Radon Concentration using Zero Inflated Negative Binomial (ZINB) Modeling

G.A. Dawodu[1*]; A.O. Mustapha[2]; and O.M. Olayiwola[1]

[1]Department of Statistics, College of Physical Sciences (COLPHYS), Federal University of Agriculture, Abeokuta, Nigeria.
[2]Department of Physics, College of Physical Sciences (COLPHYS), Federal University of Agriculture, Abeokuta, Nigeria.

E-mail: agad1963@yahoo.com
dawoduga@funaab.edu.ng

ABSTRACT

This study is geared towards being able to quantify and compare the counts of radon concentrations emitted for some days and nights in 2016 in an averagely populated office. With the use of Radon Scout, the captured count is often zero inflated and thus requires the use of an appropriate zero inflated modeling technique, which in the present case is the proposed ZINB.

The results from ZINB modeling shows that radon is more emitted; in the night more than in the day, at low temperatures, high pressures and is directly proportional to relative humidity. This result also avails the researcher the opportunity to determine a constant with which one can predict for daylight radon emission counts when given the corresponding nocturnal radon emission counts.

(Keywords: temperature, relative humidity, pressure, time, radon concentration, R, ZINB, logit, bootstrapping)

INTRODUCTION

In clean and properly ventilated offices, one may tend to think that pollution (i.e., activities of pollutants) is either totally absent or kept at an extreme minimum, more so, if the offices are spacious and contain very few items. Radon (Chege et al., 2009; Ramola et al., 2000; Seftelis et al., 2007) is ever-present, it is a radioactive element that is emitted at any time of the day in varying quantities.

Much has been said about potential risks and health hazards associated with elevated levels of residential radon (Al Zabadil et al, 2012; Chege et al, 2009; Darby et al, 2001; Fitzpatrick-Lewis et al, 2010; Singh, 2010; Smith and Oleson, 2008). The need to assess the least and maximum exposures, with respect to location, a hypothetical member of a community can experience, cannot be over-emphasized because the inhabitants of such a community will always desire to know how "friendly" their environment is with respect to "freely available", carcinogenic radioactive radon, more so, if they share their neighborhoods with rocks, and quarries, Electricity and nuclear power stations, judging from the experience at Chernobyl and Fukushima.

Even when there are no accidents, inhabitants of "rocky" environments such as those obtainable in Abeokuta, Ogun state, Nigeria have enough to worry about concerning carcinogenic radon. The fact that factors involved with Radon (i.e. Temperature, Pressure, Relative-Humidity etc.) are all quantifiable makes it seem natural for us to expect the radon emission to be a function of the levels of these variables also. Some of which are covariates while others are auxiliary variables. Through this approach, ZINB, radon "counts" will be appropriately modeled and its "dependent" variables will be equally categorized.

MATERIALS AND METHODS

Here, our equipment (i.e., the radon emission detector called Radon Scout™, made by SARAD GmBh, Germany) is placed on the table in a Dean's office in the Federal University of agriculture in Abeokuta for a "short" period, to collect the data (i.e., captured readings on; time, temperature, relative humidity, pressure, Region of Interest (RoI), the indication of relative movements of the device while in operation (Tilt), and Radon per hour) of size 250.

Observations were taken on hourly basis throughout the days and nights covering about eleven days and involving ten nights. Radon Scout measures radon concentration/counts in Bq m^{-3}, as well as Temperature in ^{o}C, and Relative humidity and Pressure in mbar.

It is insensitive to extremely low radon counts and often returns zero counts for all situations in this category. This is why any observations that were obtained through the Radon Scout will contain many zero readings, thus making the data zero-inflated. Zero-inflated data are better modeled either through zero-inflated Poisson modeling or through zero-inflated negative binomial modeling (Lambert, 1992; Soto et al., 2016; Yau et al., 2004) and R codes (Crawley, 2007), thus leading to the topic of this work.

