
WARNING
ASSUMPTIONS
A H E A D

Dealing with “some” important statistical assumptions.

1) The issue of normality (today):

- Parametric (e.g., ANOVA): assume parametrized families of probability distributions (e.g., normal defined by two parameters, i.e., mean and variance). Parameter estimates tend to be sensitive to non-normality (e.g., issue in regression slopes), but not necessarily in statistical hypothesis testing (P-values may be not as sensitive).

- Non-parametric: either distribution free (e.g., permutation tests) or ranked based tests.

Dealing with “some” important statistical assumptions.

2) The issue of homogeneity of variances (later in the course):

- Standard (e.g., ANOVAs, regressions) assume homoscedasticity.

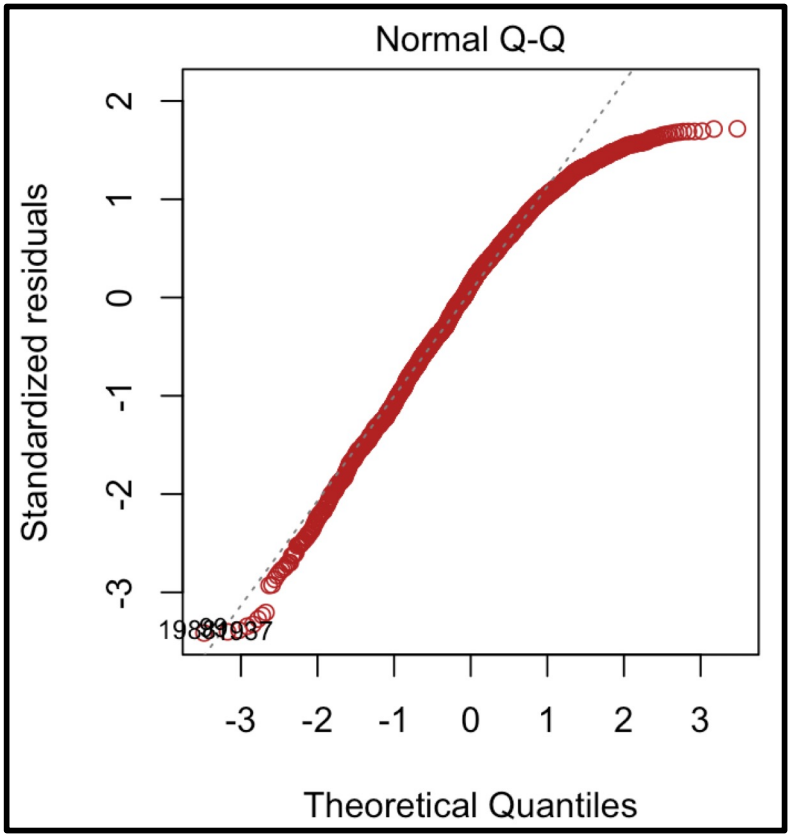
- Robust approaches (Welch’s ANOVA, Weighted least squares) are good to deal with heteroscedasticity.

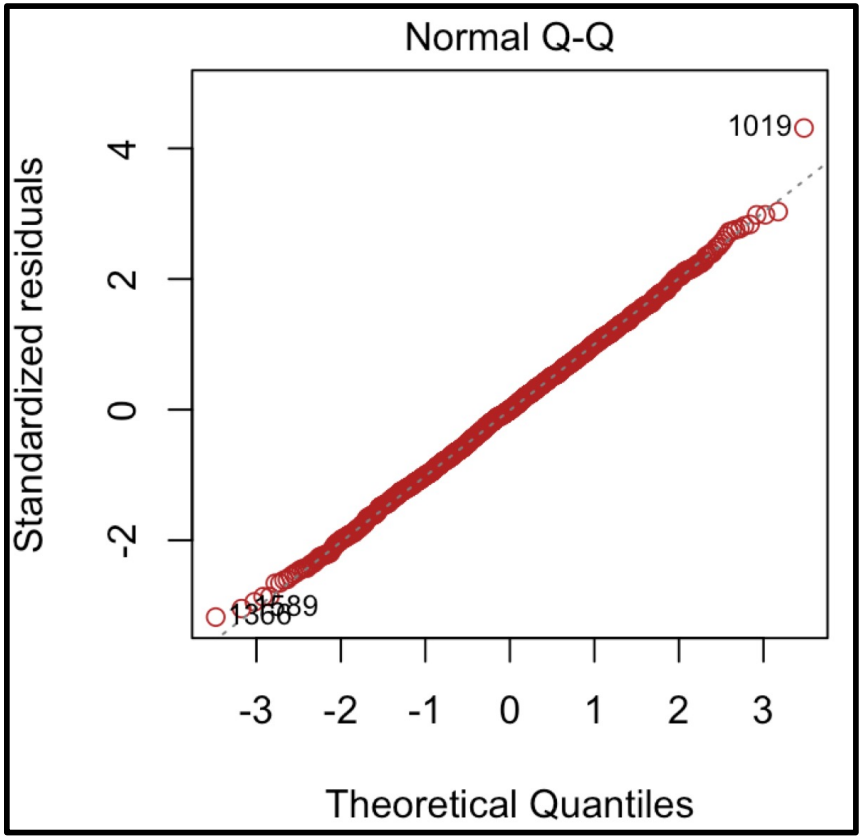
One response variable &
Multiple categorical factors (ANOVAs)

Are variables normally distributed in each combination of treatment?
(Normal QQ Plot of residuals)

NO

YES





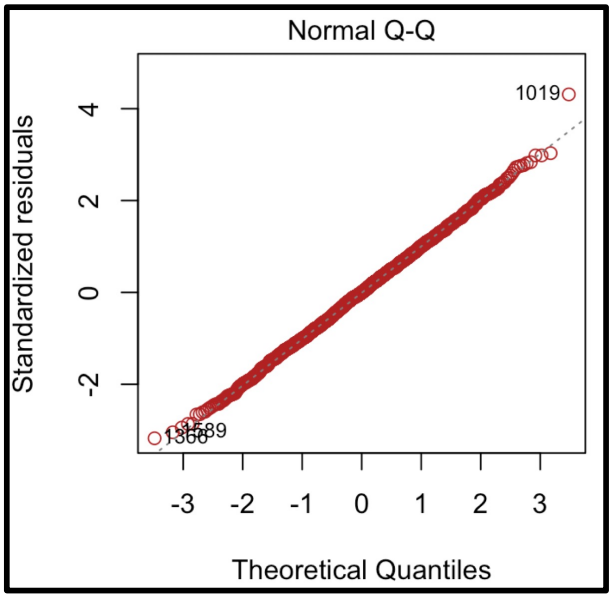
One response variable &
Multiple categorical factors

Are variables normally distributed in each combination of treatment?
(Normal QQ Plot of residuals)

NO

YES

PARAMETRIC TESTS



Are variances equal among all populations?
(Levene's test)

NO

YES

ANOVA

One response variable &
Multiple categorical factors

Are variables normally distributed in each combination of treatment?
(Normal QQ Plot of residuals)

NO

YES

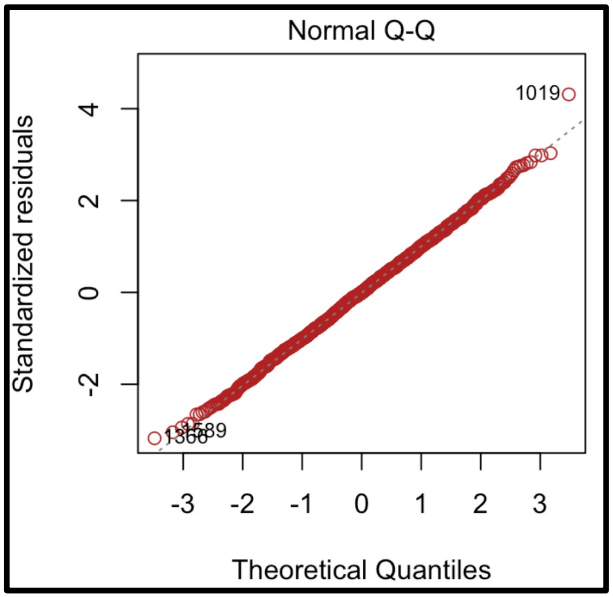
PARAMETRIC TESTS

Are variances equal among all populations?
(Levene's test)

NO

YES

ANOVA



Parametric is supposed to be about assuming parameters about the population where data were sampled; but many practitioners see as only about normality (which is not true).

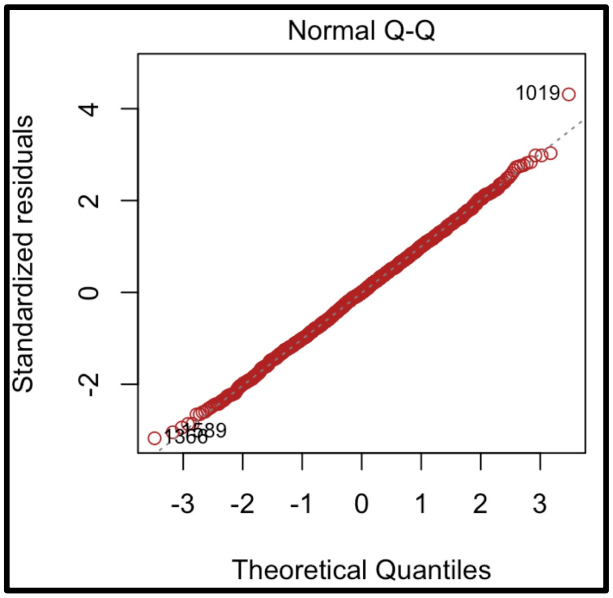
One response variable &
Multiple categorical factors

Are variables normally distributed in each combination of treatment?
(Normal QQ Plot of residuals)

NO

YES

PARAMETRIC TESTS



Are variances equal among all populations?
(Levene's test)

NO

YES

Welch's ANOVA
Weighted least squares (later in the semester)

ANOVA

transformations
(log, square root, etc)

One response variable &
Multiple categorical factors

Are variables normally distributed in each combination of treatment?
(Normal QQ Plot of residuals)

NO

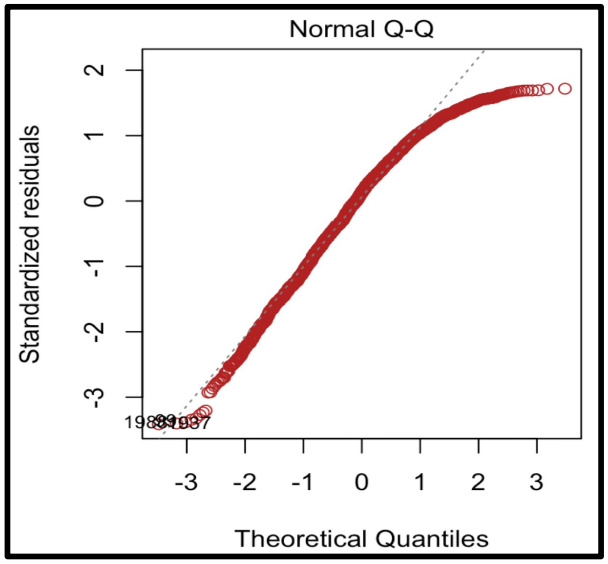
YES

Data Transformation
(rank, log, square root, Box-Cox power transformation, etc) and verify data normality again after transformation

If normal after transformation

If NOT normal after transformation

NON-PARAMETRIC TESTS



Even though parametric tests are robust against normality, we often don't know how much for the particular data at hands; the tradition is then to use non-parametric tests

One response variable &
Multiple categorical factors

Are variables normally distributed in each combination of treatment?
(Normal QQ Plot of residuals)

NO

YES

If not normal after transformation

Can we assume that variances are equal among all populations? (Levene's test)

NO

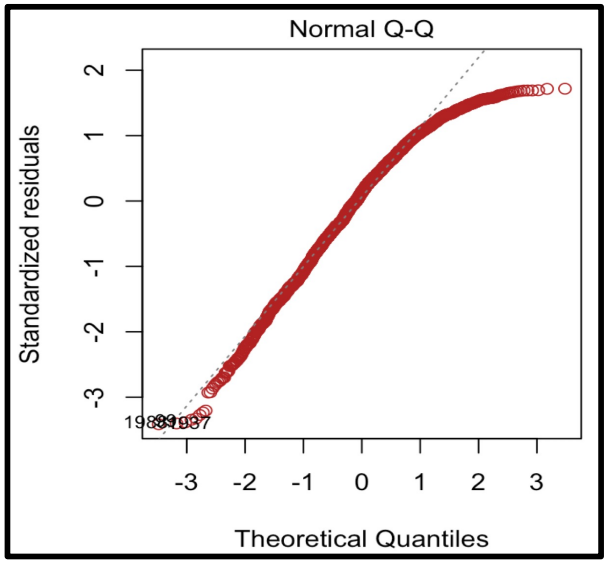
YES

ANOVA

Rank transformation

Kruskal-Wallis

NON-PARAMETRIC TESTS



One response variable &
Multiple categorical factors

Are variables normally distributed in each combination of treatment?
(Normal QQ Plot of residuals)

NO

If not normal after transformation

Rank transformation

Can we assume that variances are equal among all populations? (Levene's test)

