**Quantitative Methods in the Law**

**Statistical Analysis**

**Cause in a Mass Tort Case**

This paper will perform a statistical analysis of the claim against Syntrex for negligently allowing tolulene to leak into Winder’s water supply, which may have significantly increased tolulene and caused the liver cancer death rates in Winder to increase. To ascertain whether the liver cancer death rates increased, it is important to analyze the rate of liver cancer deaths of Winder contrasted to the rates of the United States and Georgia. (Pl. ¶ 1.) Additionally, in order for plaintiff’s experiments on laboratory rats, trying to show a causal link between tolulene and liver cancer, to be useful, the experiments should display a statistical significance between the experimental and control groups. (*Id*. ¶ 3.) Finally, the water sample tests taken in 1999 and 2005 along with the annual consumption of water may be useful in demonstrating Syntrex’s contribution of tolulene to the water supply. (*Id*. ¶¶ 2, 4, 5.)

Viewing the observed deaths in Winder, we can calculate their significance by comparing them to both the national and state rates and approximate how much Winder’s rates, as a sample, can be attributed to chance. (*Id*. ¶ 1.) Data is given for the annual deaths between 1999 and 2005 from the 10,200 Winder population, and to begin, the data must be converted into comparable rates of deaths per 100,000 of population by, first, dividing the annual numbers by 10,200 to get the per person rate of death in Winder.  *Id*.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0$÷$10,200 | 1$÷$10,200 | 1$÷$10,200 | 0$÷$10,200 | 1$÷$10,200 | 1$÷$10,200 | 3$÷$10,200 | 2$÷$10,200 | 4$÷$10,200 | 3$÷$10,200 |
| 0.00% | 0.0098% | 0.0098% | 0.00% | 0.0098% | 0.0098% | 0.02941% | 0.01961% | 0.03922% | 0.02941% |
| 1$÷$10,200 | 1$÷$10,200 | 0$÷$10,200 | 0$÷$10,200 | 1$÷$10,200 | 2$÷$10,200 | 0$÷$10,200 | 0$÷$10,200 | 1$÷$10,200 | 0$÷$10,200 |
| 0.0098% | 0.0098% | 0.00% | 0.00% | 0.0098% | 0.01961% | 0.00% | 0.00% | 0.0098% | 0.00% |

The next step will be to multiply by 100,000 to get comparable United States and Georgia rates per 100,000 of population.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0.00 | 9.80 | 9.80 | 0.00 | 9.80 | 9.80 | 29.41 | 19.60 | 39.21 | 29.41 |
| 9.80 | 9.80 | 0.00 | 0.00 | 9.80 | 19.60 | 0 | 0 | 9.80 | 0 |

We can use these values to calculate the mean annual rate of liver cancer in Winder to be 10.784, with a standard deviation of 11.14.

$$mean\_{sample}= \frac{0+9.8+9.8+0+9.8+9.8+29.41+19.61+39.22+29.41+9.8+9.8+0+0+9.8+19.61+0+0+9.8+0}{20}=10.784$$

 $SD\_{sample}= \sqrt{\frac{116+1+1+116+1+1+347+78+808+347+1+1+116+116+1+78+116+116+1+116}{20}}=11.14$

$SD+ = \sqrt{\frac{n}{n-1}}\*SD= \sqrt{\frac{20}{20-1}}\*11.14=11.425$ (SD+ is used for small samples less than 30)

The mean rate of cancer deaths in Winder is 10.784 per 100,000 people, which looks high compared to the 1.4 United States or 1.1 Georgian rates; however, the standard deviation is 11.425, and having a large spread in standard deviation can have a decreasing effect on the statistical significance. The statistical significance, or p-value, will demonstrate how likely the observed values in Winder are due to chance alone, which can first be solved by finding out the degrees of freedom and the t-value, used to compare rates of small samples to those of large population.

 $SE=\frac{11.425}{\sqrt{20}}=2.555$;

 $t=\frac{Observed Differenc –Expected Difference }{(SE)Standard Error}$

 $t\_{Georgia}=\frac{10.784-1.4 }{2.555}=3.67$; And, $t\_{United States}=\frac{10.784-1.1 }{2.555}=3.79$

 $Degrees of Freedom=n\_{Years}-1= 20-1=19$

Applying the table in the appendix A-2, a sample with 19 degrees of freedom and a t-value greater than 3.58, indicates a p-value of less than 0.1%, and compared to the national and state rates, we would expect to get a sample similar to the Winder’s 0.1% of the time or less if solely by chance. This is statistically significant and either Winder has had a run of extremely bad luck or something other than chance has caused the elevated cancer rates.

Allowing that the cancer rates in Winder are statistically significant, the next step is to establish whether the plaintiff’s experiments on rats display any affect of tolulene on liver cancer rates. The percentage of the rats getting cancer in the control group is 11.6% and the percentage in the experimental group is 15.4%. (*Id*. ¶ 3.)

