

Regional Effects of Covid-19 in California

Grant Hutchings

Abstract

Our previous analyses of Covid-19 deaths in [1] and [2] have focused on effects at the county level. We now turn our attention to group effects. California's census data is grouped by region, shown in Figure 1. The goal of this analysis is to determine whether these regions have a strong effect on Covid-19 cases, as well as how population density effects case counts. It is a reasonable assumption that counties with a high population density should have higher rates of infection due to the challenge of social distancing. This is why Montana has been able to quickly reopen businesses while New York continues to struggle. The effect of population density may serve as a useful proxy for the effect of social distancing if we assume that lower population density naturally leads to more social distancing. We find that the model most able to predict number of cases includes a population density effect as well as regional effects.

1. Data

We will be using the data from our previous analyses, adding a grouping variable for census regions, and defining a new covariate for each county; $population\ density = persons/mi^2$. Census regions are defined geographically and are shown in Figure 1. For our analysis we have chosen to reduce the groups by half. We group the counties into five regions using the following grouping: Group 1 contains counties in regions 1 and 2; Group 2, counties in region 3;

Group 3, counties in region 5; Group 4, counties in regions 4 and 6; Group 5, counties in regions 7,8,9, and 10. Group 1 corresponds to Northern California; group 2, the Bay Area; group 3, The Central Coast; group 4, The Central Valley; and group 5, Southern California. This not only reduces the complexity of our models, but also increases interpretability of regional effects. We see in Figure 2 and Figure 3 that the groups with the highest population and population density–

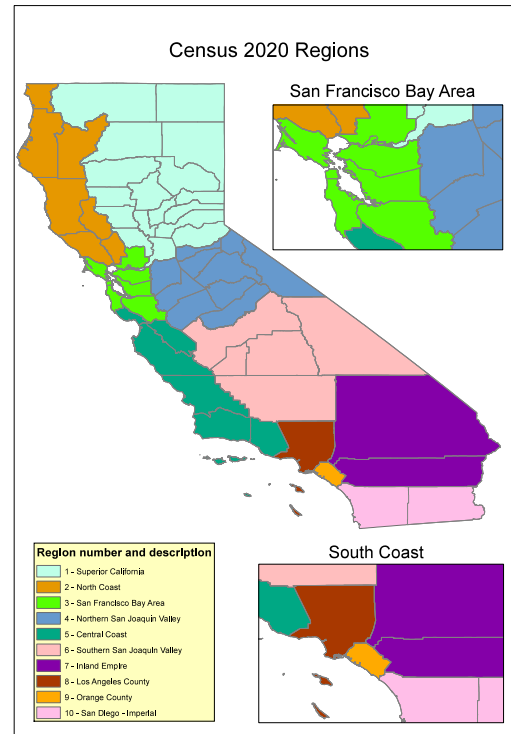


Figure 1: <https://census.ca.gov/regions/>

the Bay Area and Southern California—have the highest number of cases. We would therefore suspect that these variables have an important impact on case counts. We see that variabil-

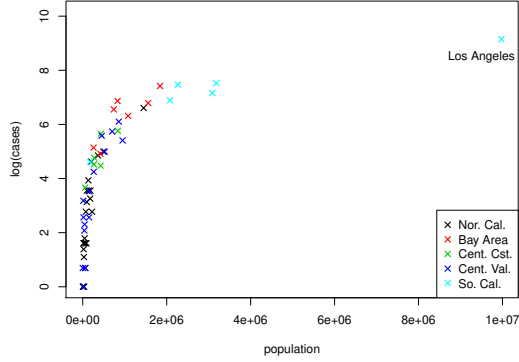


Figure 2: Log case count vs total population by county and group

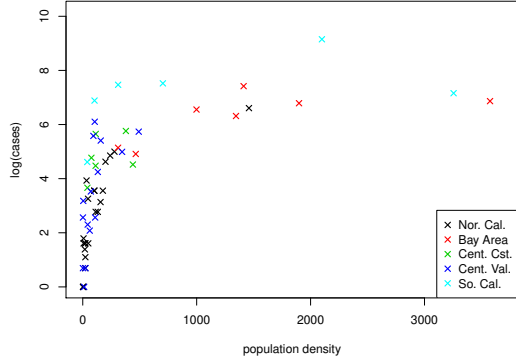


Figure 3: Log case count vs population density by county and group

ity in the number of cases is much higher in the regions with lower populations and population densities. Both figures show a similar trend with the main difference being that group clustering