The choice of ZINB over zero-inflated Poisson modeling is borne out of the fact that the zero readings are not "genuine" zeros, it is just that the Radon Scout is not "powerful" enough to "capture" those very low readings and thus returns zeros for them, hence it is appropriate to assume that those zeros are generated in a separate level or process.

The modeling assumptions are that; when the radon emission is too low, zero is recorded. However, if the emission count is just above the low limit, the Radon Scout can read the emission count which is assumed to be negative binomially distributed. Consequently, the data distribution combines the negative binomial distribution and the logit distribution. Now if we suppose that the probability of having too low radon emission is π while the probability of having readable emission count is 1 – π. Therefore, the probability distribution of the ZINB random variable can be written as:

$$P(y_i = k) = \begin{cases} \pi_i + (1-\pi_i) f(y_i), & k = 0 \\ (1-\pi_i) f(y_i), & k > 0 \end{cases}$$

(1)

Where; k is a specific radon count and π_i is the logistic link (or logit) function defined below (i.e., Equation 4) and $f(y_i)$ is the negative binomial distribution given by Equation 2:

$$f(y_i) = P(Y = y_i | \mu_i, \alpha) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}$$

(2)

The negative binomial component can include an exposure time t and a set of j regressor variables (the x's). The expression relating these quantities is Equation 3:

$$\mu_i = \exp\left(\ln(t_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_j x_{ji}\right)$$

(3)

Often, $x_{1i} \equiv 1, \quad \forall i$, in which case β_1 is called the intercept. The regression coefficients $\beta_1, \beta_2, ..., \beta_j$ are unknown parameters that are estimated from a set of data. Their estimates are symbolized as $b_1, b_2, ..., b_j$. The logistic link function π_i is given by Equation 4:

$$\pi_i = \frac{\lambda_i}{1+\lambda_i}$$

(4)

Where;

$$\lambda_i = \exp\left(\ln(t_i) + \gamma_1 z_{1i} + \gamma_2 z_{2i} + ... + \gamma_m z_{mi}\right)$$

(5)

The logistic component includes an exposure time t and a set of m regressor variables (the z's). Note that the z's and the x's may or may not include terms in common. Hence the likelihood function (L), defined over $f(y_i)$ is (Equation 6):

$$L\left(\mu \middle| f\left(y_i\right),\alpha\right)=\prod_{i=1}^{n}\exp\left(y_i\ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)-\frac{1}{\alpha}\ln\left(1+\alpha\mu_i\right)+\ln\Gamma\left(y_i+\frac{1}{\alpha}\right)-\ln\Gamma\left(y_i+1\right)-\ln\Gamma\left(\frac{1}{\alpha}\right)\right)$$

(6)

Here, we find the likelihood of the expected value, μ given the data and α which allows for dispersion. Typically, this would be expressed as a log(likelihood), denoted by script L* (i.e. Equation 7);

$$L^*\left(\mu \middle| f\left(y_i\right),\alpha\right)=\sum_{i=1}^{n}y_i\ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)-\frac{1}{\alpha}\ln\left(1+\alpha\mu_i\right)+\ln\Gamma\left(y_i+\frac{1}{\alpha}\right)-\ln\Gamma\left(y_i+1\right)-\ln\Gamma\left(\frac{1}{\alpha}\right)$$

(7)

Which can be expressed in terms of our model by replacing μ_i with $\exp\left(x_i^*\beta\right)$.

