Mammal Life History Traits as an Indicator of Genus Diversity

Daire Crawford, Kyra Fine, Ruxin Liu, Hanatu Tak

12/5/2018

Abstract

Mammals consist of over 5000 species, all of which have diverse life histories. In this study, we focus on how different mammal life history traits and their interactions affect diversification. Several statistical analyses were performed, including: Linear Regression, Anova, and Randomization tests, and it was found that four traits (gestation time, age of first reproduction, litters per year, and maximum life span) contributed significantly to diversification in eutherian mammals. A species' diversification rate has been indicated to play a role in its risk of extinction, therefore, analyzing the effect of life history traits on diversification can have potential impacts on ecology and conservation efforts.

Background

Over time, populations evolve and new species arise through a process known as speciation. Speciation occurs when two populations split from an original population into two or more species that over time become so genetically different from one another that are no longer able to interbreed (De Queiroz, 1998). This splitting of lineages, also known as cladogenesis, is the result of the various selection pressures acting on populations at any given time and can be represented in a phylogenetic tree (De Queiroz, 1998). In nature, these selection pressures act on populations on various levels of geographic isolation which lead to descendants that settle into their own ecological niche and become reproductively isolated from their ancestors.

As species arise, others will go extinct. This occurs when all members of a population of species are unable to survive due to changing environmental conditions or against competition. When extinction is common, the diversity of a group will remain low even if new species are being created at a regular rate. Selection acts to avoid extinction as genes cannot be passed down without offspring, therefore traits that reduce the risk of extinction should be selected for (Benton, 1995). Diversification (r) is the net rate at which new species form and is calculated by subtracting the speciation rate (λ) by the extinction rate (μ) (Magallon & Sanderson, 2001). As selection acts on species, different members of the population fall into their own ecological niche and over time, species diversify creating new evolutionary clades (Weins & Donoghue, 2004).

Selection also acts on the life history (or lifecycle) of species, which are the patterns of survival and reproduction events that occur over the lifetime of an individual. Life history traits consist of trade-offs between growth, survival, and reproduction which are optimized for individual species (Roff et al, 2006). For instance, early reproduction increases the chance of an adult producing offspring before death, but late reproduction improves the chance of producing healthier offspring. Different life history traits are favoured in different populations due to varying limiting resources and can contribute to diversification in several ways (Wootton, 1993). We expect species that have shorter generation times and reproduce more offspring per year to bear large population sizes. Selection, rather than genetic drift, acts stronger in these populations which leads to higher mutation rates and increased recombination events, ultimately leading to an increased rate of speciation (Wootton, 1993). As well, populations with many individuals are less vulnerable to extinction and can recover from population loss events quicker due to their short generation time.

This paper will address life history traits that contribute to rates of diversification in eutherian mammals. We expect short-living, fast growing species ('r' species), compared with species possessing opposite life history traits ('k' species) (Sakai et al, 2001), to undergo higher rates of diversification and be more robust to

extinction due to their optimized life history traits. We hypothesize that certain life history traits significantly impact the rate of diversification in different genera. Further, we expect diversification to be higher in groups that have life history traits that are common in 'r' species, as they display faster population growth and lower individual investment. The objectives of this paper are: 1) to identify significant life history traits that contribute to diversification using linear regression, randomization tests, and Anova tests and 2) to analyze the skew of the data and transform it if necessary.

This topic is of importance to us because of its impact on conservation. Species with low diversification are more prone to extinction and are less capable of recovery from drastic events. Life history traits can be used in addition to raw population numbers in identifying species at risk to drive conservation strategies forward.

Methods

Data Description

For our analysis we used a dataset describing the life history characteristics of placental non-volant mammals (Ernest, 2003), which quantifies several life history traits for eutherian mammalian species, as well as noting each species' genus and order. In order to prepare the data for analysis, we converted all missing values from the -999.99 they were stored as to NA's and removed empty rows present at the end of the dataset. We created two versions of the dataset: one filtered by genus and the other by order. We averaged all the trait values for each species and added a column listing the number of species within the datasets.

We measured diversification by the number of species contained within a genus in order to assess total diversification, as it accounted for factors that either increased speciation rate or reduced extinction rate (Cardillo et al, 2003). We chose to perform our analysis at the level of genus rather than at the level of order as the low number of orders resulted in weaker statistical power. Additionally, because we averaged the data, grouping too broadly on a phylogenetic scale would result in cases where species with vastly different life history characteristics (for example, blue whales and porpoises or rats and capybaras) would be incorrectly treated as apart of the same group. The age of the paper also made the orders less reliable since one of the orders, Insectivora, is no longer considered valid as it was composed of extremely distantly related species. Therefore, using orders for our analysis would have provided a false representation of diversification, whereas most changes that have been made to genera in recent years have been more minot, consisting of either merging them together or splitting them apart. One caveat to our analysis is that the data only addresses Eutherian mammals, so any conclusions drawn may not apply to metatherians or monotremes. The dataset is also not a complete list of all extant eutherian species so our results can only be conclusively applied to those species included in the analysis. That being said, reports of life history traits affecting diversification, and vice versa, have been shown in divergent groups such as bird (Ricklefs, 2000) and frogs (Wollenberg Valeros, 2015), suggesting that the principles underlying our analysis are widely applicable.

Data Analysis

We first chose four variables we believed would be indicative of the differences between 'r' and 'k' species and also had a low number of missing values. These variables were gestation time in months, age of first reproduction in months, average litter size and maximum lifespan in months. We looked at how they contributed to diversification by conducting a regression and a randomization test for each of the traits individually, as well as an Anova for the combined set of traits. Regression models allowed us to test whether the predictor variable had an influence on the response variable by checking the significance of the p-value. We used an α of 0.05, therefore any p-value lower than 0.05 was treated as a significant result. The predictor variables consisted of the various life history traits and the response variable was the genus size. The linear regression and Anova tests had assumptions that needed to be tested in our data in order for their results to be treated as significant, in particular, the assumption that the data was normally distributed and had constant variance.

We examined whether traits were significant by conducting an Anova on all of the life history traits and their interactions. As seen in our results, we decided to replace litter size with litters per year because the trait was more significant while representing a similar aspect of life history. We reran the Anova tests on our final four selected traits, as well as, the regression and randomization tests on litters per year.

We tested the assumptions of normality and homogeneity of variance for our regression and Anova tests by analyzing the normal Q-Q plots and Scale-Location plots. Not all the points fell evenly on the line in our Q-Q plots, indicating that our data was not normal. As well, the points were not evenly distributed in the Scale-Location plots, implying that variance heterogeneity existed in our data. Following the results of these tests, we performed transformations on the linear model by applying the square root to the predictor variables, since transformation on the x-axis can sometimes fix non-normality and non-constant variance. We then repeated the transformation with higher roots and the normality and distribution of variance did not change significantly. Regardless of our efforts, all four linear models continued to show a highly right-skewed distribution and non-evenly distributed residuals.

We also removed several genera that we believed had the potential to skew the data in drop-one analysis. These orders were: Rodentia, which contains a large number of species most of which all share 'r' characteristics; Cetacea, which have several very large and long lived species that could act as outliers; Perissodactyla, which also contains many large species, some of which have been affected by domestication; and Primates, which consists of diverse species in both 'r' and 'k' categories. Despite these removals, the data remained skewed. As such, we decided to continue using our original linear model without the transformations, as the data was skewed regardless and our conclusions still hold since they were confirmed by randomization tests, which do not assume normality and homogeneity of the data.

Results

Linear regressions were conducted for gestation period, litters per year, lifespan, and age of first reproduction with genus size as the response variable and all traits were seen to be significant with p-values below the 0.05 cutoff (Table.1 & Fig.1). We also looked at the variance of genera within each order for each of the four traits, with the graph for lifespan shown in figure 2. This figure brought our attention to the orders Cetacea, Perissodactyla, Primates, and Rodentia as they showed great variance within their order and after further review, led us to consider them for our drop-one analysis. For the drop-one analysis conducted, gestation period, litters per year, and lifespan all remained significant ($p < 0.05$) for all analyses (Table 2-5 & Fig.3). Age of first reproduction remained significant ($p < 0.05$) with the removal of Cetacea, Perissodactyla, and Primates, however, the removal of Rodentia resulted in a loss of significance for this trait (Table 2b).

For the randomization tests, after 5000 repetitions, all four traits were found to have normal distributions and high significance with p-values of 0.0012 or below (Fig.4). Interestingly, when conducting diagnostic plots in R, the points on the normal Q-Q plots did not all fall on the straight line indicating that the data was not normally distributed (Fig.5). Even after applying square root transformations and higher root transformations, the data was still highly right skewed (Fig.6-7). The Scale-Location plots for all four traits of the original linear model (Fig.8) and the model after the square transformation (Fig.9) all showed clustered points instead of evenly distributed points.

Four Anova tests were conducted, one with each trait as the first explanatory variable. For gestation period as the explanatory variable, gestation period, litters per year, the interaction between these two variables and the interaction between these two variables and life span were all found to be significant. With lifespan as the first explanatory variable, all traits except the age of first reproduction were significant. Additionally, the interactions between these three traits and interactions between lifespan and gestation were found to be significant. Analysis with age of first reproduction as the first explanatory variable showed age of first reproduction, litters per year, and their interaction to be significant. Finally, with litters per year as the explanatory variable, litters per year, gestation time and their interactions were all significant, as well as the interactions between litters per year, lifespan, and gestion period.

Discussion

From our regressions, we found significant results showing all four variables contributed to diversification. Although the data is not normally distributed and does not have homogeneity of variances, the results would still be valid because the randomization tests showed even more significant values and does not assume normality.

The regressions were found to be significant even with the removal of the four orders we believed could be causing the skewing of the data, except for the removal of Rodentia for the trait age of first reproduction. This could have been due to several possible reasons, as rodents tend to live shorter lives and have an earlier age of first reproduction compared to other orders that have longer periods of time to continually reproduce. In addition, the age of first reproduction may not matter with regards to rodents as they have a high number of species and thus may be creating a non-causal correlation. Furthermore, age of first reproduction has higher p-values overall across all traits, indicating that the trait may play the smallest role in diversification compared to the other three traits.

Our results kept with what we would expect given the mechanisms of extinction and speciation, the two components that make up diversification. Large litter sizes and frequent and early reproduction lead to a large population and when populations are large, the effects of genetic drift are reduced and the effects of selection are increased. More individuals in the population lead to increased mutation rates which creates more opportunity for adaptive change. Large populations also often coincide with large range sizes, which increases the chances of a barrier to gene flow arising within the population, leading to allopatric speciation. Large populations also provide more robustness to extinction due to the need for more individuals to be lost for the population to disappear. Similarly, low age of first reproduction and short gestation times should also increase diversification, as they provide more generations for gene frequencies to change, eventually leading to fixation, and more reproductive events implies more opportunities for new mutations to arise. Short generation times also allows populations to recover faster from disturbances, aiding in the avoidance of extinction.

