Package: powerEQTL (via r-universe)

September 10, 2024

Type Package

Title Power and Sample Size Calculation for eQTL Analysis

Version 0.1.3

Date 2017-01-24

Author Xianjun Dong [aut, ctb], Tzuu-Wang Chang [aut, ctb], Scott T. Weiss [aut, ctb], Weiliang Qiu [aut, cre]

Maintainer Weiliang Qiu <stwxq@channing.harvard.edu>

Depends R (>= 3.3.0)

Imports stats, powerMediation

Description Power and sample size calculation for eQTL analysis based on ANOVA or simple linear regression. It can also calculate power/sample size for testing the association of a SNP to a continuous type phenotype.

License GPL (>=2)

URL https://github.com/sterding/powerEQTL

LazyData TRUE

NeedsCompilation no

Repository https://sterding.r-universe.dev

RemoteUrl https://github.com/sterding/powereqtl

RemoteRef HEAD

RemoteSha 59546203e0268a67837b08306d0d1c1b34505604

Contents

`minEffectEQTL.ANOVA` *Calculation of Minimum Detectable Effect Size for EQTL Analysis Based on Un-Balanced One-Way ANOVA*

Description

Calculation of minimum detectable effect size (δ/σ) for eQTL analysis that tests if a SNP is associated to a gene probe by using un-balanced one-way ANOVA.

Usage

```
minEffectEQTL.ANOVA(MAF,
                    typeI = 0.05,
                    nTests = 2e+05,
                    myntotal = 200,
                    mypower = 0.8,
                    verbose = TRUE)
```

Arguments

`MAF`	Minor allele frequency.
`typeI`	Type I error rate for testing if a SNP is associated to a gene probe.
`nTests`	integer. Number of tests in eQTL analysis.
`myntotal`	integer. Number of subjects.
`mypower`	Desired power for the eQTL analysis.
`verbose`	logic. indicating if intermediate results should be output.

Details

The assumption of the ANOVA approach is that the association of a SNP to a gene probe is tested by using un-balanced one-way ANOVA (e.g. Lonsdale et al. 2013). According to SAS online document https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_power_a0000000982.htm, the power calculation formula is

$$power = Pr\left(F \geq F_{1-\alpha}\left(k-1, N-k\right) \middle| F \sim F_{k-1,N-k,\lambda}\right),$$

where $k = 3$ is the number of groups of subjects, N is the total number of subjects, $F_{1-\alpha}\left(k-1, N-k\right)$ is the $100(1-\alpha)$-th percentile of central F distribution with degrees of freedoms $k-1$ and $N-k$,

and $F_{k-1,N-k,\lambda}$ is the non-central F distribution with degrees of freedoms $k-1$ and $N-k$ and non-central parameter (ncp) λ. The ncp λ is equal to

$$\lambda = \frac{N}{\sigma^2} \sum_{i=1}^{k} w_i \left(\mu_i - \mu\right)^2,$$

where μ_i is the mean gene expression level for the i-th group of subjects, w_i is the weight for the i-th group of subjects, σ^2 is the variance of the random errors in ANOVA (assuming each group has equal variance), and μ is the weighted mean gene expression level

$$\mu = \sum_{i=1}^{k} w_i \mu_i.$$

The weights w_i are the sample proportions for the 3 groups of subjects. Hence, $\sum_{i=1}^{3} w_i = 1$.

We assume that $\mu_2 - \mu_1 = \mu_3 - \mu_2 = \delta$, where μ_1, μ_2, and μ_3 are the mean gene expression level for mutation homozygotes, heterozygotes, and wild-type homozygotes, respectively.

Denote p as the minor allele frequency (MAF) of a SNP. Under Hardy-Weinberg equilibrium, we have genotype frequencies: $p_2 = p^2$, $p_1 = 2pq$, and $p_0 = q^2$, where p_2, p_1, and p_0 are genotype for mutation homozygotes, heterozygotes, and wild-type homozygotes, respectively, $q = 1 - p$. Then ncp can be simplified as

$$ncp = 2pqN \left(\frac{\delta}{\sigma}\right)^2,$$

Value

minimum detectable effect size δ/σ.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Lonsdale J and Thomas J, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics, 45:580-585, 2013.