NO

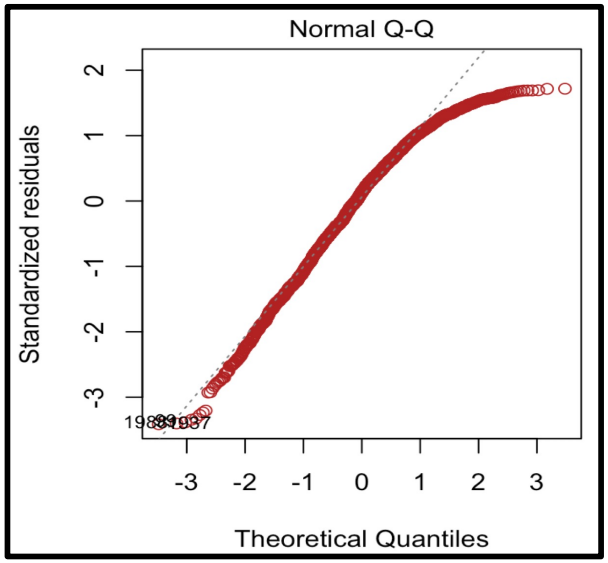
YES

Welch's ANOVA
Weighted least squares on ranks

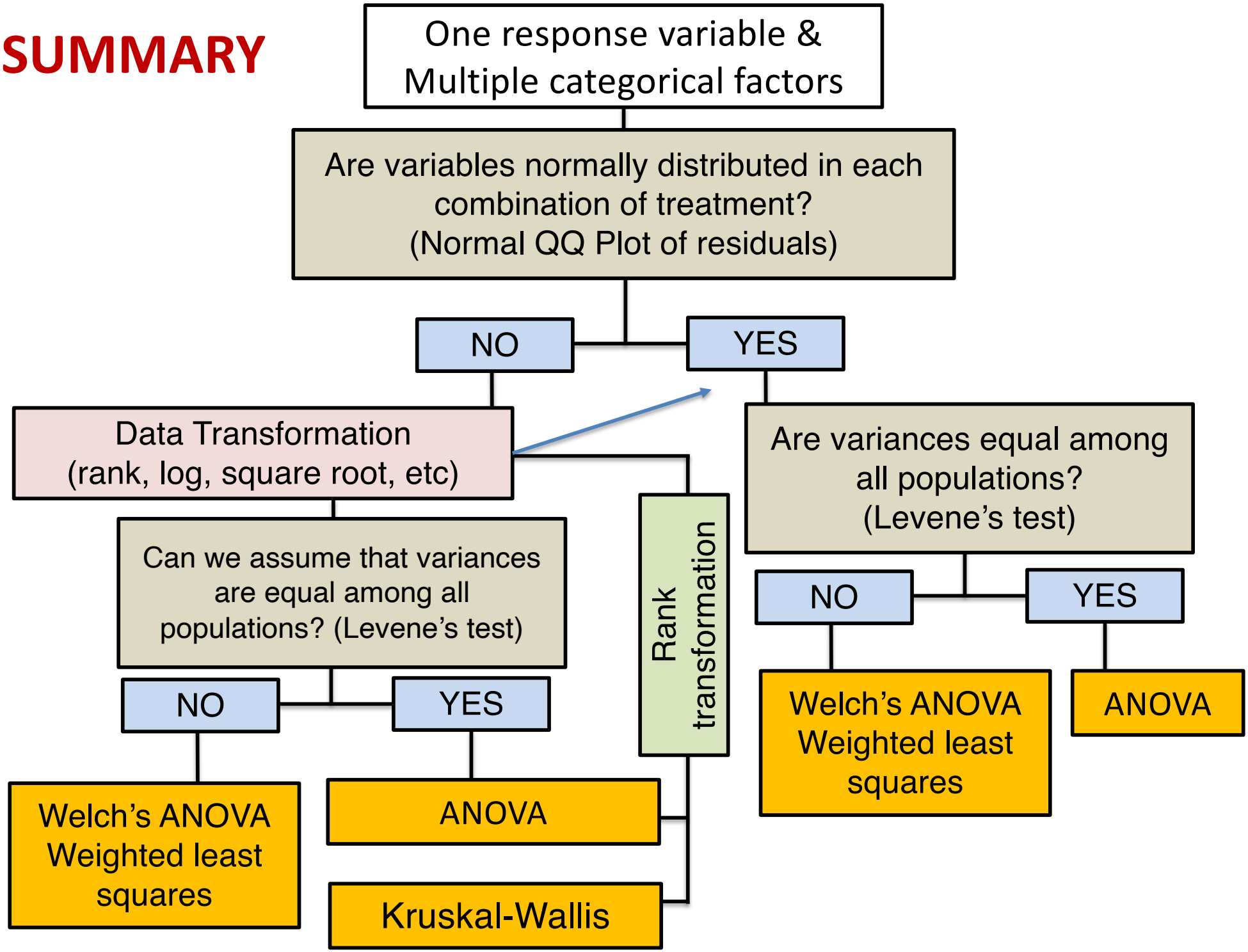
ANOVA

Kruskal-Wallis

NON-PARAMETRIC TESTS



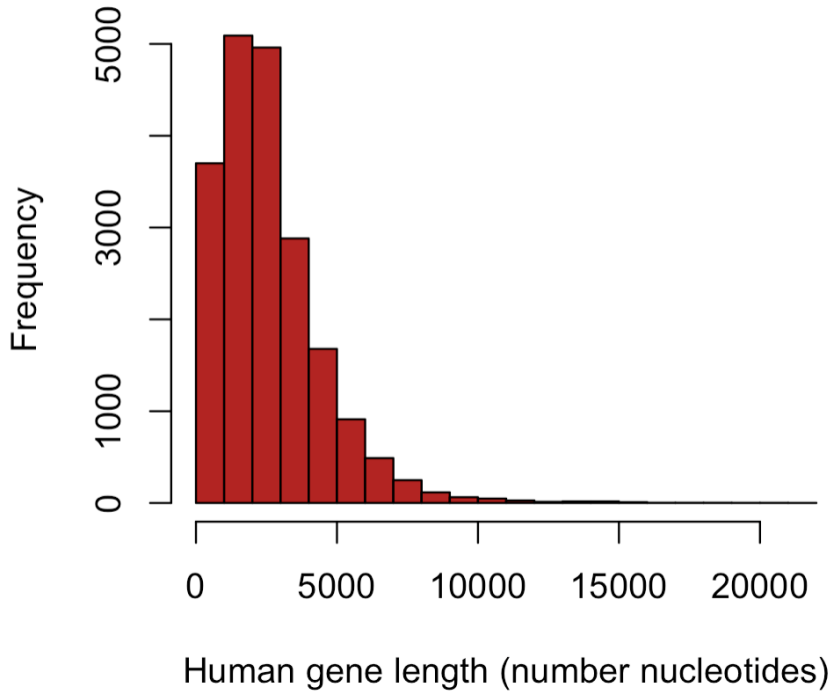
SUMMARY

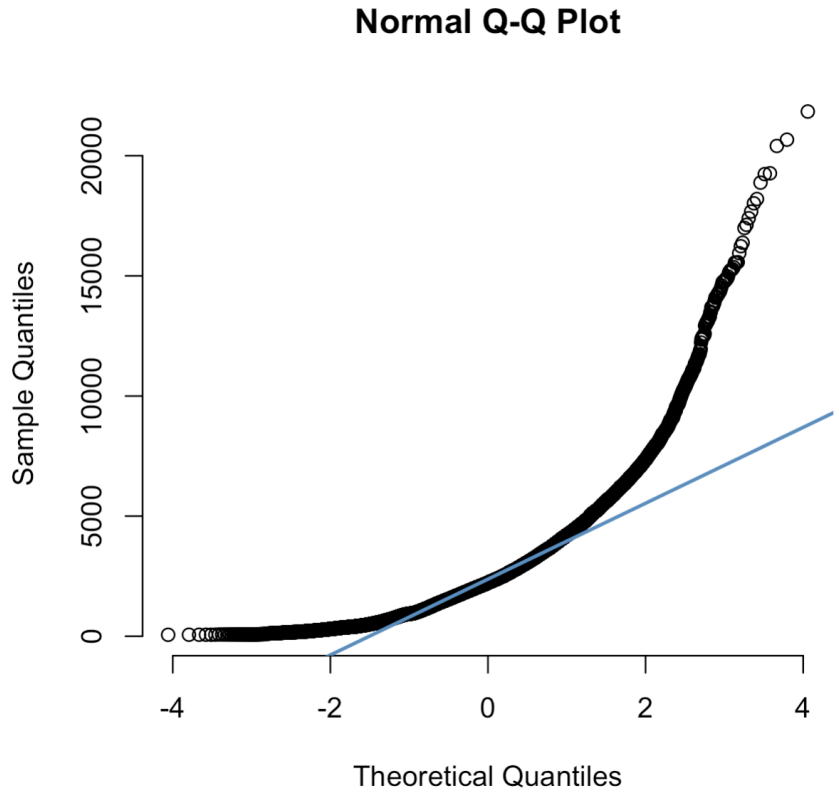

One response variable &
Multiple categorical factors
Are variables normally distributed in each combination of treatment?
(Normal QQ Plot of residuals)
NO
YES
Data Transformation
(rank, log, square root, etc)
Are variances equal among all populations?
(Levene's test)
Rank transformation
Can we assume that variances are equal among all populations? (Levene's test)
NO
YES
NO
YES
Welch's ANOVA
Weighted least squares
ANOVA
Welch's ANOVA
Weighted least squares
ANOVA
Kruskal-Wallis

The role of data transformations:

improve normality (today) &

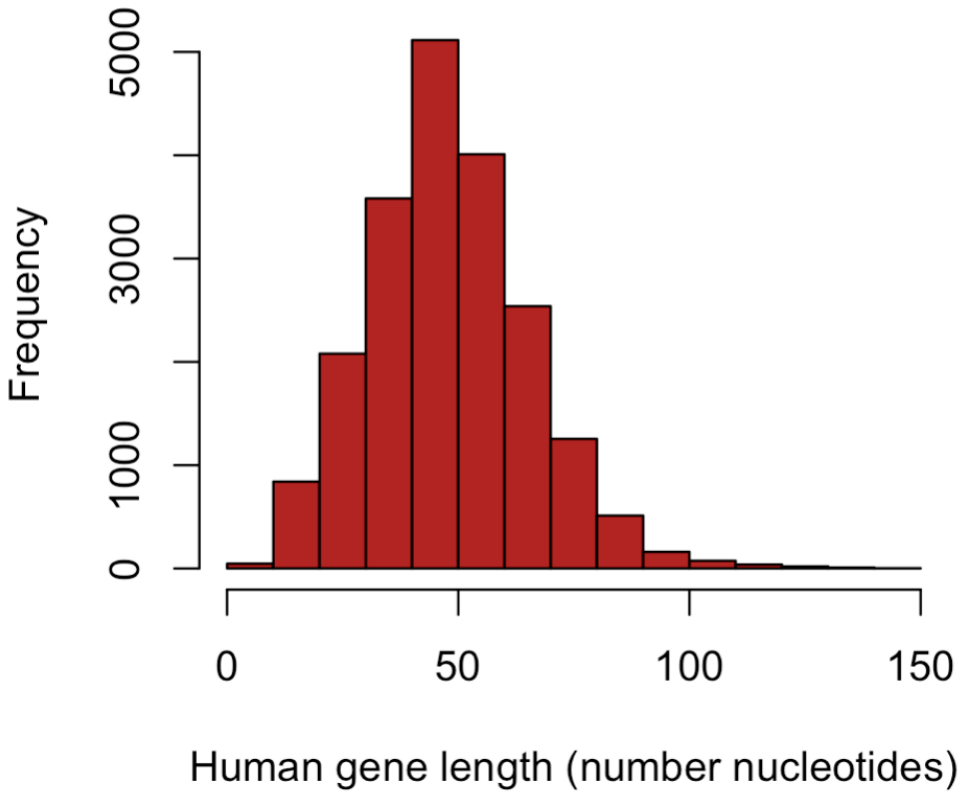
homoscedasticity (covered in another lecture)