One potential non-causal explanation of our results is the potential for phylogenetic pseudoreplication, where a suite of traits are shared by different species or genera not because the traits have any true biological relationship with each other, but because they all happened to be passed down by a common ancestor. We tried to control for this by testing if the results held even when specific orders were removed. Since we compared genera and not species, the evolutionary time scale is higher so more time has passed for any phylogenetically linked traits to be broken, but a sister group analysis would be required to conclusively dismiss this possibility.

Future studies should examine the effect of additional life history traits such as the remaining five traits from the dataset to obtain a better understanding of these traits, their interactions, and their impact on genus diversity. Furthermore, if we wanted to confirm our results, we could acquire phylogenetic tree data and perform sister group comparisons, which would also help compensate for the incomplete nature of the dataset.

Figures and tables:

Table 1a). Linear regression: genus size ~ gestation time

```
##                 Estimate Std. Error   t value      Pr(>|t|)
## (Intercept)    3.2894873  0.2224175 14.789696 2.563861e-41
## gestation_mo -0.1604144  0.0378503 -4.238127 2.684588e-05
```

Table 1b). Linear regression: genus size ~ age of first reproduction

```
##                 Estimate  Std. Error   t value      Pr(>|t|)
## (Intercept)   3.08860951 0.216330820 14.277251 2.512973e-38
## afr_mo       -0.01578231 0.005659662 -2.788561 5.521783e-03
```

Table 1c). Linear regression: genus size ~ litters per year

```
##                     Estimate Std. Error  t value     Pr(>|t|)
## (Intercept)        2.1401432  0.3166909 6.757831 5.086568e-11
## litters_per_year 0.4574881  0.1658693 2.758125 6.085564e-03
```

Table 1d). Linear regression: genus size ~ life span

```
##                     Estimate    Std. Error   t value     Pr(>|t|)
## (Intercept)   3.457079493 0.2737941065 12.626566 3.640135e-31
## life_mo      -0.003342782 0.0009367307 -3.568563 4.007741e-04
```
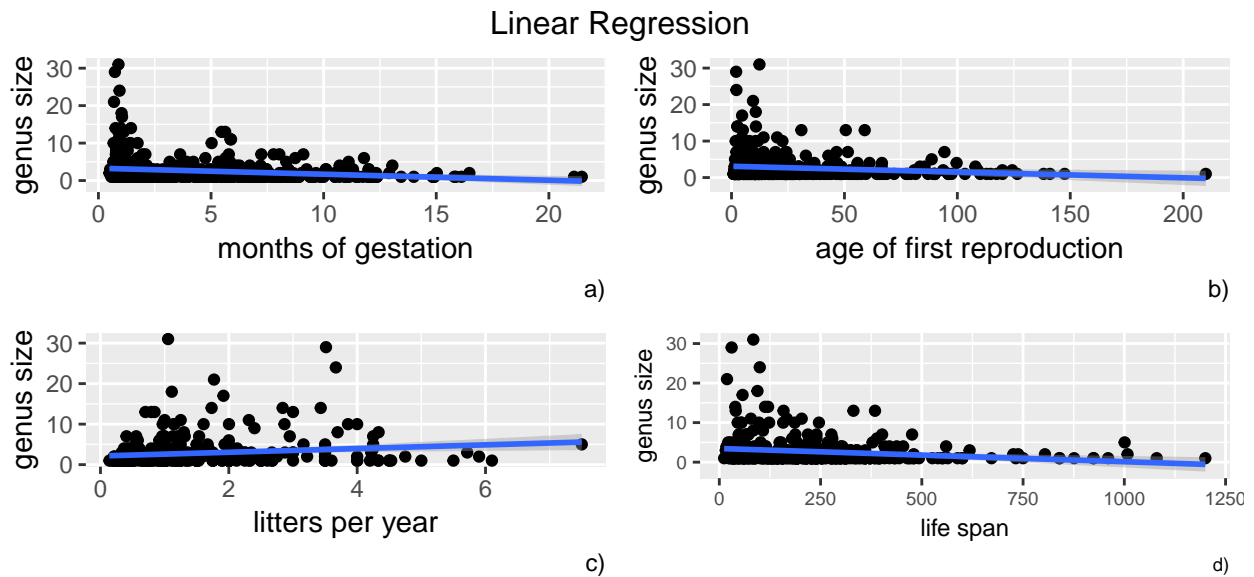

Fig.1 Linear Regression Model

a) Gestation time was highly significant with a p–value close to 0.0001 (2.684588e−05).

b) Age of first reproduction was highly significant with a p–value close to
 0.001(5.521783e−03).

c) Litters per year was highly significant with a p–value close to 0.01 (6.085564e−03).

d) Lifespan was highly significant with a p–value close to 0.001 (4.007741e−04).
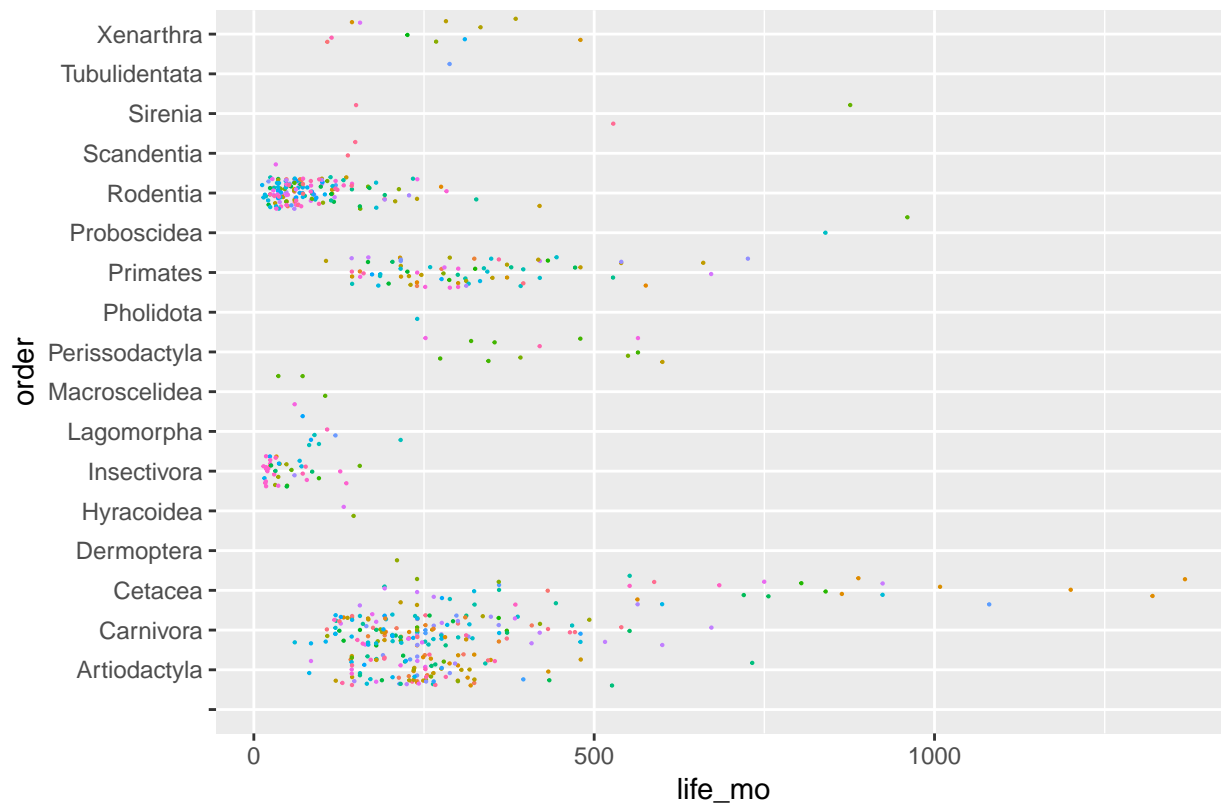
Fig.2 Graph shows variance of life span of genera within their order.

Table 2a). Drop-one analysis of Rodentia: genus size ~ gestation time (p value < 0.05)

```
##                  Estimate Std. Error  t value      Pr(>|t|)
## (Intercept)   2.67672383 0.23967173 11.16829 8.814300e-25
## gestation_mo -0.08519824 0.03394041 -2.51023 1.254326e-02
```

Table 2b). Drop-one analysis of Rodentia: genus size ~ age of first reproduction (p value > 0.05, not siginificant)

```
##                 Estimate  Std. Error    t value      Pr(>|t|)
## (Intercept)  2.586038327 0.216901124 11.922660 4.993497e-27
## afr_mo      -0.008781795 0.004678751 -1.876953 6.150806e-02
```

Table 2c). Drop-one analysis of Rodentia: genus size ~ litters per year(p value < 0.05)

```
##                   Estimate Std. Error  t value      Pr(>|t|)
## (Intercept)      1.8096330  0.2927962 6.180521 2.504553e-09
## litters_per_year 0.5269701  0.2102130 2.506839 1.280206e-02
```

Table 2d). Drop-one analysis of Rodentia: genus size ~ life span(p value < 0.05)

```
##                Estimate   Std. Error   t value      Pr(>|t|)
## (Intercept)  2.68778940 0.2591022967 10.373468 9.344256e-22
## life_mo     -0.00163917 0.0007696317 -2.129811 3.399822e-02
```

6

Drop-one test plot focussing on rodents

Drop one test

a)

b)

c)

d)
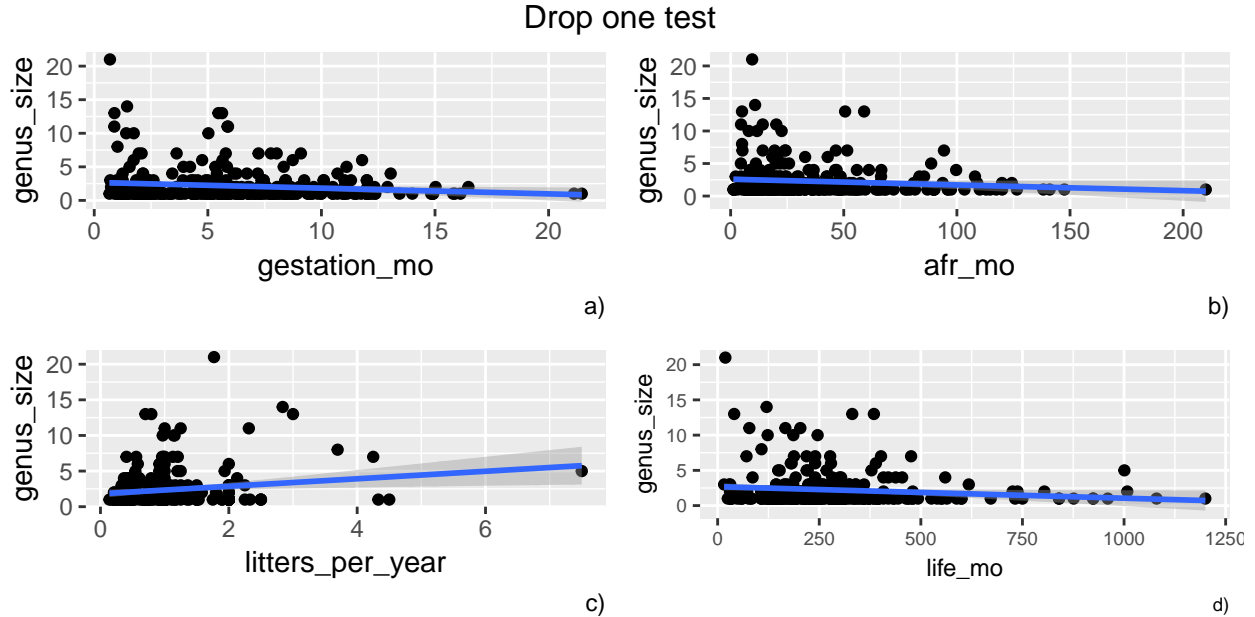
Fig.3 Drop one test for Rodentia

a) Gestation time was significant with a p–value of 0.0125.

b) Age of first reproduction was not significant with a p–value of 0.0615.

c) Litters per year was significant with a p–value 0.0128.

d) Lifespan was significant with a p–value of 0.034.