See Also

powerEQTL.ANOVA, powerEQTL.ANOVA2, ssEQTL.ANOVA, ssEQTL.ANOVA2

Examples

```
minEffectEQTL.ANOVA(
          MAF = 0.1,
          typeI = 0.05,
          nTests = 200000,
          myntotal = 234,
```

```
            mypower = 0.8,
            verbose = TRUE)
```

`minMAFeQTL.ANOVA` *Calculation of Minimum Detectable Minor Allele Frequency for EQTL Analysis Based on Un-Balanced One-Way ANOVA*

Description

Calculation of minimum detectable minor allele frequency (MAF) for eQTL analysis that tests if a SNP is associated to a gene probe by using un-balanced one-way ANOVA.

Usage

```
minMAFeQTL.ANOVA(effsize,
                 typeI = 0.05,
                 nTests = 200000,
                 myntotal = 200,
                 mypower = 0.8,
                 verbose = TRUE)
```

Arguments

`effsize`	Effect size δ/σ.
`typeI`	Type I error rate for testing if a SNP is associated to a gene probe.
`nTests`	integer. Number of tests in eQTL analysis.
`myntotal`	integer. Number of subjects.
`mypower`	Desired power for the eQTL analysis.
`verbose`	logic. indicating if intermediate results should be output.

Details

The assumption of the ANOVA approach is that the association of a SNP to a gene probe is tested by using un-balanced one-way ANOVA (e.g. Lonsdale et al. 2013). According to SAS online document https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_power_a0000000982.htm, the power calculation formula is

$$power = Pr\left(F \geq F_{1-\alpha}\left(k-1, N-k\right) \middle| F \sim F_{k-1,N-k,\lambda}\right),$$

where $k = 3$ is the number of groups of subjects, N is the total number of subjects, $F_{1-\alpha}\left(k-1, N-k\right)$ is the $100(1-\alpha)$-th percentile of central F distribution with degrees of freedoms $k-1$ and $N-k$, and $F_{k-1,N-k,\lambda}$ is the non-central F distribution with degrees of freedoms $k-1$ and $N-k$ and non-central parameter (ncp) λ. The ncp λ is equal to

$$\lambda = \frac{N}{\sigma^2}\sum_{i=1}^{k} w_i\left(\mu_i - \mu\right)^2,$$

where μ_i is the mean gene expression level for the i-th group of subjects, w_i is the weight for the i-th group of subjects, σ^2 is the variance of the random errors in ANOVA (assuming each group has equal variance), and μ is the weighted mean gene expression level

$$\mu = \sum_{i=1}^{k} w_i \mu_i.$$

The weights w_i are the sample proportions for the 3 groups of subjects. Hence, $\sum_{i=1}^{3} w_i = 1$.

We assume that $\mu_2 - \mu_1 = \mu_3 - \mu_2 = \delta$, where μ_1, μ_2, and μ_3 are the mean gene expression level for mutation homozygotes, heterozygotes, and wild-type homozygotes, respectively.

Denote p as the minor allele frequency (MAF) of a SNP. Under Hardy-Weinberg equilibrium, we have genotype frequencies: $p_2 = p^2$, $p_1 = 2pq$, and $p_0 = q^2$, where p_2, p_1, and p_0 are genotype for mutation homozygotes, heterozygotes, and wild-type homozygotes, respectively, $q = 1 - p$. Then ncp can be simplified as

$$ncp = 2pqN \left(\frac{\delta}{\sigma}\right)^2,$$

Value

minimum detectable MAF.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Lonsdale J and Thomas J, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics, 45:580-585, 2013.

