

Model Assumptions

Two standard MLR assumptions still need to hold:

- Monotone in x's
  - Clearly violated!
- Independence
  - Assume the observations are not related

Spline Fix

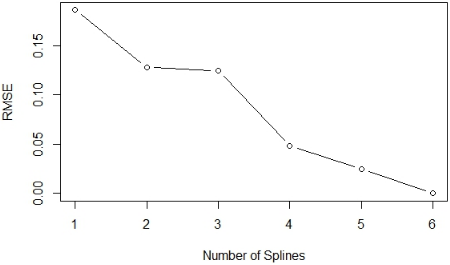

Figure 2: Six degrees of freedom minimizes RMSE, calculated without regard to population, by comparing germination ratios of all observations to model predictions

Model Fit

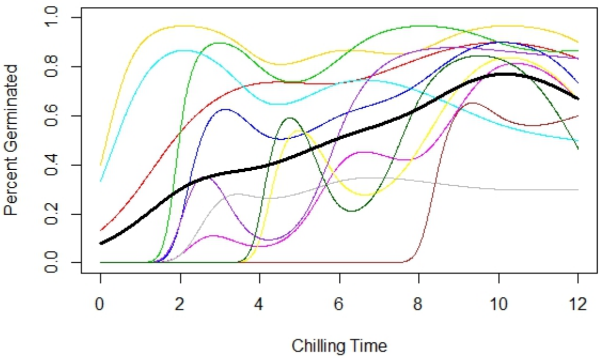

Figure 3: RMSE is essentially zero for each population and on average, as seen in the cross-validation plot

Bootstrapping

- Generate large B samples with replacement from original sample
- Retain desired statistics (maximum & difference) and build up a distribution
- Bootstrap quantile interval (.025, .975) approximate 95% confidence interval
- This technique will answer research questions 2 and 3

Ideal Chilling Time

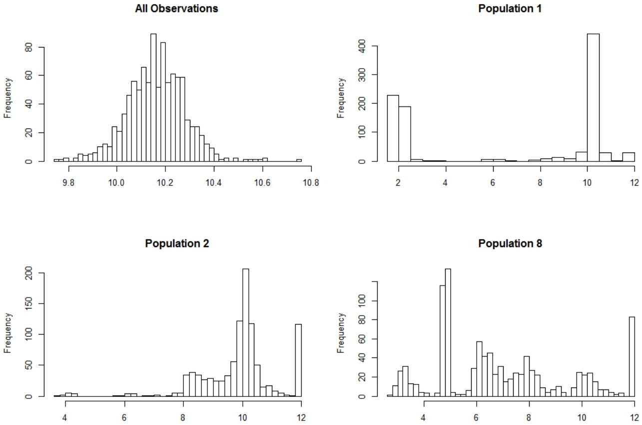

Figure 4: Note bimodal and polymodal populations

Ideal Chilling Time

	Lower	Mean	Upper
All Populations	9.92	10.17	10.39
Population 1	1.90	6.68	12.00
Population 2	7.17	9.88	12.00
Population 3	3.14	8.15	12.00
Population 4	8.86	10.04	10.92
Population 5	1.92	2.59	7.24
Population 6	6.78	10.38	12.00
Population 7	9.85	10.29	11.14
Population 8	3.13	6.91	12.00
Population 9	9.06	10.32	12.00
Population 10	7.20	9.46	12.00
Population 11	8.77	9.48	10.00

Table 1: Local modes create wide intervals and poor mean estimates in some populations

Seasonal Change Effect

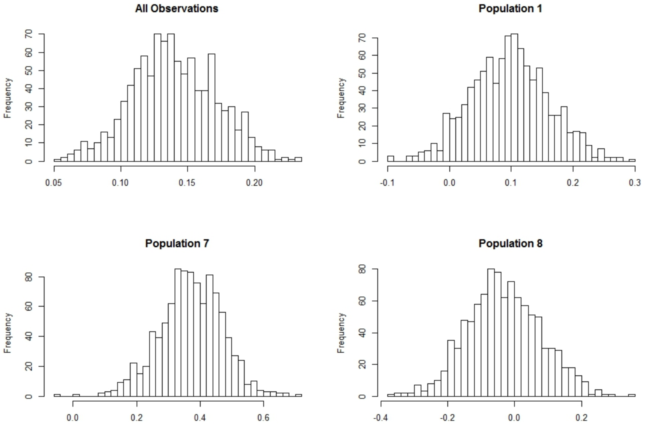

Figure 5: On average there is a decrease in germination, but individual populations report positive, or no effect

Seasonal Change Effect

	Lower	Mean	Upper
All Populations	0.08	0.14	0.20
Population 1	-0.02	0.10	0.22
Population 2	-0.10	0.06	0.21
Population 3	-0.18	-0.07	0.04
Population 4	-0.00	0.16	0.35
Population 5	-0.33	-0.13	0.10
Population 6	0.15	0.37	0.57
Population 7	0.17	0.37	0.56
Population 8	-0.24	-0.03	0.19
Population 9	0.36	0.53	0.69
Population 10	-0.17	-0.00	0.14
Population 11	-0.02	0.17	0.37

Table 2: Amounts reported are the decreases in germination as Chilling Time decreases from 10 to 8 weeks

Population Effect

In order to isolate interpretable β values for Population, fit a logistic regression model with only the interaction between Population & Chilling Time

	lower	mean	upper
β_1	0.34	0.45	0.56
β_2	0.18	0.26	0.34
β_3	0.24	0.33	0.42
β_4	0.07	0.14	0.20
β_5	0.04	0.10	0.16
β_6	-0.07	-0.01	0.05
β_7	-0.07	-0.01	0.05
β_8	-0.27	-0.20	-0.13
β_9	-0.25	-0.18	-0.11
β_{10}	0.06	0.12	0.19
β_{11}	-0.09	-0.02	0.04

Table 3: Interpret β and confidence intervals

Population Effect

- $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11}$ is tested by calculating an F statistic (58.83), p-value $< .0001$
- Difference in each β is tested by calculating a t-statistic and calculating p-values

	1	2	3	4	5	6	7	8	9	10	11
1		0.01	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2			0.13	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00
3				0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4					0.22	0.00	0.00	0.00	0.00	0.39	0.00
5						0.01	0.02	0.00	0.00	0.31	0.01
6							0.47	0.00	0.00	0.01	0.41
7								0.00	0.00	0.01	0.37
8									0.39	0.00	0.00
9										0.00	0.00
10											0.00
11											

Table 4: Non significant p-values at $\alpha = .05$ given in red

Conclusion

- There is significant variability from population to population, but overall there is a clear signal that about 10 weeks is the ideal time to chill tulips
- Decreasing Chilling Time from 10 to 8 weeks will decrease germination rate by 14% on average
- Do these 11 populations adequately represent all tulips?
- Speak with an expert: Is there a reason for the nonlinearity, or would a larger sample yield better results?