Table 3a). Drop-one analysis of Cetacia: genus size ~ gestation time (p value < 0.05)

```
##                     Estimate Std. Error    t value      Pr(>|t|)
## (Intercept)    3.2894873  0.2224175 14.789696 2.563861e-41
## gestation_mo -0.1604144  0.0378503 -4.238127 2.684588e-05
```

Table 3b). Drop-one analysis of Cetacia: genus size ~ age of first reproduction (p value < 0.05)

```
##                      Estimate  Std. Error    t value      Pr(>|t|)
## (Intercept)   3.08860951 0.216330820 14.277251 2.512973e-38
## afr_mo       -0.01578231 0.005659662 -2.788561 5.521783e-03
```

Table 3c). Drop-one analysis of Cetacia: genus size ~ litters per year (p value < 0.05)

```
##                      Estimate Std. Error  t value      Pr(>|t|)
## (Intercept)      2.1401432  0.3166909 6.757831 5.086568e-11
## litters_per_year 0.4574881  0.1658693 2.758125 6.085564e-03
```

Table 3d). Drop-one analysis of Cetacia: genus size ~ life span (p value < 0.05)

```
##                      Estimate    Std. Error    t value      Pr(>|t|)
## (Intercept)   3.457079493 0.2737941065 12.626566 3.640135e-31
## life_mo      -0.003342782 0.0009367307 -3.568563 4.007741e-04
```

Table 4a). Drop-one analysis of Perissodactyla: genus size ~ gestation time (p value < 0.05)

```
##                     Estimate Std. Error    t value      Pr(>|t|)
## (Intercept)    3.3218526 0.22541878 14.736362 5.309524e-41
## gestation_mo -0.1717437 0.03946546 -4.351748 1.642086e-05
```

Table 4b). Drop-one analysis of Perissodactyla: genus size ~ age of first reproduction (p value < 0.05)

```
##                  Estimate  Std. Error    t value      Pr(>|t|)
## (Intercept)   3.07836482  0.217747969  14.137284  1.197772e-37
## afr_mo        -0.01553569 0.005754673  -2.699665  7.209646e-03
```

Table 4c). Drop-one analysis of Perissodactyla: genus size ~ litters per year(p value < 0.05)

```
##                   Estimate Std. Error  t value      Pr(>|t|)
## (Intercept)      2.1376051  0.3234614 6.608532  1.291828e-10
## litters_per_year 0.4575106  0.1681915 2.720176  6.819680e-03
```

Table 4d). Drop-one analysis of Perissodactyla: genus size ~ life span (p value < 0.05)

```
##                  Estimate   Std. Error    t value      Pr(>|t|)
## (Intercept)   3.455761725  0.2757545691  12.532020  9.987064e-31
## life_mo       -0.003374588 0.0009543184  -3.536123  4.523986e-04
```

Table 5a). Drop-one analysis of Primate: genus size ~ gestation time (p value < 0.05)

```
##                  Estimate Std. Error    t value      Pr(>|t|)
## (Intercept)    3.2894873  0.2224175  14.789696  2.563861e-41
## gestation_mo  -0.1604144  0.0378503  -4.238127  2.684588e-05
```

Table 5b). Drop-one analysis of Primate: genus size ~ age of reproduction (p value < 0.05)

```
##                  Estimate  Std. Error    t value      Pr(>|t|)
## (Intercept)   3.08860951  0.216330820  14.277251  2.512973e-38
## afr_mo        -0.01578231 0.005659662  -2.788561  5.521783e-03
```

Table 5c). Drop-one analysis of Primate: genus size ~ litters per year (p value < 0.05)

```
##                   Estimate Std. Error  t value      Pr(>|t|)
## (Intercept)      2.1401432  0.3166909 6.757831  5.086568e-11
## litters_per_year 0.4574881  0.1658693 2.758125  6.085564e-03
```

Table 5d). Drop-one analysis of Primate: genus size ~ life span (p value < 0.05)

```
##                  Estimate   Std. Error    t value      Pr(>|t|)
## (Intercept)   3.457079493  0.2737941065  12.626566  3.640135e-31
## life_mo       -0.003342782 0.0009367307  -3.568563  4.007741e-04
```
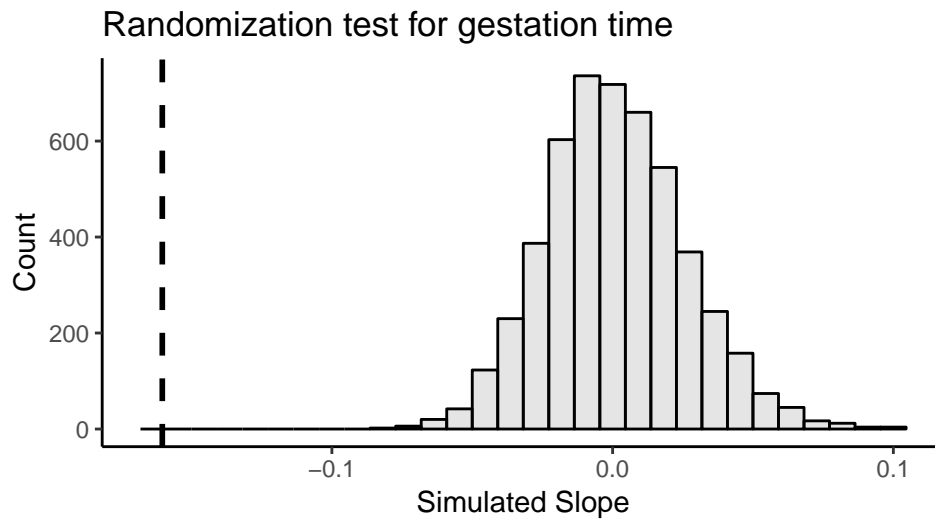
Randomization Tests

Randomization test for gestation time

Fig. 4a) Randomization test of gestation time with a highly significant p−value of 0
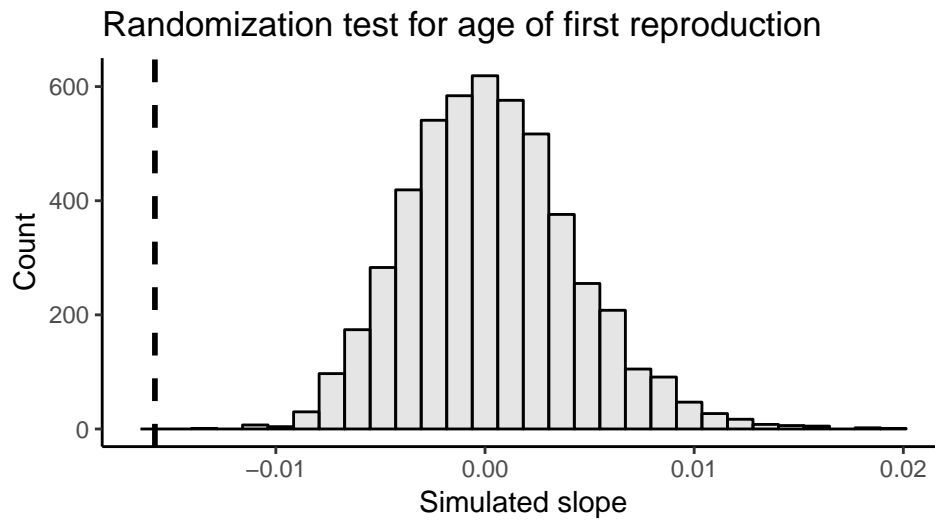
Randomization test for age of first reproduction

Fig. 4b) Randomization test for afr with a highly significant p−value of 0.0012
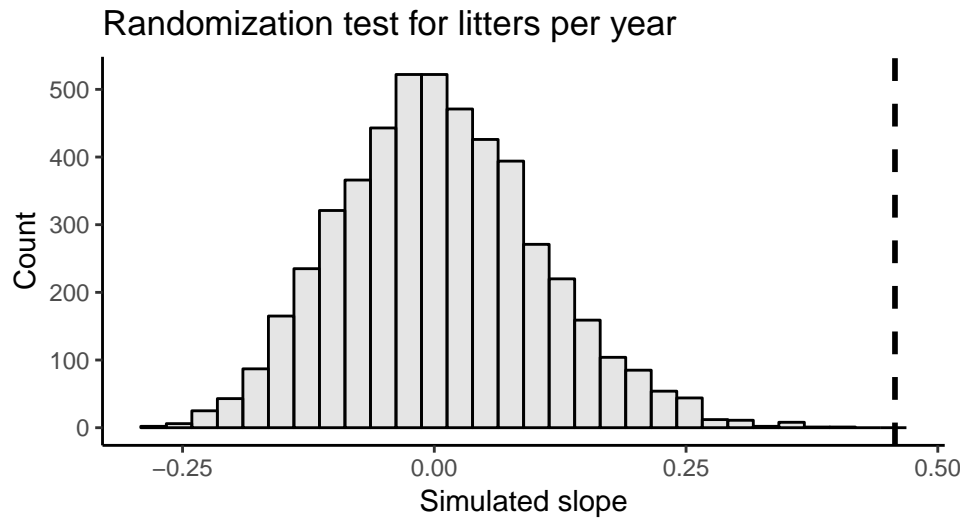
Randomization test for litters per year