See Also

powerEQTL.ANOVA, powerEQTL.ANOVA2, ssEQTL.ANOVA, ssEQTL.ANOVA2

Examples

```
minMAFeQTL.ANOVA(effsize = 1,
                 typeI = 0.05,
                 nTests = 200000,
                 myntotal = 234,
                 mypower = 0.8,
                 verbose = TRUE)
```

`minMAFeQTL.SLR` *Minimum Detectable Minor Allele Frequencey Calculation for EQTL Analysis Based on Simple Linear Regression*

Description

Minimum detectable minor allele frequency (MAF) calculation for eQTL analysis that tests if a SNP is associated to a gene probe by using simple linear regression.

Usage

```
minMAFeQTL.SLR(slope,
               typeI = 0.05,
               nTests = 200000,
               myntotal = 200,
               mypower = 0.8,
               mystddev = 0.13,
               verbose = TRUE)
```

Arguments

`slope`	Slope of the simple linear regression.
`typeI`	Type I error rate for testing if a SNP is associated to a gene probe.
`nTests`	integer. Number of tests in eQTL analysis.
`myntotal`	integer. Number of subjects.
`mypower`	Desired power for the eQTL analysis.
`mystddev`	Standard deviation of the random error term ϵ in simple linear regression.
`verbose`	logic. indicating if intermediate results should be output.

Details

To test if a SNP is associated with a gene probe, we use the simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where y_i is the gene expression level of the i-th subject, x_i is the genotype of the i-th subject, and ϵ_i is the random error term. Additive coding for genotype is used. To test if the SNP is associated with the gene probe, we test the null hypothesis $H_0 : \beta_1 = 0$.

Denote p as the minor allele frequency (MAF) of the SNP. Under Hardy-Weinberg equilibrium, we can calculate the variance of genotype of the SNP: $\sigma_x^2 = 2p(1-p)$, where σ_x^2 is the variance of the predictor (i.e. the SNP) x_i.

We then can use Dupont and Plummer's (1998) power/sample size calculation formula to calculate the minimum detectable slope, adjusting for multiple testing.

Value

The estimated minimum detectable MAF.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Dupont, W.D. and Plummer, W.D.. Power and Sample Size Calculations for Studies Involving Linear Regression. Controlled Clinical Trials. 1998;19:589-601.

See Also

powerEQTL.SLR, ssEQTL.SLR

Examples

```
minMAFeQTL.SLR(slope = 0.1299513,
               typeI = 0.05,
               nTests = 200000,
               myntotal = 176,
               mypower = 0.8,
               mystddev = 0.13,
               verbose = TRUE)
```

---

minSlopeEQTL.SLR *Minimum Detectable Slope Calculation for EQTL Analysis Based on Simple Linear Regression*

Description

Minimum detectable slope calculation for eQTL analysis that tests if a SNP is associated to a gene probe by using simple linear regression.

Usage

```
minSlopeEQTL.SLR(
  MAF,
  typeI = 0.05,
  nTests = 2e+05,
  myntotal = 200,
  mypower = 0.8,
  mystddev = 0.13,
  verbose = TRUE)
```

Arguments

MAF	Minor allele frequency.
typeI	Type I error rate for testing if a SNP is associated to a gene probe.
nTests	integer. Number of tests in eQTL analysis.
myntotal	integer. Number of subjects.
mypower	Desired power for the eQTL analysis.
mystddev	Standard deviation of the random error term ϵ in simple linear regression.
verbose	logic. indicating if intermediate results should be output.

Details

To test if a SNP is associated with a gene probe, we use the simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where y_i is the gene expression level of the i-th subject, x_i is the genotype of the i-th subject, and ϵ_i is the random error term. Additive coding for genotype is used. To test if the SNP is associated with the gene probe, we test the null hypothesis $H_0 : \beta_1 = 0$.

Denote p as the minor allele frequency (MAF) of the SNP. Under Hardy-Weinberg equilibrium, we can calculate the variance of genotype of the SNP: $\sigma_x^2 = 2p(1-p)$, where σ_x^2 is the variance of the predictor (i.e. the SNP) x_i.

