Section: Medical Sciences

A Study on the Severity of UTI – A Statistical Approach

Key Words: Colony forming units, Escherichia coli, Alcohol, Antibiotics, Natural remedies, ANOVA, t-test, p-value

Saloni Sinha*, Maitrayee Ghosh*, Anindya Bose**, Arijit Mal*, Mugdha Dutta**, Tamoghno Haldar**, Ayan Chandra**

* Dept. of Micro-Biology, St. Xavier's College, Kolkata
** Dept. of Statistics, St. Xavier's College, Kolkata

UTI (Urinary Tract Infection) is a bacterial infection caused mostly by uropathogenic *Escherichia coli* and can involve pyelonephritis (kidney), cystitis (bladder) or asymptomatic bacteriuria.Treatment with antibiotics (Livofloxacin, Amoxicillin, Gentamycin, Nitrofurantoin etc.) can cure the disease but their effectivity is largely influenced by other factors like age, sex, alcohol consumption, intake of natural remedies and even lifestyle.

A statistical analysis revealed that women are more susceptible to UTI than men for the age group 15-55 years (p value=0.0001) but for 55 and above, frequency of occurrence of infection was roughly equal in both the sexes (p value=0.042).

Again younger women were almost equally susceptible to UTI as post-menopausal women due to active sexual life, use of birth control methods etc. Also antibiotic efficiency was reduced by alcohol consumption which lowered the decrease in colony count. A p value of 0.0001 confirmed that reduction in colony count was more in non-alcoholics than in alcoholics. Treatment with natural remedies however helped to overcome the effect of alcohol to a certain extent thereby increasing the efficacy of antibiotics.

INTRODUCTION:

Urinary Tract Infection(UTI) is defined as a significant bacteriuria in the presence of symptoms. The bacteria most often seen in UTIs are of fecal origin. More than 90% of acute UTIs in patients with normal anatomic structure and functions are caused by different strains of uropathogenic *E.coli* and the rest 10% by other bacteria such as *Staphylococcus saprophyticus, Proteus mirabilis* etc.

In many cases, bacteria first travel to the urethra. When bacteria multiply, an infection can occur. An infection limited to the urethra is called urethritis. If bacteria move to the bladder and multiply, a bladder infection, called cystitis, results. If the infection is not treated promptly, bacteria may then travel further up the ureters to multiply and infect the kidneys. A kidney infection is called pyelonephritis.

Escherichia coli is the most common pathogen found in urinary tract infections (UTI). The adherence of *E. coli* to uroepithelial cells seems to be very important in the pathogenesis of the disease. Specific adhesion is mediated by bacterial proteins termed adhesins which may or may not be associated with fimbriae. The adhesin that has been most closely associated with uropathogenic *E. coli* is the P fimbria. The letter designation is derived from the ability of P fimbriae to bind specifically to the P blood group antigen which contains a D-galactose-D-galactose residue. The fimbriae bind not only to red cells but to a specific galactose disaccharide that is found on the surfaces uroepithelial cells in approximately 99% of the population. Uropathogenic strains of *E. coli* usually produce siderophores that probably play an essential role in iron acquisition for the bacteria during or after colonization.

Urinary tract infection due to *Proteus mirabilis* is not traditional and mostly reported in individuals with structural abnormalities of the urinary tract and is frequently isolated from the urine of elderly patients undergoing long-term catheterization and women with repeated UTI. To be able to invade and successfully establish on the uroepithelial cells of the host, *P.mirabilis* produces several virulence factors, e.g., pore-forming hemolysins, proticine 3, leukocidin, endotoxin, IgA and IgG proteases urease, deaminase, adhesions, polysaccharide capsules, pili/fimbrae, peritrichous flagella, the ability to form biofilms and swarming ability. The above-mentioned virulence factors enable the pathogen to overcome the different defense mechanisms of the host.