With regards to the zero-inflated negative binomial model, the expression of the likelihood function depends on whether the observed value is a zero or greater than zero. From the logistic model of $y_i > 0$ versus $y_i = 0$:

$$p=\frac{1}{1+e^{-x_i^*\beta}} \quad \textbf{(8)}$$

and

$$1-p=\frac{1}{1+e^{x_i^*\beta}} \quad \textbf{(9)}$$

Then:

$$L^*\left(\mu \middle| f\left(y_i\right),\alpha\right)=\begin{cases}\displaystyle\sum_{i=1}^{n}\left\{\ln\left(p_i\right)+\left(1-p_i\right)\left(\frac{1}{1+\alpha\mu_i}\right)^{\frac{1}{\alpha}}\right\}, & y_i=0\\ \displaystyle\sum_{i=1}^{n}\left\{\begin{array}{c}\ln\left(p_i\right)+\ln\Gamma\left(\frac{1}{\alpha}+y_i\right)-\ln\Gamma\left(y_i+1\right)-\ln\Gamma\left(\frac{1}{\alpha}\right)+\\ \left(\frac{1}{\alpha}\right)\ln\left(\frac{1}{1+\alpha\mu_i}\right)+y_i\ln\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)\end{array}\right\}, & y_i>0\end{cases}$$

(10)

However, note that R will easily allow the estimation of $\theta = \frac{1}{\alpha}$ and not α itself.

RESULTS, DISCUSSION, AND CONCLUSION

Primarily, the main interest is being able to predict the non-zero radon emission counts during the day and night. The states of the relative humidity, temperature and pressure are also very important.

The time, region of interest (RoI) and the indication of relative movements of the device while in operation (Tilt) are also present and will be used as mere covariate and auxiliary variables respectively. A summary of the data can be obtained through the run of the following R codes:

```
require(ggplot2)
require(pscl)
require(MASS)
require(boot)
radneb<-
read.csv("C:\\Users\\FUNAAB\\Desktop\\ZeroIM
\\Radon2016.csv")
radneb <- within(radneb, {
TILT <- factor(TILT)
ROI <- factor(ROI)
TIME <- factor(TIME)
 })
summary(radneb)
```

The output is Table 1.

Table 1: The Summary of the Radon Emissions Count Data (size is n=250)

TILT	ROI	TIME	TEMP	RHUMID	PRESS	RADON
0:234	0: 81	0:129	Min. :25.50	Min. :64.00	Min. : 992.0	Min. : 0.00
1: 16	1:169	1:121	1st Qu.:27.00	1st Qu.:76.00	1st Qu.: 995.0	1st Qu.: 0.00
			Median :27.50	Median :78.00	Median : 996.0	Median : 9.00
			Mean :27.45	Mean :77.03	Mean : 996.2	Mean :11.26
			3rd Qu.:28.00	3rd Qu.:80.00	3rd Qu.: 997.0	3rd Qu.:18.00
			Max. :29.50	Max. :81.00	Max. :1000.0	Max. :55.00

The histogram with the x-axis in $\log_{10}$ scale is attempted with the following R command;

```
ggplot(radneb, aes(RADON, fill = TIME)) +
  geom_histogram() +
  scale_x_log10() +
  facet_grid(TIME ~ ., margins=TRUE, scales="free_y")
```

