Quantitative Methods II
Assignment 2
September 4, 2011
Solutions

PROBLEM: VARIOUS QUESTIONS [[5]]

Please answer the following questions in complete sentences. No code is needed for this section. You do need, however, to cite your source(s).

(1) What is the most important assumption of the analysis of variance test?

The most important assumption of the analysis of variance test is the assumption of Normality (Forsberg 2011: §3.2.1).

(2) This assumption is also required in the t-test; however, why is it more important in the analysis of variance test?

The reason that the assumption of Normality is more important in the ANOVA procedure than it is in the t-test is that the ANOVA procedure is essentially repeated t-tests (Forsberg 2011: §3.2.1).

(3) What are three things you can do if the main assumption of the analysis of variance test is unreasonable in the data?

If the assumption of Normality is unreasonable for the data (and model), there are three options. The first option is to transform the dependent variable so that the new variable is more Normal. The second option is to use Monte Carlo techniques to fit the actual distribution of the data (if known). The third option is to use non-parametric tests. One should avoid non-parametric tests as they are less powerful

than parametric tests; however, if transformations do not work, and if one cannot determine the distribution of the dependent variable, then non-parametric methods are acceptable (Forsberg 2011: §3.3).

Problem: Predicting Nationalized Petroleum [[10]]

Campbell Faulkner, a student, asks "What factors influence the government's decision to nationalize its petroleum resources?" One factor that his literature review suggested was the quality of the State. From this, Faulker hypothesized that States which nationalized their petroleum production tended to have *lower* quality levels, as measured by the Failed State Index (fsi).

In other words, let us define μ_N as the average fsi measure for states with nationalized petroleum production and μ_P as the average fsi measure for states with privatized petroleum production. Our null hypothesis is

$$H_0 : \mu_N \geq \mu_P$$

A boxplot of this data (Figure **??**) suggests that there is no significant difference in fsi between the two groups. This boxplot also suggests that the variances betwen these two groups are not equal.

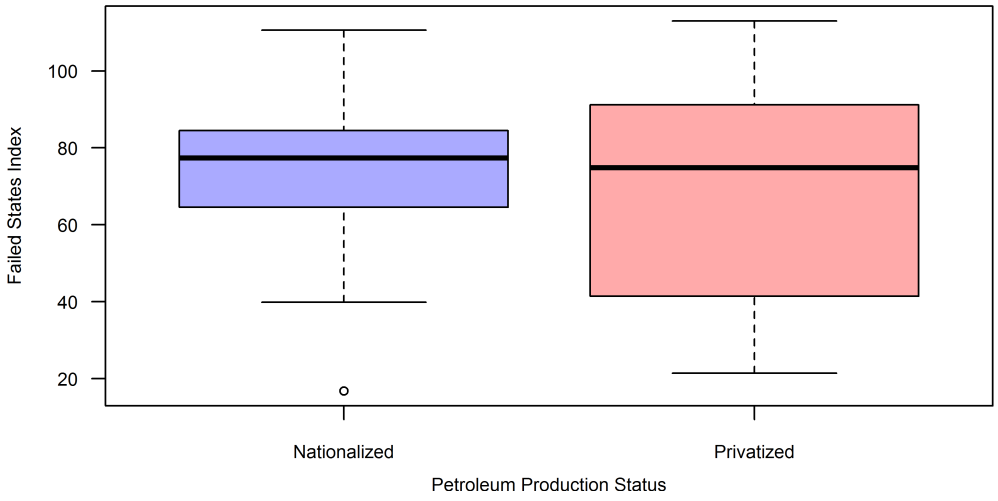

Figure 1. *A boxplot of the petroleum data. The height corresponds to the Failed States Index for the States within each category.*

As we are testing the equality of two means, we would prefer to use the parametric t-test. However, we need to test two assumptions: Normality and equal-variance.

The boxplot (Figure **??**) suggests that the assumption of Normality may not be reasonable for the Nationalized States group (there is an outlier). However, the Shapiro-Wilk test indicates that the Normality assumption is reasonable for this data, when performing the Bonferonni adjustment (minimum adjusted $p = 0.08236$). Additionally, the Bartlett test indicates that the assumption of equal variances is also reasonable ($K^2 = 2.8295; df = 1; p = 0.09255$).