The study of UTI is immensely important because of the following reasons:

1. Urinary tract infections are a serious health problem affecting millions of people each year.
2. Infections of the urinary tract are the second most common type of infection in the body.
3. Urinary tract infections (UTIs) account for about 8.3 million doctor visits each year.
4. Women are especially prone to UTIs than men by the virtue of their shortened urethra.
5. One woman in five develops a UTI during her lifetime. UTIs in men are not as common as in women but can be very serious when they do occur.

OBJECTIVES:

The main objectives of this paper are to show the following by using data support and statistical analysis:

- To show that women are more prone to UTI than men.

- To detect the decline in the efficacy of antibiotics in the presence of alcohol in UTI patients.
- To detect the effects of natural remedies towards increasing the efficacy of antibiotics in alcoholics.
- To detect the prevalence of UTI in two differently age matched group of women.

MATERIALS AND METHODS:

- **Collection of data:**

  We visited R.M. pathological laboratory and similar other laboratories of Kolkata to collect data for our analysis.

- **Processing of data:**

  A statistical analysis was performed on the basis of the data collected. The statistical methods used were Test for Proportion, Paired t-test, Box-Plot and Analysis of Variance (ANOVA) and Fisher's t-test.

METHODOLOGY:

1.1 Test for Proportion

Suppose there are two characteristics A and Band let the population proportion of individuals possessing these characteristics be p_1 and p_2 respectively. Suppose our problem is to test

H_0: $p_1 = p_2$ against H_1: $p_1 < p_2$.

Suppose we have two independently drawn random samples of sizes n_1 and n_2 from the two populations. Let f_1 and f_2 be the number of sample members possessing A and B respectively out of n_1 and n_2 individuals.

The test statistic is then given by

$$T = \frac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left\{\frac{1}{n_1}+\frac{1}{n_2}\right\}}},$$

where, $\widehat{p_1} = \frac{f_1}{n_1}$, $\widehat{p_2} = \frac{f_2}{n_2}$ and $\hat{p} = \frac{f_1+f_2}{n_1+n_2}$.

Under H_0, for moderately large sample sizes, T follows approximately a standard normal distribution.

We reject H_0 against H_1 if and only if p-value < α, α being the desired level of significance. We usually take α = 0.05. 0.01, etc.

In this test, p-value = $P\{T \leq t|H_0\}$ where t is the observed value of T on the basis of the given sample.

1.2 Paired t-test

Suppose we have n pairs of observations (x_1, y1), (x_2, y2), …, (x_n, y_n) on two variables x and y. Here our objective is to test

$$H_0 : \mu_x = \mu_y \text{ against } H_0 : \mu_x \neq \mu_y$$

where μ_x and μ_y are the population means of x and y respectively.

Let us write $z_i = x_i - y_i$, i = 1,2,…,n.

Here the test statistic is given by

$$T = \frac{\sqrt{n}\bar{z}}{s_z},$$

where $\bar{z} = \frac{1}{n}\sum z_i$ and $s_z = \sqrt{\frac{1}{n}\sum\left(z_i - \bar{z}\right)^2}$.

We reject H_0 against H_1 if and only if $t > t_{tab}$ where t is the observed value of T and t_{tab} is the tabulated value of T.

In this test, p-value = $P\{T > |t||H_0\}$ where t is the observed value of T on the basis of the given sample.

1.3 Box-Plot (or Box-Whisker Diagram)

In descriptive statistics, a box-plot (or box-whisker plot) is a convenient way of graphically depicting groups of numerical data through their five-number summaries:-

1) Sample minimum (the smallest observation) – $x_{(1)}$
2) Lower Quartile – Q1
3) Median – Q2
4) Upper Quartile – Q3

5) Sample maximum (the largest observation) – $X_{(n)}$

A box-plot may also indicate which observations, if any, might be considered as outliers (i.e. observation that appear to deviate markedly from other members of the sample in which it occurs.

