STAT 520 Homework 4

Yifan Zhu

October 2, 2017

4.1. (a) Random variable in this problem is the nitrogren concentration of well i, denoted Y_i. And the covariate is indicator variable x_{ik} for $i = 1, \cdots, n$ and $k = 1, 2, 3$, where

$$x_{ik} = \begin{cases} 1 & \text{well } i \text{ is in county } k \\ 0 & \text{otherwise} \end{cases}$$

(b) For Y_i from the same county, it follows a Gamma distribution with mean μ_k, where $\mu_k = \alpha/\beta_k$ and Y_i's are independent. The parameters used in the model are $\theta_1, \theta_2, \theta_3$, which represent the true mean for each county, and

$$\mu_i = \sum_{k=1}^{3} x_{ik}\theta_k, \quad i = 1, 2, \ldots, n$$

where $\mu_i = E(Y_i)$.

(c) The estimation of θ_k and the standard errors are shown in Table 1

Table 1: Estimation of parameter and standard error

Parameter	Estimation	SE
θ_1	4.249	0.4937
θ_2	4.577	0.5326
θ_3	4.192	0.4877

(d) We construct the simutanuous confidence interval using the Bonferroni method by constructing the 98.3% confidence interval. Then we have Table 2

Table 2: 95% simutanuous confidence interval

Parameter	Simutaneous Interval
θ_1	(3.192, 5.294)
θ_2	(3.444, 5.711)
θ_3	(3.152, 5.229)

Figure 1 shows the overlapping of three confidence intervals.

From Figure 1 we find that there is a lot overlapping between these three confidence intervals, thus we can conclude that there is no sufficient evidence that the true mean of nitrogreon concentrations are different.

(e) From the previous analysis, we conclude that $\theta_1 = \theta_2 = \theta_3 = \theta$. So we fit a glm with

$$\mu_i = \theta . i = 1, \ldots, n$$

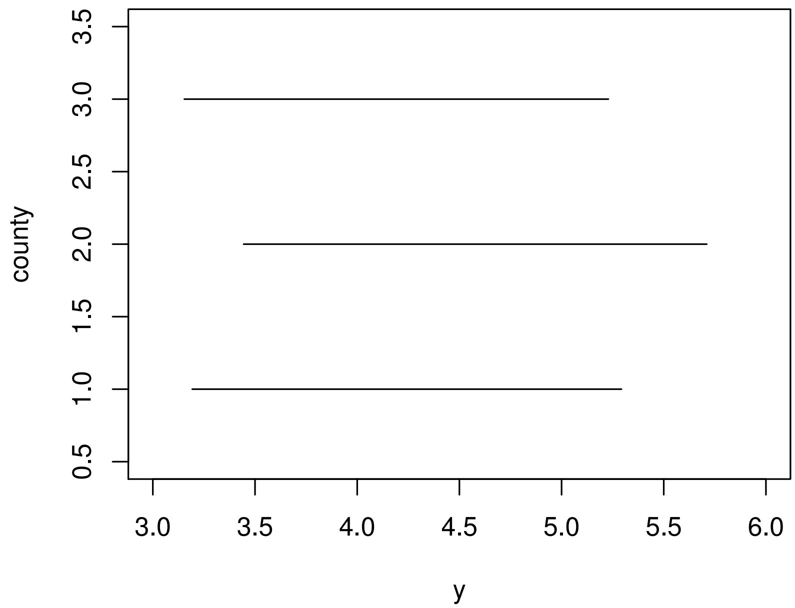

Figure 1: 95 % Simutaneous Intervals

We get $\hat{\theta} = 4.337, \hat{\phi} = 3.031$. Thus the Gamma distribution would have parameter $\alpha = 3.031, \beta = 0.6988$. The density plot is shown in Figure 2

And with those estimated parameters, we have $P(Y \leq 3) = 0.3421$ and $P(Y \geq 10) = 0.0310$.