is stronger by total population than by population density. This variability may manifest as small regional effects for Northern California and the Central Valley as there is not much clustering in case count. The Central Coast has fairly tight clustering by population, population density, and case count, so we suspect a larger regional effect. High population regions, the Bay Area and Southern California, have relatively low variability in case counts, but show higher variability in population density. Figures 2 and 3 show a strong trend in case counts with population and population density. We propose a variety of models to uncover these effects.

2. Proposed Models

We use standard bayesian linear regression techniques to fit all models, implement g-priors on the regression parameters for ease of posterior sampling, and model comparison using a marginal likelihood approach. G-priors are a common choice for bayesian linear regression as they lead to simple, closed form posterior distributions which can be samples in a Gibbs fashion. The generic linear regression

$$Y = \mathbf{1}\mu + X\beta + \epsilon, \epsilon \sim N_n(\epsilon|0, \sigma^2 I_n) \quad (1)$$

can handle g-priors provided that the columns of X are centered and the error covariance is indeed $\sigma^2 I_n$. The g-prior on β is defined as

$$p(\beta|X, \sigma^2) \sim N_p(\beta|g\sigma^2(X^T X)^{-1}) \quad (2)$$

where g is a hyperparameter controlling the amount of shrinkage in regression estimates from a posterior shrinkage factor $g/(1+g)$. We complete the hierarchy with a non-informative prior for $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$.

This leads to the full conditional distributions

$$\begin{aligned} p(\mu|Y, \sigma^2) &= N(\mu|\bar{x}, \frac{s^2}{n}) \\ p(\beta|Y, \sigma^2) &= N(\beta|\frac{g}{1+g}\hat{\beta}, \frac{g}{1+g}\sigma^2(X^T X)^{-1}) \\ p(\sigma^2|Y) &= IG(\sigma^2|\frac{n}{2}, \frac{s^2}{2} + \frac{1}{2(g+1)}\hat{\beta}^T X^T X \hat{\beta}) \end{aligned} \quad (3)$$

where $\hat{\beta}$ has been hit with the shrinkage factor. Shrinkage is a powerful tool to induce sparsity into our model. To induce more shrinkage, we choose a small value of g . Large values of g will quickly converge to shrinkage factors near one. We will discuss choice of g in section 4. For now we note that g is chosen to maximize the marginal likelihood, which is yet to be defined.

For all models we let the error covariance structure be zero mean, independent, and proportional to the population.

$$\epsilon_{ij} \sim N(0, \sigma^2 \sqrt{10^3/c_{ij}}) \quad (4)$$

If we let V be the diagonal matrix defined by equation 4, our likelihood becomes

$$Y \sim N(\mathbf{1}\mu + X\beta, \sigma^2 V) \quad (5)$$

As noted above, a requirement for g-priors is an IID error structure. To achieve this we must precondition our linear system by the inverse of a Cholesky factor of V . Let L be a Cholesky factor of the error covariance matrix V , then $L^{-1}Y = L^{-1}X\beta + L^{-1}\epsilon$ defines a new linear system with IID errors since $Var(L^{-1}\epsilon) = \sigma^2 L^{-1}V(L^T)^{-1} = \sigma^2 L^{-1}LL^T(L^T)^{-1} = \sigma^2 I_n$. For all models we provide bayesian p-values, testing the hypotheses that the parameters are significantly different from zero. P-values are computed from posterior samples as a simple proportion of samples in the null hypothesis region of the parameter space.