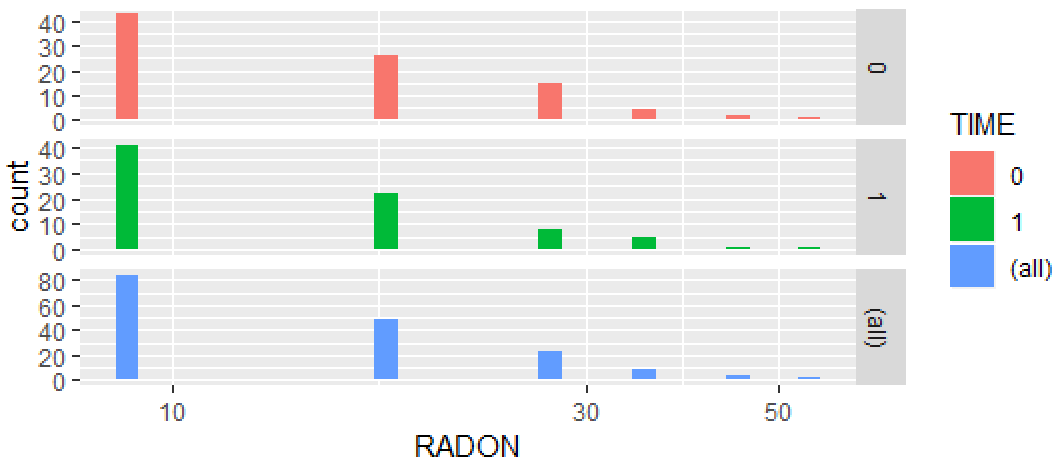

Figure 1: Showing the Histogram Plot of the Data ("0" denotes "night" while "1" denotes "day", the combination of day and night is called "all").

To continue to build the model, the variables TEMP and TIME (for temperature and time respectively) will be used to model the count in the part of negative binomial model and the variable PRESS (for pressure) in the logit part of the model. The R package "pscl" is used to run a zero-inflated negative binomial regression. To estimating the model with the variables of interest, the following codes come handy:

m1 <- zeroinfl(RADON ~ TEMP + TIME | PRESS, data = radneb, dist = "negbin", EM = TRUE)

summary(m1)

The output is Figure 2.

```
Call:
zeroinfl(formula = RADON ~ TEMP + TIME | PRESS, data = radneb, dist =
"negbin", EM = TRUE)

Pearson residuals:
    Min      1Q  Median      3Q     Max
-1.1033 -1.0316 -0.2049  0.6019  4.3329

Count model coefficients (negbin with log link):
            Estimate  Std.Error  z-value Pr(>|z|)
(Intercept)  4.73780    1.40009    3.384  0.000715 ***
TEMP        -0.06916    0.05083   -1.361  0.173644
TIME1       -0.07038    0.08386   -0.839  0.401368
Log(theta)   1.48430    0.13481   11.011  < 2e-16 ***

Zero-inflation model coefficients (binomial with logit link):
             Estimate  Std.Error  z-value  Pr(>|z|)
(Intercept) -39.25502        NA       NA        NA
PRESS         0.03866        NA       NA        NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Theta = 4.4119
Number of iterations in BFGS optimization: 1
Log-likelihood: -749.5 on 6 Df
```

Figure 2: Showing the Estimation of the Model with the Variables of Interest.

Figure 2 looks very much like the output from two Ordinary Least Squares (OLS) regressions in R. Below the model call, you will find a block of output containing negative binomial regression coefficients for each of the variables along with their respective standard errors, z-scores, and p-values for the coefficients. The second block corresponds to the inflation model. It includes logit coefficients for predicting excess zeros along with their standard errors, z-scores, and p-values.

The intercept and log(Theta) are statistically significant. A comparison of the current model with a null model without predictors using chi-squared test on the difference of log likelihoods may contain some useful information. Hence the code for the comparison starts with:

m0 <- update(m1, . ~ 1)

pchisq(2 * (logLik(m1) - logLik(m0)), df = 3, lower.tail=FALSE)

The output obtain from it is the information:

'log Lik.' 0.4842524 (df=6)

From the output above, we can see that there is a statistically significant difference between the models (i.e. m0 and m1).

Note that the model difference information above does indicate that the zero-inflated model is an improvement over a standard negative binomial regression. However, by running the corresponding standard negative binomial model and then performing a Vuong test of the two models more information may come to the fore. We use the R package (MASS) to run the standard negative binomial regression. The code:

summary(m2 <- glm.nb(RADON ~ TEMP + TIME, data = radneb))

gives the output in Figure 3.

```
Call:
glm.nb(formula = RADON ~ TEMP + TIME, data = radneb, init.theta =
0.4872650851, link = log)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8158  -1.7329  -0.1463   0.3213   1.6081

Coefficients:
            Estimate  Std.Error  z-value Pr(>|z|)
(Intercept)   6.3430     3.0961    2.049   0.0405 *
TEMP         -0.1401     0.1121   -1.250   0.2113
TIME1        -0.1741     0.1884   -0.924   0.3555
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(0.4873) family taken to be 1)

    Null deviance: 288.20  on 249  degrees of freedom
Residual deviance: 286.08  on 247  degrees of freedom
AIC: 1687.4

Number of Fisher Scoring iterations: 1

          Theta:  0.4873
          Std.Err.:  0.0522

 2 x log-likelihood:  -1679.4100
```

