Hasselt University

Master of Statistics

Computer Intensive Methods

Assignment II

Students:
Christine Jani (1747124)
Joe Gwatsvaira (1747457)
Bache Emmanuel Bache
(1747695)
Adama Kazienga (1747603)

Lecturers:
Prof.Ziv Shkedy

January 20, 2019

Question 1

(a) i **Non-parametric bootstrap**

Three estimators of the mean real per capita disposable income are considered, which are the Sample mean ($\hat{\theta}_1$), median ($\hat{\theta}_2$) and trimmed mean ($\hat{\theta}_3$); comparing these estimates using the bias, variance and mean squared error(MSE). We do not know the distribution F_n from which the data was drawn, we estimate it via the sample i.e Empirical distribution (F_n) by taking bootstrap samples of size n (random samples with replacement) from $\hat{F}_n$ i.e non-parametric bootstrap. To obtain a bootstrap estimate of the MSE of $\hat{\theta}_i, \{i = 1, 2, 3\}$ we take a large number, B=1000, of bootstrap samples from $\hat{F}_n$ (taking random samples with replacement from the observed samples). For each of these samples we calculate the estimator $\hat{\theta}^*_{bi}$, $b = 1, \ldots, B$ and $\{i = 1, 2, 3\}$; then proceed to calculate the bootstrap estimates of MSE, bias and variance as:

1. $M\hat{S}E(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}^*_b - \hat{\theta})^2$

2. $\hat{Var}(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\theta}^*_b - \bar{\hat{\theta}}^*)^2$ where $\bar{\hat{\theta}}^* = \frac{1}{B} \sum_{b=1}^{B} \theta^*_b$

3. $B\hat{ias}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}^*_b - \hat{\theta})$

The results shown on the Table 1 indicate that the preferred estimator is the one with the best compromise between the bias and variance; that is the one with a smallest mean squared error (MSE) implying $\hat{\theta}_1$ is the best estimator.

	Estimates			
Estimator	**Estimate**	**Bias**	**Variance**	**MSE**
$\hat{\theta}_1$	1106.76	-7.17	18845.90	18897.30
$\hat{\theta}_2$	695.67	17.46	31804.56	32109.39
$\hat{\theta}_3$	879.99	2.82	29547.45	29555.40

Table 1: Bootstrap Estimates of Bias, Variance and MSE

(a) ii **Jacknife Method**

There are some variants of bootstrap such as the Jackknife (leave one observation out each time) and sub sampling (only sub sample $n - 1$ out of n samples). In other words the Jacknife is based on the amount of information each unique observed value contain about the per capita disposable income. The algorithm for the Jacknife: Remove the jth observation where $j = \{j = 1, ..., n\}$, compute the estimates of interest $\hat{\theta}(j)$ as described in [(a)i]. $\hat{\theta}_{(j)}$ is computed on all observations excluding j^{th} observation. This is called the Jackknife estimate. Repeat for each $j = 1, ..., n$, and find the mean $\hat{\theta}_{(.)}$ of the n values of $\hat{\theta}_{(j)}$

The Jackknife estimate of bias, variance and Mean Squared Error are defined as

1. $B\hat{ias}(\theta)_{jack} = (n - 1) \left(\sum_{j=1}^{n} \frac{\hat{\theta}_{(j)}}{n} - \hat{\theta} \right)$

2. $\hat{Var}(\theta)_{jack} = \frac{n-1}{n} \sum_{j=1}^{n} (\hat{\theta}_{(j)} - \hat{\theta}_{(.)})^2$

3. $M\hat{S}E(\theta)_{jack} = \hat{Var}(\theta) + (\hat{bias}(\theta)_{jack})^2$

	Estimates			
Estimator	Estimate	Bias	Variance	MSE
$\hat{\theta}_1$	1106.758	0	19636.42	19636.42
$\hat{\theta}_2$	695.665	0	52732.23	52732.23
$\hat{\theta}_3$	898.1566	889.8358	30826.72	822634.4

Table 2: Jacknife Estimates of Bias, Variance and MSE

The sample mean $\hat{\theta}_1$ is chosen as the best estimator of the mean per capita disposable income which is consistent with the findings presented in [(a)i].