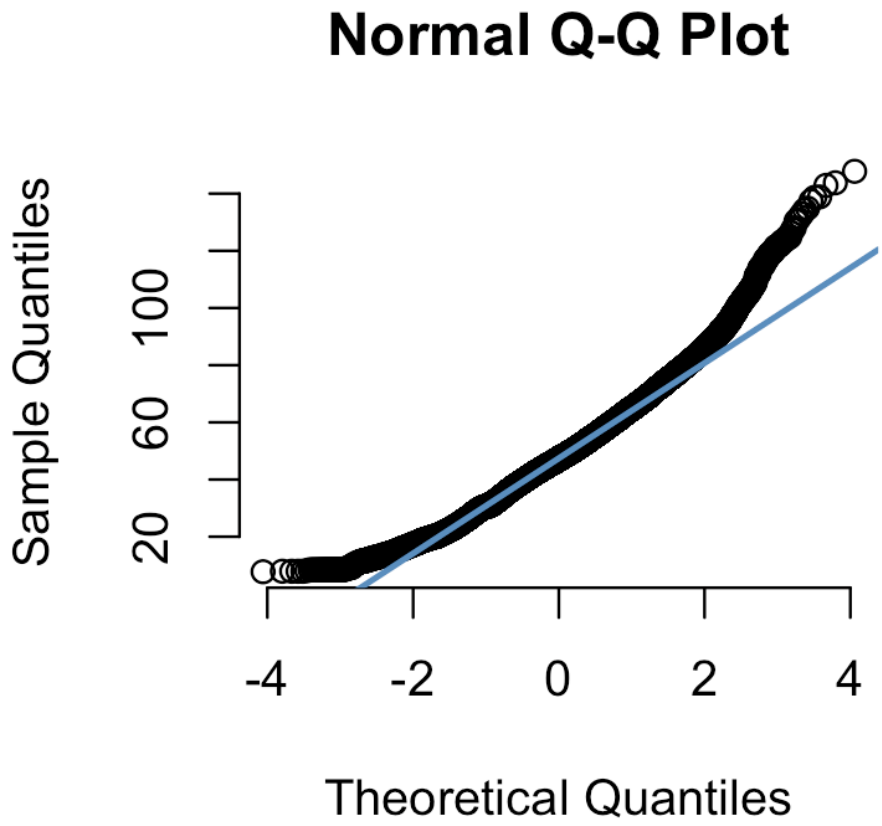




The role of data transformations:
improve normality &
homoscedasticity (another lecture)

square-root transformation



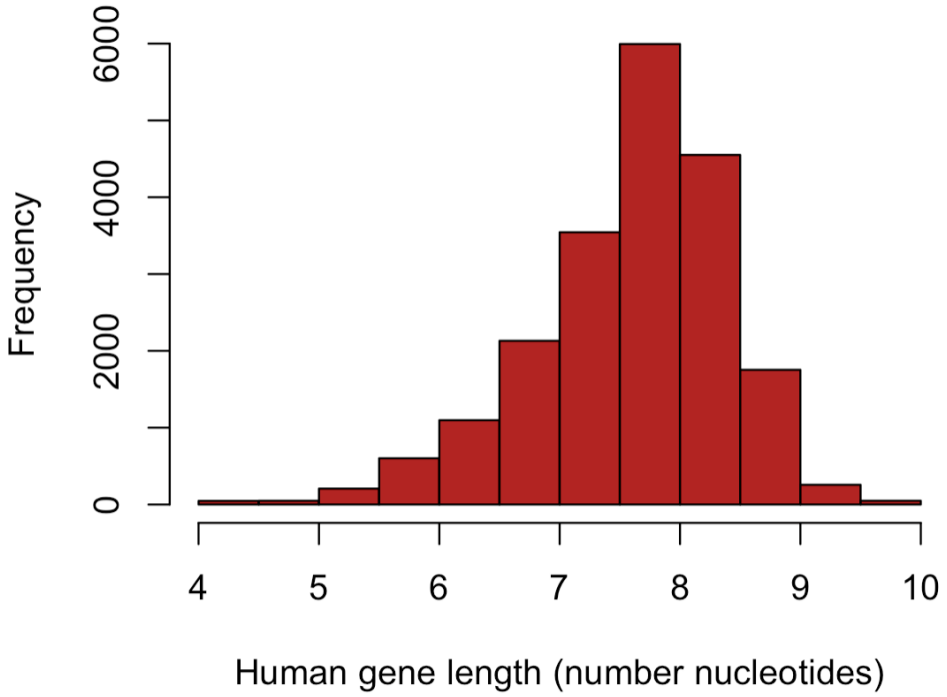


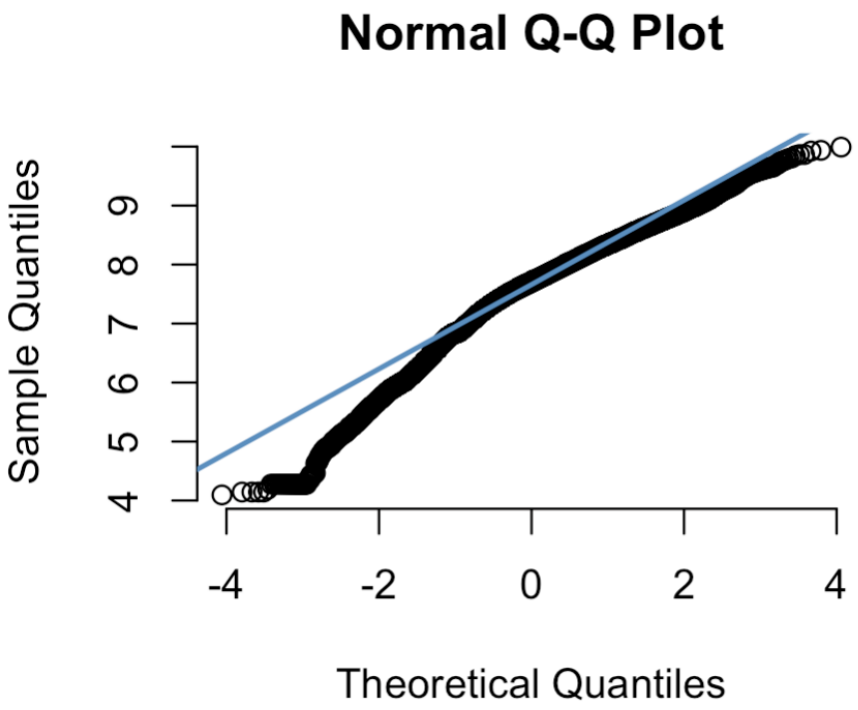
The role of data transformations:

improve normality &

homoscedasticity (another lecture)

log transformation





A few words on data transformation

One size may not fit all:

1) One transformation may help approximate normality, but another transformation may be required to approximate homoscedasticity (e.g., log(sqrt(data))).

2) One transformation may negate (reverse) the other – the one that makes the data approximate homoscedasticity may make data non-longer normal.

3) If data are complex (e.g., several predictors in a regression model), it may not be possible that one single transformation will allow data to behave properly under assumptions.

Possible solution: focus on analytical solutions (many covered in this course) and not always transformations; or combine different transformation.

A few words on data transformation

3) If data are complex (e.g., several predictors in a regression model), it may not be possible that one single transformation will allow data to behave properly under assumptions.

Possible solution: focus on analytical solutions (many covered later in the semester) and not always transformations; or combine different transformation.

The R Package trafo for Transforming Linear Regression Models

Lily Medina
Humboldt Universität zu Berlin

Piedad Castro
Humboldt Universität zu Berlin

Ann-Kristin Kreutzmann
Freie Universität Berlin

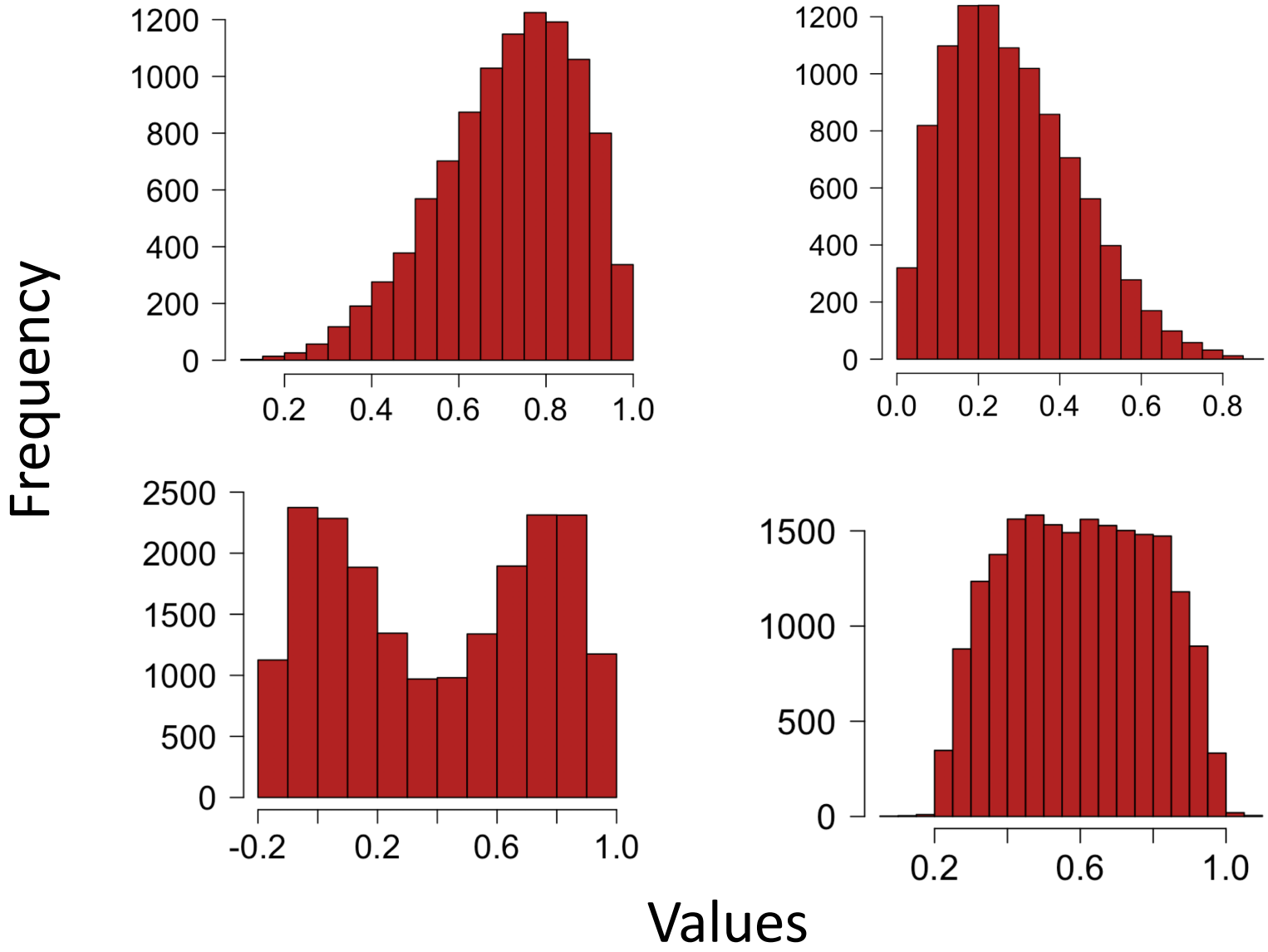
Natalia Rojas-Perilla
Freie Universität Berlin

Abstract

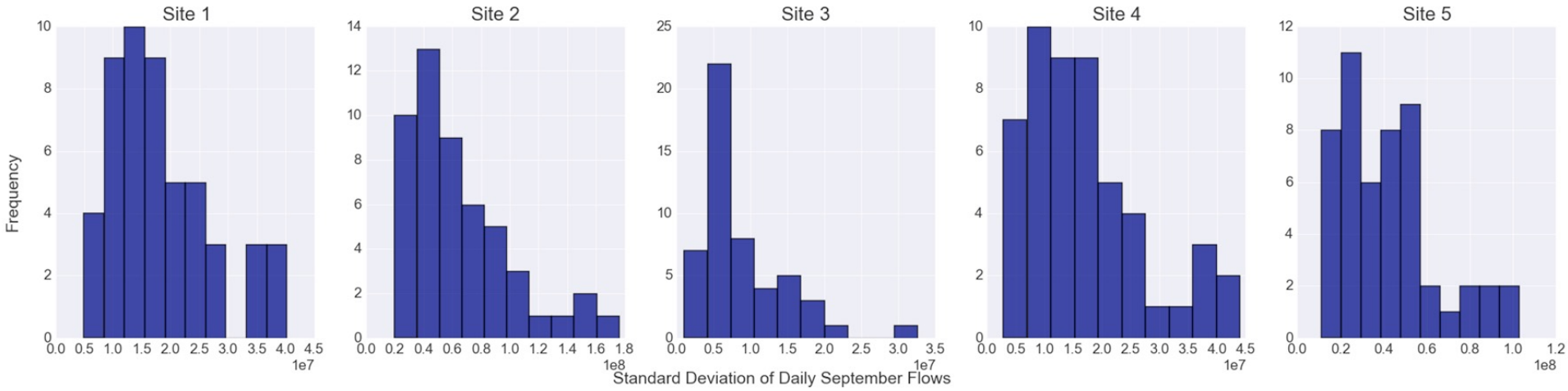
The linear regression model has been widely used for descriptive, predictive, and inferential purposes. This model relies on a set of assumptions, which are not always fulfilled when working with empirical data. In this case, one solution could be the use of more complex regression methods that do not strictly rely in the same assumptions. However, in order to improve the validity of model assumptions, transformations are a simpler approach and enable the user to keep using the well-known linear regression model. But how can a user find a suitable transformation? The R package **trafo** offers a simple user-friendly framework for selecting a suitable transformation depending on the user needs. The collection of selected transformations and estimation methods in the package **trafo** complement and enlarge the methods that are existing in R so far.

The effects of non-normality on statistical inference

Dealing with non-normality in statistical inference - hypothesis testing



Dealing with non-normality in statistical inference – hypothesis testing



Non-normal distributions have many shapes and would be quite hard to develop sampling distributions for all these different shapes

(though it can and has been done in more advanced analysis)

The effects of non-normality on statistical test

Parametric tests assuming normality (e.g., t-test & ANOVA) are often robust against non-normality; but depending on the type of non-normality (shape), parametric tests can have type I errors different (often greater) from alpha; and also low power (increased type II error).