The box-plot is interpreted as follows:

- The box itself contains the middle 50% of the data. The upper edge (hinge) of the box indicates the 75th percentile of the data set, and the lower hinge indicates the 25th percentile. The range of the middle two quartiles is known as the inter-quartile range.
- The line in the box indicates the median value of the data.
- If the median line within the box is not equidistant from the hinges, then the data is skewed.
- The ends of the vertical lines or "whiskers" indicate the minimum and maximum data values, unless outliers are present in which case the whiskers extend to a maximum of 1.5 times the inter-quartile range.
- The points outside the ends of the whiskers are outliers or suspected outliers.

Box-plots have the following strengths:

- Graphically display a variable's location and spread at a glance.
- Provide some indication of the data's symmetry and skewness.
- Unlike many other methods of data display, box-plots show outliers.
- By using a box-plot for each categorical variable side-by-side on the same graph, one can quickly compare data sets.

1.4 Analysis of Variance (ANOVA)

The total variation present in a set of observable quantities may, under certain circumstances, be partitioned into a number of components associated with the nature of classification of the data. The systematic procedure for achieving this is called the Analysis of Variance (ANOVA).

➢ Fixed Effects Model

Let us consider the linear model,

$y_{i} = \mu_{i+} e_i$ where μ_i is the true value and e_i is the error.

Now, it is possible that there may be association between errors of successive measurements, but we shall assume that the errors e_i are always independent random variables. These are also expected to have zero expectations and to be homoscedastic. We shall call a model in which all the effects are unknown constants, which we call parameters, **a Fixed - Effect model**.

➢ ANOVA One-way classified Data (Fixed Effects Model)

Let, there be n observations, classified into k classes $A_1, A_2, \ldots, A_k$, the number of observation in the i^{th} class be n_i. Let y_{ij} be the j^{th} observation in the i^{th} class, i = 1,2,...,k and j = 1,2,...,n_i. The scheme of classification is given below:

Class

A_1 A_2.....A_k

y_{11} y_{21}.....y_{k1}

y_{12} y_{22}..... y_{k2}

y_{1n_1} y_{2n_2}.....y_{kn_k}

This is called **One - Way Classification** as we are investigating the effect of one factor only.

Linear Model –

Let, $y_{ij} = \mu_i + e_{ij}$ [For i = 1 (1) k and j = 1 (1) n_i]
here y_{ij} is the observed value, μ_i is the true value and e_{ij} is the error-component.

Let us write, $\mu_i = \mu + \alpha_i$; where $\mu = \sum_{i=1}^{k} n_i\mu_i / n$; $n = \sum_{i=1}^{k} n_i$.

where, μ : general/mean effect and α_i: additional effect due to i^{th} factor level.

Thus, $y_{ij} = \mu + \alpha_i + e_{ij}$ for i = 1 (1) k and j = 1 (1) n_i.

- **Assumptions :** $\sum_{i=1}^{k} n_i\alpha_i = 0$ and $e_{ij} \sim N(0,\sigma^2)$
- **Hypothesis to be tested :**

We have to test, $H_0 = \mu_1 = \mu_2 = \ldots\ldots\ldots\ldots = \mu_k$

against H_1: all μ_i's are not equal.

equivalently, H_0: $\alpha_1 = \alpha_2 = \ldots\ldots\ldots\ldots = \alpha_k = 0$

vs H_1: all α_i's are not zero.

- **ANOVA Table:**

Sources of Variation	Sum of Squares	d.f	MS	F (obs.)
Factor A	SSA	k-1	MSA = SSA/k-1	F_0 = MSA/MSE

Error	SSE	n-k	MSE = SSE/n-k	
Total	TSS	n-1	-	-

- **Critical Region:** $F_0 > F_\alpha$; k-1, n-k.

1.5 Fisher's t-test:

Suppose we have two random samples drawn independently from two populations with means μ_1 and μ_2 respectively. Here we are to test the hypothesis $H_o : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. Here the test statistic is given by

$$t = \frac{\overline{x_1} - \overline{x_2}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Here we reject H_0 against H_1 if the p-value of the test is less than α, the level of significance of the test.