(b) **Bootstrap Hypothesis testing**

The basic idea of hypothesis testing is to compare the observed value of a test statistic, say $\hat{\theta}$, with the distribution that follows null hypothesis if it was true. The null is then rejected if $\hat{\theta}$ is sufficiently extreme relative to this distribution. However, the distribution of the test statistic we use is not known. We therefore have to compare $\hat{\theta}$ with a distribution that is only approximately correct. We generate a large number of simulated values of the test statistic and compare $\hat{\theta}$ with the empirical distribution function. We observe a random sample $\mathbf{X} = \{x_1, ..., x_n\}$ and we wish to test whether the mean of the population equals to some predetermined value θ_0 i.e $H_0 : \theta = \theta_0$. The test is based on the approximate distribution of the test statistic $t(x)^{obs} = \frac{\theta - \theta_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$. The empirical distribution $\hat{F}$ is not an appropriate estimation, because it does not obey H_0 (sampling should be done under the null). We can use the empirical distribution of the points: $\tilde{x}_i = x_i - \bar{X} + 700$ $i = 1, ..., 50$, which has mean 700. For the non parametric bootstrap B=1000 are taken from the empirical distribution $\hat{F}(\tilde{x})$. For parametric bootstrap sampling was done from a normal distribution thus we assume that $\tilde{X} \sim N(\bar{\tilde{X}} = 700, S_{\tilde{x}}^2)$ where $\bar{\tilde{X}} = 700$ and $S_{\tilde{x}}^2$ is the plug in estimates of $\sigma_{\tilde{x}}$. We sample B=1000 samples $\{\tilde{x}_1^*, ..., \tilde{x}_n^*\}$ with replacement from $\hat{F}(\tilde{x})$ and $N(= 700, S_{\tilde{x}}^2)$ for non parametric and parametric bootstrap respectively for each sample compute $t(\tilde{x}^*) = \frac{\tilde{\theta} - \theta_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$ compute the Monte Carlo

$$p - value = \frac{1 + (\text{number of } t(\tilde{x}^*) > t(x)^{obs})}{B + 1}$$

The p values are given by 0.00699 and 0.00099 for parametric and non parametric bootstrap hypothesis test respectively. At 5% level of significance, there is sufficient evidence to reject the null hypothesis that $\theta = 700$ in favour of the alternative hypothesis $H_1 > 700$.

The distribution of the statistics Figure 1 under the null are approximately bell shaped and symmetric around zero.
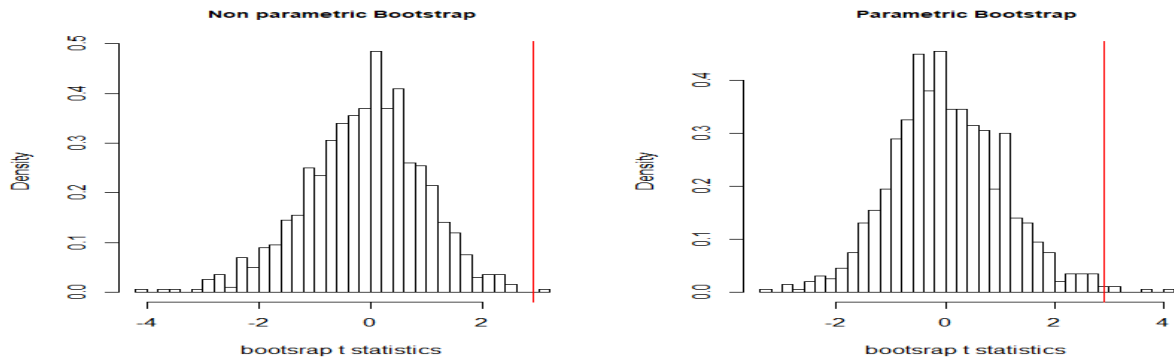
Figure 1: Histogram of bootstrap t statistics

Question 2

We are interested in estimating the confidence intervals for the differences in mean increase in hours of sleep ($\mu_1 - \mu_2$) for two drugs given to patients. The algorithm for the boostrap is: we first calculate the statistic of interest from the original sample i.e $\hat{\theta} = \hat{\mu}_1 - \hat{\mu}_2$. For non parametric bootstrap we take $B = 1000$ resamples of pairs (x_1, x_2) since data are measured twice for each patient from the sample with replacement at each re-sample we calculate $\hat{\theta}_b$

1. **Bootstrap Percentile interval**
   Sort the bootstrap statistics $\hat{\theta}_b$ in rank order and estimate the confidence interval as the 2.5% and 97.5% of the bootstrap statistic distribution. Need "a lot of replications" to get a reliable interval because you're reaching out into the tails of the distribution