Figure 3: Showing the Output that Contains the Deviances and Akaike Information Criterion (AIC) to Enable the Comparison with other Models.

Now, by invoking the Vuong test, using the code "vuong(m1, m2)", the output is contained in Figure 4.

```
Vuong Non-Nested Hypothesis Test-Statistic:
(test-statistic is asymptotically distributed N(0,1) under the
 null that the models are indistinguishable)
-------------------------------------------------------------
              Vuong z-statistic       H_A         p-value
Raw                    14.99052 model1 > model2 < 2.22e-16
AIC-corrected          14.65807 model1 > model2 < 2.22e-16
BIC-corrected          14.07270 model1 > model2 < 2.22e-16
```

Figure 4: Showing that the Results Obtained through "raw" is Not Significantly Different when "correct" for AIC and Bayesian Information Criterion (BIC).

Further, confidence intervals for the parameters and the "exponentiated" parameters using bootstrapping can be easily obtained. For the negative binomial model, these would be incident risk ratios, for the zero inflation models, the odds ratios.

The R's "boot" package can also give some interesting results. First, we get the coefficients from our original model to use as start values for the model to speed up the time it takes to estimate. Then we write a short function that takes data and indices as input and returns the parameters we are interested in.

Finally, we pass that to the boot function and do 1200 replicates, using "snow" to distribute across four cores. Note that you should adjust the number of cores to whatever your machine admits. Also, for final results, one may wish to increase the number of replications to help ensure stable results. The code:

"dput(round(coef(m1, "count"), 4))"

gives the output;

structure(c(4.7378, -0.0692, -0.0704), .Names = c("(Intercept)", "TEMP", "TIME1"))

Whilst the code:

"dput(round(coef(m1, "zero"), 4))"

gives the output:

structure(c(-39.255, 0.0387), .Names = c("(Intercept)", "PRESS"))

It is time to invoke the bootstrap function, the codes:

```
f <- function(data, i) {
 require(pscl)
 m <- zeroinfl(RADON ~ TEMP + TIME | PRESS,
        data = data[i, ], dist = "negbin",
        start = list(count = c(4.7378, -0.0692, -0.0704), zero = c(-39.255, 0.0387)))
 as.vector(t(do.call(rbind, coef(summary(m)))[, 1:2]))
 }
set.seed(10)
(res <- boot(radneb, f, R = 1200, parallel = "snow", ncpus = 4))
```

Gives the output in Figure 5.

```
ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = radneb, statistic = f, R = 1200, parallel = "snow",
    ncpus = 4)


Bootstrap Statistics :
         original        bias     std.error
t1*    4.73773459 -0.0671728912  1.325070042
t2*    1.40008851  0.0054044975  0.111779783
t3*   -0.06916052  0.0023225829  0.047750767
t4*    0.05083320  0.0002006901  0.004088204
t5*   -0.07037445  0.0009346581  0.096049903
t6*    0.08386250 -0.0001360772  0.004160592
t7*    1.48429766  0.0278558151  0.116993119
t8*    0.13480569  0.0013036618  0.005491511
t9*  -39.25610834 -5.4165054563 75.951543034
t11*   0.03866154  0.0054320788  0.076234095
WARNING: All values of t10* are NA
WARNING: All values of t12* are NA
```

Figure 5: Showing that the Values for "seeds" 10 and 12 are Generally "Not Available" (NA).

The results are alternating parameter estimates and standard errors. That is, the first row has the first parameter estimate from our model and so forth.

