Sex: 1=Male, 2=Female

1. Once the data is "clean", perform a summary analysis of the three discrete variables (sex, lab, and agegroup). For the variables alkphos, cammol and phosmmol, report the mean, median, standard deviation, min and max broken down by sex. Also summarize the variables alkphos, cammol and phosmmol in a similar way with the factor variable as lab.

|  |  |
| --- | --- |
| **Male** | **Female** |
| 91 | 87 |
| Lab |  |
| **Lab 1** | **Lab 2** | **Lab 3** | **Lab 4** | **Lab 5** | **Lab 6** |
| 88 | 42 | 16 | 14 | 11 | 6 |
| Age Group |  |  |  |  |  |
| **Group 1** | **Group 2** | **Group 3** | **Group 4** | **Group 5** |
| 56 | 70 | 38 | 11 | 3 |

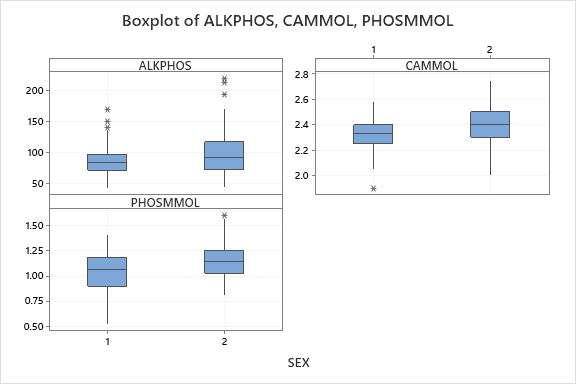
**Statistics**

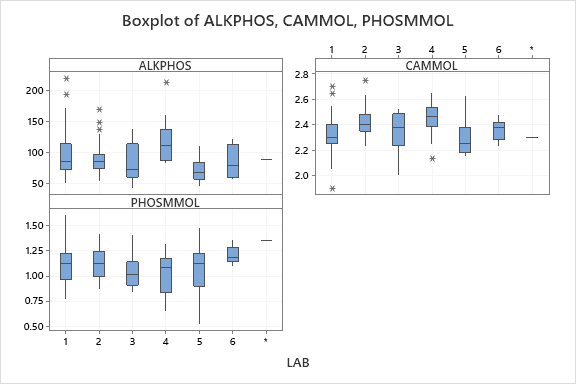
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **SEX** | **N** | **N\*** | **Mean** | **StDev** | **Minimum** | **Median** | **Maximum** |
| ALKPHOS | 1 | 91 | 0 | 85.81 | 22.99 | 42.00 | 83.00 | 168.00 |
|  | 2 | 87 | 0 | 99.22 | 35.74 | 43.00 | 91.00 | 219.00 |
|  |  |  |  |  |  |  |  |  |
| CAMMOL | 1 | 91 | 0 | 2.3181 | 0.1217 | 1.9000 | 2.3300 | 2.5800 |
|  | 2 | 86 | 1 | 2.3938 | 0.1403 | 2.0000 | 2.4000 | 2.7500 |
|  |  |  |  |  |  |  |  |  |
| PHOSMMOL | 1 | 91 | 0 | 1.0593 | 0.1825 | 0.5200 | 1.0700 | 1.4200 |
|  | 2 | 86 | 1 | 1.1486 | 0.1600 | 0.8100 | 1.1450 | 1.6100 |

**Statistics**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **LAB** | **N** | **N\*** | **Mean** | **StDev** | **Minimum** | **Median** | **Maximum** |
| ALKPHOS | 1 | 88 | 0 | 94.80 | 31.34 | 50.00 | 85.50 | 219.00 |
|  | 2 | 42 | 0 | 88.93 | 24.10 | 54.00 | 85.50 | 168.00 |
|  | 3 | 16 | 0 | 83.38 | 30.94 | 42.00 | 72.50 | 138.00 |
|  | 4 | 14 | 0 | 118.43 | 36.58 | 83.00 | 111.00 | 213.00 |
|  | 5 | 11 | 0 | 70.64 | 18.78 | 45.00 | 67.00 | 111.00 |
|  | 6 | 6 | 0 | 84.3 | 26.3 | 57.0 | 78.5 | 122.0 |
|  | \* | 1 | 0 | 89.000 | \* | 89.000 | 89.000 | 89.000 |
|  |  |  |  |  |  |  |  |  |
| CAMMOL | 1 | 87 | 1 | 2.3154 | 0.1324 | 1.9000 | 2.3000 | 2.7000 |
|  | 2 | 42 | 0 | 2.4224 | 0.1005 | 2.2300 | 2.4000 | 2.7500 |
|  | 3 | 16 | 0 | 2.3519 | 0.1478 | 2.0000 | 2.3750 | 2.5300 |
|  | 4 | 14 | 0 | 2.4450 | 0.1372 | 2.1300 | 2.4650 | 2.6500 |
|  | 5 | 11 | 0 | 2.3018 | 0.1586 | 2.1500 | 2.2500 | 2.6300 |
|  | 6 | 6 | 0 | 2.3600 | 0.0874 | 2.2300 | 2.3750 | 2.4800 |
|  | \* | 1 | 0 | 2.3000 | \* | 2.3000 | 2.3000 | 2.3000 |
|  |  |  |  |  |  |  |  |  |
| PHOSMMOL | 1 | 88 | 0 | 1.1053 | 0.1740 | 0.7700 | 1.1300 | 1.6100 |
|  | 2 | 41 | 1 | 1.1293 | 0.1575 | 0.8700 | 1.1300 | 1.4200 |
|  | 3 | 16 | 0 | 1.0363 | 0.1583 | 0.8400 | 1.0200 | 1.4200 |
|  | 4 | 14 | 0 | 1.0343 | 0.1946 | 0.6500 | 1.0850 | 1.3200 |
|  | 5 | 11 | 0 | 1.0845 | 0.2616 | 0.5200 | 1.1300 | 1.4900 |
|  | 6 | 6 | 0 | 1.2100 | 0.0899 | 1.1000 | 1.1900 | 1.3600 |
|  | \* | 1 | 0 | 1.3600 | \* | 1.3600 | 1.3600 | 1.3600 |