One challenge is to separate normality from heteroscedasticity issues (even in simulations).

The other challenge is when samples come all from populations with different distributions (even though they could have the same means, i.e., H_0 is true).

The effects of non-normality on statistical test

Parametric tests assuming normality (e.g., t-test & ANOVA) are often robust against non-normality; but depending on the type of non-normality (shape of the distribution), parametric tests can have type I errors (false positives) that differ (often greater) from alpha; and low power (increased type II error; false negatives).

Br J Math Stat Psychol. 2013 May;66(2):224-44. doi: 10.1111/j.2044-8317.2012.02047.x. Epub 2012 May 24.

The impact of sample non-normality on ANOVA and alternative methods.

Lantz B[1].

Author information

Abstract

In this journal, Zimmerman (2004, 2011) has discussed preliminary tests that researchers often use to choose an appropriate method for comparing locations when the assumption of normality is doubtful. The conceptual problem with this approach is that such a two-stage process makes both the power and the significance of the entire procedure uncertain, as type I and type II errors are possible at both stages. A type I error at the first stage, for example, will obviously increase the probability of a type II error at the second stage. Based on the idea of Schmider et al. (2010), which proposes that simulated sets of sample data be ranked with respect to their degree of normality, this paper investigates the relationship between population non-normality and sample non-normality with respect to the performance of the ANOVA, Brown-Forsythe test, Welch test, and Kruskal-Wallis test when used with different distributions, sample sizes, and effect sizes. The overall conclusion is that the Kruskal-Wallis test is considerably less sensitive to the degree of sample normality when populations are distinctly non-normal and should therefore be the primary tool used to compare locations when it is known that populations are not at least approximately normal.

The effects of non-normality on statistical test

Parametric tests assuming normality (e.g., t-test & ANOVA) are often robust against non-normality; but depending on the type of non-normality (shape), parametric tests can have type I errors different (often greater) from alpha and also low power (increased type II error).

What happens if the Type I error probability (rate) is ***greater*** than alpha? i.e., increase number of False Positives.

The effects of non-normality on statistical test

Parametric tests assuming normality (e.g., t-test & ANOVA) are often robust against non-normality; but depending on the type of non-normality (shape), parametric tests can have type I errors different (often greater) from alpha and also low power (increased type II error).

What happens if the Type I error probability (rate) is ***greater*** than alpha? i.e., increase number of False Positives.

What happens if the Type I error probability (rate) is ***smaller*** than alpha? decrease False Positives but also decrease True Positives (i.e., lower statistical power).

Type I versus Type II errors – the "common" view

A **Type** I **error (false positive)** is an **error** in every sense of the word. A conclusion is drawn that the null hypothesis is false when, in fact, it is true.

Therefore, **Type** I errors are generally considered more serious than **Type II** errors (false negatives).

Type II errors are often considered as "oh well, we were not able to detect an effect"…perhaps increase sample size!

Adapted from http://davidmlane.com/hyperstat/A2917.html

Type I versus Type II errors – the "common" view

A **Type** I **error (false positive)** is an **error** in every sense of the word. A conclusion is drawn that the null hypothesis is false when, in fact, it is true.

Therefore, **Type** I errors are generally considered more serious than **Type II** errors (false negatives). Type II errors are often considered as "oh well, we were not able to detect an effect"…perhaps increase sample size!

Adapted from http://davidmlane.com/hyperstat/A2917.html

When committing a type I error, you are stating that something that is false to be true.

CONFUSING: When committing at type II error, you are NOT stating that something that is true to be false (you are just not discovering something new).

Non-parametric tests based on ranks are those that can handle non-normal data

These are the main tests traditionally used in Biology for comparing samples:

1) For comparing two samples (analogue of the parametric two sample t-test) – ***The Mann–Whitney U-test*** (also known as the Mann–Whitney–Wilcoxon test, the Wilcoxon rank-sum test, or the Wilcoxon two-sample test).

Non-parametric tests based on ranks are those that can handle non-normal data

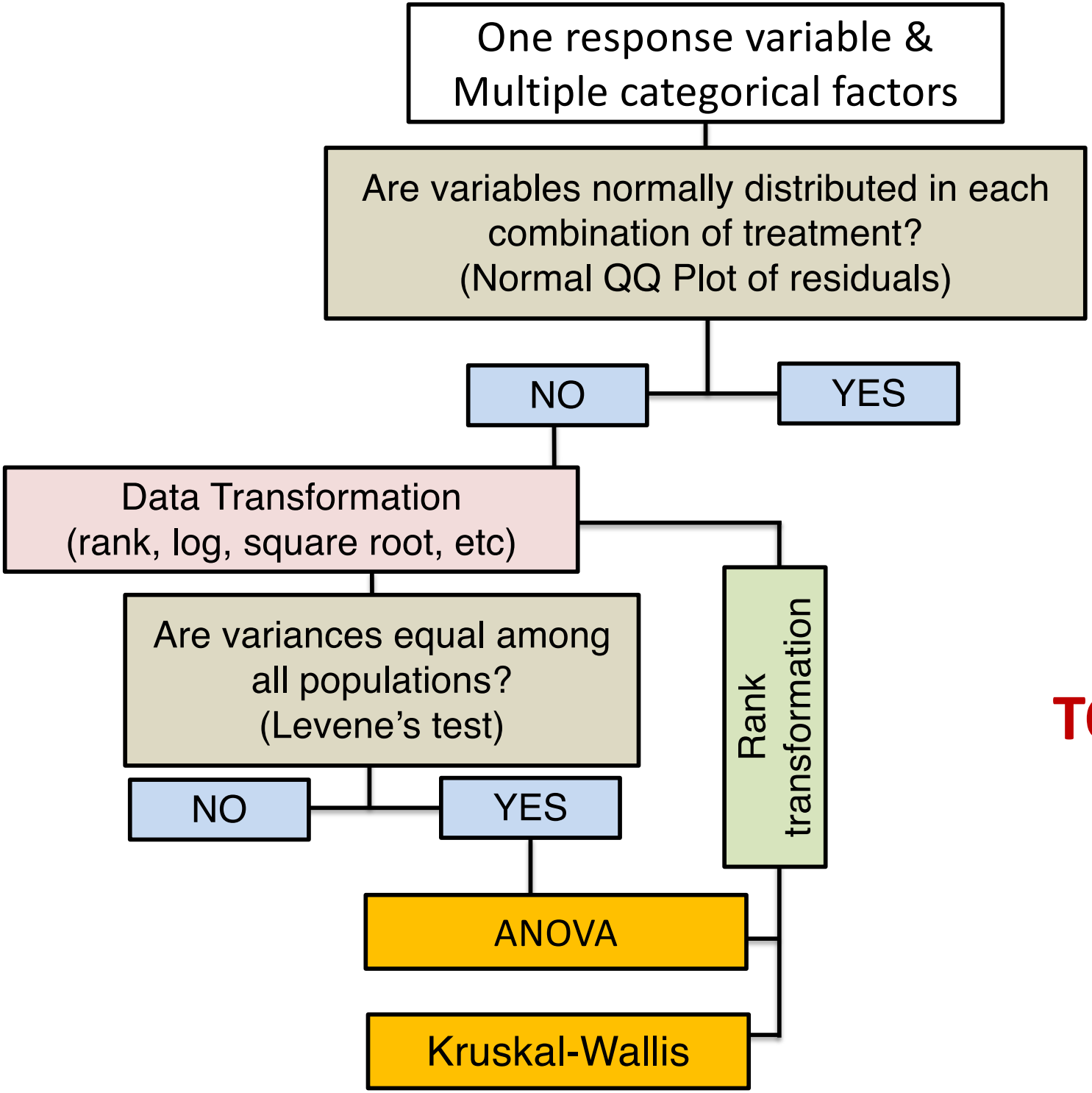
These are the main tests traditionally used in Biology for comparing samples:

1) For comparing two samples (analogue of the parametric two sample t-test) – ***The Mann–Whitney U-test*** (also known as the Mann–Whitney–Wilcoxon test, the Wilcoxon rank-sum test, or the Wilcoxon two-sample test).

2) For comparing multiple samples (analogue of the parametric ANOVA) – ***The Kruskal-Wallis test*** (generalization of the U-test)

The P-value for the ***The Mann–Whitney U-test and the The Kruskal-Wallis test*** is mathematically the same; as such, we will cover only the latter.

Note: remember that $t^2 = F$; we often cover t-tests (and not only ANOVAs) in courses for two main reasons – [1] one sample t-tests; [2] understand the nature of post-hoc testing (e.g., post-hoc pairwise comparisons of means after ANOVA and because there is a t-test dealing with samples when their populations differ in their variances).


One response variable &
Multiple categorical factors
Are variables normally distributed in each combination of treatment?
(Normal QQ Plot of residuals)
NO
YES
Data Transformation
(rank, log, square root, etc)
Are variances equal among all populations?
(Levene's test)
NO
YES
Rank transformation
ANOVA
Kruskal-Wallis

TODAY

Many non-parametric tests are based on rank transformations

gene	class	F_{ST}
CVJ5	DNA	-0.006
CVB1	DNA	-0.005
6Pgd	protein	-0.005
Pgi	protein	-0.002
CVL3	DNA	0.003
Est-3	protein	0.004
Lap-2	protein	0.006
Pgm-1	protein	0.015
Aat-2	protein	0.016
Adk-1	protein	0.016
Sdh	protein	0.024
Acp-3	protein	0.041
Pgm-2	protein	0.044
Lap-1	protein	0.049
CVL1	DNA	0.053
Mpi-2	protein	0.058
Ap-1	protein	0.066
CVJ6	DNA	0.095
CVB2m	DNA	0.116
Est-1	protein	0.163

Example: F_{ST} is a measure of the amount of geographic variation in a genetic polymorphism. Here, McDonald et al. (1996) compared two populations of the American oyster regarding the F_{ST} based on six anonymous DNA polymorphisms (variation in random bits of DNA of no known function) and compared them to F_{ST} values on 13 proteins.

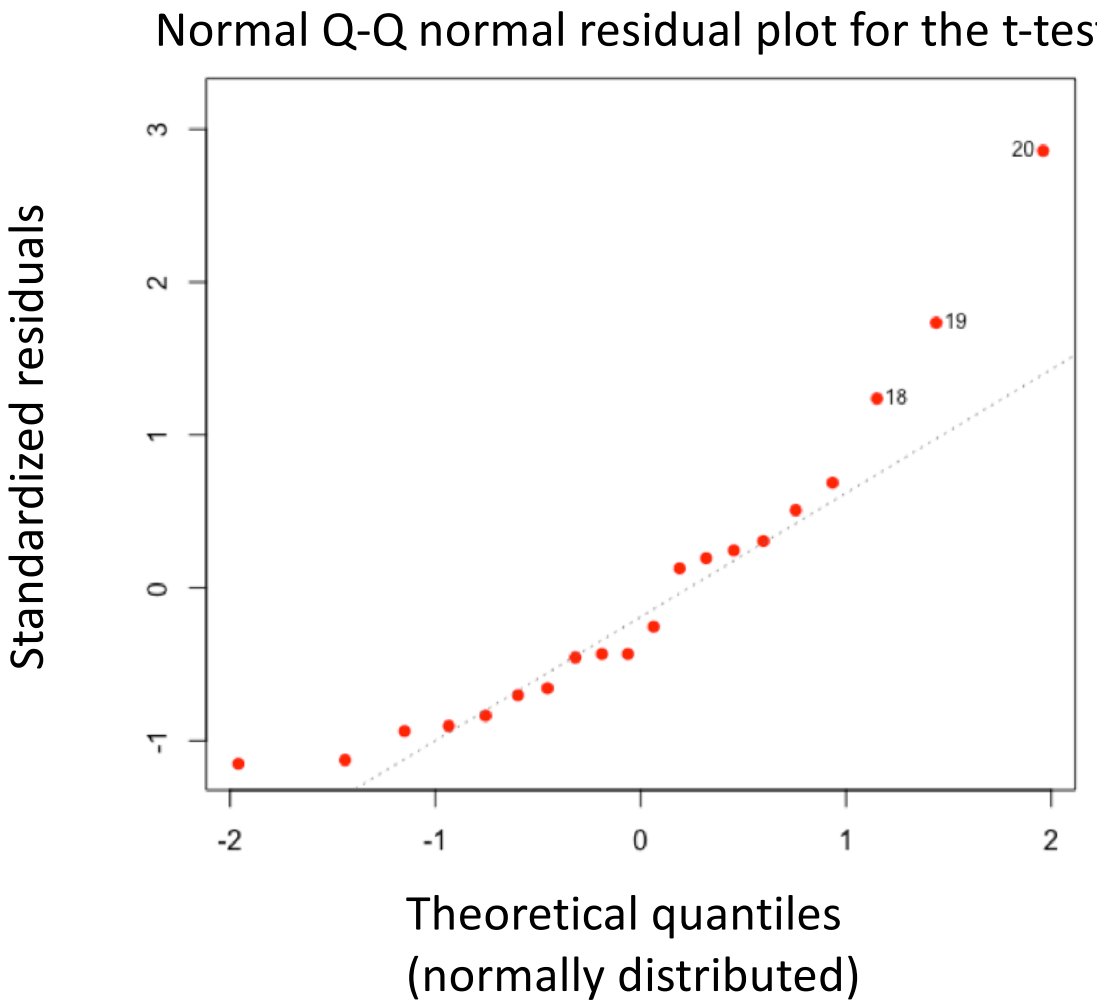
Question: Do protein differ in F_{ST} values in contrast to anonymous DNA polymorphisms?

