Using Bootstrap to Verify Normal Assumptions in Statistical Inference for Treatment Difference

Ruji Yao, Amarjot Kaur, Qing Li, Anjela Tzontcheva

Merck & Co., Inc., Rahway, NJ, USA

In a clinical trial, it is common to use an ANOVA model to compare treatment effects between an active and placebo group, while adjusting for interesting covariates. The estimated means and variances are based on the underlying data and the model settings, but for statistical inference, the distribution of estimators is also needed. In most cases, in order to get the critical value for statistical inference, we often assume that the residuals of the model fitting are normally distributed or claim that the sample sizes are large enough to apply the central limit theorem. In practice, in order to check these assumptions, we often use normality tests, such as the Kolmogorov-Smirnov test [1]; however, somewhat crucially, this test does not provide direct information on the actual distribution of the estimators. In this presentation, we introduce a new method for normality verification which is straight forward and can be used to decide whether a data transformation is necessary. In particular, we first utilize Bootstrapping to construct an empirical bivariate distribution on the estimated least squares (LS) means of data from active and placebo groups [2]. We then compare this with the bivariate distribution using the same LS means from an ANOVA model with a normality assumption.

Introduction and Idea

In a clinical trial, it is common to use an ANOVA model to compare treatment effect. LS mean estimators from ANOVA gave us the treatment effect and with bivariate normal assumptions on the estimators we can perform statistical inference on treatment effect. But how can we be sure the assumption is correct or accurate enough? When we are worried about this normal assumption on the estimators for a very skewed data, a data transformation, like square root or long-normal transformation, often used [3]. Sometimes, clinicals are also interested in a function of LS means and a bootstrap often used for statistical inference for such a function of estimated LS means if there is no analytic formula for its distribution. During the bootstrapping, an empirical bivariate distribution is automatically created and then used for different statistical inferences. If number of bootstrap runs is big, say 10,000, this empirical bivariate distribution could be very close to the actual distribution of estimated LS means and therefore could be used to verify the normality assumption.

Issue and Examples

In a clinical trial, a variable DMS, rescue medication, was used, which was known as very skewed with a lot of zeros. In the statistical analysis plan, it was pre-specified that if there were more than 30% zeros, the zero-inflated log-normal model would be used, instead of an ANOVA model since normal assumptions might be violated for statistical inference. It was a very reasonable choice since "log-normal" would take care of positive skewed data and "zero-inflated" would use the information within all zeros and avoid to set up an arbitrary small value for zeros in log-transformation. But this arbitrary choice might have some impact on the comparison of treatment effect. Since the actual DMS data had more than 50% of zeros, zero-inflated log-normal model was used. But as we all know that it is hard to interpret the treatment difference with this model.

The question we had is how much violation of normal assumptions on these 2 estimated LS means since there is no perfectly normal distributed data in real world?

In table 1 below, we showed how to verify normal assumption and it seems that statistical inference with ANOVA model should be acceptable.

Table 1. Using Bootstrap to verify normality for results of ANOVA on DMS

Variable	treatment	method	median	mean	std	95% CI
DMS	active	ANOVA model		1.83	0.162	(1.51, 2.15)
		Bootstrap	1.83	1.83	0.151	(1.54, 2.13)
		model/bootstrap		1.00	1.07	(0.98, 1.01)
		Bootstrap	mean +/- 1.96*std			(1.53, 2.13)
	placebo	ANOVA model		3.17	0.163	(2.85, 3.49)
		Bootstrap	3.17	3.17	0.183	(2.82, 3.53)
		model/bootstrap		1.00	0.89	(1.01, 0.99)
		Bootstrap	mean +/- 1.96*std			(2.81, 3.53)
	difference	ANOVA model		-1.34	0.215	(-1.76, -0.92)
		Bootstrap	-1.34	-1.34	0.217	(-1.77, -0.92)
		model/bootstrap		1.00	0.99	(0.99, 1.00)
		Bootstrap	mean +/- 1.96*std			(-1.77, -0.91)

Model/bootstrap is to show how close of the ratio of 2 values around 1.

Analysis of variance (ANOVA) model was used, which included fixed effects of treatment, baseline asthma status (yes, no), age group (<12 years, ≥ 12 years), pollen season, and pollen region nested within pollen season.

Bootstrap used the same model with 10,000 runs and median, mean, std are from SAS PROC UNIVARIATE, and 95%CI is (250^{th} , 9750^{th}), based on 10,000 sorted bootstrap results.

Note for bivariate normal distribution, the estimated correlation coefficient corr from the ANOVA model is 0.123, which can be derived from above table

0.215^2 = 0.162^2 – 2*corr*0.162*0.163 + 0.163^2.

Method to verify normal assumption:

1. median and mean of bootstrap should be equal or very close (symmetry).

2. Compare mean, std and 95% CI between Bootstrap and ANOVA model

3. Exam how close the ratio of 2 methods equal to 1.

4. From Bootstrap data, compare 95%CIs between mean +/- 1.96*std and (250^{th} , 9750^{th})

In clinical trial, statistical method for analysis is preferred to be decided in details before database is locked or a clear criteria is set up before database is locked. The example below used the same DMS data as in previous example, but it was before database is locked, i.e. the treatment code is dummy.