 $\%\_{Control Group} \frac{Number of Deaths\_{CP}}{Number of rats\_{CG}}= \frac{20}{173}=11.6\%$

 $\%\_{Experimental Group} \frac{Number of Deaths\_{EP}}{Number of rats\_{EG}}= \frac{34}{221}=15.4\%$

Similar to finding the t-value for Winder’s liver cancer rates, we will find the z-value for comparing two large samples, using the following method.

**Step 1: Determine SD of the two samples:**

 $SD\_{CG }= \sqrt{11.6\%×(1-11.6\%)}= 0.320$

 $SD\_{EG }= \sqrt{15.4\%×(1-15.4\%)}= 0.361$

 **Step 2: Determine SE of two samples.**

 $SE\_{CG} = \sqrt{Number of Rats}× SD\_{CG }=\sqrt{173}×0.320=4.21$

 $SE\_{EG} = \sqrt{Number of Rats}× SD\_{CG }=\sqrt{221}×0.361=5.37$

 **Step 3: Determine SE for the Percentage of two samples.**

 $SE for \%\_{CG} = \frac{SE\_{CG}}{N\_{CG}}=\frac{4,21}{173}=2.43$

 $SE for \%\_{EG} = \frac{SE\_{EG}}{N\_{EG}}=\frac{5.37}{221}=2.43$

 **Step 4: Determine SE for the Difference.**

 $SE for the Difference= \sqrt{\left(SE for \%\_{CG}\right)^{2}+\left(SE for \%\_{EG}\right)^{2}}=$

 $\sqrt{\left(2.43\right)^{2}+\left(2.43\right)^{2}}= 0.344$

 **Step 5: find z-value**

 $z= \frac{Observed\_{Difference}-Expected\_{Difference}}{SE for the Difference}$

 $z= \frac{\left(15.4\%-11.6\%\right)- 0}{0.344}=1.11$

To be statistically significant the p-value is ordinarily desired to be less than 5%, and looking at the table in appendix A-1, any z-value of less than 1.65, like our 1.11, will give a p-value of greater than 5%. The greater rate of liver cancer in the experimental group compared to the control group is not statistically significant, and the higher rate of cancer could be attributed to tolulene; however, the experiment has not ruled out that the increased rate was cause by mere chance. The plaintiff should try to find more favorable evidence for the connection.

Finally, assuming the plaintiff can find better evidence for tolulene’s effect on liver cancer, it will be important to examine Syntrex’s contribution of tolulene to the total amount of toluene in Winder’s water system. Syntrex’s contribution of tolulene, from the facts of the case, totals 220 kilograms for all the years in question, which Syntrex claims it can reduce to 20%. (*Id*. ¶ 4.) If Syntrex is the only source of tolulene, we would expect that Winder’s ground water contains at most 220 kilograms of tolulene and anything over 220 kilograms would have had to come from somewhere other than Syntrex’s storage. From the facts, approximately 900 million liters are consumed from ground wells annually, and given that 220 kilograms is the highest concentration, the highest concentration in one annual bucket of water would have to be 244.44 micrograms per liter, assuming that there is no tolulene in the other years. (*Id*. ¶ 5.)

$$\frac{220 kilograms}{900,000,000 liters of water}=2.44×-10^{7}=244.44 migrograms per liter$$

A 1999 test, of 25 samples, showed a mean presence of 328 micrograms of tolulene per liter with a standard deviation of 175 micrograms. Looking at the normal distribution table in appendix A-3, we can determine with 95% confidence rate having two standard errors above and below the mean. (*Id*. ¶ 2.)

 $n \left(numberof samples taken\right)=25, SD(Standard Deviation)=175$

 $SE of Average= \frac{SD}{\sqrt{n}}$; $\frac{175}{\sqrt{25}} =35$; $Twe SEs= SE\*2=70$

Based on the 1999 samples, within 95% confidence, the level of tolulene in the 1999 water supply is 328 ± 70 micrograms, with 258 micrograms at the lower end, which is 38 micrograms more than the maximum amount of tolulene Syntrex could have leaked into the system. (*Id*. ¶ 2.) This does not account for tolulene in the other years, and another test in 2005, using the calculation above, showed a 95% confidence of 150 ± 36 micrograms. The plaintiff has a high hurdle showing Syntrex’s contribution to negligence since the base line amounts from just one year has already exceed the maximum amount Syntrex could have leaked. Most other assumptions only dilute the amount of tolulene like the 900 million liters drawn in other years, the amount leaving the system or still in the ground, and the 20% maximum amount of leakage Syntrex says it can prove. Nevertheless, some degree of high concentration could be due to samples taken next to the Syntrex plant, depending on how many samples were taken near Syntrex. It would be important to know how distributed the samples from the two years were across the whole town. If the samples are shown to be concentrated around Syntrex, it would be important to know amount of cancer deaths near those wells compared to the rest of Winder.

