Dog Bite and Full-Moon Relation

It was believed for a long time by medical practitioners that the full moon influenced the expression of medical conditions including fevers, rheumatism, epilepsy and bipolar disorder – in fact, the antiquated term "lunatic" derives from the word lunar, i.e., of the moon. In the late 1990's a (tongue in cheek) study was undertaken to test if the full moon induced dogs to become more aggressive, with a resulting increased likelihood of biting people. In addition to being a little bit of fun, examining a problem like this through the lense of data science is an instructive example on how quantitative methods can be used to answer "folk-lore" questions/hypotheses.

The file dogbites.fullmoon.csv contains the daily number of admissions to hospital of people being bitten by dogs from 13th of June, 1997 through to 30th of June, 1998 . It also contains a second column indicating whether the day in question was a full moon or not. Use this data to answer the following questions. As we know that the Poisson distribution is not a good fit to the daily dog-bite data: instead, for this question we will use a normal distribution as it provides an improved fit to the data due to its increased flexibility, while accepting this assumption is also not necessarily correct; to quote the famous statistician G.E.P.Box: "all models are wrong – but some are more useful than others".

1) Calculate an estimate of the average number of dog-bites for days on which there was a full moon. Calculate a 95% confidence interval for this estimate using the t-distribution:

Assignment - 2

Q → 1

(1.)

→ As shown above,

→ the estimator (average number of dog-bites for days on which there was a full moon): $\hat{\mu}_{\text{Full-Moon}} = \boxed{4.231}$

* 95% confidence interval using t-distribution,

$$= \left(\hat{\mu}_{\text{Full-Moon}} - t_{\alpha/2,\, n-1} \cdot \frac{\hat{\sigma}_{FM}}{\sqrt{n}} ,\ \hat{\mu}_{\text{Full-moon}} + t_{\alpha/2,\, n-1} \cdot \frac{\hat{\sigma}_{FM}}{\sqrt{n}} \right)$$

→ As from R script above,

→ $\alpha = 0.05$, $n = 13$, $\hat{\sigma}_{FM} = 2.5545$, $t_{\frac{0.05}{2},\, 12} = 2.1788$

* Substituting,

$$\rightarrow CI = \left(4.231 - 2.1788 \times \frac{2.5545}{\sqrt{12}} ,\ 4.231 + 2.1788 \times \frac{2.5545}{\sqrt{12}} \right)$$

$$= (4.231 - 2.1788 \times 0.7374 ,\ 4.231 + 2.1788 \times 0.7374)$$

$$= (4.231 - 1.6067 ,\ 4.231 + 1.6067)$$

$$= \boxed{(2.6243 ,\ 5.8377)}$$

Researchers asked the question: do dogs bite more on the full moon? Using the provided data and the approximate method for difference in means with unknown variances presented in Lecture 4, calculate the estimated mean difference in mean dog bite occurences between full moon days and non-full moon days, and a 95% confidence interval for this difference

* substituting,

$\rightarrow CLI = \left(4.231 - 2.1788 \times \frac{2.5545}{\sqrt{13}}, \; 4.231 + 2.1788 \times \frac{2.5545}{\sqrt{13}}\right)$

$= (4.231 - 2.1788 \times 0.7085, \; 4.231 + 2.1788 \times 0.7085)$

$= (4.231 - 1.5437, \; 4.231 + 1.5437)$

$= \boxed{(2.6873, 5.7747)}$

⇒ Hence, We are 95% confident that the value of the Population mean lies between 2.6873 and 5.7747.