Fig. 4c) Randomization test for litters per year with a highly significant p−value of 0
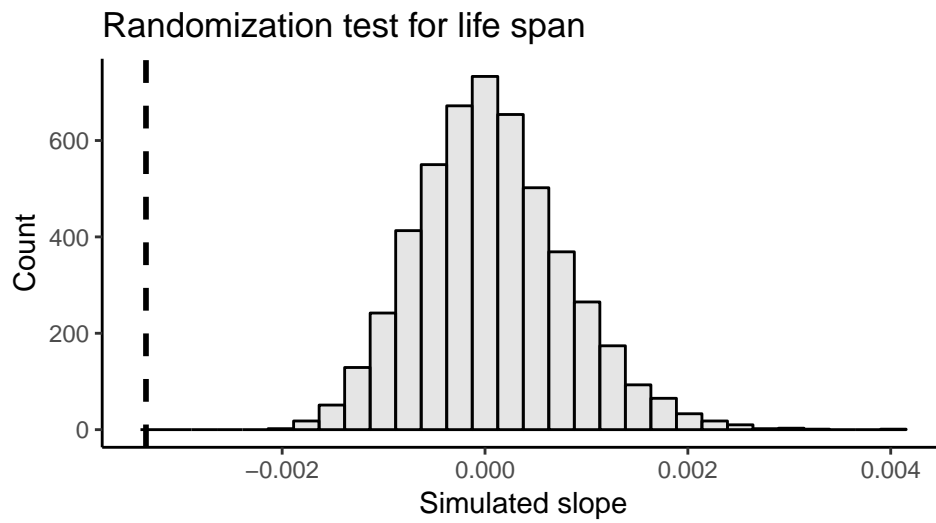
Randomization test for life span

Fig. 4d) Randomization test for life span with a highly significant p−value of 0.0002
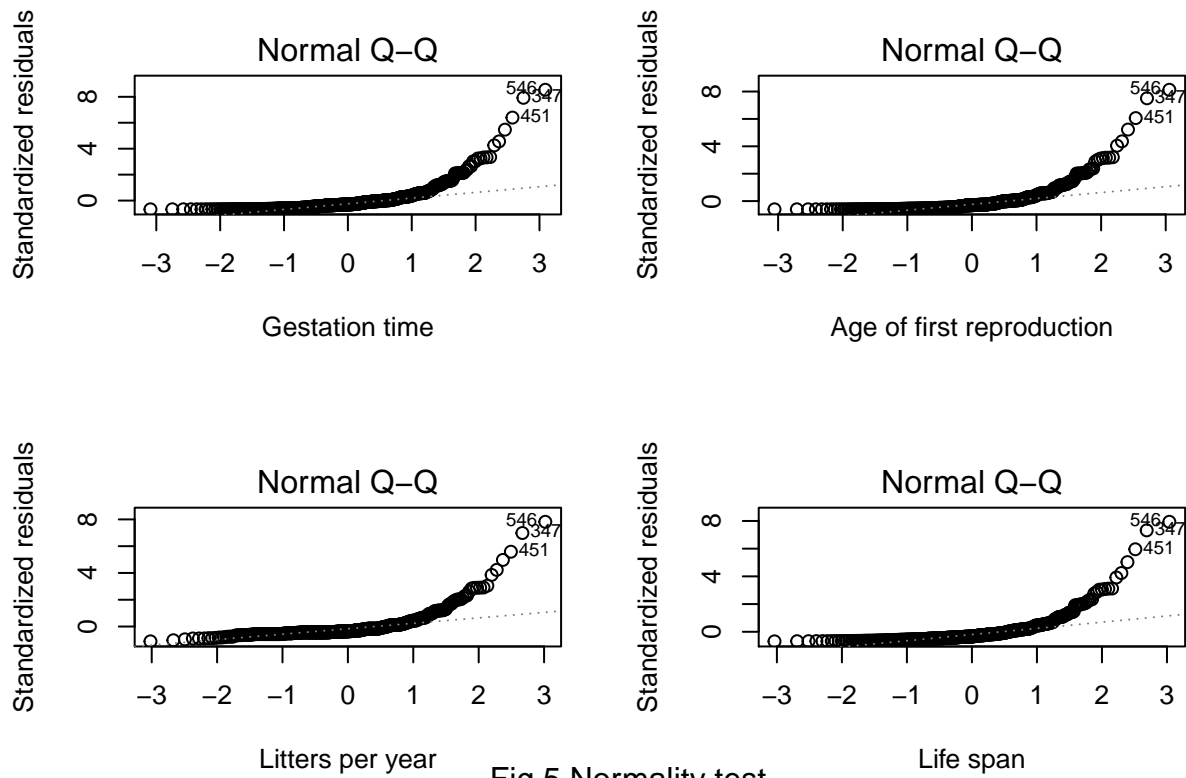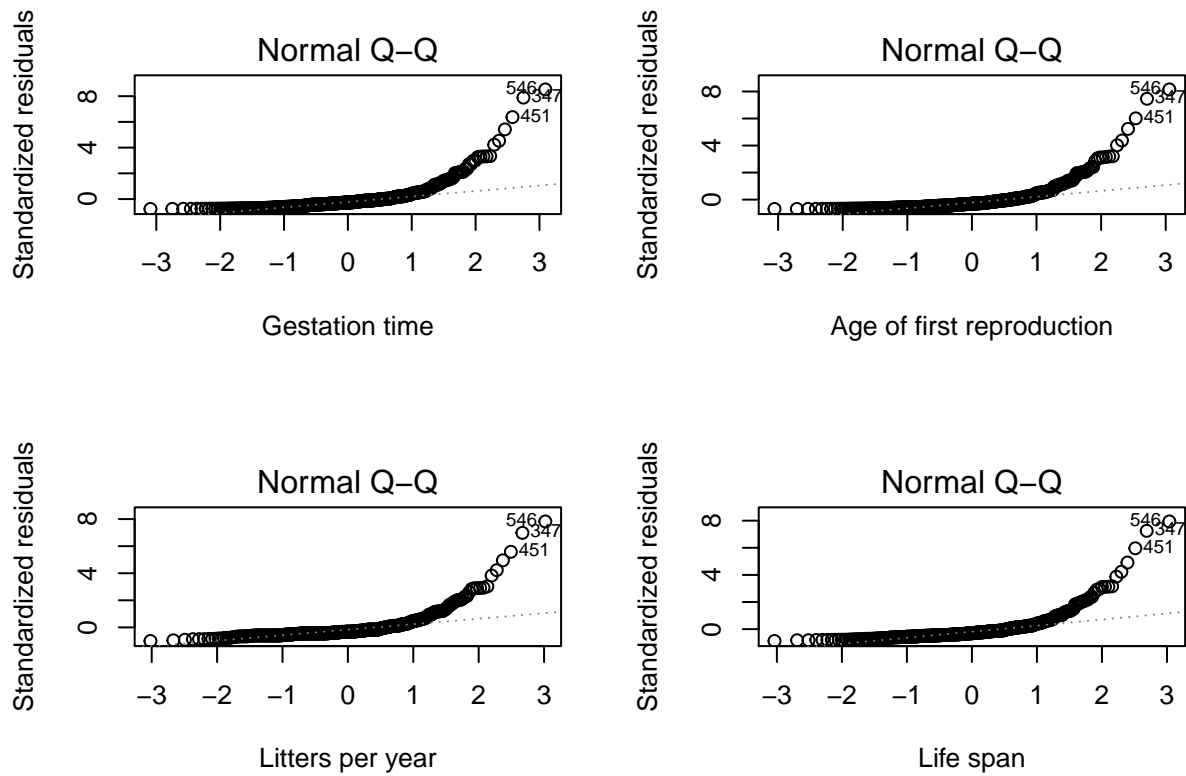
Fig.5 Normality test

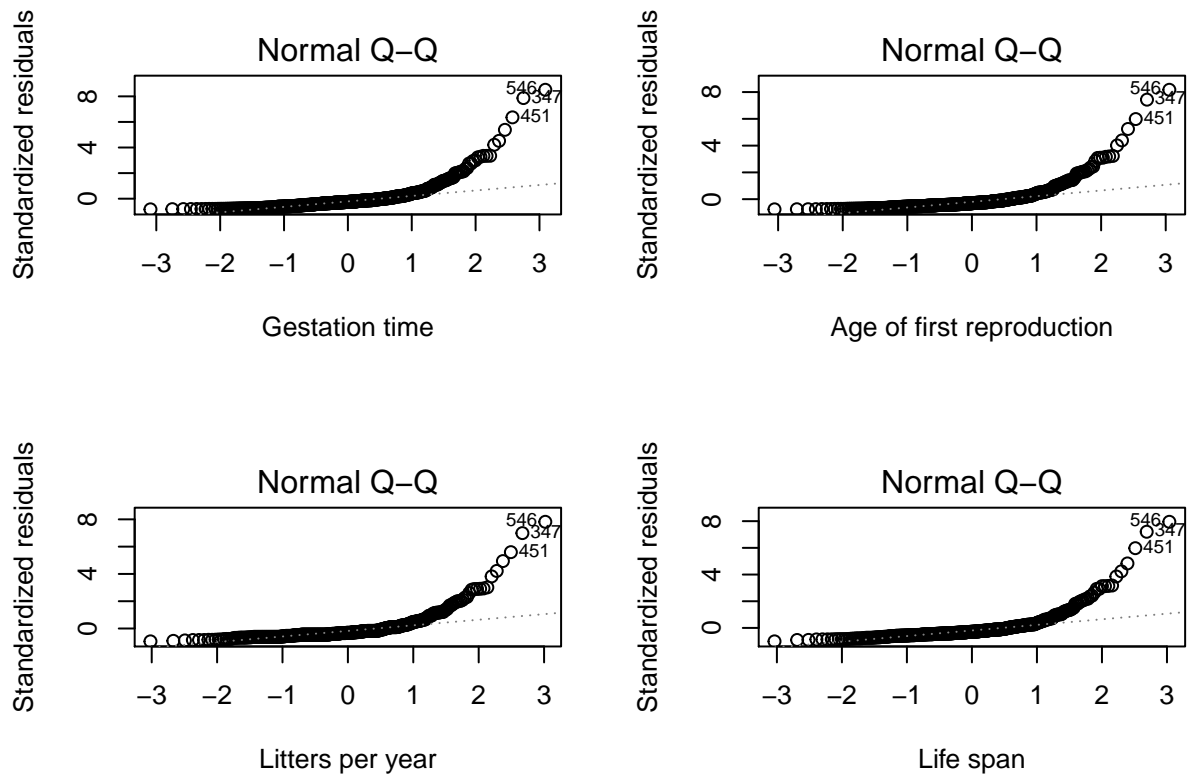Fig.6 Normality test after square root transformation of x