The second column starts with the standard error for the first parameter. The third column contains the bootstrapped standard errors, which are considerably larger than those estimated by "zeroinfl". Now we can get the confidence intervals (CI) for all the parameters.

We start on the original scale with percentile and bias adjusted CIs. We also compare these results with the regular confidence intervals based on the standard errors. The basic parameter estimates with percentile and bias adjusted CIs are determined using the codes:

"parms <- t(sapply(c(1, 3, 5, 7), function(i) {

out <- boot.ci(res, index = c(i, i + 1), type = c("perc", "bca"), h = exp)

with(out, c(Est = t0, pLL = percent[4], pUL = percent[5],

bcaLL = bca[4], bcaLL = bca[5])) }))"

The output (Figure 6), after adding the row names, using the code "row.names(parms)<-names(coef(m1))", is

```
                        Est       pLL          pUL     bcaLL      bcaLL
count_(Intercept) 3.9394864 1.6706640    7.7989834 2.1304808 10.2452309
count_TEMP        0.2197509 0.1165403    0.3393411 0.1371516  0.3742786
count_TIME1       2.4086158 1.1184525    6.4420052 0.7971518  5.2996704
zero_(Intercept)  0.3733061 0.2686740    0.6398457 0.2352317  0.5269119
zero_PRESS        4.9685866 1.4218768 3227.9195801 0.8304531 39.9129615
```

Figure 6: Showing the Confidence Intervals (CIs) for all the Parameters.

And by comparing the contents of Figure 6 with that of the Normal based approximation by using the code “confint(m1)” , the output is contained in Figure 7:

```
                        2.5 %      97.5 %
count_(Intercept)   1.9936804  7.48192551
count_TEMP         -0.1687942  0.03046824
count_TIME1        -0.2347432  0.09399176
zero_(Intercept)          NaN         NaN
zero_PRESS                NaN         NaN
```

Figure 7: Showing the Normal Based Approximations, but the CIs for Zero_(Intercept) and Zero_PRESS are Not Available.

The bootstrapped confidence intervals are considerably wider than the normal based approximation. Now, one can estimate the incident risk ratio (IRR) for the negative binomial model and odds ratio (OR) for the logistic (zero inflation) model. This is done using almost identical code as before, but passing a transformation function to the h argument of “boot.ci”, in this case, ”exp” to exponentiate.

To better understand our model, we can compute the expected radon emission count for different combinations of our predictors. In fact, since we are working with essentially categorical predictors, we can compute the expected values for all combinations using the “expand.grid” function to create all combinations and then the “predict” function to do it. Finally, we create a graph (Figure 8).

CONCLUSION

This attempt does not exhibit all the latent characteristics involved with the emission of radon, it however, informs the researcher that the said emission is more potent during the course of the night than the day. The emission is higher at lower temperatures either during the night or the day, keeping pressure high (indicated by the factor(PRESS) at 4 in Figure 8) and irrespective of the state of variables like, TILT and ROI, but because of the strong relationship existing between temperature and relative humidity, temperature could be used to determine the relationship between radon emission count and relative humidity.

In the present work, the higher the relative humidity, the higher the radon emission count, irrespective of whether it is taken during the night or the day. If one excludes the TILT and ROI in the meantime, an extract of the data is Table 2.