2.1 Intercept Only

We first consider a baseline model which includes only an intercept term.

$$y_{ij} = \mu + \epsilon_{ij} \quad (6)$$

We can clearly see from Figure 2 and Figure 3 that there exists trends and variability in the data that an intercept only model is far too simplistic to model accurately. We will see that this model is universally the poorest in all model comparison criteria, but nonetheless we implement it as a baseline. As we defined our linear regression with a location parameter μ , this model has no other covariates and thus no design matrix X . In order to use g-priors for this model, we must consider μ as part of the β vector and the associated design matrix as a singular column of 1's. Since the design matrix only needs to be centered with respect to additional covariates beyond a location parameter, there is no problem implementing g-priors.

If we test the hypothesis $H_0 : \mu > 0$ vs $H_a : \mu \leq 0$ we find a posterior probability $Pr(\mu > 0) = 1$, indicating that the data is significantly different from zero and a location parameter is significant.

2.2 Density Effect

Our first level of added complexity is the addition of the population density covariate.

$$y_{ij} = \mu + \beta d_{ij} + \epsilon_{ij} \quad (7)$$

This model should be an improvement over the intercept only model, but still lacks the ability to capture group effects which we believe to be important due to clustering in the data.

We find posterior probabilities $Pr(\mu > 0) = 1$, $Pr(\beta > 0) = 1$. Figure 4 shows posterior samples of β – the population density effect.

This confirms our theory that population density is positively correlated with cases as we have extremely strong evidence that $\beta > 0$.

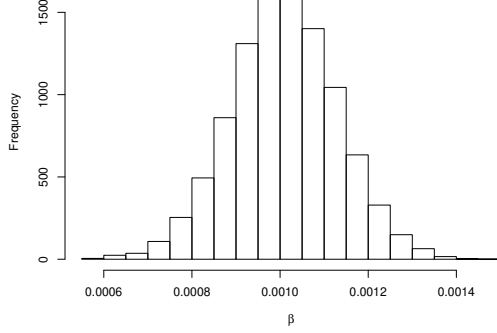


Figure 4: Samples of population density effect from model 2

To obtain posterior predictive samples, we use the same distribution as the likelihood, where the parameters are sampled from their posterior distributions.

$$\tilde{Y} \sim N(\mu^{(s)} + X\beta^{(s)}, \sigma^{2(s)}V) \quad (8)$$

for each posterior sample $s=1, \dots, S$.

Figure 5 shows posterior predictive distributions by county. True data are shown as colored X's with colors indicating groups as in Figures 2 and 3. Predictions are fairly good but tend to over-predict the true data values. We will see that the addition of group effects will somewhat alleviate this prediction problem.

2.3 Group Effect

We now consider a model without the population density effect but with a group effect η_j .

$$y_{ij} = \mu + \eta_j + \epsilon_{ij}, \quad \sum_{j=1}^5 \eta_j = 0 \quad (9)$$

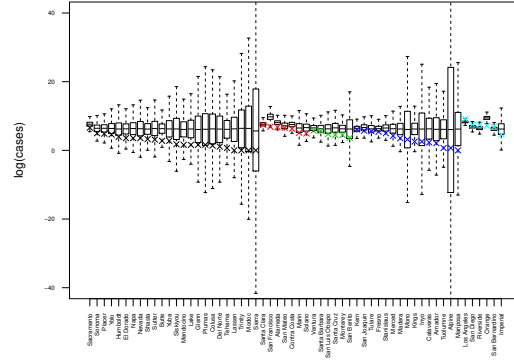


Figure 5: Model 2 posterior predictive samples

Parameter estimates and posterior probabilities summarize in Table 2.3.

Table 1: Model 3 summary

	$\hat{\beta}$	H_0	p-val
μ	6.12	$Pr > 0$	1
η_1	-1.04	$Pr < 0$	0.9959
η_2	0.71	$Pr > 0$	0.9876
η_3	-1.08	$Pr < 0$	0.9898
η_4	-0.74	$Pr < 0$	0.9745
η_5	2.15	$Pr > 0$	1

We find that all parameters are significantly different from zero indicating that group effects are important. It makes sense that groups 2 and 5—the Bay Area and Southern California—would have a strong positive effect. Since we have constrained our model so the effects sum to zero, groups 1, 3, and 4 have negative effects.