RESULTS:

Women are more prone to UTIs than men because in females the urethra is much closer to the anus. Its length is 4cm as compared to males where the urethra is about 20 cm. Furthermore, women lack the bacteriostatic properties of prostatic secretions. Among the elderly, UTI frequency is roughly equal in women and men. This is due, in part, to an enlarged prostate in older men. As the gland grows, it obstructs the urethra, leading to increased frequency of urinary retention.

We performed a test for proportion to compare the proportions of male and female UTI patients belonging to the age group 15-55 years. The null hypothesis was 'male and females are equally susceptible to UTI' and it was tested against the alternative hypothesis 'male and females are not equally susceptible to UTI'. The p-value of the test was found to be 0.0001. Thus, it was found that females are much more susceptible than males towards UTI.

Another test for proportion to compare the proportions of male and female UTI patients aged 55 years and above was also performed. The p-value has been found to be 0.042. Hence at 5% level of significance we may conclude that these two groups are more or less equally prone to UTI.

The patients were divided into two groups – one consuming alcohol and the other not consuming, both the groups being treated with the same antibiotics. We have recorded the reduction in colony count for these two groups after an interval of 30days. We then performed a

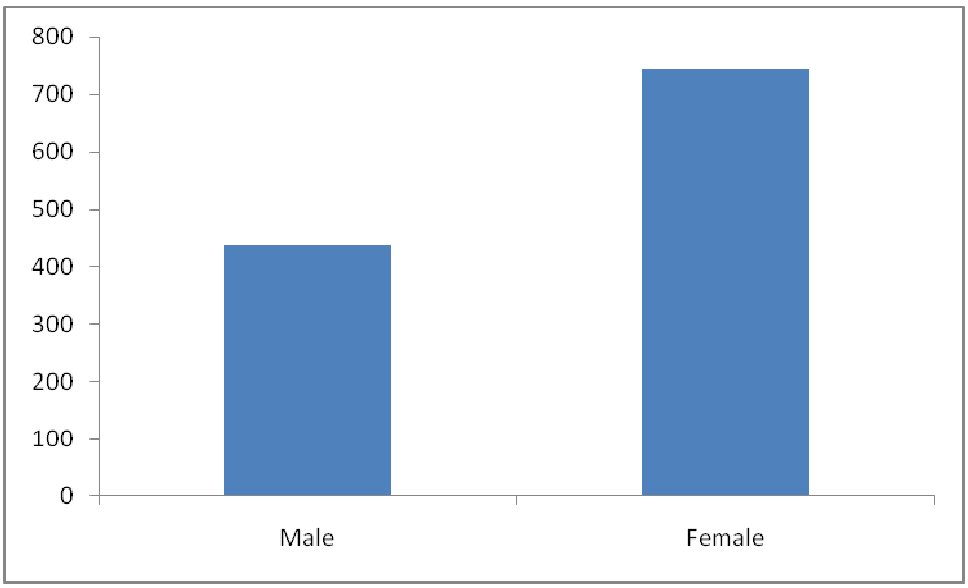
t- test to see whether the decrease in the average colony count is same for the two groups or is more for the non-alcoholic group. The p-value has been found to be 0.0001. The, at 5% level of significance, we may conclude that the reduction in colony count of the non-alcoholic group is significantly more than their counterparts.