The results are in Table 2 below and very similar in term of normality verification, compared with Table 1. So we might decide that log-transformation is not necessary even before database is locked.

Table 2. Using Bootstrap to verify normality for results of ANOVA on DMS

(use DUMMY treatment code for data before DBL)

Variable	treatment	method	median	mean	std	95% CI
DMS	active	ANOVA model		2.43	0.165	(2.11, 2.75)
		Bootstrap	2.43	2.43	0.160	(2.12, 2.74)
		model/bootstrap		1.00	1.03	(1.00,1.00)
		Bootstrap	mean +/- 1.96*std			(2.12, 2.74)
	placebo	ANOVA model		2.56	0.166	(2.23, 2.88)
		Bootstrap	2.56	2.56	0.179	(2.22, 2.92)
		model/bootstrap		1.00	0.93	(1.00 , 0.99)
		Bootstrap	mean +/- 1.96*std			(2.21, 2.91)
	difference	ANOVA model		-0.128	0.218	(-0.556, 0.301)
		Bootstrap	-0.130	-0.129	0.218	(-0.553, 0.292)
		model/bootstrap		0.99	1.00	(1.01,1.03)
		Bootstrap	mean +/- 1.96*std			(-0.556, 0.298)

Model/bootstrap is to show how close of the ratio of 2 values around 1.

Analysis of variance (ANOVA) model was used, which included fixed effects of treatment, baseline asthma status (yes, no), age group (<12 years, ≥ 12 years), pollen season, and pollen region nested within pollen season.

Bootstrap used the same model with 10,000 runs and median, mean, std are from SAS PROC UNIVARIATE, and 95%CI is (250^{th} , 9750^{th}), based on 10,000 sorted bootstrap results.

We all know that we can used Kolmogorov-Smirnov test to test the normality of residuals if we have residuals [1]. But sometimes, there is no residuals at all from model fitting, like the logistic regression model below. The fitting is on the logit data and 95% CI is also based on bivariate normal assumptions of the estimated LS mean in logit and there is no residuals in logit.

The results in Table 3 below showed that we can use Bootstrap method to verify the normality assumption even there is no fitting residuals [4].

Table 3. Using Bootstrap to verify normality for results of Logistic model on DMS

(set DMS = 1 if DMS > 0)

Variable	treatment	method	median	mean	std	95% CI
Modified DMS.	active	Logistic model		0.337	0.0966	(0.148, 0.527)
		Bootstrap	0.340	0.341	0.0981	(0.146, 0.538)
		model/bootstrap		0.99	0.98	(1.01, 0.98)
		Bootstrap	mean +/- 1.96*std			(0.149, 0.533)
	placebo	Logistic model		-0.311	0.0944	(-0.496, -0.126)
		Bootstrap	-0.313	-0.314	0.0963	(-0.503, -0.128)
		model/bootstrap		0.99	0.98	(0.99, 0.98)
		Bootstrap	mean +/- 1.96*std			(-0.503, -0.125)
	difference	Logistic model		0.648	0.133	(0.388, 0.909)
		Bootstrap	0.654	0.654	0.135	(0.387, 0.922)
		model/bootstrap		0.99	0.99	(1.00, 0.99)
		Bootstrap	mean +/- 1.96*std			(0.389, 0.919)

Model/bootstrap is to show how close of the ratio of 2 values around 1.

Modified DMS is derived from the original DMS as to set response =1 if response > 1.

Logistic Model was used, which included fixed effects of treatment, baseline asthma status (yes, no), age group (<12 years, ≥ 12 years), pollen season, and pollen region nested within pollen season.

Bootstrap used the same model with 10,000 runs and median, mean, std are from SAS PROC UNIVARIATE, and 95%CI is (250^{th} , 9750^{th}), based on 10,000 sorted bootstrap results.

Conclusion:

We proposed a method for normality verification based on the bootstrap. It is non-parametric and easy to understand and can be used even without model fitting residuals. When facing very skewed data, statistical inference, based on normal assumption, could be challenged and the analysis based on transformation may be necessary. But this method gives us a chance to verify the normality assumptions directly with an empirical distribution of estimators from bootstrap as benchmark [5] and a more intuitive and confident decision would be available, even before database is locked.

References:

[1] Lilliefors, Hubert W. "On the Kolmogorov-Smirnov test for normality with mean and variance unknown." Journal of the American statistical Association 62.318 (1967): 399-402.

[2] Bickel, Peter J., and David A. Freedman. "Asymptotic normality and the bootstrap in stratified sampling." The annals of statistics (1984): 4

[3] Beran, Rudolf, and Muni S. Srivastava. "Bootstrap tests and confidence regions for functions of a covariance matrix." The Annals of Statistics 13.1 (1985): 95-115.

[4] Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato. "Central limit theorems and bootstrap in high dimensions." The Annals of Probability 45.4 (2017): 2309-2352.

[5] Mammen, Enno. When does bootstrap work?: asymptotic results and simulations. Vol. 77. Springer Science & Business Media, 2012.