Zero F_{ST} = no genetic variation (panmictic)
negative F_{ST} = more genetic variation within populations than between the two populations being compared.
positive F_{ST} = more variation between populations than within the two populations being compared.

http://www.biostathandbook.com/kruskalwallis.html

Data from McDonald et al. (1996)

F_{st} data highly non-normal, so transformation is advised; let's apply the rank transformation



Many non-parametric tests are based on rank transformations

gene	class	F_{ST}	Rank	Rank
CVJ5	DNA	-0.006	1	
CVB1	DNA	-0.005	2.5	
6Pgd	protein	-0.005		2.5
Pgi	protein	-0.002		4
CVL3	DNA	0.003	5	
Est-3	protein	0.004		6
Lap-2	protein	0.006		7
Pgm-1	protein	0.015		8
Aat-2	protein	0.016		9.5
Adk-1	protein	0.016		9.5
Sdh	protein	0.024		11
Acp-3	protein	0.041		12
Pgm-2	protein	0.044		13
Lap-1	protein	0.049		14
CVL1	DNA	0.053	15	
Mpi-2	protein	0.058		16
Ap-1	protein	0.066		17
CVJ6	DNA	0.095	18	
CVB2m	DNA	0.116	19	
Est-1	protein	0.163		20

$(2+3)/2=2.5$

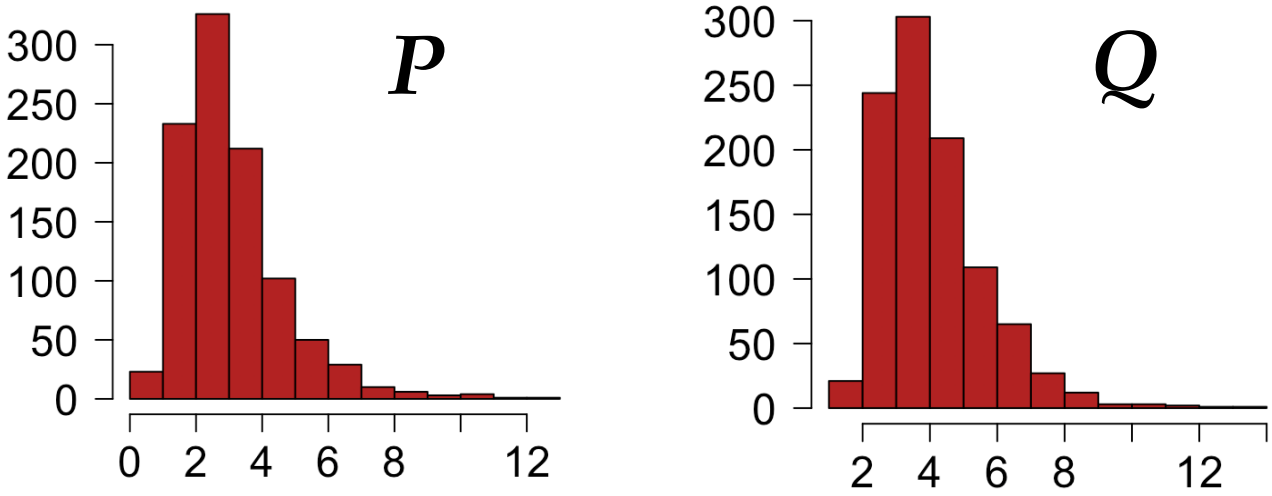
$(9+10)/2=9.5$

http://www.biostathandbook.com/kruskalwallis.html

Data from McDonald et al. (1996)

We want to know whether samples come from statistical populations that vary in their ranks

What is the probability that a randomly sampled observation from population ***P*** is greater (or smaller) in rank than a randomly sampled observation from ***Q***?
If the probability is small, then the samples come from different populations!



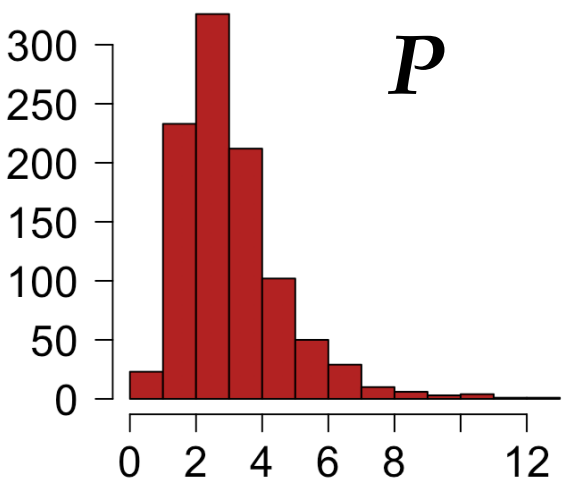
Varga and Delanay (1998)

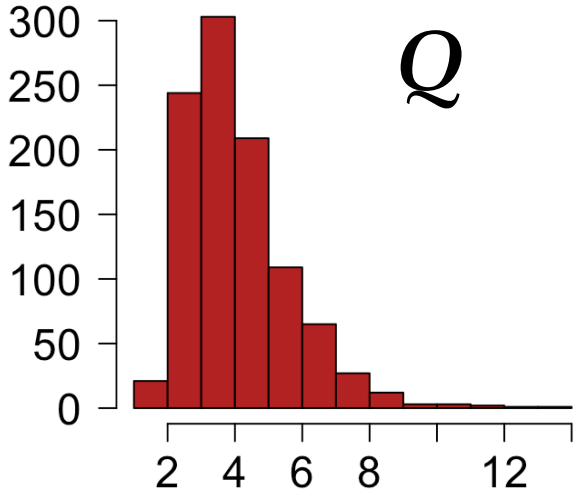
Original values for each population

We want to know whether samples come from statistical populations that vary in their ranks – example from two large samples

What is the probability that a randomly sampled observation from population ***P*** is greater (or smaller) in rank than a randomly sampled observation from ***Q***?
If the probability is small, then the samples come from different populations!

Varga and Delanay (1998)

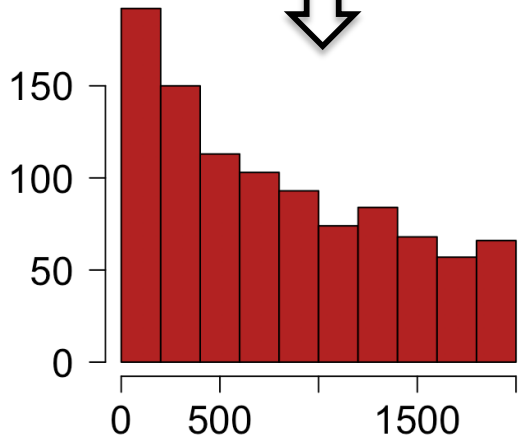


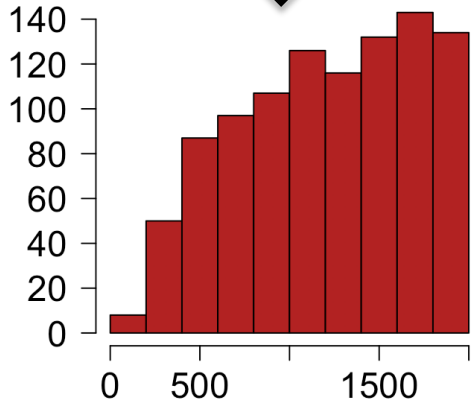


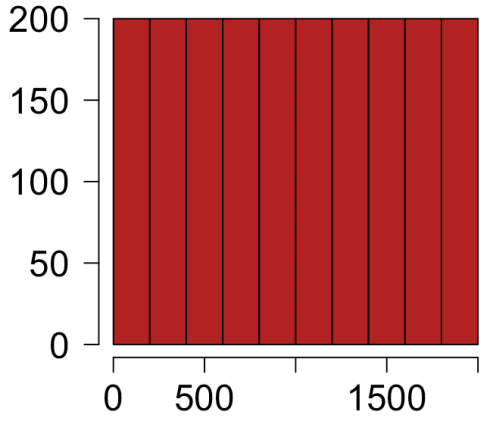
Original values for each population

Two distributions of ranks combined (always uniform)

rank-transformation



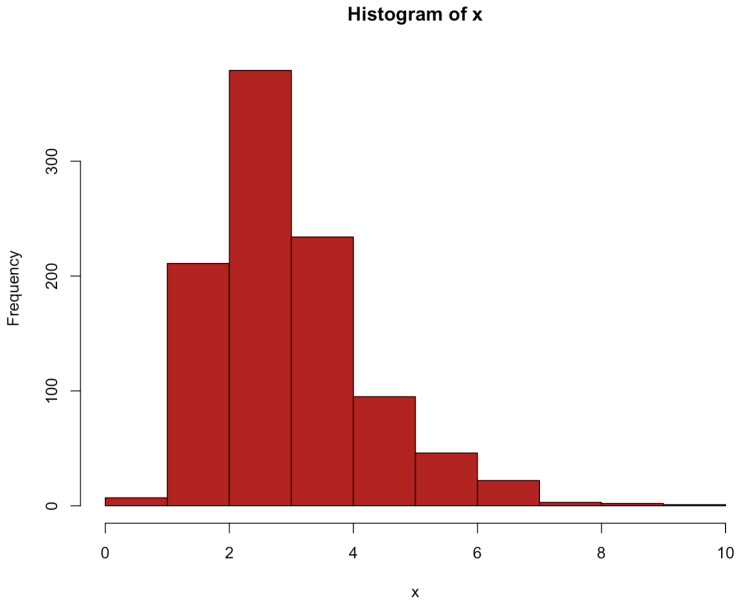


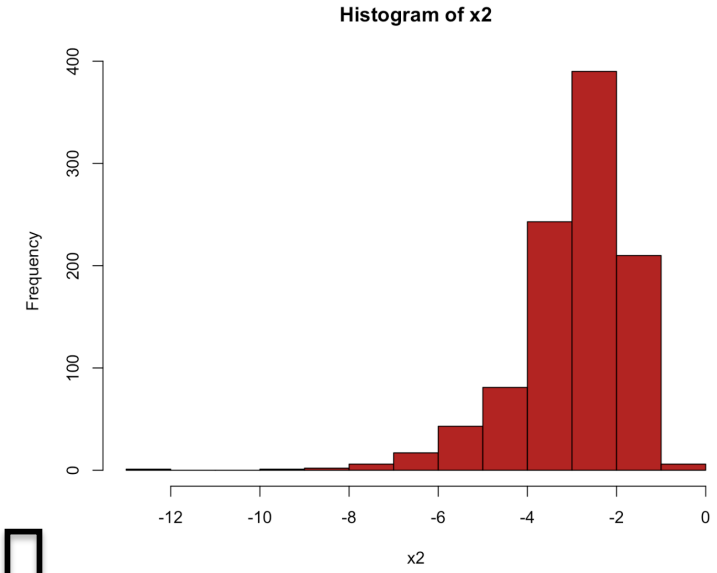


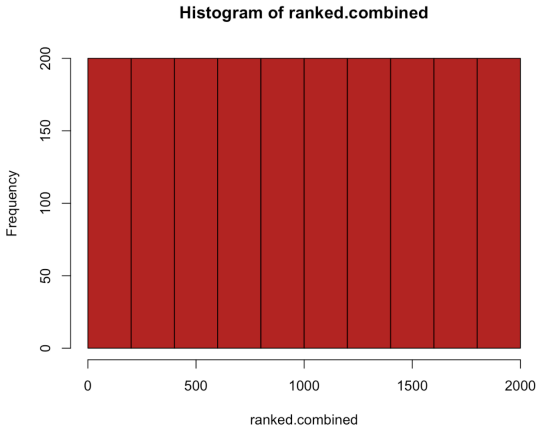
```r
x <- rlnorm(1000,1,0.4)
hist(x,col="friebrick")
x2 <- -rlnorm(1000,1,0.4)
hist(x2,col="friebrick")

ranked.combined <- rank(c(x,x2))
hist(ranked.combined,col="friebrick")
```

Two distributions of ranks combined (always uniform)