2. **Bootstrap t interval**
   We compute the original sample estimate of θ given by $\hat{\theta}$ and the corresponding standard deviation $S.\hat{e}(\theta)$. Then bootstrap samples (X_i^*, Y_i^*) are generated by re-sampling from the original sample and calculate a second stage bootstrap statistic $\hat{\theta}^{**}$. This is repeated to obtain 1000 statistics $\hat{\theta}^{**}$ and find their standard deviation $S.e^{**}$. Repeat these steps B times and calculate $t = \frac{\hat{\theta}_i^{**} - \hat{\theta}}{S.e_i^{(**)}}$ sort the bootstrap t statistic in rank order estimate the 2.5% and 97.5% percentiles of the bootstrap t statistics Estimate the bootstrap confidence Interval as $[\hat{\theta} - t_{2.5\%}S.\hat{e}(\hat{\theta}); \hat{\theta} + t_{97.5\%}S.\hat{e}(\hat{\theta})]$.

3. **BCa Interval** Calculate the statistic of interest $\hat{\theta}$ from the original sample. We obtain the bootstrap estimates $\hat{\theta}^*$ in the same way as the percentile method. We estimate the bias in the $\hat{\theta}^*s$ as the proportion of the $\hat{\theta}^*s$ that exceed the sample estimate $\hat{\theta}$ we denote it $P_b = P(\hat{\theta}^* > \hat{\theta})$. Convert P_b to a standard normal using the inverse quantile normal distribution function thus we have $b = \phi^{-1}(P_b)$. Subtract twice that bias from the standard normal 95% limits thus $L = 1.96 - 2b$ and $U = 1.96 - 2b$. Convert these corrected standard normal limits to the proportions using the normal probability distribution function thus $P_L = \phi(L)$ and $P_U = \phi(U)$.

2

Like the percentile estimates the confidence limits are the corresponding P_L^{th} and P_U^{th} quantiles of the bootstrap statistics. The bootstrap bias-corrected accelerated (BCa) interval is a modification of the percentile method that adjusts the percentiles to correct for bias and skewness. This interval should always be used over the percentile and t intervals if software provides it. The distribution of the bootstrap estimates of the difference in means $\hat{\theta}^*$, Figure 2 is symmetric around the observed difference $\hat{\theta} = -1.58$ The confidence intervals are shown in Table 3. The naive and the percentile confidence intervals are symmetric around $\hat{\theta}$ while the bootstrap t and BCa are asymmetric
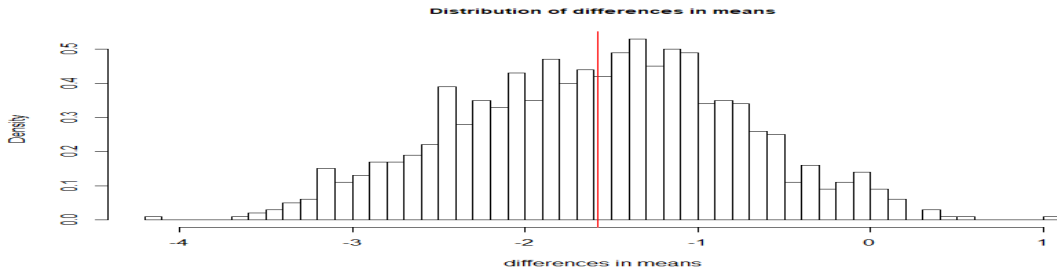

Figure 2: Distribution of bootstrap statistics

95% Confidence Interval			
Naive	t interval	percentile	Bca interval
[-2.459;-0.700]	[-2.204;0.071]	[-2.400;0.-0.959]	[-3.135,0.043]

Table 3: Bootstrap Confidence Intervals

Question 3

(a) The ratio of mean Hamilton depression scores in the first and the second visits i.e the parameter of interest $\hat{\theta} = \frac{\mu_x}{\mu_y}$.