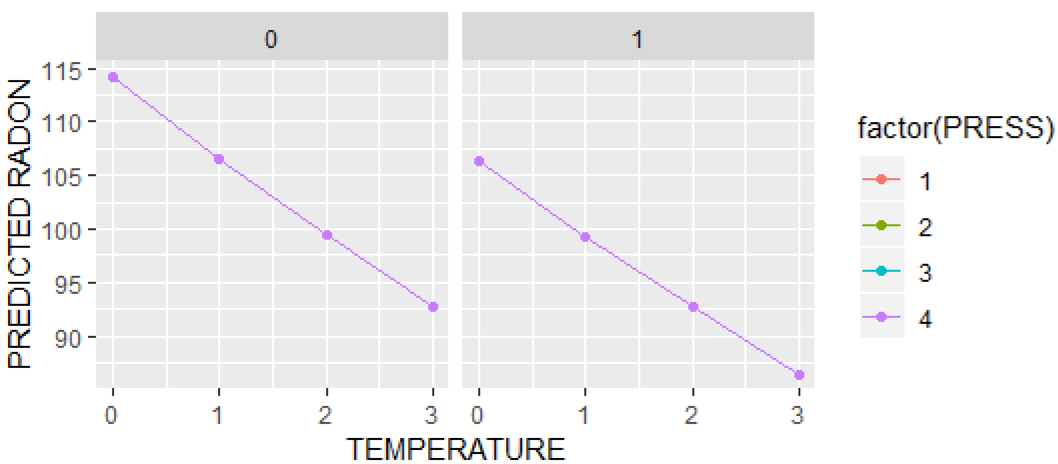

Figure 8: Showing the Graph of Prediction with “0” for “night” and “1” for “day” at highest values for Pressure as the Temperature Increases from the lowest “0” to the highest “3”.

Table 2: An Extract of the Data with TILT and ROI Removed.

	TEMP	HUMID	PRESS	TIME	RADON
1	26	85	998	1	18
2	26	85	997	1	9
3	26	85	997	1	55
4	26	85	996	1	18
5	26	85	997	1	27
6	26	85	997	1	36

From Table 2, the column for pressure is seen to contain values in the level 4 (i.e., an intrinsic class 750<PRESS<1000, the first of such classes is 0<PRESS<250 which happens to be level 1).

Figure 1 is actually a combination of three histograms, that is, the histogram of "RADON scales" versus "RADON count" for night readings (coded as 0), an equivalent of it for daylight (coded as 1) and another equivalent of it for the entire data (which is referred to as "all").

Figure 2 contains the results when; Negative-Binomial and Binomial distributions (with their respective links) are used to model the data. Both gave good results with Negative-Binomial distribution performing better than Binomial distribution.

The contents of Figure 3 further authenticate the fact that Negative-Binomial (i.e. "nb") is good enough for the data. For instance, the generalized linear model (i.e. "glm") version of nb was fitted for daylight (coded 1) and the result shows that it is adequate for the data and that the probability that this is happening by "chance" is 0.0405 < 0.05 (i.e. significantly low).

The Vuong Non-nested hypothesis test statistics, in Figure 4, is a certificate that the models (i.e. "nb" and "glm.nb") are good enough for fitting the data, both are asymptotically standard normal (i.e. N(0,1)) and they do not need further "polishing". Nb (denoted as model1) is somewhat better than glm.nb (denoted as model2) their AICs and BICs are approximately equal and they compare very well with that of the raw if fitted with standard normal (this is our expectation, as stated in the "popular" Central Limits Theorem (CLT)).

Figure 5 "inspects" the possibility of fitting the data through the use of ordinary nonparametric bootstrapping method but the result shows that the method is not adequate for the data. In

Figure 6, "nb" (i.e. "m1") stipulates that to fit the daylight radon count data, the variables "TIME" and "TEMP" are the most needed whilst to fit the night radon count data, the variable "PRESS" is mostly required.

Figure 7 gives the confidence intervals for all the mostly needed variables and the result of the work of R's "ggplotter", on the mostly needed variables, is shown in Figure 8.

Finally, a means of using the night's non-zero radon emission count to predict the corresponding non-zero day's radon emission count and vice-versa has been determined as follows; If the researcher has the highest night's radon emission count then to obtain an approximation of the corresponding day's radon emission count, amounts to multiplying it with the constant, $\frac{11}{12}$.

If, on the other hand, he has the highest day's radon emission count, then he needs to multiply it with $\frac{12}{11}$ to obtain an approximation of the equivalent night's count.