We then can use Dupont and Plummer's (1998) power/sample size calculation formula to calculate the minimum detectable slope, adjusting for multiple testing.

Value

The estimated minimum detectable slope.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Dupont, W.D. and Plummer, W.D.. Power and Sample Size Calculations for Studies Involving Linear Regression. Controlled Clinical Trials. 1998;19:589-601.

See Also

powerEQTL.SLR, ssEQTL.SLR

Examples

```
minSlopeEQTL.SLR(
  MAF = 0.1,
  typeI = 0.05,
  nTests = 2e+05,
  myntotal = 176,
  mypower = 0.8,
  mystddev = 0.13,
  verbose = TRUE)
```

`powerEQTL.ANOVA` *Power Calculation for EQTL Analysis Based on Un-Balanced One-Way ANOVA*

Description

Power calculation for eQTL analysis that tests if a SNP is associated to a gene probe by using un-balanced one-way ANOVA.

Usage

```
powerEQTL.ANOVA(MAF,
                typeI = 0.05,
                nTests = 2e+05,
                myntotal = 200,
                mystddev = 0.13,
                deltaVec = c(0.13, 0.13),
                verbose = TRUE)
```

Arguments

`MAF`	Minor allele frequency.
`typeI`	Type I error rate for testing if a SNP is associated to a gene probe.
`nTests`	integer. Number of tests in eQTL analysis.
`myntotal`	integer. Number of subjects.
`mystddev`	Standard deviation of gene expression levels in one group of subjects. Assume all 3 groups of subjects (mutation homozygote, heterozygote, wild-type homozygote) have the same standard deviation of gene expression levels.
`deltaVec`	A vector having 2 elements. The first element is equal to $\mu_2 - \mu_1$ and the second elementis equalt to $\mu_3 - \mu_2$, where μ_1 is the mean gene expression level for the mutation homozygotes, μ_2 is the mean gene expression level for the heterozygotes, and μ_3 is the mean gene expression level for the wild-type gene expression level.
`verbose`	logic. indicating if intermediate results should be output.

Details

The assumption of the ANOVA approach is that the association of a SNP to a gene probe is tested by using un-balanced one-way ANOVA (e.g. Lonsdale et al. 2013). According to SAS online document https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_power_a0000000982.htm, the power calculation formula is

$$power = Pr\left(F \geq F_{1-\alpha}\left(k-1, N-k\right) \middle| F \sim F_{k-1,N-k,\lambda}\right),$$

where $k = 3$ is the number of groups of subjects, N is the total number of subjects, $F_{1-\alpha}\left(k-1, N-k\right)$ is the $100(1-\alpha)$-th percentile of central F distribution with degrees of freedoms $k-1$ and $N-k$, and $F_{k-1,N-k,\lambda}$ is the non-central F distribution with degrees of freedoms $k-1$ and $N-k$ and non-central parameter (ncp) λ. The ncp λ is equal to

$$\lambda = \frac{N}{\sigma^2}\sum_{i=1}^{k} w_i\left(\mu_i - \mu\right)^2,$$

where μ_i is the mean gene expression level for the i-th group of subjects, w_i is the weight for the i-th group of subjects, σ^2 is the variance of the random errors in ANOVA (assuming each group has equal variance), and μ is the weighted mean gene expression level

$$\mu = \sum_{i=1}^{k} w_i\mu_i.$$

The weights w_i are the sample proportions for the 3 groups of subjects. Hence, $\sum_{i=1}^{3} w_i = 1$.

Value

power of the test after Bonferroni correction for multiple testing.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Lonsdale J and Thomas J, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics, 45:580-585, 2013.

See Also

minEffectEQTL.ANOVA, powerEQTL.ANOVA2, ssEQTL.ANOVA, ssEQTL.ANOVA2

Examples

```
powerEQTL.ANOVA(
          MAF = 0.1,
          typeI = 0.05,
          nTests = 200000,
```

```
            myntotal = 234,
            mystddev = 0.13,
            deltaVec = c(0.13, 0.13))
```

---

powerEQTL.ANOVA2 *Power Calculation for EQTL Analysis Based on Un-Balanced One-Way ANOVA*

Description

Power calculation for eQTL analysis that tests if a SNP is associated to a gene probe by using un-balanced one-way ANOVA (assuming Hardy-Weinberg equilibrium).