Because the data (and model) did not violate either assumption, we can use the equal-variance, two independent sample t-test. According to this test, we cannot reject the null hypothesis at the $\alpha = 0.05$ level. Thus, we conclude that the Failed States Index for nationalized States is not less than that for privatized States ($t = 0.5736; df = 56; p = 0.7157$).

Problem: Biomes and the Mean Fire Return Interval [[10]]

Lindsey Rao, a student, asks "Does the biome type affect the mean fire return interval?" Rao believes, and the literature review supports her, that the mean fire return interval does vary across the five biomes. Thus, we test the null hypothesis of

$$H_0 : \mu_{STR} = \mu_{TBF} = \mu_{TCF} = \mu_{TS} = \mu_{XER}$$

A boxplot of the data (Figure **??**) suggests that the mean fire return intervals *do* differ across the five biomes. It also suggests that the variance across the biomes is not constant (a fact relevant to our choice of statistical test).

As we are testing the equality of means across more than two groups, we would like to use the analysis of variance test. However, there are two important assumptions that need testing: Normality and equal-variance.

The Shapiro-Wilk test indicates that the measurements in each of the groups are reasonably Normally distributed (minimum unadjusted $p = 0.1833$). Unfortunately, Bartlett's

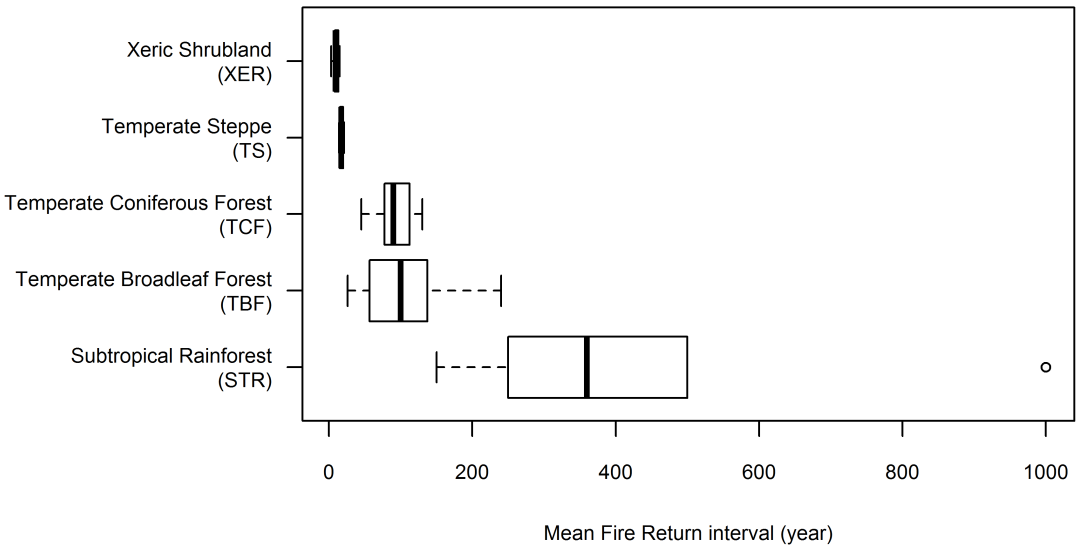

Figure 2. *A horizontal boxplot of the biome data. The horizontal distance corresponds to the mean fire return interval for each of the five biomes.*

test concurs with our initial conclusions based on the boxplots: the variances are not equal across groups ($K^2 = 70.4287; df = 4; p \ll 0.0001$). As such, we cannot use the analysis of variance procedure. Thus, we need to use the Kruskal-Wallis test — a non-parametric test.

According to the Kruskal-Wallis test, the null hypothesis can be safely rejected at the $\alpha = 0.05$ level. Thus, we conclude that the mean fire return interval differs across these five biomes ($\chi^2 = 25.0651; df = 4; p \ll 0.0001$).