Figure 6 shows posterior predictive samples from model 3. The under-prediction problem of model 2 is better here but the distributions are very similar so it seems a marginal improve-

ment. Even if the improvement is small, it indicates that group effects may be more important than the population density effect.

As with the model 2, this model is an improvement over the intercept only model. The clustering described in section 1 indicates that group effects are significant, and this model confirms that. But this model lacks the ability to capture a population density effect, which model 2 shows to be significant. We therefore find this model to be informative but lacking. With the knowledge from models 2 and 3 that β and η_j are both significant we can combine them into a more complex model.

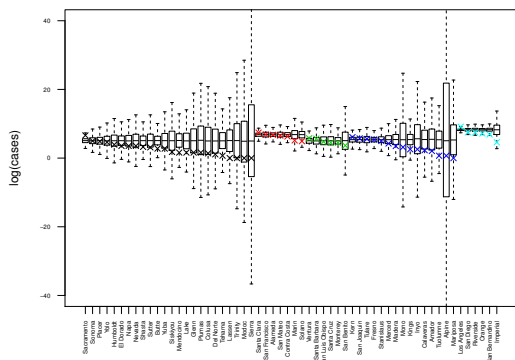


Figure 6: Model 3 posterior predictive samples

2.4 Density and Group Effects

Finally we consider a model that includes both group effects and a population density effect.

$$y_{ij} = \mu + \eta_j + \beta d_{ij} + \epsilon_{ij}, \quad \sum_{j=1}^5 \eta_j = 0 \quad (10)$$

Parameter estimates and posterior p-values are summarized in Table 2.4. We find that the β , while still significant, has been almost

completely overshadowed by the group effects η_j with an estimated value of only 6×10^{-4} . This gives more credibility to the hypothesis that group effects are more important than a population density effect. We see that the inclusion of β has shrunk all η_j towards zero, with a very noticeable impact on the group 2 effect. The estimate for η_2 and its p-value have shrunk significantly indicating that much of group 2's effect can be captured by including a the β parameter.

Table 2: Model 4 summary

	$\hat{\beta}$	H_0	p-val
μ	5.9417	$Pr > 0$	1
β	6e-04	$Pr > 0$	1
η_1	-0.8093	$Pr < 0$	0.9911
η_2	0.1892	$Pr > 0$	0.7435
η_3	-0.5465	$Pr < 0$	0.909
η_4	-0.3292	$Pr < 0$	0.8401
η_5	1.4959	$Pr > 0$	1

Figure 7 shows the posterior predictive distributions from model 4. The difference in predictive accuracy between model 4 and model 3 is difficult to discern from the figure, but we will later see a quantification how model 4 outperforms model 3 in predictive accuracy. Posterior p-values indicate that a mix of a population density effect and some – but maybe not all – group effects will produce the best model with these parameters. In the following section we will compare these models more carefully.

3. Model Comparison

3.1 Marginal Likelihood

Marginal likelihood is one way to compare models using g-priors. It is easy to compute and is

more interested in prediction, we may wish prefer model 3 to model 2.

Table 4: Posterior Predictive Loss Criteria

D_1	4629.20
D_2	2301.45
D_3	1836.59
D_4	1426.59

4. Choosing the hyperparameter g

There are many possible methods for choosing a value of g . We choose to take the Empirical Bayes approach. For each model we choose the value of g that maximizes the marginal distribution as defined in equation 11. This is a very simple but effective way of choosing g , as it is very reasonable to want to maximize the marginal likelihood of your model.

Figure 8 shows the marginal distribution for model 4 over a range of g values. The marginal distribution for model 4 is optimized at $g = 630$. This value of g is quite high in the sense that the shrinkage factor $\frac{g}{1+g}$ as seen in the posterior distribution for β is very near to 1. We see similarly high optimal g values for the other models. Model 1 uses $g = 588$, model 2 has $g = 1069$, and model 3 has $g = 536$; all of which produce very little shrinkage in the coefficients, so the g -prior does not have a large effect on inference and does not induce sparsity. For our use g -priors turn out to be more of a convenience tool for making posterior sampling straightforward.