Usage

```
powerEQTL.ANOVA2(effsize,
                 MAF,
                 typeI = 0.05,
                 nTests = 2e+05,
                 myntotal = 200,
                 verbose = TRUE)
```

Arguments

effsize	effect size δ/σ, where $\delta = \mu_2 - \mu_1 = \mu_3 - \mu_2$, μ_1, μ_2, μ_3 are the mean gene expression level of mutation homozygotes, heterozygotes, and wild-type homozygotes, and σ is the standard deviation of gene expression levels (assuming each genotype group has the same variance).
MAF	Minor allele frequency.
typeI	Type I error rate for testing if a SNP is associated to a gene probe.
nTests	integer. Number of tests in eQTL analysis.
myntotal	integer. Number of subjects.
verbose	logic. indicating if intermediate results should be output.

Details

The assumption of the ANOVA approach is that the association of a SNP to a gene probe is tested by using un-balanced one-way ANOVA (e.g. Lonsdale et al. 2013). According to SAS online document https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_power_a0000000982.htm, the power calculation formula is

$$power = Pr\left(F \geq F_{1-\alpha}\left(k-1, N-k\right) \middle| F \sim F_{k-1,N-k,\lambda}\right),$$

where $k = 3$ is the number of groups of subjects, N is the total number of subjects, $F_{1-\alpha}\left(k-1, N-k\right)$ is the $100(1-\alpha)$-th percentile of central F distribution with degrees of freedoms $k-1$ and $N-k$,

and $F_{k-1,N-k,\lambda}$ is the non-central F distribution with degrees of freedoms $k-1$ and $N-k$ and non-central parameter (ncp) λ. The ncp λ is equal to

$$\lambda = \frac{N}{\sigma^2} \sum_{i=1}^{k} w_i \left(\mu_i - \mu\right)^2,$$

where μ_i is the mean gene expression level for the i-th group of subjects, w_i is the weight for the i-th group of subjects, σ^2 is the variance of the random errors in ANOVA (assuming each group has equal variance), and μ is the weighted mean gene expression level

$$\mu = \sum_{i=1}^{k} w_i \mu_i.$$

The weights w_i are the sample proportions for the 3 groups of subjects. Hence, $\sum_{i=1}^{3} w_i = 1$.

We assume that $\mu_2 - \mu_1 = \mu_3 - \mu_2 = \delta$, where μ_1, μ_2, and μ_3 are the mean gene expression level for mutation homozygotes, heterozygotes, and wild-type homozygotes, respectively.

Denote p as the minor allele frequency (MAF) of a SNP. Under Hardy-Weinberg equilibrium, we have genotype frequencies: $p_2 = p^2$, $p_1 = 2pq$, and $p_0 = q^2$, where p_2, p_1, and p_0 are genotype for mutation homozygotes, heterozygotes, and wild-type homozygotes, respectively, $q = 1 - p$. Then ncp can be simplified as

$$ncp = 2pqN \left(\frac{\delta}{\sigma}\right)^2,$$

Value

power of the test after Bonferroni correction for multiple testing.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Lonsdale J and Thomas J, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics, 45:580-585, 2013.

See Also

minEffectEQTL.ANOVA, powerEQTL.ANOVA, ssEQTL.ANOVA, ssEQTL.ANOVA2

Examples

```
powerEQTL.ANOVA2(effsize = 1,
                 MAF = 0.1,
                 typeI = 0.05,
                 nTests = 2e+05,
                 myntotal = 234,
                 verbose = TRUE)
```

`powerEQTL.SLR` *Power Calculation for EQTL Analysis Based on Simple Linear Regression*

Description

Power calculation for eQTL analysis that tests if a SNP is associated to a gene probe by using simple linear regression.