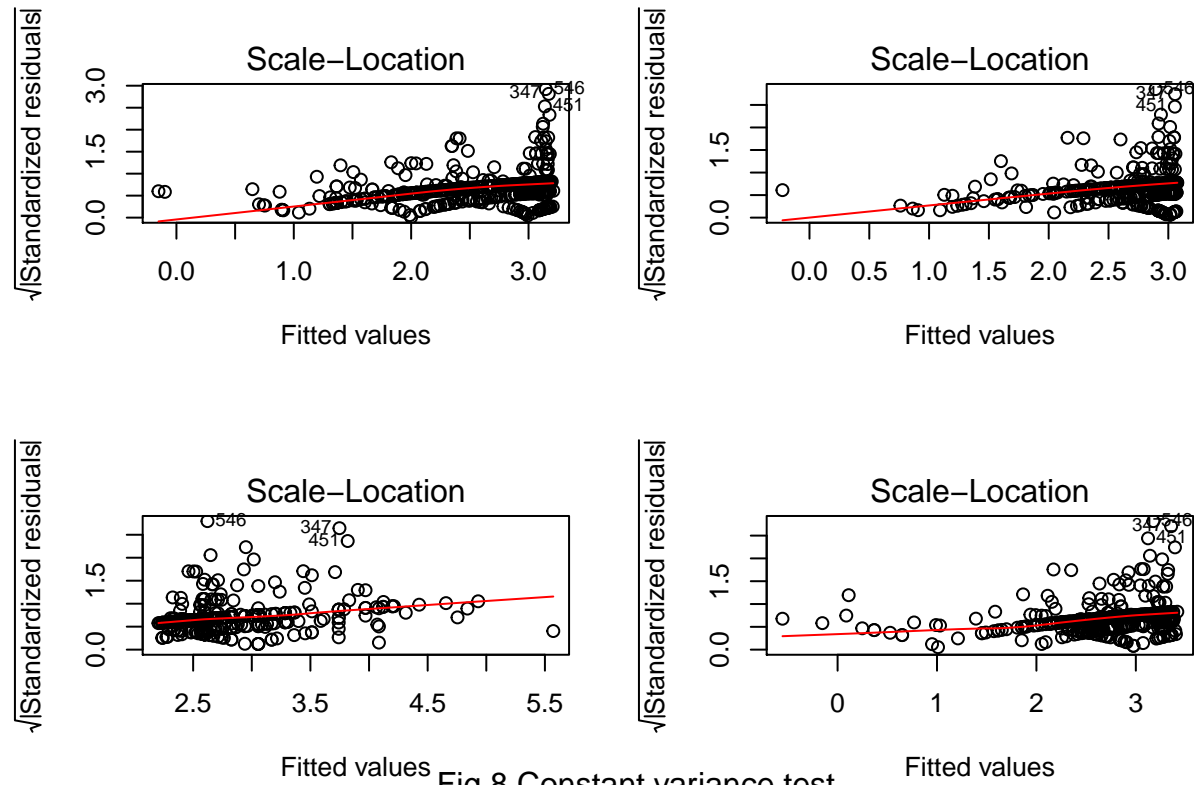Fig.7 Normality test after the fourth root transformation of x
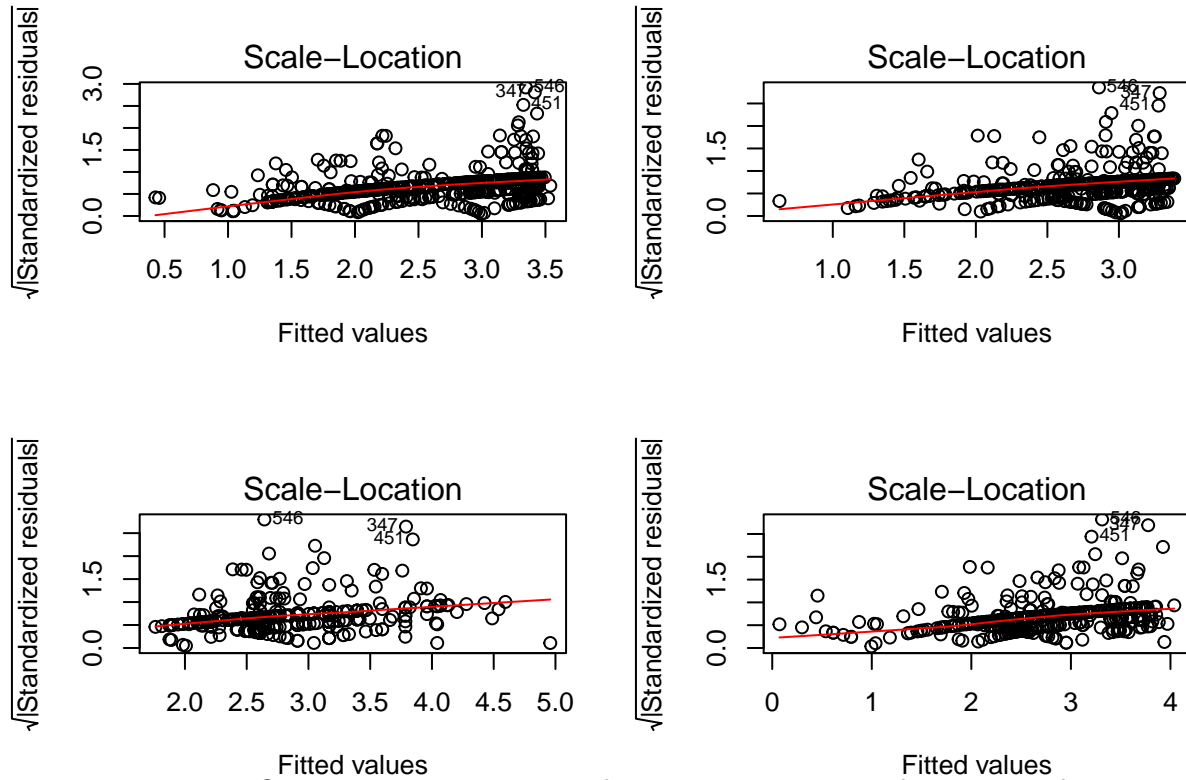
Fig.8 Constant variance test

Fig.9 Constant variance test after square root transformation of x

Note: In Table 6, gestation_mo is the gestation time, afr_mo is the age of first reproduction, litters_per_year is the litters per year and life_mo is the life span. All lines highlighted with red show a siginificant result (p value < 0.05).

Table. 6a) Anova test using gestation time as the first explanatory variable

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gestation_mo	1	359.1233661	359.1233661	23.6821743	0.0000019
litters_per_year	1	69.3845978	69.3845978	4.5755256	0.0332998
life_mo	1	2.8458077	2.8458077	0.1876651	0.6652018
afr_mo	1	11.2044757	11.2044757	0.7388724	0.3907603
gestation_mo:litters_per_year	1	150.4774373	150.4774373	9.9231441	0.0018095
gestation_mo:life_mo	1	0.3574912	0.3574912	0.0235745	0.8780834
litters_per_year:life_mo	1	10.8669415	10.8669415	0.7166139	0.3979824
gestation_mo:afr_mo	1	10.8143210	10.8143210	0.7131439	0.3991256
litters_per_year:afr_mo	1	2.6946584	2.6946584	0.1776976	0.6736840
life_mo:afr_mo	1	0.1496319	0.1496319	0.0098674	0.9209439
gestation_mo:litters_per_year:life_mo	1	61.4637052	61.4637052	4.0531871	0.0450483
gestation_mo:litters_per_year:afr_mo	1	1.6979742	1.6979742	0.1119719	0.7381617
gestation_mo:life_mo:afr_mo	1	16.3080644	16.3080644	1.0754255	0.3006211
litters_per_year:life_mo:afr_mo	1	6.2763046	6.2763046	0.4138871	0.5205315
gestation_mo:litters_per_year:life_mo:afr_mo	1	2.8677214	2.8677214	0.1891102	0.6639944
Residuals	279	4230.8369931	15.1642903	NA	NA

Table. 6b) Anova test using age of first reproduction as the first explanatory variable

13

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
afr_mo	1	205.484984	205.484984	13.5505836	0.0002787
litters_per_year	1	140.046937	140.046937	9.2353110	0.0025990
life_mo	1	40.693284	40.693284	2.6834942	0.1025203
gestation_mo	1	56.333042	56.333042	3.7148486	0.0549457
afr_mo:litters_per_year	1	89.297518	89.297518	5.8886711	0.0158722
afr_mo:life_mo	1	1.051112	1.051112	0.0693149	0.7925312
litters_per_year:life_mo	1	42.202435	42.202435	2.7830142	0.0963906
afr_mo:gestation_mo	1	1.308244	1.308244	0.0862714	0.7691903
litters_per_year:gestation_mo	1	33.299273	33.299273	2.1959005	0.1395069
life_mo:gestation_mo	1	8.201900	8.201900	0.5408693	0.4626904
afr_mo:litters_per_year:life_mo	1	28.806441	28.806441	1.8996234	0.1692247
afr_mo:litters_per_year:gestation_mo	1	24.182587	24.182587	1.5947062	0.2077094
afr_mo:life_mo:gestation_mo	1	11.076622	11.076622	0.7304412	0.3934735
litters_per_year:life_mo:gestation_mo	1	21.680398	21.680398	1.4297008	0.2328285
afr_mo:litters_per_year:life_mo:gestation_mo	1	2.867721	2.867721	0.1891102	0.6639944
Residuals	279	4230.836993	15.164290	NA	NA

Table. 6c) Anova test using litters per year as the first explanatory variable

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
litters_per_year	1	316.3729421	316.3729421	20.8630233	0.0000074
gestation_mo	1	112.1350218	112.1350218	7.3946765	0.0069520
life_mo	1	2.8458077	2.8458077	0.1876651	0.6652018
afr_mo	1	11.2044757	11.2044757	0.7388724	0.3907603
litters_per_year:gestation_mo	1	150.4774373	150.4774373	9.9231441	0.0018095
litters_per_year:life_mo	1	10.6547394	10.6547394	0.7026204	0.4026221
gestation_mo:life_mo	1	0.5696934	0.5696934	0.0375681	0.8464540
litters_per_year:afr_mo	1	0.1565319	0.1565319	0.0103224	0.9191478
gestation_mo:afr_mo	1	13.3524474	13.3524474	0.8805191	0.3488711
life_mo:afr_mo	1	0.1496319	0.1496319	0.0098674	0.9209439
litters_per_year:gestation_mo:life_mo	1	61.4637052	61.4637052	4.0531871	0.0450483
litters_per_year:gestation_mo:afr_mo	1	1.6979742	1.6979742	0.1119719	0.7381617
litters_per_year:life_mo:afr_mo	1	7.3879976	7.3879976	0.4871971	0.4857616
gestation_mo:life_mo:afr_mo	1	15.1963713	15.1963713	1.0021156	0.3176666
litters_per_year:gestation_mo:life_mo:afr_mo	1	2.8677214	2.8677214	0.1891102	0.6639944
Residuals	279	4230.8369931	15.1642903	NA	NA