Since the data consist of measurements taken from the same individual we consider taking B=1000 re-samples of pairs (x_i^*, y_i^*) with replacement from the original sample (x_i, y_i). At each bootstrap we calculate the parameter of interest $\hat{\theta}^* = \frac{\hat{\mu_{x^*}}}{\hat{\mu_{y^*}}}$. The classical, percentile, BCa interval and normal confidence intervals are computed using the boot function in R. Table 4 shows the results. All the intervals provided look similar. None of the intervals include 1 and all the lower limits are larger then 1. This means at the first measurement, the patients had a higher depression score than at the second measurement.
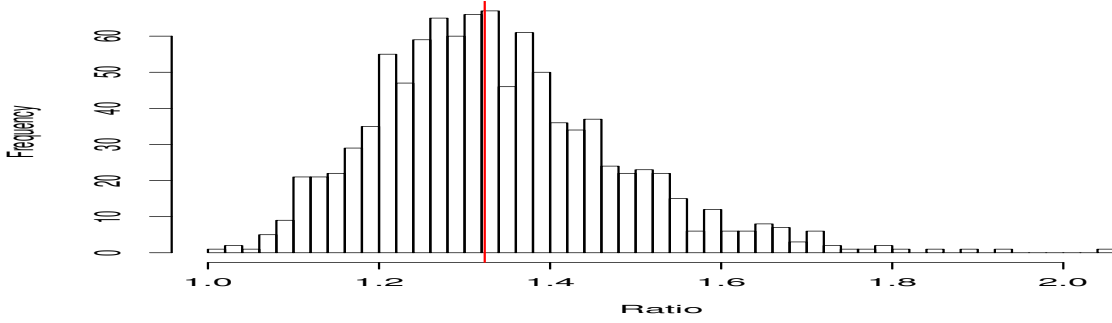
Figure 3: Non parametric bootstrap for Ratio of means

95% Confidence Interval		
Normal	Percentile	BCa
[1.018;1.597]	[1.099;1.690]	[1.095;1.669]

Table 4: Confidence Intervals

(b) The hypothesis to be tested is $H_0 : \hat{\theta} = \frac{\mu_x}{\mu_y} = 1$, which reduces to a test for equality of group means for paired data. However, paired t test makes sense when the difference is consistent. The first measurement values might bounce around, but the difference between first measurement and second measurement is a consistent measure of what happened. With some kinds of data, the difference between first measurement and second measurement is not a consistent measure of effect. Instead, the differences are larger when the first measurement values are larger. In this case, the ratio (first measurement/first measurement) may be a much more consistent way to quantify the effect of the treatment. Analyzing ratios can lead to problems because ratios are intrinsically asymmetric – all decreases are expressed as ratios between zero and one; all increases are expressed as ratios greater than 1. Instead it makes more sense to look at the logarithm of ratios. Then no change is zero (the logarithm of 1), increases are positive and decreases are negative.

A ratio t test averages the logarithm of the ratio of first measurement/first measurement and then tests the null hypothesis that the population mean of that set of logarithms is really zero.

Resampling is done under the null hypothesis lets assume this is a test for equality of two distribution means and combine the samples. We denote the combined samples by $\mathbf{Z}$ then calculate the sample mean $\bar{Z}$ for the combined sample. Define the following transformed variables $\tilde{\mathbf{x}}_i = x_i - \bar{x} + \bar{z}$ and $\tilde{\mathbf{y}}_i = y_i - \bar{y} + \bar{z}$ for $i = 1, 2, 3$. The transformed $\tilde{\mathbf{x}} \ \tilde{\mathbf{y}}$ share a common mean. We perform a parametric bootstrap we sample (x_i*, y_i^*) form a bivariate normal distribution $(X_i, Y_i) \sim N(\boldsymbol{\mu}_{\tilde{x},\tilde{y}}, \Sigma_{\tilde{x},\tilde{y}})$

Now to perform the Ratio test for equality of means we do the following steps:

4

1 Transform all the values to their logarithm: $x_i = log(x_i)$, $y_i = log(y_i)$ for original samples and $\tilde{x}_i = log(\tilde{x}_i)$, $\tilde{y}_i = log(\tilde{y}_i)$ for the transformed x and y

2 Perform a paired t test on the logarithms i.e for the log transformed original samples we calculate $\hat{\mu}_x$, $\hat{\mu}_y$ and S_x and S_y and calculate the t.test of interest assuming unequal sample variances $t^{obs} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$. Draw B bootstrap samples with replacement $(\tilde{x}^*, \tilde{y}^*)$ from $\{\tilde{x}_1, ..., \tilde{x}_3\}$ and $\{\tilde{y}_1, ..., \tilde{y}_3\}$ (log transformed). and evaluate the statistic t^* for each sample $t_b^* = \frac{\bar{\tilde{X}}^* - \bar{\tilde{Y}}^*}{\sqrt{\frac{\tilde{S}_1^{2*}}{n_1} + \frac{\tilde{S}_2^{2*}}{n_2}}}$

This test is done using the ***ttestratio()*** function in R package ***mratios***.