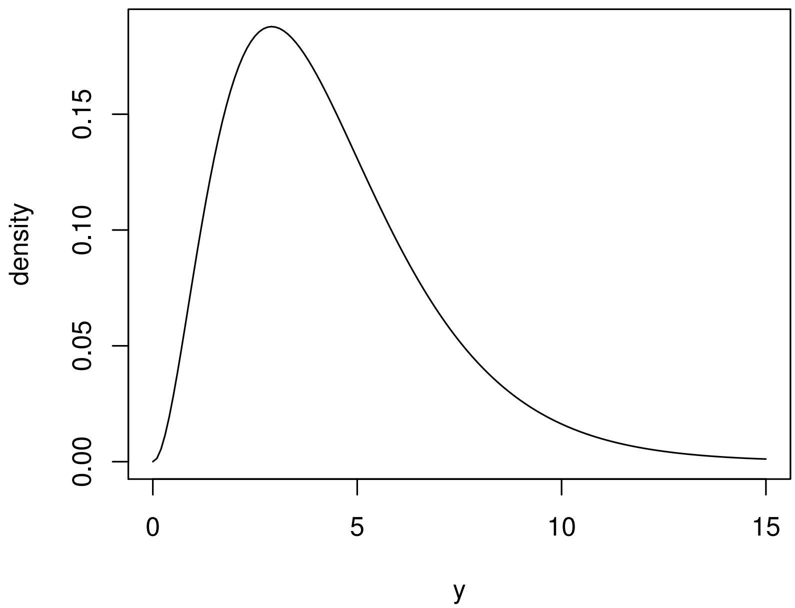

Figure 2: Density of the estimated common gamma distribution

4.2. (a) The random variables are Y_{ik}, where $i = 1, \ldots, n_k$ and $k = 1, 2, 3$. n_k is the number if wells sampled in the k-th county, and Y_{ik} is the nitrate concentration of the i-th well in county k.

(b) In order to conduct a likelihood analysis, we fit two models. One reduced model with 3 different gamma distributions with different parameters (α_k, β_k) and a reduced model with only one gamma distribution with parameter (α, β).

Let $f(y|\alpha, \beta)$ be the pdf of gamma distribution with parameters α and β, then the log likelihood for the full model is

$$\ell_F = \sum_{k=1}^{3} \sum_{i=1}^{n_k} \log(f(y_{ik}|\alpha_k, \beta_k))$$

And for the reduced model, log likelihood is

$$\ell_R = \sum_{k=1}^{3} \sum_{i=1}^{n_k} \log(f(y_{ik})|\alpha, \beta)$$

(c) The MLE's for the parameters in the full and reduced model are shown in Table 3.

Table 3: Estimation and Standard Error for parameters

Parameter	Estimation	SE
α_1	1.778	0.4635
β_1	0.419	0.1260
α_2	3.578	1.0193
β_2	0.821	0.2382
α_3	6.603	1.822
β_3	1.575	0.4517
α	3.010	0.4669
β	0.694	0.1171

(d) When we plug in MLE into these two models, we can get ℓ_F and ℓ_R mentioned in (b). Then we have the value of Λ which follows asymptotically a Chi-square distribution with 2 degrees of freedom. And

$$\Lambda = -2(\ell_R - \ell_F) = 12.212$$

That gave us a p-value $p = 0.01584$. Thus we reject the reduced model and conclude that the nitrate concentrations in wells for different counties follows different distributions.

In order to compare with the analysis we did in 4.1, we calculated the estimation for each $\hat{\mu}_k = \frac{\hat{\alpha}_k}{\hat{\beta}_k}$ and the their standard error using Delta method. And then we produced the 95% simultaneous confidence intervals. The results are shown in Table

Table 4: Estimation, standard error and 95% simultaneous interval for μ_k

Parameter	Estimation	SE	95% simultaneous interval
μ_1	4.243	0.6364	(2.8892, 5.5968)
μ_2	4.577	0.4722	(3.5725, 5.5815)
μ_3	4.192	0.3262	(3.4981, 4.8859)

Figure 3 shows the overlapping of these 3 simultaneous intervals, from which we can see there is still a lot a overlapping although our conclusion is these three distributions are different.