This claim is borne out of the following four observations on Figure 8:

1. At location "0", the highest night reading is approximately equal to 114 while its daylight equivalent is approximately 106 (i.e., $\frac{11}{12}(114) = 104.5$ which is, barring experimental errors, close to 106).

2. At location "1", the highest night reading is approximately equal to 107 while its daylight equivalent is approximately 98 (i.e. $\frac{11}{12}\left(107\right) = 98$ which is an exact value).

3. At location "2", the highest night reading is approximately equal to 99 while its daylight equivalent is approximately 93 (i.e.

$\frac{11}{12}(99) = 90.75$ which is, barring experimental errors, close to 93).

4. At location "3", the highest night reading is approximately equal to 93 while its daylight equivalent is approximately 86 (i.e. $\frac{11}{12}(93) = 85$ which is close to 86).

The converse is also true if one desires to go from highest daylight radon count to its night equivalent.

REFERENCES

1. Al Zabadi, H., S. Musmar, S. Issa, N. Dwaikat, and G. Saffarini. 2012. "Exposure Assessment of Radon in the Drinking Water Supplies: A Descriptive Study in Palestine". *BMC Research Notes*. 5:29. http://www.biomedcentral.com/1756-0500/5/29.

2. Chege, M.W., I.V.S. Rathore, S.C. Chhabra, and A.O. Mustapha. 2009. "The Influence of Meteorological Parameters on Indoor Radon in Selected Traditional Kenyan Dwellings". *J. Radiol. Prot*. 29:95 – 103.

3. Crawley, M.J. 2007. *The R Book.* John Wiley and Sons: West Sussex, UK.

4. Darby, S., H. Deo, R. Doll, and E. Whitley. 2001. "A Parallel Analysis of Individual and Ecological Data on Residential Radon and Lung Cancer in South-West England". *J. R. Statist. Soc. A*. 164, part 1: 193 – 203.

5. Fitzpatrick-Lewis, D., J. Yost, D. Ciliska, and S. Krishnaratne. 2010. "Communication about Environmental Health Risks: A Systematic Review". *Environmental Health*. 9:67. http://www.ehjournal.net/content/9/1/67.

6. Lambert, D. 1992. "Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing". *Technometrics*. 34(1).

7. Ramola, R.C., M.S. Kandari, M.S. Negi, and V.M. Choubey. 2000. "A Study of Diurnal Variation of Indoor Radon Concentrations". *Journal of Health Physics*. 35(2): 211-216.

8. Seftelis, L., G. Nicolaou, S. Trassanidis, and F.N. Tsagas. 2007. "Diurnal Variation of Radon Progeny". *Journal of Environmental Radioactivity*. 97: 116-123.

9. Singh, J., H. Singh, S. Singh, and B.S. Bajwa. 2010. "Measurement of Soil Gas Radon and its Correlation with Indoor Radon around some Areas of Upper Siwaliks, India". *J. Radiol. Prot*. 30: 63 – 71.

10. Smith., B.J. and J.J. Oleson. 2008. "Geostatistical Hierarchical Model for Temporally Integrated Radon Measurements". American Statistical Association and the International Biometric Society. *Journal of Agricultural, Biological, and Environmental Statistics*, 13(2):140–158. DOI: 10.1198/108571108X312896.

11. Yau, K.K.W., A.H. Lee, and P.J.W. Carrivick. 2004. "Modeling Zero-Inflated Count Series with Application to Occupational Health". *Computer Methods and Programs in Biomedicine*. 74: 47-52.

SUGGESTED CITATION

Dawodu, G.A., A.O. Mustapha, and O.M. Olayiwola. 2020. "Radon Concentration using Zero Inflated Negative Binomial (ZINB) Modeling". *Pacific Journal of Science and Technology*. 21(1):52-62.

Pacific Journal of Science and Technology