

# Package ‘ZINQ’

August 8, 2020

**Type** Package

**Title** Zero-Inflated Quantile Approach for Microbiome Association Testing

**Version** 1.0

**Author** Wodan Ling and Michael C. Wu

**Maintainer** Wodan Ling <wling@fredhutch.org>

**Description** This package conducts univariate association test between microbiome data (un-normalized or normalized by any method) and clinical variables (dicrete or quantitative), supporting both unadjusted and adjusted tests.

**License** GPL (>=2)

**Encoding** UTF-8

**LazyData** true

**NeedsCompilation** no

**Imports** quantreg, MASS

**RoxygenNote** 7.1.1

**Suggests** knitr,  
rmarkdown

**VignetteBuilder** knitr

## R topics documented:

Sample_Data . . . . .	1
ZINQ_combination . . . . .	2
ZINQ_tests . . . . .	3
<b>Index</b>	<b>6</b>

---

Sample_Data	<i>Example data, normalized abundance of two taxa and covariates</i>
-------------	--

---

## Description

A dataset containing two taxa which have typical abundance profiles highlighting power of ZINQ, including rarefied abundance of taxon 1, CSS normalized abundance of taxon 2, the clinical variable of interest, and several covariates.

**Usage**

Sample\_Data

**Format**

A data frame with 531 rows and 6 variables:

**rarefied\_taxon1** rarefied abundance of taxon 1  
**CSS\_taxon2** CSS normalized abundance of taxon 2  
**X** the clinical variable of interest, binary variable  
**Z1** covariate 1, continuous variable  
**Z2** covariate 2, continuous variable  
**Z3** covariate 3, continuous variable

---

ZINQ_combination	<i>Combine the marginal p-values</i>
------------------	--------------------------------------

---

**Description**

Combine the marginal p-values

**Usage**

```
ZINQ_combination(
  input,
  method = "MinP",
  taus = c(0.1, 0.25, 0.5, 0.75, 0.9),
  M = 10000
)
```

**Arguments**

input	An output from <a href="#">ZINQ_tests</a> .
method	Combination method, "MinP" for MinP test, "Cauchy" for Cauchy combination test; default is "MinP".
taus	A grid of quantile levels, must be a subset or equal to that from input; default is c(0.1, 0.25, 0.5, 0.75, 0.9).
M	The number of MC draws from the joint distribution of quantile rank-scores when method is "MinP"; default is 10000.

**Details**

- Please choose 'MinP' or 'Cauchy' for method, no other options.
- taus must be a subset or equal to the grid used to produce input.

**Value**

A pvalue, the final p-value of ZINQ

## References

- Ling, W. et al. (2020+). Powerful and robust non-parametric association testing for microbiome data via a zero-inflated quantile approach (ZINQ).
- He, Z. et al. (2017). Unified sequence-based association tests allowing for multiple functional annotations and meta-analysis of noncoding variation in metabochip data. *The American Journal of Human Genetics* 101(3), 340–352.
- Lee, S. et al. (2012). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 13(4), 762–775.
- Liu, Y., Xie, J. (2019). Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, 1–18

## Examples

```
library(quantreg)
library(MASS)
n = 300
p <- function(x0, gam0=0.75, gam1=-0.15){
  lc = gam0 + gam1*x0
  exp(lc) / (1 + exp(lc))
}
x = c(rep(0, n), rep(1, n))
w = 0.5 + 1.5*x + (1+0.15*x)*rchisq(2*n,df=1)
b = rbinom(2*n, 1, p(x))
y = w*b
dat = data.frame(y, x)

result = ZINQ_tests(formula.logistic=y~x, formula.quantile=y~x, C="x", data=dat)
ZINQ_combination(result, method="Cauchy")
```

---

ZINQ\_tests

---

*Marginal tests for the logistic and quantile regression components*


---

## Description

Marginal tests for the logistic and quantile regression components

## Usage

```
ZINQ_tests(
  formula.logistic,
  formula.quantile,
  C,
  y_CorD = "C",
  data,
  taus = c(0.1, 0.25, 0.5, 0.75, 0.9),
  seed = 2020
)
```

## Arguments

<code>formula.logistic</code>	The full model of logistic regression, e.g., $Y \sim X + Y + Z$ , where $Y$ is zero-inflated.
<code>formula.quantile</code>	The full model of quantile regression, can be different from <code>formula.logistic</code> .
<code>C</code>	The name(s) of clinical variable(s) of interest, e.g., "Condition" or <code>c("Condition", "Batch")</code> .
<code>y_CorD</code>	An indicator: use "D" if $Y$ is count, a perturbation from $U(0, 1)$ will be added to the response; use "C" if $Y$ is continuous; default is "C".
<code>data</code>	A data.frame: better cleaned and processed, use numeric for $Y$ and binary covariates, use factor for multi-class discrete covariates.
<code>taus</code>	A grid of quantile levels, e.g., 0.5 for the median, 0.75 for the 3rd quartile; default is <code>c(0.1, 0.25, 0.5, 0.75, 0.9)</code> .
<code>seed</code>	A seed for perturbation when <code>y_CorD</code> is "D"; default is 2020.

## Details

- Compositional data is regarded as continuous, determined by its support.
- `taus` is a tuning parameter that does not have an efficient selection process yet, try from coarsened to fine grids (e.g., `seq(0.1, 0.9, by=0.2)` to `seq(0.1, 0.9, by=0.1)`), or try adding more extreme levels (e.g., `c(0.25, 0.5, 0.75)` to `c(0.1, 0.25, 0.5, 0.75, 0.9)`), with a goal to keep type I error controlled and boost the power; for common taxa, start from the default; for rare taxa, start from `c(0.25, 0.5, 0.75)`.
- Quantile rank-score test corrected for zero-inflation is used for the quantile regression component.
- If `C` is a single continuous or binary covariate, Wald test is used for the logistic regression component, else Rao's score test is used

## Value

A list

- `pvalue.logistic` - A single p-value from the logistic regression component.
- `pvalue.quantile` - A `length(taus)` by 1 vector, a sequence of p-values from the quantile regression component.
- `Sigma.hat` - A `df x length(taus)` by `df x length(taus)` matrix, where `df` is the dimension of `C`, the covariance matrix of quantile rank-scores.
- `zerorate` - The proportion of zeroes in  $Y$ .
- `taus` - The grid of quantile levels used.

## References

- Ling, W. et al. (2020+). Powerful and robust non-parametric association testing for microbiome data via a zero-inflated quantile approach (ZINQ)
- Machado, J.A.F., Silva, J.S. (2005). Quantiles for counts. *Journal of the American Statistical Association* 100(472), 1226–1237.

**Examples**

```
library(quantreg)
n = 300
p <- function(x0, gam0=0.75, gam1=-0.15){
  lc = gam0 + gam1*x0
  exp(lc) / (1 + exp(lc))
}
x = c(rep(0, n), rep(1, n))
w = 0.5 + 1.5*x + (1+0.15*x)*rchisq(2*n,df=1)
b = rbinom(2*n, 1, p(x))
y = w*b
dat = data.frame(y, x)

ZINQ_tests(formula.logistic=y~x, formula.quantile=y~x, C="x", data=dat)
```

# Index

\*Topic **datasets**

Sample\_Data, [1](#)

Sample\_Data, [1](#)

ZINQ\_combination, [2](#)

ZINQ\_tests, [2](#), [3](#)