Usage

```
powerEQTL.SLR(
  MAF,
  typeI = 0.05,
  nTests = 2e+05,
  slope = 0.13,
  myntotal = 200,
  mystddev = 0.13,
  verbose = TRUE)
```

Arguments

`MAF`	Minor allele frequency.
`typeI`	Type I error rate for testing if a SNP is associated to a gene probe.
`nTests`	integer. Number of tests in eQTL analysis.
`slope`	Slope β_1 of the simple linear regression $$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$ where y_i is the gene expression level of the i-th subject, x_i is the genotype of the i-th subject, and ϵ_i is the random error term. Additive coding for genotype is used.
`myntotal`	integer. Number of subjects.
`mystddev`	Standard deviation of the random error term ϵ in simple linear regression.
`verbose`	logic. indicating if intermediate results should be output.

Details

To test if a SNP is associated with a gene probe, we use the simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where y_i is the gene expression level of the i-th subject, x_i is the genotype of the i-th subject, and ϵ_i is the random error term. Additive coding for genotype is used. To test if the SNP is associated with the gene probe, we test the null hypothesis $H_0 : \beta_1 = 0$.

Denote p as the minor allele frequency (MAF) of the SNP. Under Hardy-Weinberg equilibrium, we can calculate the variance of genotype of the SNP: $\sigma_x^2 = 2p(1-p)$, where σ_x^2 is the variance of the predictor (i.e. the SNP) x_i.

We then can use Dupont and Plummer's (1998) power/sample size calculation formula to calculate the minimum detectable slope, adjusting for multiple testing.

Value

power of the test after Bonferroni correction for multiple testing.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Dupont, W.D. and Plummer, W.D.. Power and Sample Size Calculations for Studies Involving Linear Regression. Controlled Clinical Trials. 1998;19:589-601.

See Also

ssEQTL.SLR, minSlopeEQTL.SLR

Examples

```
powerEQTL.SLR(
  MAF = 0.1,
  typeI = 0.05,
  nTests = 2e+05,
  slope = 0.13,
  myntotal = 176,
  mystddev = 0.13,
  verbose = TRUE)
```

---

ssEQTL.ANOVA *Sample Size Calculation for EQTL Analysis Based on Un-Balanced One-Way ANOVA*

Description

Sample size calculation for eQTL analysis that tests if a SNP is associated to a gene probe by using un-balanced one-way ANOVA.

Usage

```
ssEQTL.ANOVA(
  MAF,
  typeI = 0.05,
  nTests = 2e+05,
  mypower = 0.8,
  mystddev = 0.13,
  deltaVec = c(0.13, 0.13))
```

Arguments

MAF	Minor allele frequency.
typeI	Type I error rate for testing if a SNP is associated to a gene probe.
nTests	integer. Number of tests in eQTL analysis.
mypower	Desired power for the eQTL analysis.
mystddev	Standard deviation of gene expression levels in one group of subjects. Assume all 3 groups of subjects (mutation homozygote, heterozygote, wild-type homozygote) have the same standard deviation of gene expression levels.
deltaVec	A vector having 2 elements. The first element is equal to $\mu_2 - \mu_1$ and the second elementis equalt to $\mu_3 - \mu_2$, where μ_1 is the mean gene expression level for the mutation homozygotes, μ_2 is the mean gene expression level for the heterozygotes, and μ_3 is the mean gene expression level for the wild-type gene expression level.