Let's see that "manually" using R code

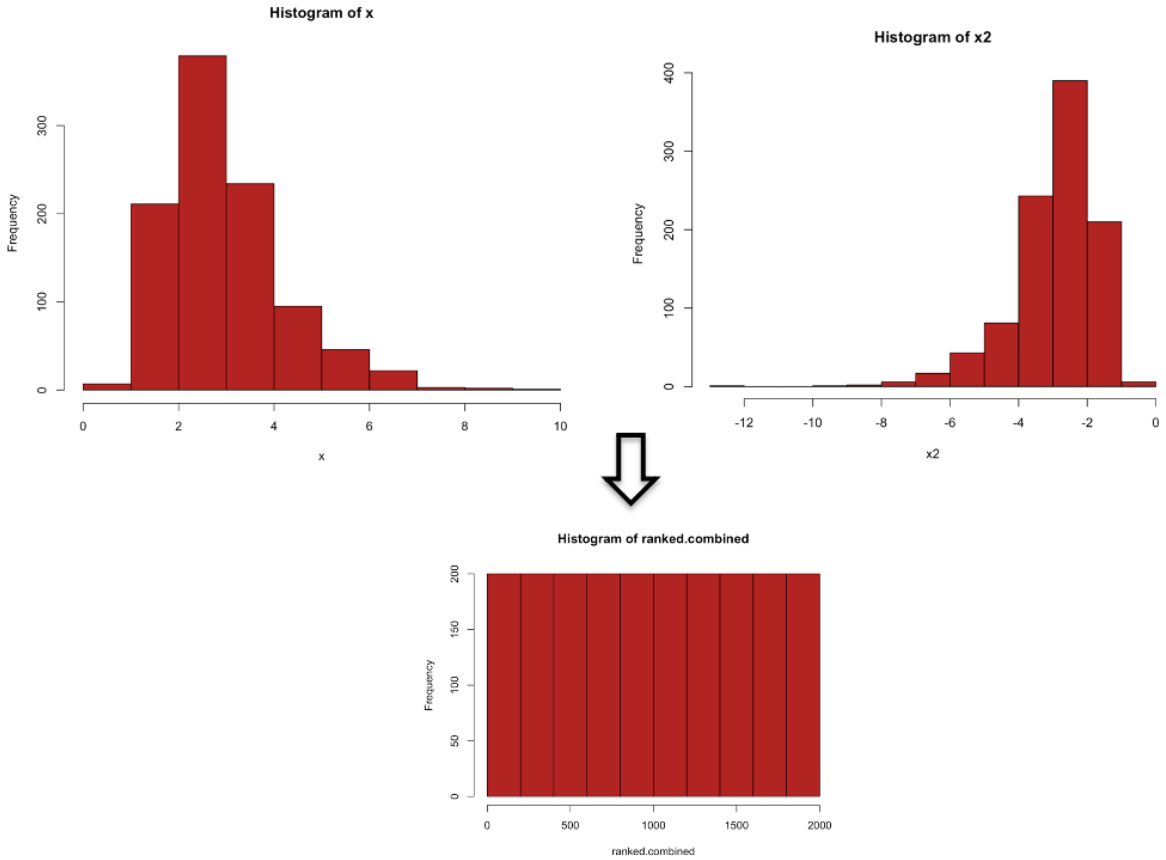






Ranked-based statistical tests remove the natural ways we think about the original units of the variables of interest

and they also reduce statistical power to detect true differences, i.e., increase type II error (false negatives).



Rank based tests

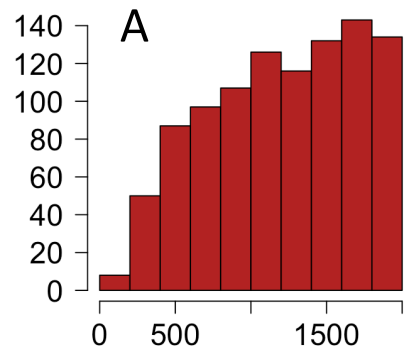
Kruskal-Wallis test
(akin to one-factorial ANOVA but based on ranks)

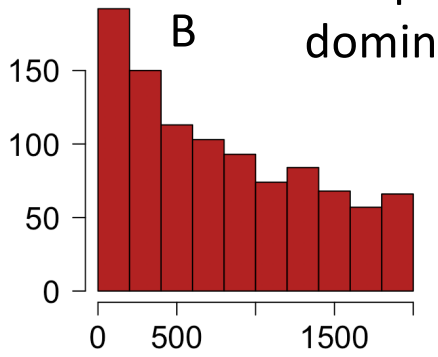
Ho: no population from where the samples were taken stochastically dominates another population (stochastic homogeneity).

Ha: at least one population from where the sample was taken stochastically dominates another population (stochastic heterogeneity).

⇩

Which sample? Post-hoc tests (based on ranks)





Sample A stochastically dominates sample B

Kruskal-Wallis test
(akin to one-factorial ANOVA but based on ranks)

H_0: no population from where the samples were taken stochastically dominates another population (stochastic homogeneity).

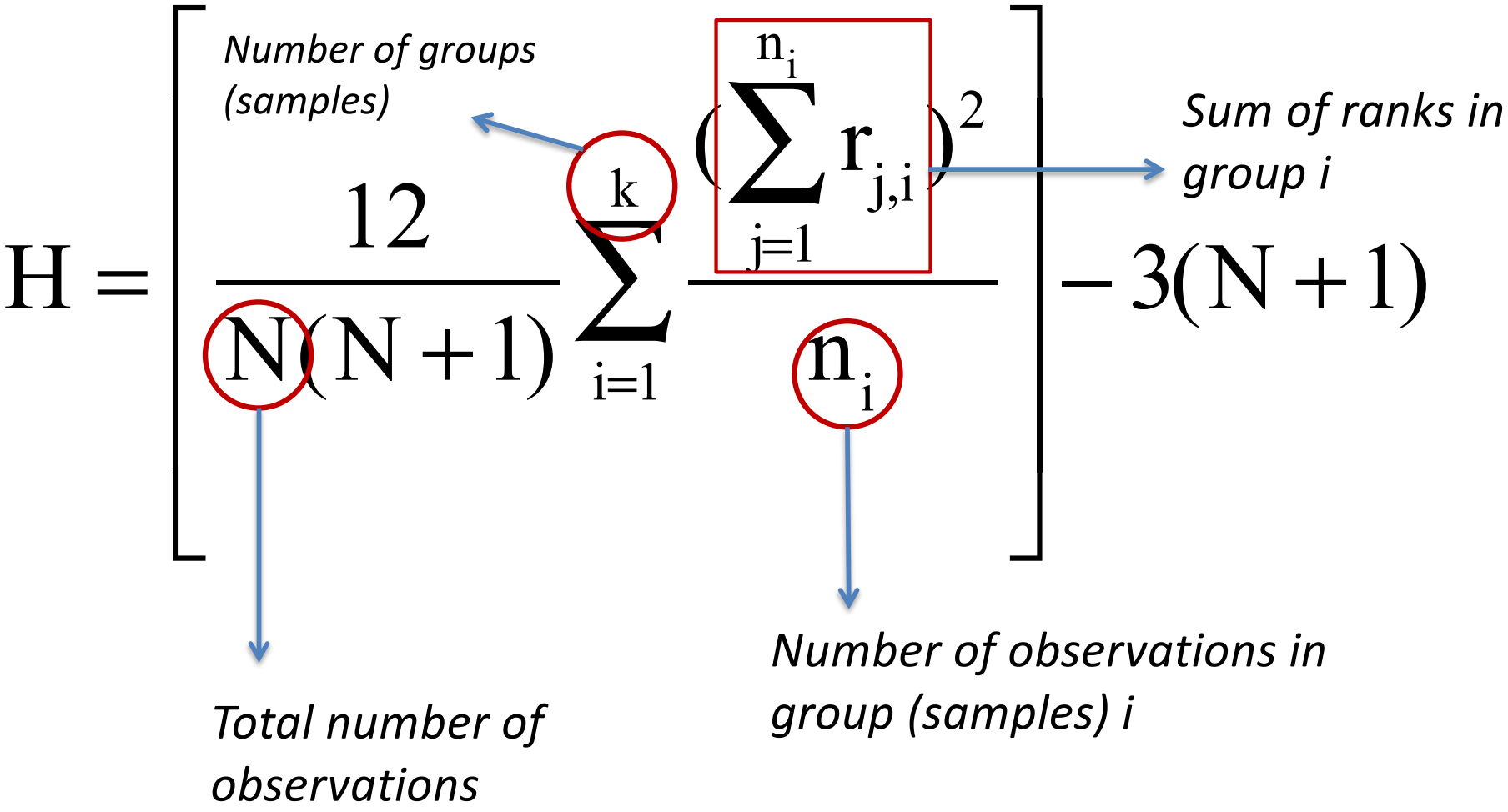
H_A: at least one population from where the sample was taken stochastically dominates another population (stochastic heterogeneity).

F_{STs} data

H_0 : DNA and protein do not stochastically dominate each other in their F_{STs}.

H_A : Either DNA or protein stochastically dominate each other in their F_{STs}.

Kruskal-Wallis test – statistic H

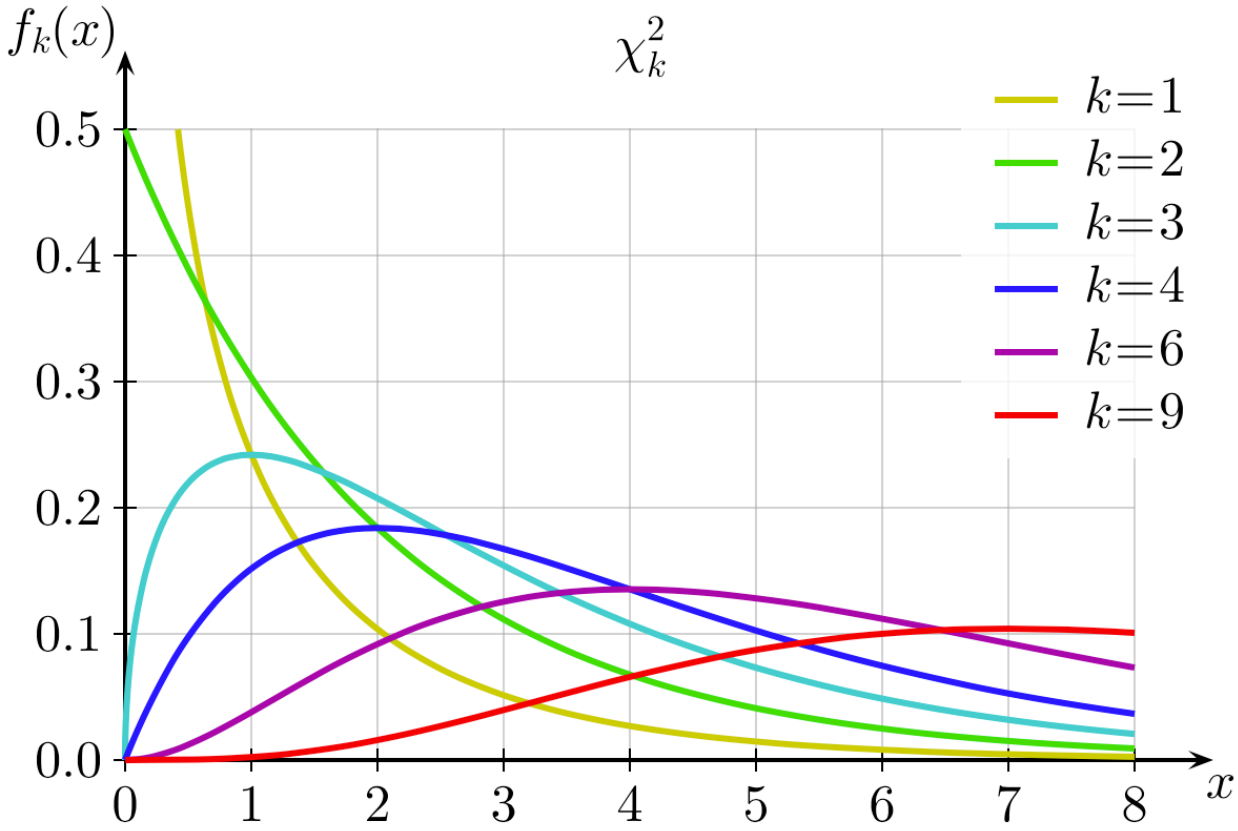


No need to memorize or understand this formula (F much more important) – but I think is relevant to understand that statisticians spend serious time on these formulae (or formulas).

Kruskal-Wallis test – statistic H

No need to memorize or understand this formula (keep your "energy" for F if you want to).