5. Additional Model

We have shown group effects to be very important. We therefore consider one final model with

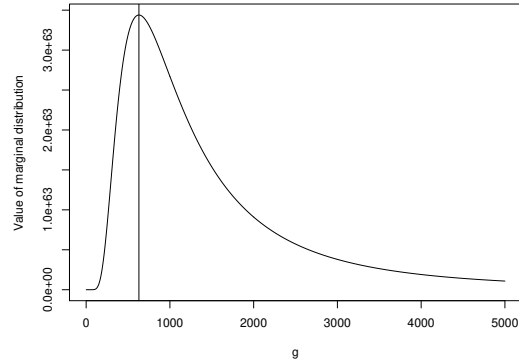


Figure 8: Marginal distribution maximized over g

group specific population density effects.

$$y_{ij} = \mu + \eta_j + \beta_j d_{ij} + \epsilon_{ij}, \quad \sum_{j=1}^5 \eta_j = 0 \quad (13)$$

This model performs poorly with respect to marginal likelihood with $\log(p(Y|M5)) = 138.77$. This is worse than models 2,3, and 4; likely due to the fact that marginal likelihood penalizes complexity quite harshly and we have added 4 more parameters than model 4. We say marginal likelihood penalizes complexity harshly because $(1+g)^{(n-1-p)/2}$ in the numerator of equation 11 is a quickly decreasing function of p . We find that this model actually performs best under the posterior predictive loss criteria with a value of 1277.32. PPLC does not penalize increasing complexity as much as the marginal likelihood. As model size grows, the complexity penalty $p/(1+p)$ goes to 1, so we are not surprised that this criteria has gone down. Table 5 indicates that regional population density effects are likely significant. If we are not yet fully convinced that group effects are important, model 5 gives further evidence in support

of that hypothesis. We could possibly improve model 5 by removing some parameters, but we will not consider further variable selection here as we have given sufficient evidence for the importance of group effects.

Table 5: Model 5 parameter estimates and posterior probabilities.

			p-val
μ	5.8148	$Pr > 0$	1
β_1	0.0018	$Pr > 0$	0.9995
β_2	5e-04	$Pr > 0$	0.9412
β_3	0.0036	$Pr > 0$	0.8906
β_4	0.0021	$Pr > 0$	0.8445
β_5	4e-04	$Pr > 0$	0.9931
η_1	-1.2782	$Pr < 0$	0.9946
η_2	0.5478	$Pr > 0$	0.8584
η_3	-0.8802	$Pr < 0$	0.8956
η_4	-0.3673	$Pr < 0$	0.7626
η_5	1.9779	$Pr > 0$	1

6. Conclusion

We hope to have convinced the reader that regional grouping effects are very significant. Including group effects increases in sample predictive power even under the penalized PPLC. Models 3, 4, and 5 have the lowest posterior predictive loss and all include group effects η_j . It is important to note that PPLC for model 3 is lower than model 2 even though model 3 has a higher complexity penalty. This indicates that group effects are probably more important than population density effects. This is rather surprising as a-priori we expected population density to be a reasonable proxy for social distancing, which should have a large effect on cases. Even

though β , the population density parameter may not be as influential as the group effects, we still find it to be significant. In both models 2 and 4 we find the posterior probability of that β is greater than zero is 1. In model 5, we split the population density effect into 5 regional effects and still find high posterior probabilities of significance. We conclude with the assertion that model 5 is the best of the models analyzed here as it allows for region specific effects for all parameters. Further analysis may be able to refine model 5, reducing complexity with minimal loss in predictive accuracy.

References

- [1] Hutchings, G. (2020), "Analysis of Covid-19 cases in California Counties"
- [2] Hutchings, G. (2020), "Prediction of Covid-19 Cases and Deaths in California"
- [3] ALAN E. GELFAND, SUJIT K. GHOSH, Model choice: A minimum posterior predictive loss approach, *Biometrika*, Volume 85, Issue 1, March 1998, Pages 1–11, <https://doi.org/10.1093/biomet/85.1.1>