Details

The assumption of the ANOVA approach is that the association of a SNP to a gene probe is tested by using un-balanced one-way ANOVA (e.g. Lonsdale et al. 2013). According to SAS online document https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_power_a0000000982.htm, the power calculation formula is

$$power = Pr\left(F \geq F_{1-\alpha}\left(k-1, N-k\right) \middle| F \sim F_{k-1,N-k,\lambda}\right),$$

where $k = 3$ is the number of groups of subjects, N is the total number of subjects, $F_{1-\alpha}\left(k-1, N-k\right)$ is the $100(1-\alpha)$-th percentile of central F distribution with degrees of freedoms $k-1$ and $N-k$, and $F_{k-1,N-k,\lambda}$ is the non-central F distribution with degrees of freedoms $k-1$ and $N-k$ and non-central parameter (ncp) λ. The ncp λ is equal to

$$\lambda = \frac{N}{\sigma^2}\sum_{i=1}^{k} w_i\left(\mu_i - \mu\right)^2,$$

where μ_i is the mean gene expression level for the i-th group of subjects, w_i is the weight for the i-th group of subjects, σ^2 is the variance of the random errors in ANOVA (assuming each group has equal variance), and μ is the weighted mean gene expression level

$$\mu = \sum_{i=1}^{k} w_i\mu_i.$$

The weights w_i are the sample proportions for the 3 groups of subjects. Hence, $\sum_{i=1}^{3} w_i = 1$.

Value

sample size required for the eQTL analysis to achieve the desired power.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Lonsdale J and Thomas J, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics, 45:580-585, 2013.

See Also

`minEffectEQTL.ANOVA`, `powerEQTL.ANOVA`, `powerEQTL.ANOVA2`, `ssEQTL.ANOVA2`

Examples

```
ssEQTL.ANOVA(MAF = 0.1,
       typeI = 0.05,
       nTests = 200000,
       mypower = 0.8,
       mystddev = 0.13,
       deltaVec = c(0.13, 0.13))
```

---

ssEQTL.ANOVA2 *Sample Size Calculation for EQTL Analysis Based on Un-Balanced One-Way ANOVA*

Description

Sample size calculation for eQTL analysis that tests if a SNP is associated to a gene probe by using un-balanced one-way ANOVA.

Usage

```
ssEQTL.ANOVA2(
  effsize,
  MAF,
  typeI = 0.05,
  nTests = 2e+05,
  mypower = 0.8
)
```

Arguments

effsize	effect size δ/σ, where $\delta = \mu_2 - \mu_1 = \mu_3 - \mu_2$, μ_1, μ_2, μ_3 are the mean gene expression level of mutation homozygotes, heterozygotes, and wild-type homozygotes, and σ is the standard deviation of gene expression levels (assuming each genotype group has the same variance).
MAF	Minor allele frequency.
typeI	Type I error rate for testing if a SNP is associated to a gene probe.
nTests	integer. Number of tests in eQTL analysis.
mypower	Desired power for the eQTL analysis.

Details

The assumption of the ANOVA approach is that the association of a SNP to a gene probe is tested by using un-balanced one-way ANOVA (e.g. Lonsdale et al. 2013). According to SAS online document https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_power_a0000000982.htm, the power calculation formula is

$$power = Pr\left(F \geq F_{1-\alpha}\left(k-1, N-k\right) \middle| F \sim F_{k-1,N-k,\lambda}\right),$$

where $k = 3$ is the number of groups of subjects, N is the total number of subjects, $F_{1-\alpha}\left(k-1, N-k\right)$ is the $100(1-\alpha)$-th percentile of central F distribution with degrees of freedoms $k-1$ and $N-k$, and $F_{k-1,N-k,\lambda}$ is the non-central F distribution with degrees of freedoms $k-1$ and $N-k$ and non-central parameter (ncp) λ. The ncp λ is equal to

$$\lambda = \frac{N}{\sigma^2}\sum_{i=1}^{k} w_i\left(\mu_i - \mu\right)^2,$$

where μ_i is the mean gene expression level for the i-th group of subjects, w_i is the weight for the i-th group of subjects, σ^2 is the variance of the random errors in ANOVA (assuming each group has equal variance), and μ is the weighted mean gene expression level

$$\mu = \sum_{i=1}^{k} w_i\mu_i.$$

The weights w_i are the sample proportions for the 3 groups of subjects. Hence, $\sum_{i=1}^{3} w_i = 1$.