(e) From the previous analysis, our conclusion is that there is sufficient evidence that the nitrate concentration in wells from different counties are different. The probabilities are given in Table 5. Figure 4 shows the pdf's of three gamma distributions.

Table 5: Probabilities

County	$P(Y < 3)$	$P(Y > 10)$
1	0.4292	0.0587
2	0.2786	0.0286
3	0.2480	0.0032

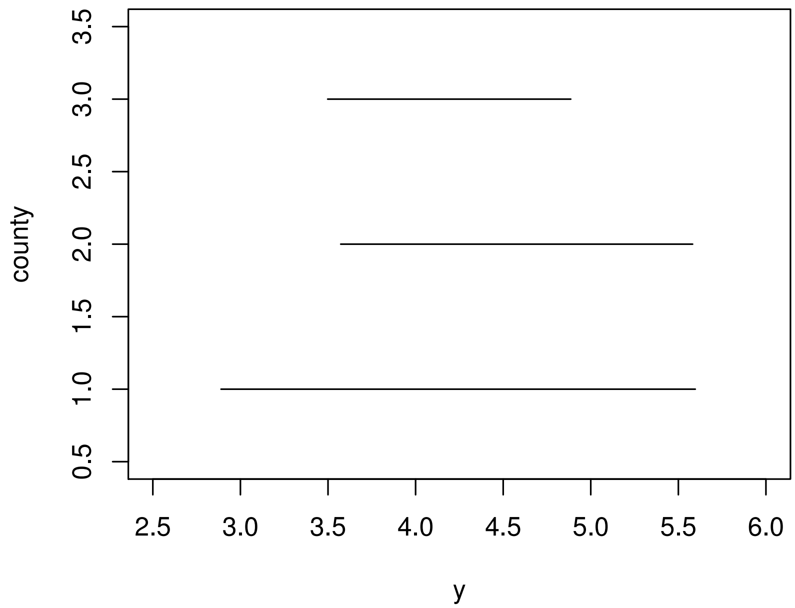

Figure 3: 95 % Simutaneous Intervals

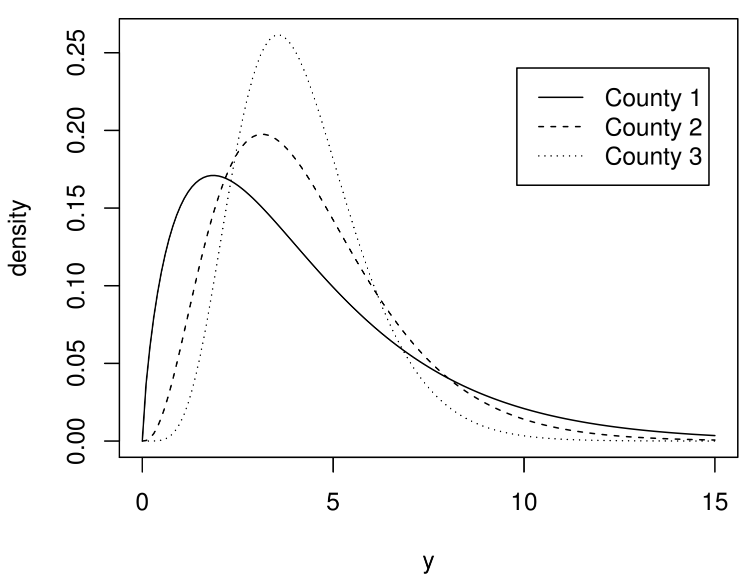

Figure 4: Density of estimated gamma distributions for three counties

4.3. In these two analyses, we get two different conclusions for different test. For the first one using simultaneous intervals, we conclude that there is no sufficient evidence that the means of nitrate concentration in wells of 3 counties are different. For the second analysis, we conclude that there is sufficient evidence that the distributions of nitrate concentration for 3 counties are different. The different conclusions are resulted from different hypothesis and different tests conducted. In the second analysis, when we construct the simultaneous intervals like the first analysis, we get a similar conclusion about the mean of nitrate concentration.