(2) For estimated mean difference,

→ mean (estimated) of dog bites ~~on the~~ for days on which there was a full moon, $\hat{\mu}_{FM} = 4.231$

→ estimated mean of dog bites for days on which there was not a full moon, $\hat{\mu}_{NFM} = 4.515$

⇒ Estimated mean difference in mean dog bite occurences between full moon days and non-full moon days:

$= \hat{\mu}_{FM} - \hat{\mu}_{NFM} = 4.231 - 4.515 = \boxed{-0.284}$

→ So, the difference of estimators is 0.284

* The 95% confidence interval of estimated mean difference is given by:

$$= \left(-0.284 \pm Z_{\frac{0.05}{2}} \cdot \sqrt{\frac{\hat{\sigma}^2_{FM}}{n_{FM}} + \frac{\hat{\sigma}^2_{NFM}}{n_{NFM}}}\right)$$

→ As we know, $Z_{\frac{0.05}{2}} = 1.96$, $\hat{\sigma}_{FM} = 2.5545$, $\hat{n}_{FM} = 13$

$\hat{\sigma}_{NFM} = 3.5669$, $\hat{n}_{NFM} = 365$

* 95% confidence level interval:

$$= \left(-0.284 - 1.96 \times \sqrt{\frac{(2.5545)^2}{13} + \frac{(3.5669)^2}{365}},\ -0.284 + 1.96 \times \sqrt{\frac{(2.5545)^2}{13} + \frac{(3.5669)^2}{365}} \right)$$

$$= \left(-0.284 - 1.96 \times \sqrt{0.5019 + 0.0348},\ -0.284 + 1.96 \times \sqrt{0.5019 + 0.0348} \right)$$

$$= \left(-0.284 - 1.96 \times \sqrt{0.5367},\ -0.284 + 1.96 \times \sqrt{0.5367} \right)$$

$$= (-0.284 - 1.4359,\ -0.284 + 1.4359)$$

$$= \boxed{(-1.7199,\ 1.1519)}$$

⇒ Hence, we are 95% confident that population mean difference in dog bite occurences between full moon days and non-full moon days lie between −1.7199 and 1.1519. That also means that we cannot rule out the possiblity that the popular mean difference could be zero which means there could be same frequency of dogbites on full moon and non-full moon days.

Test the hypothesis that dogs bite more frequently on full moon days than on non-full moon days. Write down explicitly the hypothesis you are testing, and then calculate a p-value using the approximate hypothesis test for differences in means with unknown variances presented in Lecture 5. What does this p-value suggest about the behaviour of dogs on full moon days vs non-full moon days? Show working as required.

(3.)

→ We want to test that dogs bite more frequently on full moon days than on non-full moon days.

⇒ H_0: mean estimator difference between on full moon and non-full moon is less than or equal to zero

$$\Rightarrow \hat{\mu}_{FM} - \hat{\mu}_{NFM} \leq 0$$

$$\Rightarrow H_A: \hat{\mu}_{FM} - \hat{\mu}_{NFM} > 0$$

⇒

z value of the data above is,

$$\rightarrow z_0 = \frac{\hat{\mu}_{FM} - \hat{\mu}_{NFM}}{\sqrt{\frac{\hat{\sigma}^2_{FM}}{n_{FM}} + \frac{\hat{\sigma}^2_{NFM}}{n_{NFM}}}}$$

$$\rightarrow z_0 = \frac{4.231 - 4.515}{\sqrt{\frac{(2.5545)^2}{13} + \frac{(3.5669)^2}{365}}}$$

$$\rightarrow z_0 = \frac{-0.284}{\sqrt{0.5367}} = \frac{-0.284}{1.4359} = -0.1978$$

* Therefore, P-value is: P = ~~Rspb~~ 1 − pnorm (z < −0.1978)

$$\rightarrow P \simeq \boxed{0.4216}$$

⇒ As, $P > 0.1$, we have very weak or no evidence against null hypothesis that means mean estimator of dog bite is more frequent on full moon than on non-full moon days. ~~will be rejected~~ won't be rejected. So, the null hypothesis can't

——— continued ———

be rejected. So, there can be the case where dog bite occurs more frequently on non full moon day than full moon day or these two might be equal too.

$\boxed{1[3]}$