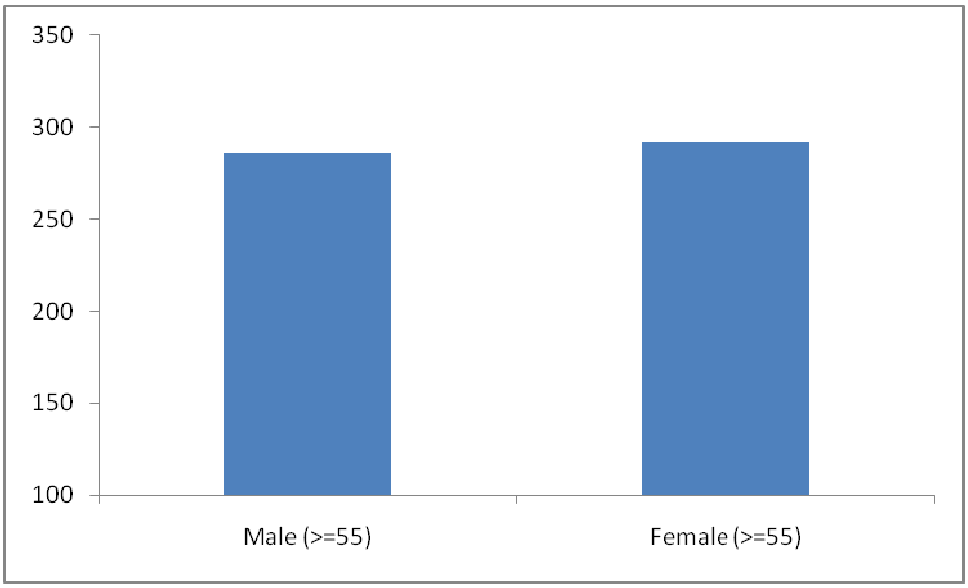
Next, we treated the alcoholic group with natural remedies like bromelain extracted from pineapple(in the form of tablets) as well as cranberry juice which are vitamin C enriched products and effective antioxidants. We conducted the above mentioned t – test again after a time duration of 30days. The p-value has now been found to be 0.0721. This shows at 5% level of significance that there is no significant difference between the two groups as far as the reduction in colony count is concerned. Combining the results obtained in the two t-tests we may then conclude that the natural remedies have increased the efficiency of the antibiotics overcoming the effect of alcohol.

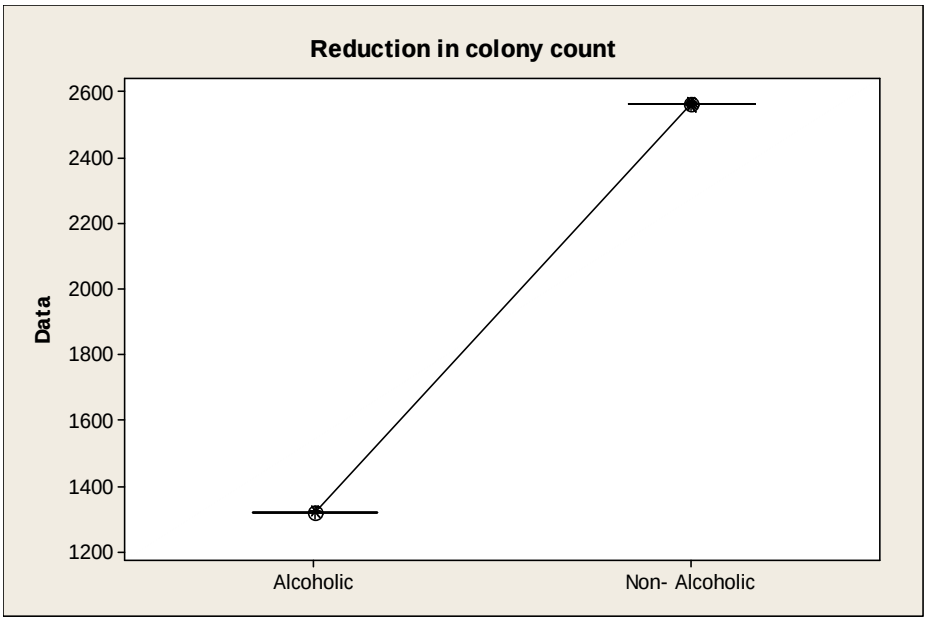
We divided the female patients into two groups – one consisting of young women and the other approaching their menopause. For younger women sexual intercourse, diaphragm and/or spermicidal exposure, a history of UTI and recent antimicrobial exposure have been identified as important host characteristics associated with UTI. For older institutionalized women, urinary catheterization and functional status are the most important risk factors associated with UTI. We have conducted a test of proportion to see whether UTI is more susceptible to any one of these groups. The p-value of the test has been found to be 0.1221. Hence we conclude at 5% level of significance that both the groups are equally susceptible to UTI. Although not directly evaluated in this study, the reduced levels of estrogenic hormones present after menopause appear to contribute to the occurrence of UTI in postmenopausal women.