Monte carlo p-value is given by;

$$p = \frac{1 + (\text{number of } t^* > tobs)}{B + 1} = 0.012$$

Reject H_0 and conclude there is sufficient evidence at 5% level of significance to say that the ratio of the means is significantly different from 1. i.e the mean Hamilton depression scale factor is different at the two visits.

Figure 4 shows the distribution of bootstrap test statistics which is symmetric around zero and the observed t statistic is shown by the red line.
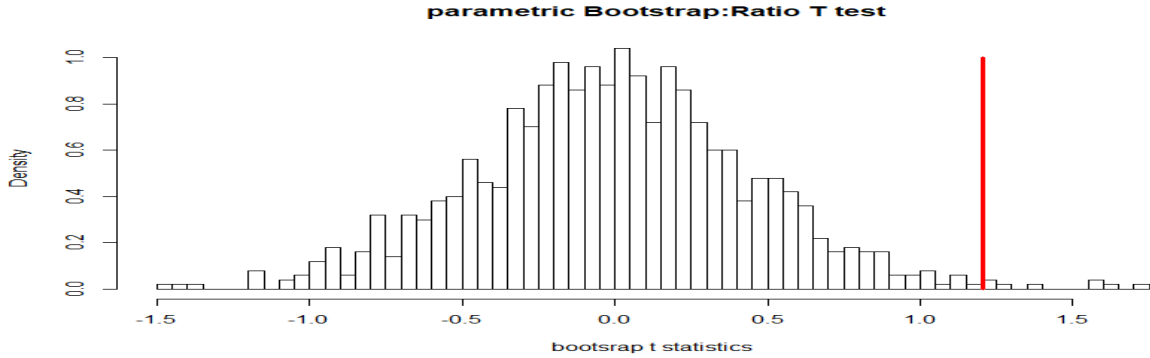

Figure 4: Parametric bootstrap t statistics for Ratio T test

Question 4

(a) We wish to test if one of the two treatments reduce the number of depression events for patient randomized in the two groups. This is achieved through testing the hypotheses $H_0 : \mu_A = \mu_B$ versus $H_0 : \mu_A < \mu_B$. We test the null hypothesis assuming unequal variances. Since resampling is done under the null hypothesis lets assume this is a test for equality of two distribution means and combine the

samples. We denote the combined samples by $\mathbf{Z}$ and the corresponding empirical distribution by $\hat{F}_0$. Under the null $\hat{F}_0$ provides a non parametric estimate for the common population that gives rise to both $\mathbf{X}$ and $\mathbf{Y}$. We calculate the sample mean $\bar{Z}$ for the combined sample. We define the following transformed variables $\tilde{\mathbf{x}}_i = x_i - \bar{x} + \bar{z}$ and $\tilde{\mathbf{y}}_i = y_i - \bar{y} + \bar{z}$ for $i = 1, 2, 3$. The empirical distributions of $\tilde{\mathbf{x}}$ $\tilde{\mathbf{y}}$ share a common mean. For parametric bootstrap we sample x^* from $F_0 = Po(\bar{\tilde{X}})$ and For parametric bootstrap we sample y^* from $F_0 = Po(\bar{\tilde{Y}})$ since our data is on counts (weekly number of depression events) For the original samples we calculate $\hat{\mu}_A$, $\hat{\mu}_B$ and S_A and S_B and calculate the t.test of interest assuming unequal sample variances $t^{obs} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$. We draw R=1000 bootstrap samples with replacement $(\tilde{x}^*, \tilde{y}^*)$ from $\{\tilde{x}_1, ..., \tilde{x}_3\}$ and $\{\tilde{y}_1, ..., \tilde{y}_3\}$. and evaluate the statistic t^* for each sample $t_b^* = \frac{\bar{X}^* - \bar{Y}^*}{\sqrt{\frac{S_1^{2*}}{n_1} + \frac{S_2^{2*}}{n_2}}}$

Monte carlo p-value is given by; $p = \frac{1 + (\text{number of } t^* > tobs)}{B + 1}$