Table. 6d) Anova test using life span as the first explanatory variable

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
life__mo	1	276.540163	276.540163	18.2362747	0.0000268
gestation__mo	1	94.060546	94.060546	6.2027661	0.0133370
afr__mo	1	5.757652	5.757652	0.3796849	0.5382743
litters__per__year	1	66.199886	66.199886	4.3655117	0.0375793
life__mo:gestation__mo	1	69.432558	69.432558	4.5786883	0.0332395
life__mo:afr__mo	1	4.352958	4.352958	0.2870532	0.5925420
gestation__mo:afr__mo	1	11.704684	11.704684	0.7718583	0.3803988
life__mo:litters__per__year	1	55.113532	55.113532	3.6344287	0.0576233
gestation__mo:litters__per__year	1	31.915531	31.915531	2.1046505	0.1479752
afr__mo:litters__per__year	1	2.841219	2.841219	0.1873625	0.6654554
life__mo:gestation__mo:afr__mo	1	2.364448	2.364448	0.1559221	0.6932407
life__mo:gestation__mo:litters__per__year	1	72.063896	72.063896	4.7522103	0.0300966
life__mo:afr__mo:litters__per__year	1	3.322211	3.322211	0.2190812	0.6401063
gestation__mo:afr__mo:litters__per__year	1	7.995494	7.995494	0.5272580	0.4683706
life__mo:gestation__mo:afr__mo:litters__per__year	1	2.867721	2.867721	0.1891102	0.6639944
Residuals	279	4230.836993	15.164290	NA	NA

References:

Benton, M. J. (1995). Diversification and extinction in the history of life. Science, 268(5207), 52-58.

Cardillo, M., Huxtable, J. S., & Bromham, L. (2003). Geographic range size, life history and rates of diversification in Australian mammals. Journal of evolutionary biology, 16(2), 282-288.

De Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation.

Ernest, S. M. (2003). Life history characteristics of placental nonvolant mammals: ecological archives E084-093. Ecology, 84(12), 3402-3402.

Magallon, S., & Sanderson, M. J. (2001). Absolute diversification rates in angiosperm clades. Evolution, 55(9), 1762-1780.

Ricklefs, R. E. (2000). Density dependence, evolutionary optimization, and the diversification of avian life histories. The Condor, 102(1), 9-22.

Roff, D. A., Heibo, E., & Vøllestad, L. A. (2006). The importance of growth and mortality costs in the evolution of the optimal life history. Journal of evolutionary biology, 19(6), 1920-1930.

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, D. M., Molofsky, J., With, K. A., . . . & McCauley, D. E. (2001). The population biology of invasive species. Annual review of ecology and systematics, 32(1), 305-332.

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. Trends in ecology & evolution, 19(12), 639-644.

Wollenberg Valero, K. C. (2015). Evidence for an intrinsic factor promoting landscape genetic divergence in Madagascan leaf-litter frogs. Frontiers in genetics, 6, 155.

Wootton, R. J. (1993). The evolution of life histories: theory and analysis. Reviews in Fish Biology and Fisheries, 3(4), 384-385.

R codes

Initial data exploration, before datasets for order and genus levels were created

```r
mammals <- Mammal_lifehistories_v2 # Load in the data set

# Explore possible relationship between traits
qplot(data = mammals, x = newborn.g., y = max..life.mo.)

qplot(data = mammals, x = litter.size, y = max..life.mo.) +
  coord_cartesian(xlim = c(0,50), ylim = c(0,1000))
```

Cleaning data and creating genus and order level datasets + more exploratory analysis

```r
# Change the column names to make them easier to read
colnames(mammals)[colnames(mammals)=="mass.g."] <- "mass_g"
colnames(mammals)[colnames(mammals)=="gestation.mo."] <- "gestation_mo"
colnames(mammals)[colnames(mammals)=="newborn.g."] <- "newborn_g"
colnames(mammals)[colnames(mammals)=="weaning.mo."] <- "weaning_mo"
colnames(mammals)[colnames(mammals)=="wean.mass.g."] <- "wean_g"
colnames(mammals)[colnames(mammals)=="AFR.mo."] <- "afr_mo"
colnames(mammals)[colnames(mammals)=="max..life.mo."] <- "life_mo"
colnames(mammals)[colnames(mammals)=="litter.size"] <- "litter_size"
colnames(mammals)[colnames(mammals)=="litters.year"] <- "litters_per_year"

# Converting all values Of -999.00 to NA
mammals <- mammals %>%
  mutate(mass_g = ifelse(mass_g == -999.00, NA, mass_g),
         gestation_mo = ifelse(gestation_mo == -999.00, NA, gestation_mo),
         newborn_g = ifelse(newborn_g == -999.00, NA, newborn_g),
         weaning_mo = ifelse(weaning_mo == -999.00, NA, weaning_mo),
         wean_g = ifelse(wean_g == -999.00, NA, wean_g),
         afr_mo = ifelse(afr_mo == -999.00, NA, afr_mo),
         life_mo = ifelse(life_mo == -999.00, NA, life_mo),
         litter_size = ifelse(litter_size == -999.00, NA, litter_size),
         litters_per_year = ifelse(litters_per_year == -999.00, NA, litters_per_year))

write.csv(mammals, "mammals_clean.csv") # Create a new data with the cleaned data

mammals %>%
  group_by(order) %>%
  tally()

# Clean the empty rows at the end of the data
mammals <- head(mammals, 1440)

mammals2 <- mammals %>%
  group_by(order) %>%
  summarize(order_size = n_distinct(species),
            mass_g = mean(mass_g, na.rm = TRUE),
            gestation_mo = mean(gestation_mo, na.rm = TRUE),
            newborn_g = mean(newborn_g, na.rm = TRUE),
```

```
          weaning_mo = mean(weaning_mo, na.rm = TRUE),
          wean_g = mean(wean_g, na.rm = TRUE),
          afr_mo = mean(afr_mo, na.rm = TRUE),
          life_mo = mean(life_mo, na.rm = TRUE),
          litter_size = mean(litter_size, na.rm = TRUE),
          litters_per_year = mean(litters_per_year, na.rm = TRUE))
```

```
# Explore possible relationship between order size and traits
qplot(data = mammals2, x = order_size, y = mass_g)
```

```
qplot(data = mammals2, x = order_size, y = gestation_mo)
```

```
qplot(data = mammals2, x = order_size, y = newborn_g)
```

```
qplot(data = mammals2, x = order_size, y = weaning_mo)
```

```
qplot(data = mammals2, x = order_size, y = wean_g)
```

```
qplot(data = mammals2, x = order_size, y = afr_mo)
```

```
qplot(data = mammals2, x = order_size, y = life_mo)
```

```
qplot(data = mammals2, x = order_size, y = litter_size)
```

```
qplot(data = mammals2, x = order_size, y = litters_per_year)
```

```
mammals3 <- mammals %>%
  group_by(Genus) %>%
  summarize(genus_size = n_distinct(species),
            mass_g = mean(mass_g, na.rm = TRUE),
            gestation_mo = mean(gestation_mo, na.rm = TRUE),
            newborn_g = mean(newborn_g, na.rm = TRUE),
            weaning_mo = mean(weaning_mo, na.rm = TRUE),
            wean_g = mean(wean_g, na.rm = TRUE),
            afr_mo = mean(afr_mo, na.rm = TRUE),
            life_mo = mean(life_mo, na.rm = TRUE),
            litter_size = mean(litter_size, na.rm = TRUE),
            litters_per_year = mean(litters_per_year, na.rm = TRUE))
```

```
write.csv(mammals2, "mammals_order.csv")
write.csv(mammals3, "mammals_genus.csv")
```

Statistical analysis for four traits

Linear regression

```
library(tidyverse)
library(ggplot2)
# Linear model: genus size ~ gestation time
ggplot(mammals_g, aes(x = gestation_mo, y = genus_size)) +
    geom_point() +
    geom_smooth(method = "lm") + ylab('genus size') +
    xlab('months of gestation') # Plot the linear regression model
```

```r
result_ges <- lm(genus_size ~ gestation_mo, data = mammals_g) # Fit a linear model
summary(result_ges) # Check whether the p value is siginificant
```

```r
# Linear model: genus size ~ age of first reproduction
ggplot(mammals_g, aes(x = afr_mo, y = genus_size)) +
    geom_point() +
    geom_smooth(method = "lm") + ylab('genus size') +
    xlab('age of first reproduction') # Plot the linear regression model

result_afr <- lm(genus_size ~ afr_mo, data = mammals_g) # Fit a linear model
summary(result_afr) # Check whether the p value is siginificant
```

```r
# Linear model: genus size ~ litters per year
ggplot(mammals_g, aes(x = litters_per_year, y = genus_size)) +
    geom_point() +
    geom_smooth(method = "lm") + ylab('genus size') +
    xlab('litters per year') # Plot the linear regression model

result_litter <- lm(genus_size ~ litters_per_year,
                    data = mammals_g) # Fit a linear model
summary(result_litter) # Check whether the p value is siginificant
```

```r
# Linear model: genus size ~ life span
ggplot(mammals_g, aes(x = life_mo, y = genus_size)) +
    geom_point() +
    geom_smooth(method = "lm") + ylab('genus size') +
    xlab('life span') # Plot the linear regression model

result_life<- lm(genus_size ~ life_mo, data = mammals_g) # Fit a linear model
summary(result_life) # Check whether the p value is siginificant
```

```r
# Linear model: genus size ~ litter size (the original chosen trait)
result_litter_size <- lm(genus_size ~ litter_size,
                        data = mammals_g) # Fit a linear model
summary(result_litter_size) # Check whether the p value is siginificant
```

Anova tests

```r
# Multifactor for looking at which factors we might might want to focus on
anova(lm(genus_size~ mass_g + newborn_g + wean_g + afr_mo + litters_per_year +
        gestation_mo + life_mo + litter_size + weaning_mo,
        data = mammals_g))
# Comapring the p values, litter_size was replaced with litters_per_year since
# the latter was more significant while still representing similar aspects of life history
```

```r
# Anova test using gestation time as the first explanatory variable
anova(lm(genus_size ~ gestation_mo*litters_per_year*life_mo*afr_mo,
        data = mammals_g))
```

```r
# Anova test using age of first reproduction as the first explanatory variable
anova(lm(genus_size ~ afr_mo*litters_per_year*life_mo*gestation_mo,
        data = mammals_g))
```

```r
# Anova test using litters per year as the first explanatory variable
anova(lm(genus_size ~ litters_per_year*gestation_mo*life_mo*afr_mo,
         data = mammals_g))

# Anova test using life span as the first explanatory variable
anova(lm(genus_size ~ life_mo*gestation_mo*afr_mo*litters_per_year,
         data = mammals_g))
```