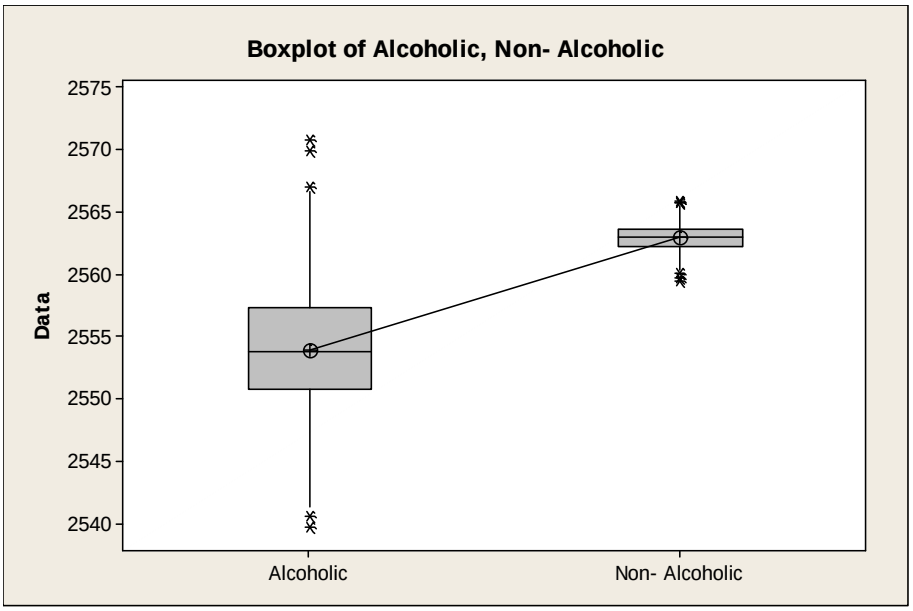
The patients have been divided into three groups – the first group was treated with antibiotics only, the second group consumed alcohol and treated with antibiotics and the third group consumed alcohol, treated with antibiotics and also was given natural remedies. We performed an Analysis of Variance (ANOVA) to see whether the reduction in the colony count is same for all the three groups. We have found that Group 1 shows the maximum reduction while it is the minimum for Group 2.

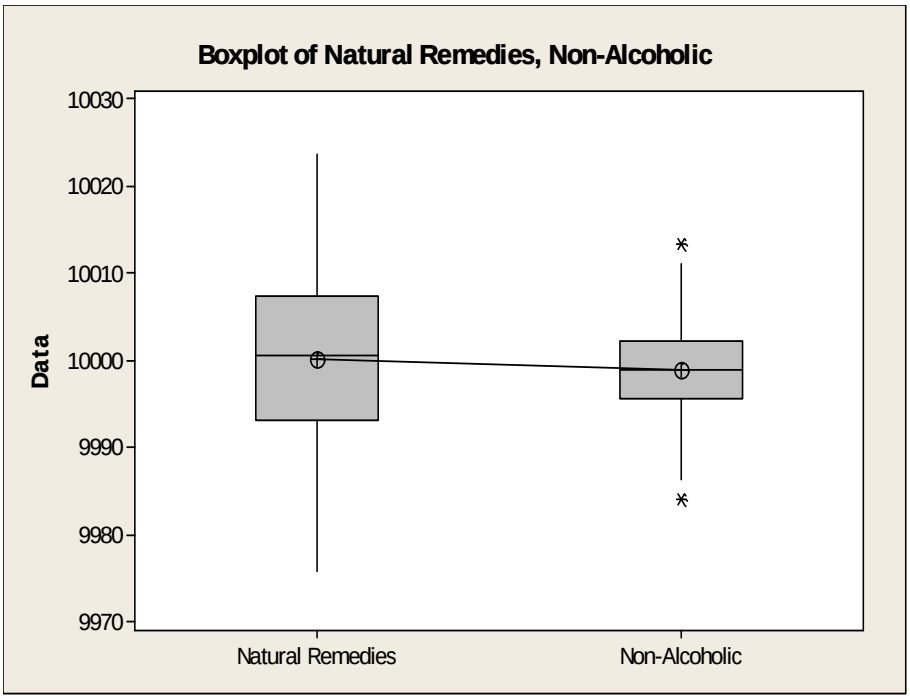
GRAPHS AND TABLES:

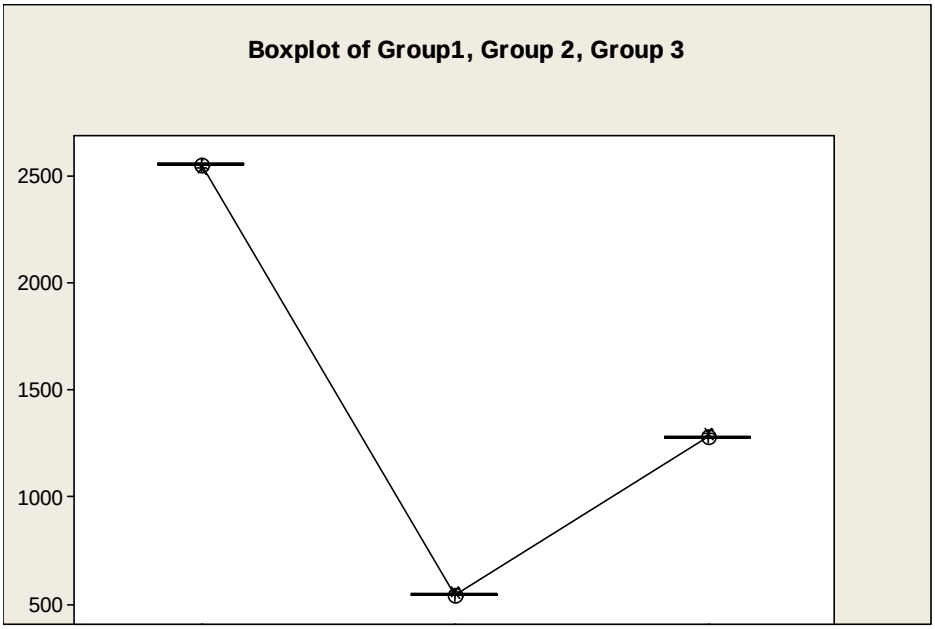

800
700
600
500
400
300
200
100
0
Male
Female


350
300
250
200
150
100
Male (>=55)
Female (>=55)


Reduction in colony count
Data
2600
2400
2200
2000
1800
1600
1400
1200
Alcoholic
Non- Alcoholic


Boxplot of Alcoholic, Non- Alcoholic
Data
2575
2570
2565
2560
2555
2550
2545
2540
Alcoholic
Non- Alcoholic





Source	DF	SS	MS	F	P
Factor	2	959257587	479628793	18442115.20	0.000
Error	1440	37450	26		
Total	1442	959295037			

REFERENCES:

Antibiotics for the prevention of urinary tract infection: a systematic review of randomized controlled trials (2001). G Williams, A Lee.

Urinary Tract Infections and renal papillary necrosis in alcoholism (1986). TO Pitts.

Risk factors for recurrent urinary tract infection in young women (2000). Scholes, TM Hooton, PL Roberts

Recurrent urinary tract infections in post-menopausal women(2000). R Raz, Y Gennesin, J Wasser.