But I think is relevant to understand that statisticians spend serious time on those.



$$H = \left[\frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{n_i} r_{j,i} \right)^2}{n_i} \right] - 3(N+1)$$

Number of groups (samples)

Sum of ranks in group i

Total number of observations

Number of observations in group (samples) i

Equations also demonstrate the work others do to make test statistics (H here) to be contrastable to existing probability distributions (chi-square in this case)

Kruskal-Wallis test – statistic H

gene	class	F_{ST}	Rank	Rank
CVJ5	DNA	-0.006	1	
CVB1	DNA	-0.005	2.5	
6Pgd	protein	-0.005		2.5
Pgi	protein	-0.002		4
CVL3	DNA	0.003	5	
Est-3	protein	0.004		6
Lap-2	protein	0.006		7
Pgm-1	protein	0.015		8
Aat-2	protein	0.016		9.5
Adk-1	protein	0.016		9.5
Sdh	protein	0.024		11
Acp-3	protein	0.041		12
Pgm-2	protein	0.044		13
Lap-1	protein	0.049		14
CVL1	DNA	0.053	15	
Mpi-2	protein	0.058		16
Ap-1	protein	0.066		17
CVJ6	DNA	0.095	18	
CVB2m	DNA	0.116	19	
Est-1	protein	0.163		20
Sum			***60.5***	***149.5***

$$H = \left[\frac{12}{20(20+1)} * \sum_{i=1}^{2} \frac{(\sum_{j=1}^{n_i} r_{j,i})^2}{n_i} \right] - 3(20+1)$$

$$H = \left[\frac{12}{20(20+1)} * \left(\frac{60.5^2}{6} + \frac{149.5^2}{14} \right) \right] - 3(20+1)$$

$$H = \left[0.029 * (610.04 + 1596.45) \right] - 63 =$$

$$H = 0.0425$$

Kruskal-Wallis test – statistic H

gene	class	F_{ST}	Rank	Rank
CVJ5	DNA	-0.006	1	
CVB1	DNA	-0.005	2.5	
6Pgd	protein	-0.005		2.5
Pgi	protein	-0.002		4
CVL3	DNA	0.003	5	
Est-3	protein	0.004		6
Lap-2	protein	0.006		7
Pgm-1	protein	0.015		8
Aat-2	protein	0.016		9.5
Adk-1	protein	0.016		9.5
Sdh	protein	0.024		11
Acp-3	protein	0.041		12
Pgm-2	protein	0.044		13
Lap-1	protein	0.049		14
CVL1	DNA	0.053	15	
Mpi-2	protein	0.058		16
Ap-1	protein	0.066		17
CVJ6	DNA	0.095	18	
CVB2m	DNA	0.116	19	
Est-1	protein	0.163		20
Sum			***60.5***	***149.5***

$$H = \left[0.029 * (610.04 + 1596.45)\right] - 63 =$$

$$H = 0.0425$$

Correction for ties

Number of ties

Number of values from a set of ties

$$C_H = 1 - \frac{\sum_{i=1}^{n_T} (T_i^3 - T_i)}{N^3 - N}$$

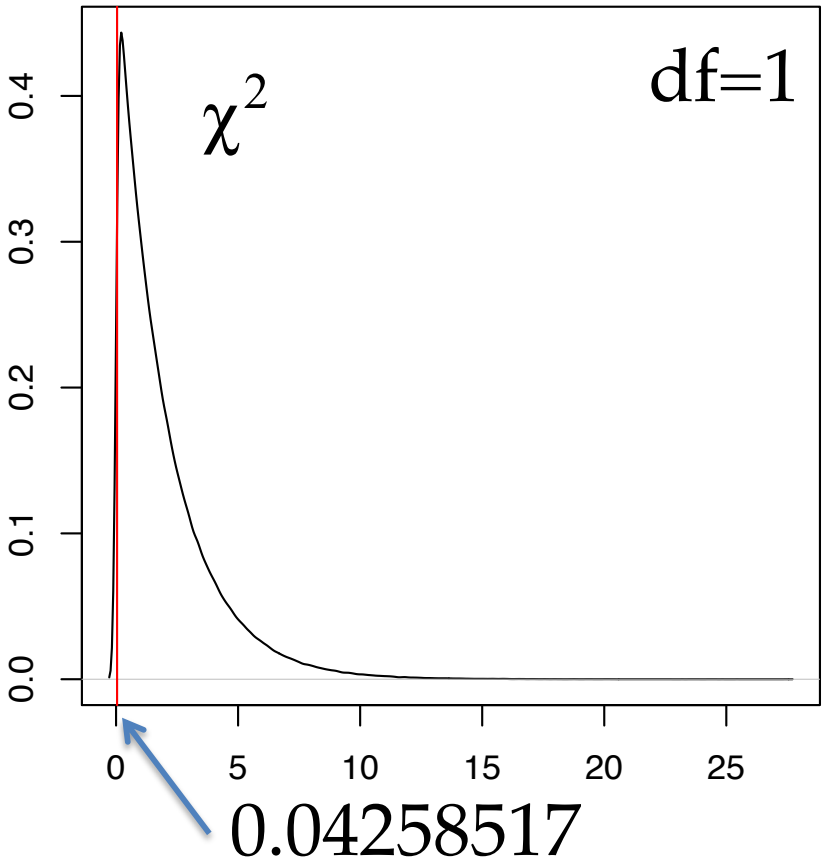
$$C_H = 1 - \frac{\sum_{i=1}^{2} (T_i^3 - T_i)}{20^3 - 20} = 1 - \frac{(2^3 + 2) + (2^3 + 2)}{20^3 - 20} = 0.998$$

$$H_c = H / C_H = 0.0425 / 0.998 = 0.04258517$$

Kruskal-Wallis test – statistic H

$$H_c = H / C_H = 0.0425 / 0.998 = 0.04258517$$

For small samples sizes (n <= 5), a special H distribution needs to be used (though R does not have it and uses the standard X^2); if n > 5, then H follows a chi-square distribution with (k-1) degrees of freedom (df=2-1=1)



P=0.8365;
probability of finding by chance an H_c greater than the observed when assuming that H_0 is true.

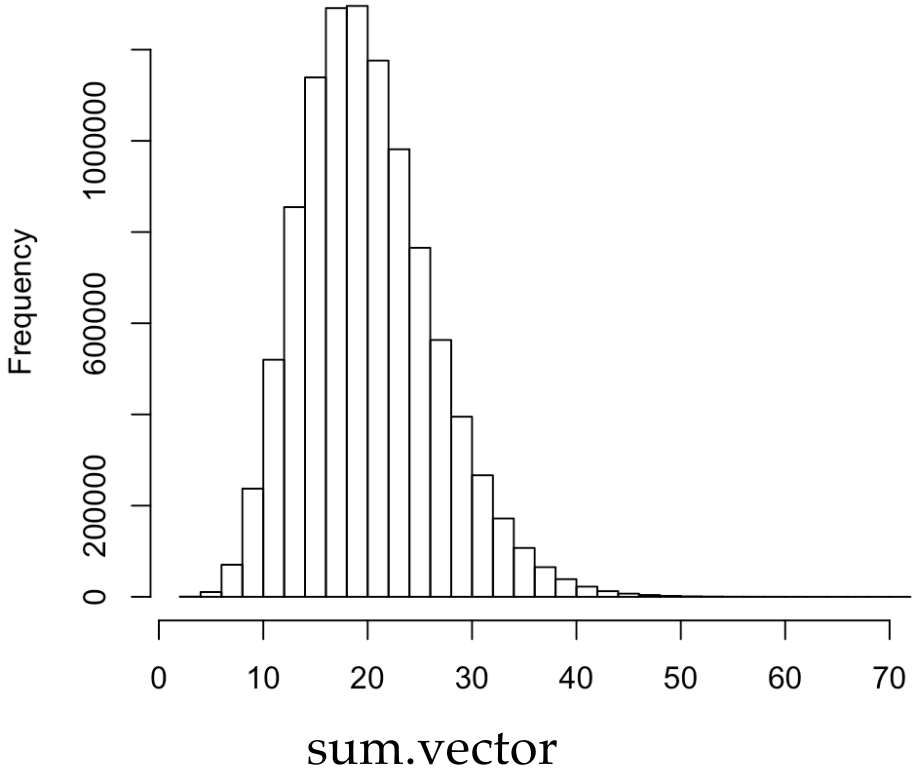
Fun fact: **The chi-square distribution is the distribution of the sum of squared standard normal deviates.**

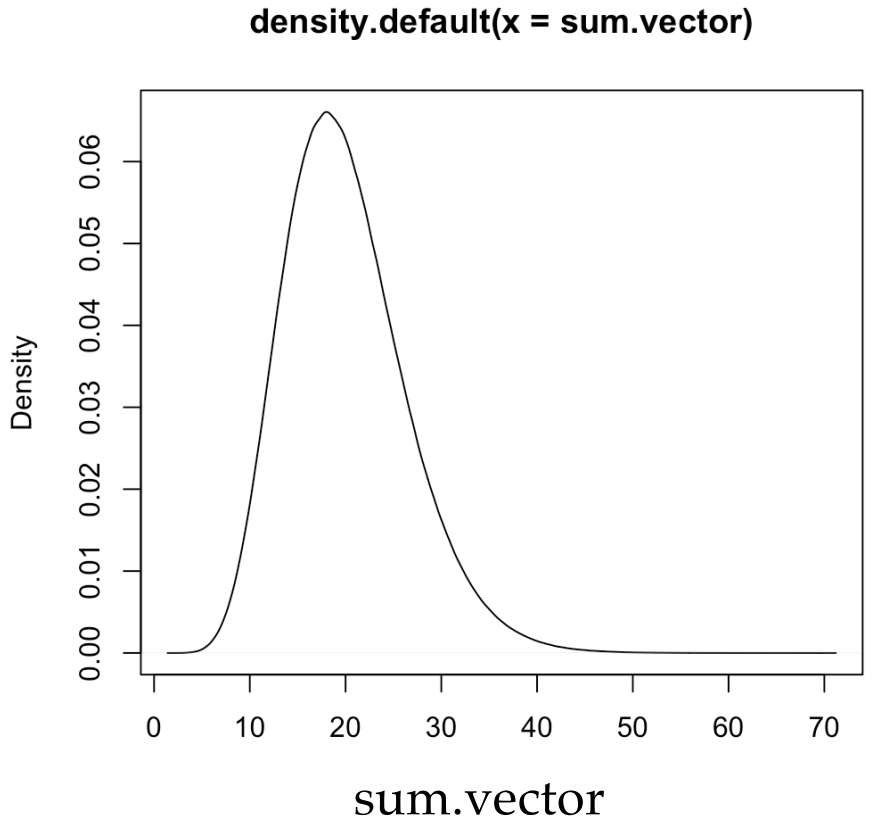
Good place to generate more intuition about statistical distributions!

R code to generate the chi-square computationally *versus* analytically for 20 degree of freedom

```
> samples <- replicate(1000000, rnorm(n=20))
> sum2.vector <- apply(samples^2,2,sum)
> qchisq(.95, df=20)
[1] 31.41043
> quantile(sum2.vector,  probs = 0.95)
     95%
31.38769
```

The chi-square distribution is the distribution of the sum of squared standard normal deviates.



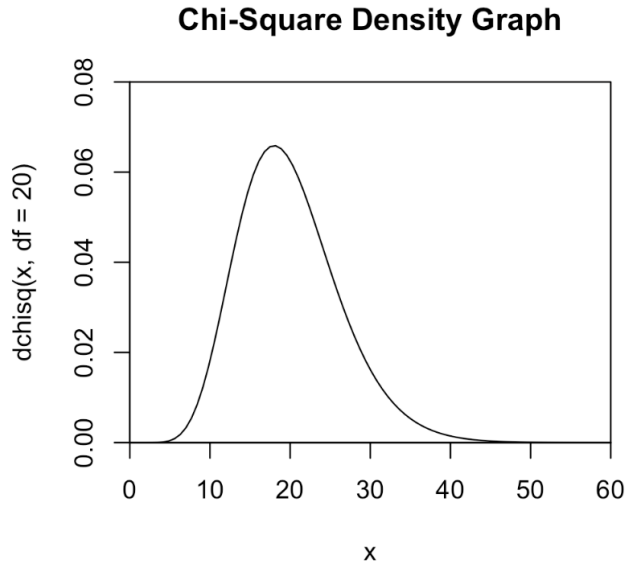


computational approach

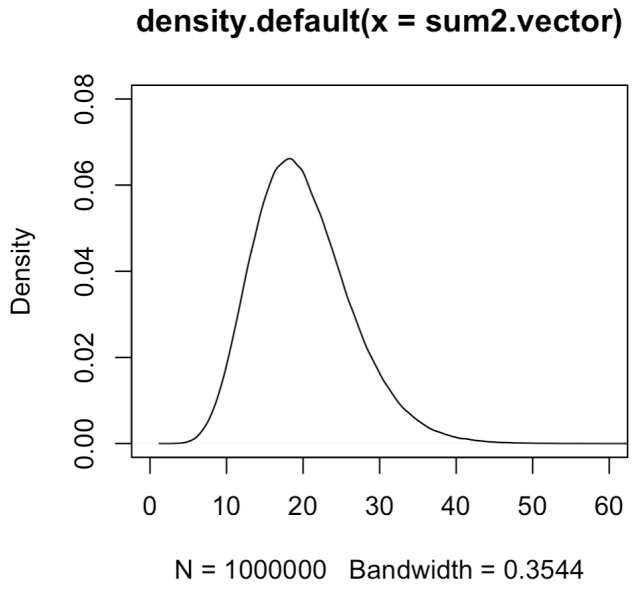
```r
samples <- replicate(1000000, rnorm(n=20))
sum2.vector <- apply(samples^2,2,sum)
plot(density(sum2.vector),xlim=c(0,60),ylim=c(0,0.08))
```

analytical approach

```r
x <- rchisq(100000000, df=20)
curve(dchisq(x, df=20), col='black', main = "Chi-Square Density Graph",
      from=0,to=70,yaxs="i",xaxs="i",xlim=c(0,60),ylim=c(0,0.08))
```



=



The chi-square distribution is the distribution of the sum of squared standard normal deviates.

fun fact: The F distribution is the ratio of two (scaled) chi-square distributed values. The scaling is done by appropriate division of degrees of freedom.

A general solution to rank-based tests

Kruskal-Wallis test is equivalent (close enough) to an ANOVA on ranks

Ho: no sample stochastically dominates another sample (stochastic homogeneity).

Ha: at least one sample stochastically dominates one other sample (stochastic heterogeneity).

*"**Stochastic homogeneity** is equivalent to the equality of the expected values of the **rank sample means**. This finding implies that the null hypothesis of stochastic homogeneity can be tested by an ANOVA performed on the rank transforms, which is essentially equivalent to doing a Kruskal-Wallis* H *test."*

Varga and Delanay (1998)

Journal of Educational and Behavioral Statistics
Summer 1998, Vol. 23, No. 2, pp. 170–192

The Kruskal-Wallis Test and Stochastic Homogeneity

András Vargha
Eötvös Loránd University
Harold D. Delaney
University of New Mexico

Kruskal-Wallis test = ANOVA on ranks

Kruskal-Wallis:

Ho: no sample stochastically dominates another sample (stochastic homogeneity).