Randomization test

```r
# Randomization test for gestation time
set.seed(16)
reshuffled_ges <- mammals_c
reshuffled_ges$gestation_mo <- sample(reshuffled_ges$gestation_mo,
                                      size = nrow(reshuffled_ges),
                                      replace = FALSE)
slope_real_ges <- coef(result_ges)["gestation_mo"]
simulated_slopes_ges <- list()
nreps = 5000 # 5000 iterations
for(i in 1:nreps){

    reshuffled_ges <- mammals_c
    reshuffled_ges$gestation_mo <- sample(reshuffled_ges$gestation_mo,
                                          size = nrow(reshuffled_ges),
                                          replace = FALSE)
    # Calculate the slope
    genus_temp_ges <- reshuffled_ges %>%
        group_by(Genus) %>%
        summarize(genus_size = n_distinct(species),
                  mass_g = mean(mass_g, na.rm = TRUE),
                  gestation_mo = mean(gestation_mo, na.rm = TRUE),
                  newborn_g = mean(newborn_g, na.rm = TRUE),
                  weaning_mo = mean(weaning_mo, na.rm = TRUE),
                  wean_g = mean(wean_g, na.rm = TRUE),
                  afr_mo = mean(afr_mo, na.rm = TRUE),
                  life_mo = mean(life_mo, na.rm = TRUE),
                  litter_size = mean(litter_size, na.rm = TRUE),
                  litters_per_year = mean(litters_per_year, na.rm = TRUE))

    reshuffled_res_ges <- lm(genus_size ~ gestation_mo, data = genus_temp_ges)
    slope_sim_ges <- coef(reshuffled_res_ges)["gestation_mo"]

    # Append simulated slope to list
    simulated_slopes_ges[i] <- slope_sim_ges
}
# Unlist simulated means list into numeric vector
simulated_slopes_ges <- unlist(simulated_slopes_ges)
ggplot() +
    ylab("Count") + xlab("Simulated Slope") +
    geom_histogram(aes(x = simulated_slopes), bins = 30,
                   fill = "grey", alpha = 0.4, colour = "black") +
    geom_vline(xintercept = slope_real, size = 1,
```

```r
                       linetype = "dashed", colour = "black") +
      theme_classic() + ggtitle("Months of Gestation")
  ggsave("Randomization test of gestation time.png")
```

```r
# Caculate p value for randomization test of gestation time
abs_simulated_slopes_ges <- abs(simulated_slopes_ges)
exceed_count_ges <- length(abs_simulated_slopes_ges[abs_simulated_slopes_ges >=
                                               abs(slope_real_ges)])
p_val_ges <- exceed_count_ges / nreps
p_val_ges
```

```r
# Randomization test for age of first reproduction
set.seed(16)
reshuffled_afr <- mammals_c
reshuffled_afr$afr_mo <- sample(reshuffled_afr$afr_mo,
                                size = nrow(reshuffled_afr),
                                replace = FALSE)
slope_real_afr <- coef(result_afr)["afr_mo"]
simulated_slopes_afr <- list()
nreps = 5000 # 5000 iterations
for(i in 1:nreps){

    reshuffled_afr <- mammals_c
    reshuffled_afr$afr_mo <- sample(reshuffled_afr$afr_mo,
                                    size = nrow(reshuffled_afr),
                                    replace = FALSE)

    genus_temp_afr <- reshuffled_afr %>%
          group_by(Genus) %>%
          summarize(genus_size = n_distinct(species),
                    mass_g = mean(mass_g, na.rm = TRUE),
                    gestation_mo = mean(gestation_mo, na.rm = TRUE),
                    newborn_g = mean(newborn_g, na.rm = TRUE),
                    weaning_mo = mean(weaning_mo, na.rm = TRUE),
                    wean_g = mean(wean_g, na.rm = TRUE),
                    afr_mo = mean(afr_mo, na.rm = TRUE),
                    life_mo = mean(life_mo, na.rm = TRUE),
                    litter_size = mean(litter_size, na.rm = TRUE),
                    litters_per_year = mean(litters_per_year, na.rm = TRUE))

    reshuffled_res_afr <- lm(genus_size ~ afr_mo, data = genus_temp_afr)
    slope_sim_afr <- coef(reshuffled_res_afr)["afr_mo"]

    simulated_slopes_afr[i] <- slope_sim_afr
}

simulated_slopes_afr <- unlist(simulated_slopes_afr)
ggplot() +
    ylab("Count") + xlab("Simulated slope") +
    geom_histogram(aes(x = simulated_slopes), bins = 30,
                   fill = "grey", alpha = 0.4, colour = "black") +
    geom_vline(xintercept = slope_real, size = 1,
               linetype = "dashed", colour = "black") +
    theme_classic() + ggtitle("Age of first reproduction")
```

```r
    ggsave("Randomization test of age of first reproduction.png")

# Caculate p value for randomization test of age of first reproduction
abs_simulated_slopes_afr <- abs(simulated_slopes_afr)
exceed_count_afr <- length(abs_simulated_slopes_afr[abs_simulated_slopes_afr >=
                                                    abs(slope_real_afr)])
p_val_afr <- exceed_count_afr / nreps
p_val_afr

# Randomization test for litters per year
set.seed(16)
reshuffled_litter <- mammals_c
reshuffled_litter$litters_per_year <- sample(reshuffled_litter$litters_per_year,
                                              size = nrow(reshuffled_litter),
                                              replace = FALSE)
slope_real_litter <- coef(result_litter)["litters_per_year"]
simulated_slopes_litter <- list()
nreps = 5000 # 5000 iterations
for(i in 1:nreps){

    reshuffled_litter <- mammals_c
    reshuffled_litter$litters_per_year <- sample(reshuffled_litter$litters_per_year,
                                                 size = nrow(reshuffled_litter),
                                                 replace = FALSE)

    genus_temp_litter <- reshuffled_litter %>%
        group_by(Genus) %>%
        summarize(genus_size = n_distinct(species), mass_g = mean(mass_g, na.rm = TRUE),
                  gestation_mo = mean(gestation_mo, na.rm = TRUE),
                  newborn_g = mean(newborn_g, na.rm = TRUE),
                  weaning_mo = mean(weaning_mo, na.rm = TRUE),
                  wean_g = mean(wean_g, na.rm = TRUE),
                  afr_mo = mean(afr_mo, na.rm = TRUE),
                  life_mo = mean(life_mo, na.rm = TRUE),
                  litter_size = mean(litter_size, na.rm = TRUE),
                  litters_per_year = mean(litters_per_year, na.rm = TRUE))

    reshuffled_res_litter <- lm(genus_size ~ litters_per_year, data = genus_temp_litter)
    slope_sim_litter <- coef(reshuffled_res_litter)["litters_per_year"]

    simulated_slopes_litter[i] <- slope_sim_litter
}

simulated_slopes_litter <- unlist(simulated_slopes_litter)
ggplot() +
    ylab("Count") + xlab("Simulated slope") +
    geom_histogram(aes(x = simulated_slopes), bins = 30,
                   fill = "grey", alpha = 0.4, colour = "black") +
    geom_vline(xintercept = slope_real, size = 1,
               linetype = "dashed", colour = "black") +
    theme_classic() + ggtitle("Litters per year")
  ggsave("Randomization test of litters per year.png")
```

```r
# Caculate p value for randomization test of litters per year
abs_simulated_slopes_litter <- abs(simulated_slopes_litter)
exceed_count_litter <- length(abs_simulated_slopes_litter[abs_simulated_slopes_litter >=
                                          abs(slope_real_litter)])
p_val_litter <- exceed_count_litter / nreps
p_val_litter
```

```r
# Randomization test for life span
set.seed(16)
reshuffled_life <- mammals_c
reshuffled_life$life_mo <- sample(reshuffled_life$life_mo,
                                          size = nrow(reshuffled_life),
                                 replace = FALSE)
slope_real_life <- coef(result_life)["life_mo"]
simulated_slopes_life <- list()
nreps = 5000 # 5000 iterations
for(i in 1:nreps){

    reshuffled_life <- mammals_c
    reshuffled_life$life_mo <- sample(reshuffled_life$llife_mo,
                                              size = nrow(reshuffled_life),
                                     replace = FALSE)

    genus_temp_life <- reshuffled_life %>%
            group_by(Genus) %>%
            summarize(genus_size = n_distinct(species), mass_g = mean(mass_g, na.rm = TRUE),
                    gestation_mo = mean(gestation_mo, na.rm = TRUE),
                    newborn_g = mean(newborn_g, na.rm = TRUE),
                    weaning_mo = mean(weaning_mo, na.rm = TRUE),
                    wean_g = mean(wean_g, na.rm = TRUE),
                    afr_mo = mean(afr_mo, na.rm = TRUE),
                    life_mo = mean(life_mo, na.rm = TRUE),
                    litter_size = mean(litter_size, na.rm = TRUE),
                    litters_per_year = mean(litters_per_year, na.rm = TRUE))

    reshuffled_res_life<- lm(genus_size ~ life_mo, data = genus_temp_life)
    slope_sim_life<- coef(reshuffled_res_life)["life_mo"]

    simulated_slopes_life[i] <- slope_sim_life
}

simulated_slopes_life <- unlist(simulated_slopes_life)
ggplot() +
    ylab("Count") + xlab("Simulated slope") +
    geom_histogram(aes(x = simulated_slopes), bins = 30,
                fill = "grey", alpha = 0.4, colour = "black") +
    geom_vline(xintercept = slope_real, size = 1,
             linetype = "dashed", colour = "black") +
    theme_classic() + ggtitle("Life span")
  ggsave("Randomization test of life span.png")
```

```r
# Caculate p value for randomization test of life span
abs_simulated_slopes_life <- abs(simulated_slopes_life)
exceed_count_life <- length(abs_simulated_slopes_life[abs_simulated_slopes_life >=
```

```
                                                   abs(slope_real_life)])
p_val_life <- exceed_count_life / nreps
p_val_life
```