Permutation

Permutation provides an efficient approach to testing when the data do not conform to the distributional assumptions of the statistical method one wants to use (e.g. normality). Permutation testing is applicable to very small samples, like non-parametric tests. Consider sample of m observations from treatment 1 and n observations from treatment 2. Assume that under the null hypothesis there is no difference between the effect of treatment 1 and treatment 2. Then any permutation of the observations between the two treatments has the same chance to occur as any other permutation. The steps for a two-treatment permutation test: Compute the difference between the mean of observed data, called it D_{obs}. Create a vector of $n_1 + n_2$ observations. Select at random experimental units to one of the two treatments with n_1 units assigned to treatment 1 and n_1 units assigned to treatment 2. Permute the $n_1 + n_2$ observations between the two treatments so that there are n_1 observations for treatment 1 and n_2 observations for treatment 2. Number of possibilities $\frac{(n_1 + n_2)!}{n_1! n_2!}$. For each permutation of the data, calculate the difference between the mean of treatment 1 and mean of treatment 2, called it, D_p. For the lower tailed test, compute p-value as the proportion of D_p less than or equal D_{obs} i.e

$$p - value = \frac{number\ of\ D_p \leq D_{obs}}{\frac{(n_1 + n_2)!}{n_1! n_2!}}$$

Conclusion We note that the conclusions drawn after doing same tests on samples of size $n_1 = 3$ and $n_2 = 10$ are different. P values in Table 5 show that the classical, Bootstrap (parametric and non parametric) and permutation tests reject the null hypothesis when $n = 3$. However the power of the test is increased when the sample size is increased to $n = 10$. The Histograms of the test statistics Figure 5 when $n = 3$ the boostrap tests does not approximate well the distribution of the test

statistic but the permutation test performs it better. Increasing the sample size to $n = 10$ gives a better approximation of the test statistic under the null hypothesis, however we do notget a significant p-value.

	P-values		
Sample size per group	Non Parametric	Parametric	Permutation
3	0.002	0.005	0.001
10	0.131	0.109	0.0869

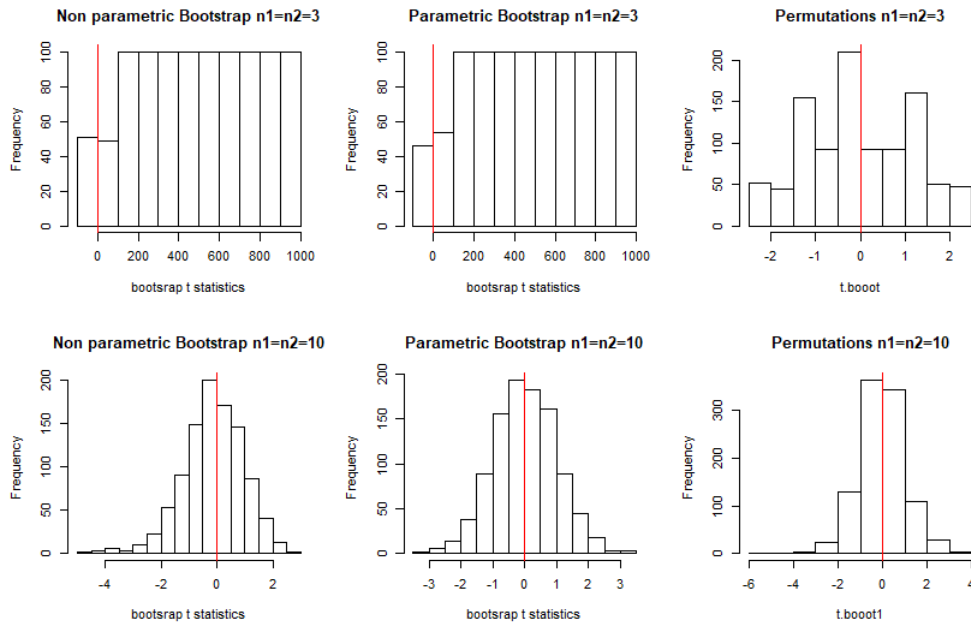Table 5: P values for different tests

Figure 5: Bootstrap and permutation statistics

References

[Gemechis Djira, 2018] Gemechis Djira, Mario Hasler, D. G. F. S. (2018). Ratios of coefficients in the general linear model. https://cran.r-project.org/web/packages/mratios/mratios.pdf. [Online; accessed 10-January-2019].

[Software, 2017] Software, G. (2017). Paired vs. ratio t tests. https://www.graphpad.com/guides/prism/7/statistics/stat$_paired_or_ratio_{tt}est.htm?toc = 0printWindow/. [Online; accessed 15 - January - 2019].$