Ha: at least one sample stochastically dominates one other sample (stochastic heterogeneity).

Varga and Delanay (1998)

ANOVA:

Ho: no mean differences in ranked values

Ha: at least one sample differs in mean ranked values from another sample

Kruskal-Wallis test = ANOVA on ranks

```
> Fst.values <- c(-0.006,-0.005,-0.005,-0.002,0.003,0.004,
                  0.006,0.015,0.016,0.016,0.024,0.041,0.044,
                  0.049,0.053,0.058,0.066,0.095,0.116,0.163)
> Fst.rank <- rank(Fst.values)
> hist(Fst.rank,col="firebrick")
> Fst.group <- c(1,1,2,2,1,2,2,2,2,2,2,2,2,1,2,2,1,1,2)
> kruskal.test(Fst.values~Fst.group)
```

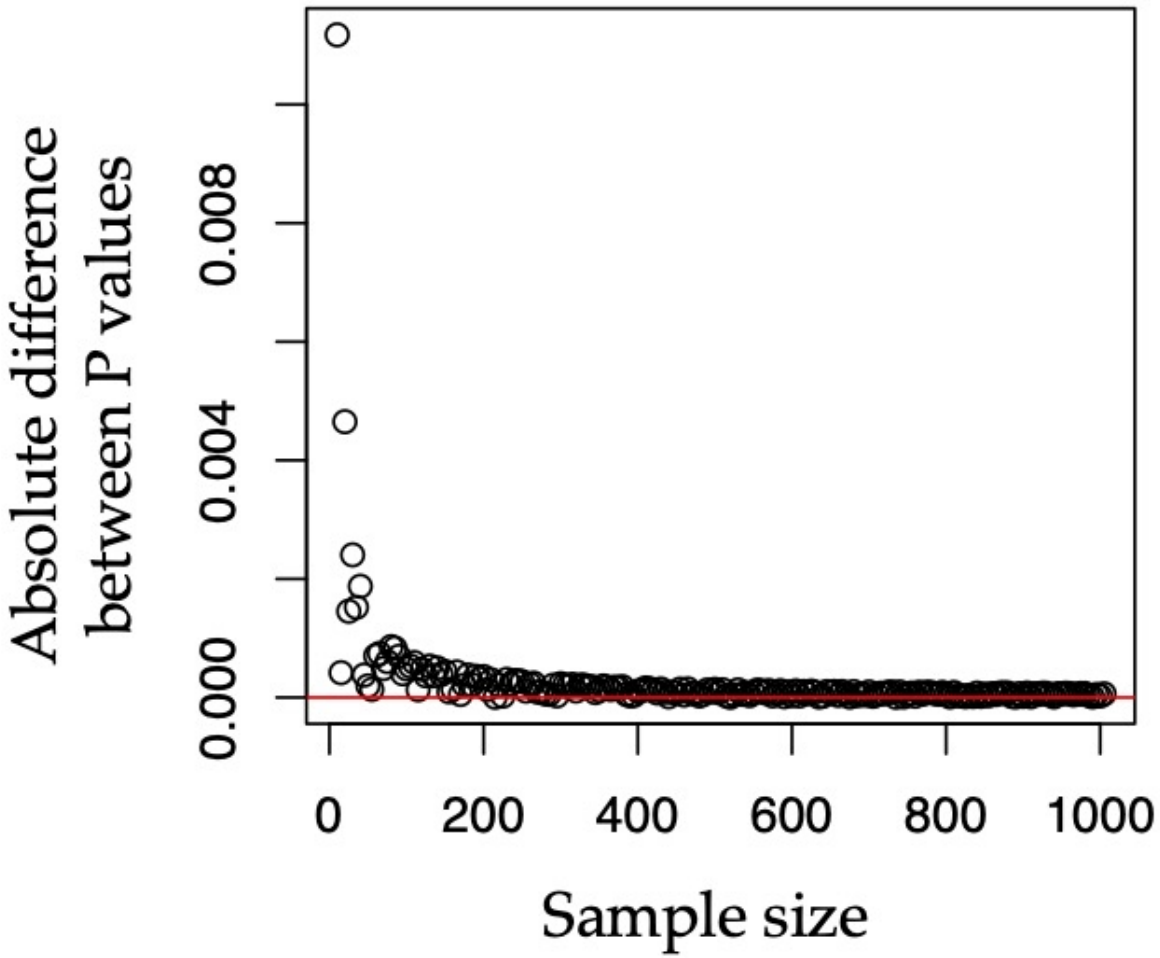
```
> kruskal.test(Fst.values~Fst.group)

    Kruskal-Wallis rank sum test

data:  Fst.values by Fst.group
Kruskal-Wallis chi-squared = 0.042581, df = 1, p-value = 0.8365
```

```
> summary(aov(Fst.rank~Fst.group))
            Df Sum Sq Mean Sq F value Pr(>F)
Fst.group    1    1.5    1.49    0.04  0.843
Residuals   18  662.5   36.81
```


They are slightly different, no?

Kruskal-Wallis test = ANOVA on ranks

Kruskal-Wallis and ANOVA are “asymptotically equivalent” (i.e., the two functions "eventually" become "essentially **equal**") and so P-values are exactly the same for very large samples and they do not differ by much for small sample size.



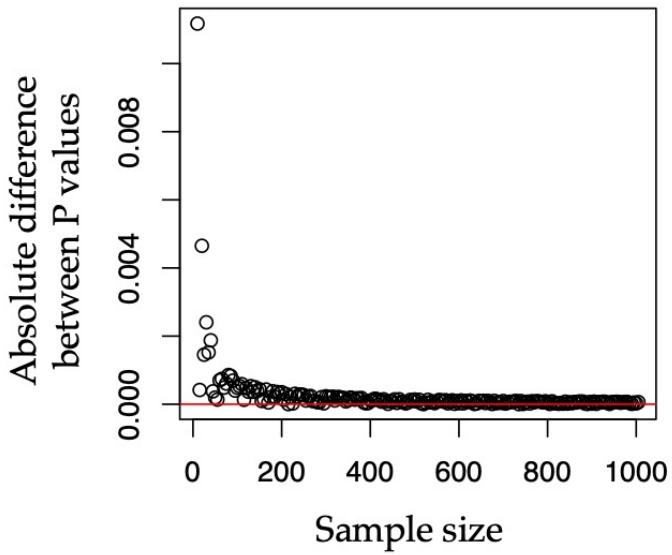
Two sample Kruskal-Wallis P-values (chi-square based) and F-based P values)

Kruskal-Wallis and ANOVA are “asymptotically equivalent” and so P-values are the same for very large samples and they do not differ by much for small sample size. Using R code to demonstrate the asymptotic equivalence.

```r
n.simul <- 200
Pvector <- matrix(0,n.simul,2)
n <- 10
n.vector <- matrix(0,n.simul,1)
for (i in 1:n.simul){
  groups <- c(rep(1,n),rep(2,n))
  x <- rnorm(n*2)
  Pvector[i,1] <- kruskal.test(x~groups)$p.value
  Pvector[i,2] <- anova(lm(rank(x)~groups))$'Pr(>F)'[1]
  n <- n + 10
  n.vector[i] <- n
}

plot(n.vector/2,abs(Pvector[,1]-Pvector[,2]))
abline(h=0,col="red")
```

Kruskal-Wallis and ANOVA are “asymptotically equivalent”

```r
n.simul <- 200
Pvector <- matrix(0,n.simul,2)
n <- 10
n.vector <- matrix(0,n.simul,1)
for (i in 1:n.simul){
        groups <- c(rep(1,n),rep(2,n))
        x <- rnorm(n*2)
        Pvector[i,1] <- kruskal.test(x~groups)$p.value
        Pvector[i,2] <- anova(lm(rank(x)~groups))$'Pr(>F)'[1]
        n <- n+10
        n.vector[i] <- n
}
plot(n.vector/2,abs(Pvector[,1]-Pvector[,2]))
abline(h=0,col="red")
```



Kruskal-Wallis test = ANOVA on ranks

Kruskal-Wallis and ANOVA are "asymptotically equivalent" and so P-values are exactly the same for very large samples and they do not differ by much for small sample size.

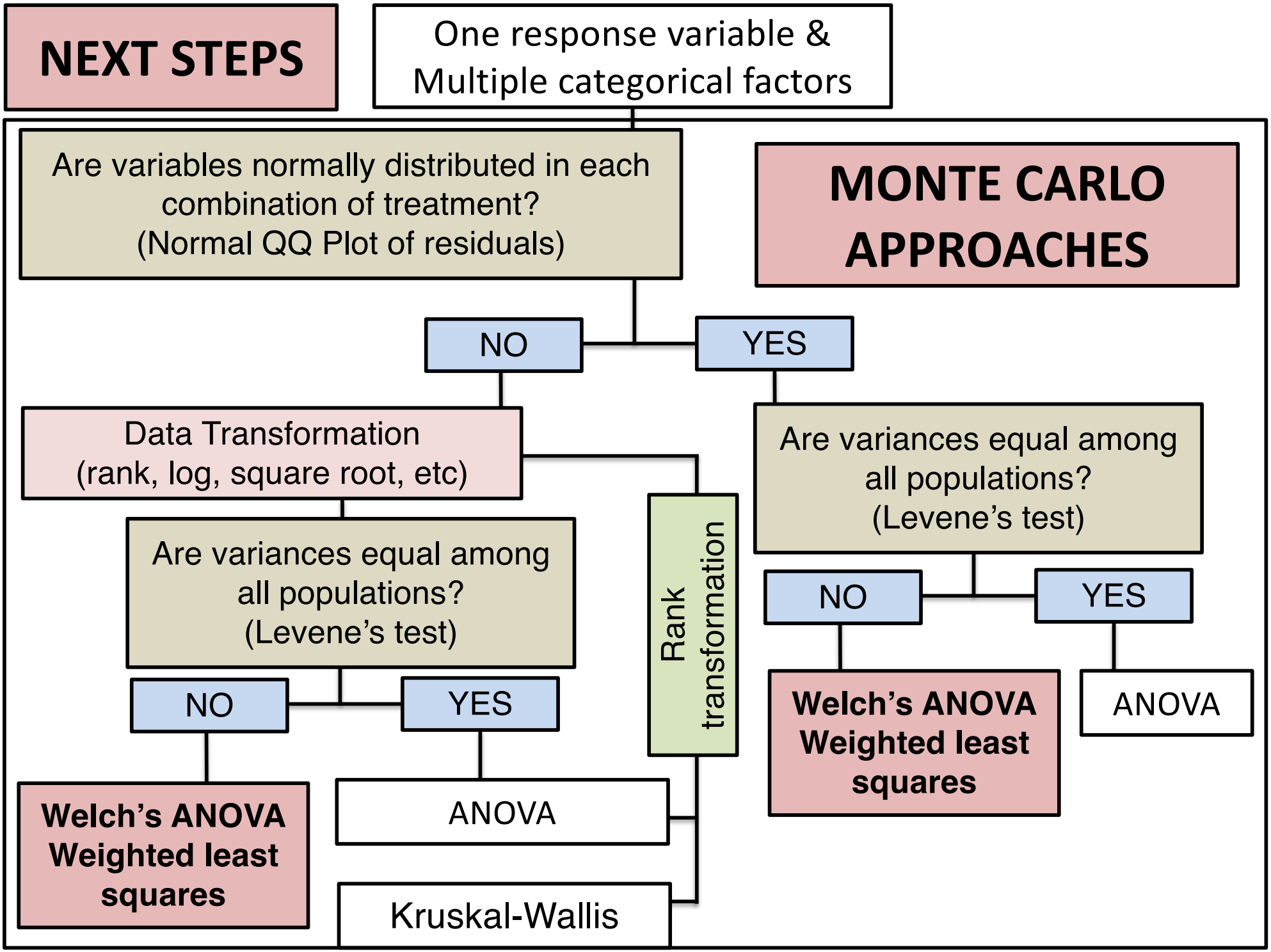
Because of the equivalence, we can then expand non-parametric analysis based on ranks to any multi-factorial ANOVAs, regressions, MANOVA, ANCOVA, etc

NOTE: Non-parametric tests are those that can handle non-normal data

There is a common misunderstanding in the statistical literature and among practitioners, including many biostatistics books, that non-parametric tests can also handle differences in variances among samples.

THIS IS NOT TRUE! They are also affected by variance differences among groups/treatments (i.e., homoscedasticity).

Test variance differences in ranks (almost never done in the literature)!


NEXT STEPS
One response variable & Multiple categorical factors
Are variables normally distributed in each combination of treatment? (Normal QQ Plot of residuals)
MONTE CARLO APPROACHES
NO
YES
Data Transformation (rank, log, square root, etc)
Are variances equal among all populations? (Levene's test)
Rank transformation
NO
YES
Welch's ANOVA Weighted least squares
ANOVA
Are variances equal among all populations? (Levene's test)
NO
YES
Welch's ANOVA Weighted least squares
ANOVA
Kruskal-Wallis