To summarize the statistical analysis of the data in the case, the rate of liver cancer deaths in Winder has a high statistical significance compared to the national and state rates, and the high rates in Winder can be reasonably attributed to something other than chance. (*Id*. ¶ 1.) Also, the cancer causing properties of tolulene can’t be statistically proven by the plaintiff’s rat experiment alone, and this test is with a concentration of 5000 micrograms of tolulene on small rats, almost 20 times larger than any other assumed concentration in the Winder water supply. (*Id*. ¶ 3.) Finally, the water supply test from 25 samples in 2005 and 20 samples in 1999 show a much higher concentration than Syntrex could have contributed to the water supply. A high concentration of the two sets of samples could be attributed to samples taken near Syntrex and not widely distributed across the rest of the city’s water supply. (*Id*. ¶¶ 2, 4, 5.) If this is the case, it would be important to know the cancer deaths from areas near these wells compared to the rest of the city.

Appendix A-1: Distribution



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Z | Area |  | z | area |  | z | area |
| 0.00 | 0.00% |  | 1.50 | 86.64% |  | 3.00 | 99.730% |
| 0.15 | 11.92% |  | 1.65 | 90.11% |  | 3.15 | 99.837% |
| 0.30 | 23.58% |  | 1.80 | 92.81% |  | 3.30 | 99.903% |
| 0.45 | 34.73% |  | 1.95 | 94.88% |  | 3.45 | 99.944% |
| 0.60 | 45.15% |  | 2.10 | 96.43% |  | 3.60 | 99.968% |
| 0.75 | 54.67% |  | 2.25 | 97.56% |  | 3.75 | 99.982% |
| 0.90 | 63.19% |  | 2.40 | 97.56% |  | 3.90 | 99.990% |
| 1.05 | 70.63% |  | 2.55 | 98.36% |  | 4.05 | 99.995% |
| 1.20 | 76.99% |  | 2.70 | 98.92% |  | 4.20 | 99.997% |
| 1.35 | 82.30% |  | 2.85 | 99.31% |  | 4.35 | 99.999% |

Appendix A-2: T-Table



With 19 degrees of freedom, a t value that is greater than 1.73 will give us a p value greater than 5% and we can reject the null.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Degrees of Freedom | 10% | 5% | 1% | 0.50% | 0.10% |
| 1 | 3.08 | 6.31 | 31.82 | 63.66 | 318.31 |
| 2 | 1.89 | 2.92 | 6.96 | 9.92 | 22.33 |
| 3 | 1.64 | 2.35 | 4.54 | 5.84 | 10.21 |
| 4 | 1.53 | 2.13 | 3.75 | 4.60 | 7.17 |
| 5 | 1.48 | 2.02 | 3.36 | 4.03 | 5.89 |
| 6 | 1.44 | 1.94 | 3.14 | 3.71 | 5.21 |
| 7 | 1.41 | 1.89 | 3.00 | 3.50 | 4.79 |
| 8 | 1.40 | 1.86 | 2.90 | 3.36 | 4.50 |
| 9 | 1.38 | 1.83 | 2.82 | 3.25 | 4.30 |
| 10 | 1.37 | 1.81 | 2.76 | 3.17 | 4.14 |
| 11 | 1.36 | 1.80 | 2.72 | 3.11 | 4.02 |
| 12 | 1.36 | 1.78 | 2.68 | 3.05 | 3.93 |
| 13 | 1.35 | 1.77 | 2.65 | 3.01 | 3.85 |
| 14 | 1.35 | 1.76 | 2.62 | 2.98 | 3.79 |
| 15 | 1.34 | 1.75 | 2.60 | 2.95 | 3.73 |
| 16 | 1.34 | 1.75 | 2.58 | 2.92 | 3.69 |
| 17 | 1.33 | 1.74 | 2.57 | 2.90 | 3.65 |
| 18 | 1.33 | 1.73 | 2.55 | 2.88 | 3.61 |
| 19 | 1.33 | 1.73 | 2.54 | 2.86 | 3.58 |
| 20 | 1.33 | 1.72 | 2.53 | 2.85 | 3.55 |

Appendix A-3



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| z | area |  | z | area |  | z | area |
| 0.00 | 0.00% |  | 1.00 | 68.27% |  | 2.00 | 95.450% |
| 0.10 | 7.97% |  | 1.10 | 72.87% |  | 2.10 | 96.427% |
| 0.20 | 15.85% |  | 1.20 | 76.99% |  | 2.20 | 97.219% |
| 0.30 | 23.58% |  | 1.30 | 80.64% |  | 2.30 | 97.855% |
| 0.40 | 31.08% |  | 1.40 | 83.85% |  | 2.40 | 98.360% |
| 0.50 | 38.29% |  | 1.50 | 86.64% |  | 2.50 | 98.758% |
| 0.60 | 45.15% |  | 1.60 | 86.64% |  | 2.60 | 99.068% |
| 0.70 | 51.61% |  | 1.70 | 89.04% |  | 2.70 | 99.307% |
| 0.80 | 57.63% |  | 1.80 | 91.09% |  | 2.80 | 99.489% |
| 0.90 | 63.19% |  | 1.90 | 92.81% |  | 2.90 | 99.627% |

Words: – 448 (tables and equations) = 1341