We assume that $\mu_2 - \mu_1 = \mu_3 - \mu_2 = \delta$, where μ_1, μ_2, and μ_3 are the mean gene expression level for mutation homozygotes, heterozygotes, and wild-type homozygotes, respectively.

Denote p as the minor allele frequency (MAF) of a SNP. Under Hardy-Weinberg equilibrium, we have genotype frequencies: $p_2 = p^2$, $p_1 = 2pq$, and $p_0 = q^2$, where p_2, p_1, and p_0 are genotype for mutation homozygotes, heterozygotes, and wild-type homozygotes, respectively, $q = 1 - p$. Then ncp can be simplified as

$$ncp = 2pqN\left(\frac{\delta}{\sigma}\right)^2,$$

Value

sample size required for the eQTL analysis to achieve the desired power.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Lonsdale J and Thomas J, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics, 45:580-585, 2013.

See Also

minEffectEQTL.ANOVA, powerEQTL.ANOVA, powerEQTL.ANOVA2, ssEQTL.ANOVA

Examples

```
ssEQTL.ANOVA2(
  effsize = 1,
  MAF = 0.1,
  typeI = 0.05,
  nTests = 2e+05,
  mypower = 0.8
)
```

---

ssEQTL.SLR *Sample Size Calculation for EQTL Analysis Based on Simple Linear Regression*

Description

Sample size calculation for eQTL analysis that tests if a SNP is associated to a gene probe by using simple linear regression.

Usage

```
ssEQTL.SLR(
  MAF,
  typeI = 0.05,
  nTests = 2e+05,
  slope = 0.13,
  mypower = 0.8,
  mystddev = 0.13,
  n.lower = 2.01,
  n.upper = 1e+30,
  verbose = TRUE)
```

Arguments

MAF	Minor allele frequency.
typeI	Type I error rate for testing if a SNP is associated to a gene probe.
nTests	integer. Number of tests in eQTL analysis.
slope	Slope β_1 of the simple linear regression $$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$ where y_i is the gene expression level of the i-th subject, x_i is the genotype of the i-th subject, and ϵ_i is the random error term. Additive coding for genotype is used.
mypower	Desired power for the eQTL analysis.
mystddev	Standard deviation of the random error term ϵ.
n.lower	integer. Lower bound of the total number of subjects.
n.upper	integer. Upper bound of the total number of subjects.
verbose	logic. indicating if intermediate results should be output.

Details

To test if a SNP is associated with a gene probe, we use the simple linear regression

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where y_i is the gene expression level of the i-th subject, x_i is the genotype of the i-th subject, and ϵ_i is the random error term. Additive coding for genotype is used. To test if the SNP is associated with the gene probe, we test the null hypothesis $H_0 : \beta_1 = 0$.

Denote p as the minor allele frequency (MAF) of the SNP. Under Hardy-Weinberg equilibrium, we can calculate the variance of genotype of the SNP: $\sigma_x^2 = 2p(1-p)$, where σ_x^2 is the variance of the predictor (i.e. the SNP) x_i.

We then can use Dupont and Plummer's (1998) power/sample size calculation formula to calculate the minimum detectable slope, adjusting for multiple testing.

Value

sample size required for the eQTL analysis to achieve the desired power.

Author(s)

Xianjun Dong <XDONG@rics.bwh.harvard.edu>, Tzuu-Wang Chang <Chang.Tzuu-Wang@mgh.harvard.edu>, Scott T. Weiss <restw@channing.harvard.edu>, Weiliang Qiu <stwxq@channing.harvard.edu>

References

Dupont, W.D. and Plummer, W.D.. Power and Sample Size Calculations for Studies Involving Linear Regression. Controlled Clinical Trials. 1998;19:589-601.

See Also

powerEQTL.SLR, minSlopeEQTL.SLR

Examples

```
ssEQTL.SLR(
  MAF = 0.1,
  typeI = 0.05,
  nTests = 2e+05,
  slope = 0.13,
  mypower = 0.8,
  mystddev = 0.13,
  n.lower = 2.01,
  n.upper = 1e+30,
  verbose = TRUE)
```

Index