Drop-one test focussing on Rodentia

```r
# Create a data set without the order: Rodentia
genus_no_rod <- mammals_c %>%
  filter(order != "Rodentia") %>%
  group_by(Genus) %>%
  summarize(genus_size = n_distinct(species),
            mass_g = mean(mass_g, na.rm = TRUE),
            gestation_mo = mean(gestation_mo, na.rm = TRUE),
            newborn_g = mean(newborn_g, na.rm = TRUE),
            weaning_mo = mean(weaning_mo, na.rm = TRUE),
            wean_g = mean(wean_g, na.rm = TRUE),
            afr_mo = mean(afr_mo, na.rm = TRUE),
            life_mo = mean(life_mo, na.rm = TRUE),
            litter_size = mean(litter_size, na.rm = TRUE),
            litters_per_year = mean(litters_per_year, na.rm = TRUE))
```

```r
# Linear regression of gestation time after droping order Rodentia
genus_no_rod %>%
  ggplot(aes(x = gestation_mo, y = genus_size)) +
  geom_point() +
  geom_smooth(method = "lm")

lm_ges_no_rod <- lm(genus_size ~ gestation_mo,
                    data = genus_no_rod)
summary(lm_ges_no_rod)
```

```r
# Linear regression of age of first reproduction after droping order Rodentia
genus_no_rod %>%
  ggplot(aes(x = afr_mo, y = genus_size)) +
  geom_point() +
  geom_smooth(method = "lm")

lm_afr_no_rod <- lm(genus_size ~ afr_mo,
                    data = genus_no_rod)
summary(lm_afr_no_rod)
```

```r
# Linear regression of litters per year after droping order Rodentia
genus_no_rod %>%
  ggplot(aes(x = litters_per_year, y = genus_size)) +
  geom_point() +
  geom_smooth(method = "lm")

lm_litter_no_rod <- lm(genus_size ~ litters_per_year,
                    data = genus_no_rod)
summary(lm_litter_no_rod)
```

```r
# Linear regression of life span after droping order Rodentia
genus_no_rod %>%
```

```
  ggplot(aes(x = life_mo, y = genus_size)) +
  geom_point() +
  geom_smooth(method = "lm")

lm_life_no_rod <- lm(genus_size ~ life_mo,
                     data = genus_no_rod)
summary(lm_life_no_rod)
```

Drop-one test focussing on Cetacia

```
# Create a data set without the order: Cetacia
genus_no_cet <- mammals_c %>%
  filter(order != "Cetacia") %>%
  group_by(Genus) %>%
  summarize(genus_size = n_distinct(species),
            mass_g = mean(mass_g, na.rm = TRUE),
            gestation_mo = mean(gestation_mo, na.rm = TRUE),
            newborn_g = mean(newborn_g, na.rm = TRUE),
            weaning_mo = mean(weaning_mo, na.rm = TRUE),
            wean_g = mean(wean_g, na.rm = TRUE),
            afr_mo = mean(afr_mo, na.rm = TRUE),
            life_mo = mean(life_mo, na.rm = TRUE),
            litter_size = mean(litter_size, na.rm = TRUE),
            litters_per_year = mean(litters_per_year, na.rm = TRUE))

# Linear regression of gestation time after droping order Cetacia
lm_ges_no_cet <- lm(genus_size ~ gestation_mo,
                    data = genus_no_cet)
summary(lm_ges_no_cet)

# Linear regression of age of first reproduction after droping order Cetacia
lm_afr_no_cet <- lm(genus_size ~ afr_mo,
                    data = genus_no_cet)
summary(lm_afr_no_cet)

# Linear regression of litters per year after droping order Cetacia
lm_litter_no_cet <- lm(genus_size ~ litters_per_year,
                       data = genus_no_cet)
summary(lm_litter_no_cet)

# Linear regression of life span after droping order Cetacia
lm_life_no_cet <- lm(genus_size ~ life_mo,
                     data = genus_no_cet)
summary(lm_life_no_cet)
```

Drop-one test focussing on Perissodactyla

```
# Create a data set without the order: Perissodactyla
genus_no_per <- mammals_c %>%
  filter(order != "Perissodactyla") %>%
  group_by(Genus) %>%
  summarize(genus_size = n_distinct(species), mass_g = mean(mass_g, na.rm = TRUE),
```

```r
            gestation_mo = mean(gestation_mo, na.rm = TRUE),
            newborn_g = mean(newborn_g, na.rm = TRUE),
            weaning_mo = mean(weaning_mo, na.rm = TRUE),
            wean_g = mean(wean_g, na.rm = TRUE),
            afr_mo = mean(afr_mo, na.rm = TRUE),
            life_mo = mean(life_mo, na.rm = TRUE),
            litter_size = mean(litter_size, na.rm = TRUE),
            litters_per_year = mean(litters_per_year, na.rm = TRUE))
```

```r
# Linear regression of gestation time after droping order Perissodactyla
lm_ges_no_per <- lm(genus_size ~ gestation_mo,
                    data = genus_no_per)
summary(lm_ges_no_per)
```

```r
# Linear regression of age of first reproduction after droping order Perissodactyla
lm_afr_no_per <- lm(genus_size ~ afr_mo,
                    data = genus_no_per)
summary(lm_afr_no_per)
```

```r
# Linear regression of litters per year after droping order Perissodactyla
lm_litter_no_per <- lm(genus_size ~ litters_per_year,
                    data = genus_no_per)
summary(lm_litter_no_per)
```

```r
# Linear regression of life span after droping order Perissodactyla
lm_life_no_per <- lm(genus_size ~ life_mo,
                    data = genus_no_per)
summary(lm_life_no_per)
```

Drop-one test focussing on Primate

```r
# Create a data set without the order: Primate
genus_no_prim <- mammals_c %>%
  filter(order != "Primate") %>%
  group_by(Genus) %>%
  summarize(genus_size = n_distinct(species), mass_g = mean(mass_g, na.rm = TRUE),
            gestation_mo = mean(gestation_mo, na.rm = TRUE),
            newborn_g = mean(newborn_g, na.rm = TRUE),
            weaning_mo = mean(weaning_mo, na.rm = TRUE),
            wean_g = mean(wean_g, na.rm = TRUE),
            afr_mo = mean(afr_mo, na.rm = TRUE),
            life_mo = mean(life_mo, na.rm = TRUE),
            litter_size = mean(litter_size, na.rm = TRUE),
            litters_per_year = mean(litters_per_year, na.rm = TRUE))
```

```r
# Linear regression of gestation time after droping order Primate
lm_ges_no_prim <- lm(genus_size ~ gestation_mo,
                     data = genus_no_prim)
summary(lm_ges_no_prim)
```

```r
# Linear regression of age of first reproduction after droping order Primate
lm_afr_no_prim <- lm(genus_size ~ afr_mo,
                     data = genus_no_prim)
summary(lm_afr_no_prim)
```

```r
# Linear regression of litters per year after droping order Primate
lm_litter_no_prim <- lm(genus_size ~ litters_per_year,
                        data = genus_no_prim)
summary(lm_litter_no_prim)

# Linear regression of life span after droping order Primate
lm_life_no_prim <- lm(genus_size ~ life_mo,
                      data = genus_no_prim)
summary(lm_life_no_prim)
```

Testing normality by creating the Normal Q-Q plot

```r
par(mfrow = c(2,2)) # Show the plots 2 * 2
# The second one of diagnostic plots is the Normal Q-Q plot
plot(lm(genus_size ~ gestation_mo, data = mammals_g),
     which = 2, xlab = "Gestation time")
plot(lm(genus_size ~ afr_mo, data = mammals_g),
     which = 2, xlab = "Age of first reproduction")
plot(lm(genus_size ~ litters_per_year, data = mammals_g),
     which = 2, xlab = "Litters per year")
plot(lm(genus_size ~ life_mo, data = mammals_g),
     which = 2, xlab = "Life span")
```

Testing normality after doing square root transformation to x

```r
par(mfrow = c(2,2))
# Take square root of all x values
plot(lm(genus_size ~ sqrt(gestation_mo), data = mammals_g),
     which = 2, xlab = "Gestation time")
plot(lm(genus_size ~ sqrt(afr_mo), data = mammals_g),
     which = 2, xlab = "Age of first reproduction")
plot(lm(genus_size ~ sqrt(litters_per_year), data = mammals_g),
     which = 2, xlab = "Litters per year")
plot(lm(genus_size ~ sqrt(life_mo), data = mammals_g),
     which = 2, xlab = "Life span")
```

Testing normality after doing higher root transformation to x

```r
# Take the forth roots of all x values
higer_power_gestation <- mammals_g$gestation_mo ^ (1/4)
higer_power_afr <- mammals_g$afr_mo ^ (1/4)
higer_power_litters<- mammals_g$litters_per_year ^ (1/4)
higer_power_lifespan <- mammals_g$life_mo ^ (1/4)

par(mfrow = c(2,2))
plot(lm(mammals_g$genus_size ~ higer_power_gestation),
     which = 2, xlab = "Gestation time")
```

```r
plot(lm(mammals_g$genus_size ~ higer_power_afr),
     which = 2, xlab = "Age of first reproduction")
plot(lm(mammals_g$genus_size ~ higer_power_litters),
     which = 2, xlab = "Litters per year")
plot(lm(mammals_g$genus_size ~ higer_power_lifespan),
     which = 2, xlab = "Life span")
```

Testing constant variance by creating Scale-Location graph

```r
par(mfrow = c(2,2))
# The third one of diagnostic plots is the Scale-Location plot
plot(lm(genus_size ~ gestation_mo, data = mammals_g), which = 3)
plot(lm(genus_size ~ afr_mo, data = mammals_g), which = 3)
plot(lm(genus_size ~ litters_per_year, data = mammals_g), which = 3)
plot(lm(genus_size ~ life_mo, data = mammals_g), which = 3)
```

Testing constant variance after doing square root transformation to x

```r
par(mfrow = c(2,2))
# Take square root of all x values
plot(lm(genus_size ~ sqrt(gestation_mo), data = mammals_g),
     which = 3)
plot(lm(genus_size ~ sqrt(afr_mo), data = mammals_g),
     which = 3)
plot(lm(genus_size ~ sqrt(litters_per_year), data = mammals_g),
     which = 3)
plot(lm(genus_size ~ sqrt(life_mo), data = mammals_g),
     which = 3)
```

Testing normality after dropping 4 orders

```r
library(dplyr)
# Create a new data without 4 orders: Rodentia, Cetacea, Primates and Perissodactyla
drop_4_orders <- mammals_o %>%
  filter(order != "Rodentia" & order != "Cetacea" &
           order != "Primates" & order != "Perissodactyla")

par(mfrow = c(2,2))
plot(lm(order_size ~ gestation_mo, data = drop_4_orders),
     which = 2, xlab = "Gestation time")
plot(lm(order_size ~ afr_mo, data = drop_4_orders),
     which = 2, xlab = "Age of first reproduction")
plot(lm(order_size ~ litters_per_year, data = drop_4_orders),
     which = 2, xlab = "Litters per year")
```

```
plot(lm(order_size ~ life_mo, data = drop_4_orders),
     which = 2, xlab = "Life span")
```

Graph shows diversification of life span in genera within their order

```
ggplot(mammals_c, aes(x = life_mo, y = order, color = Genus)) +
    geom_point(size = 0.1, position = 'jitter') +
    theme(legend.position = "none")
```