1. Then construct side by side box plots of the variables alkphos, cammol, and phosmmol with the factor variable as sex. Then construct side by side box plots of the alkphos, cammol, and phosmmol continuous variables with the factor variable as lab.





1. Do you believe a significant difference exists in alkphos, cammol, or phosmmol levels with respect to sex? Why or why not? Do you believe a significant difference exists in alkphos, cammol, or phosmmol levels with respect to lab? Why or why not?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **AD statistic** | **p-value** | **Normality (α=0.05)** | **Levene Test** | **p-value** |
| Alkaline Phosphate | 3.176 | <0.005 | Cannot be assumed |  |  |
| Calcium | 0.441 | 0.287 | Can be assumed | 2.00 | 0.159 |
| Inorganic Phosphorous | 0.693 | 0.069 | Can be assumed | 3.20 | 0.075 |

Tests for Each Variable:

Alkaline Phosphate: I decided to use a non-parametric Kruskal-Wallis test because we could not assume the variable’s normality, but the confidence interval for the skew statistic does include 0 (-64.06, 66.66).

Calcium: I decided to use a parametric two-sample t-test for the gender by group with equal variances and a 1-way ANOVA for the lab by variable.

Inorganic Phosphorous: I decided to use a parametric two-sample t-test for the gender by group with equal variances and a 1-way ANOVA for the lab by variable.

Results by Gender:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Test Statistic** | **p-value** | **Decision** |
| Alkaline Phosphate | H-value = 6.22 | 0.013 | Reject H0 |
| Calcium | t-value = -3.84 | 0.00017 | Reject H0 |
| Inorganic Phosphorous | t-value = -3.45 | 0.001 | Reject H0 |

I believe that there is sufficient evidence to conclude that there is a significant difference between males and females regarding his/her alkaline phosphate, calcium, and inorganic phosphorous levels. Women have a median of 8 (IU/L) alkaline phosphate more than men, 0.0757 (mmol/L) calcium more than men on average, and 0.0893 (mmol/L) inorganic phosphorous more than men on average.

Results by Lab:

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Test Statistic** | **p-value** | **Decision** |
| Alkaline Phosphate | H-value = 18.88 | 0.002 | Reject H0 |
| Calcium | t-value = -3.84 | 0.00017 | Reject H0 |
| Inorganic Phosphorous | t-value = -3.45 | 0.001 | Reject H0 |

There is sufficient evidence to conclude, at the 5% level of significance, that a significant difference exists in alkaline phosphate, calcium, and inorganic phosphorous levels between the six labs.

1. Conduct a blocked ANOVA analysis using laboratory as the blocking variable for each of the three variables, separately, and gender as the treatment variable. Explain the results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **SEX F-stat** | **SEX p-value** | **LAB F-stat** | **LAB p-value** |
| AlkPhos | 12.17 | 0.001 | 4.73 | <0.001 |
| CamMol | 9.81 | 0.002 | 4.67 | <0.001 |
| PhosMol | 16.26 | <0.001 | 2.59 | 0.027 |

The randomized block design takes account of known factors that affect outcome/response but are not of primary interest. In this case, which lab the sample came from is the blocking factor, but we are testing whether gender differences exist, taking into account the lab. At the 5% level of significance, there is sufficient evidence to conclude that a gender difference exists between males and females regarding his/her alkaline phosphate levels, calcium levels, and inorganic phosphorous levels. In particular, women, on average, have higher levels of alkaline phosphate (95.8, 79.80), higher levels of calcium (2.39, 2.33), and higher levels of inorganic phosphorous (1.14, 1.03) than men, accounting for the lab from which the tests came from.

1. Suppose Mr. and Mrs. Contrarian are married, and Mrs. Contrarian has lower calcium than Mr. Contrarian. She refuses to believe the results of the study that men tend to have lower calcium than women because she has lower calcium than her husband. Using your results to question #3, explain to Mrs. Contrarian the flaw in her thinking.

While Mrs. Contrarian’s individual case is a valid result, the overall conclusion to a more general population is the one that holds. Looking at the boxplot from question 3, it is evident that women tend to have a higher level of calcium (mmol/L) than men do. Mrs. Contrarian is confusing the idea that individual results, which can lead to high bias, are more valid than randomized samples from the more general population and seems to be under the observer bias.

1. One of the objectives of this research was to propose a reference range of values that are to be considered “normal” for calcium, inorganic phosphorus, and alkaline phosphatase. Looking at the results for cammol alone for each of the labs, explain why a single reference range is so difficult to establish.

It is difficult to establish a reference range for each of the measurements because a difference exists between not only each of the two genders, but also between each lab. This will lead to variation in the confidence interval used to compute a range of plausible values, since an interval would need to be created for each gender and for each lab, which would ultimately lead to less power. This is because each interval would decrease the power of the test itself, and would require a correction for each interval, by the formula given by the Bonferroni correction (α/n).