# Reanalysis of 11-Carlisle

Simon Schwab*

26/08/2020

# Contents

# Reference

Carlisle, A., Selwood, L., Hinds, L. A., Saunders, N., Habgood, M., Mardon, K., & Weisbecker, V. (2017). Testing hypotheses of developmental constraints on mammalian brain partition evolution, using marsupials. Scientific Reports, 7(1), 4241. https://doi.org/10.1038/s41598-017-02726-9

# Notes from reading methods section

- Dependent variable: brain partition size of olfactory bulb
- Independent variable: species (3 levels)

---

*University of Zurich, simon.schwab@uzh.ch

- Macropus eugenii
- Trichosurus vulpecula
- Monodelphis domestica
- Covariate: whole brain volume
- Design: 1-way ANCOVA with species as IV and whole brain volume as covariate

# Reading data

Data is loaded, reshaped if necessary, and factors are specified.

```r
PATH = file.path(path.expand("~"), "Data", "ancova") # ANCOVA project folder
data = read.csv(file.path(PATH, "dataPrimaryStudies", "11-Carlisle", "11-Carlisle-original.txt"),
                sep = "", header = TRUE)

data$Species.name = NA
data$Species.name[data$Species == 1] = "Macropus eugenii"
data$Species.name[data$Species == 2] = "Trichosurus vulpecula"
data$Species.name[data$Species == 3] = "Monodelphis domestica"

data$Species      = as.factor(data$Species)
data$Species.name = as.factor(data$Species.name)

# first reported analyses are only with juveniles
data = subset(data, subset = Juvenile.Adult == "juvenile")
```

# Descriptives

## Dependant variable

Number of samples and mean (SD) in levels of the independent variables.

```r
tab.dv = array(NA, dim=c(nlevels(data$Species.name),3))
tab.dv[,1] = levels(data$Species.name)
tab.dv[,2] = summary(data$Species.name)
tab.dv[,3] = tapply(data$Olfactory_Bulb, data$Species.name,
                    function (x) sprintf("%0.2f (%0.2f)", mean(x), sd(x)))
colnames(tab.dv) = c("group", "n", "mean (SD)")
print(tab.dv)
```

```
##      group                   n    mean (SD)
## [1,] "Macropus eugenii"      "13" "52.84 (86.42)"
## [2,] "Monodelphis domestica" "11" "26.72 (26.77)"
## [3,] "Trichosurus vulpecula" "11" "61.17 (93.79)"
```

```r
ggplot(data, aes(y=log(Olfactory_Bulb), x=Species.name, color=Species.name)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  xlab("species") + ylab("olfactory bulb volume (log)") + ggtitle("Dependant variable")
```

**Dependant variable**
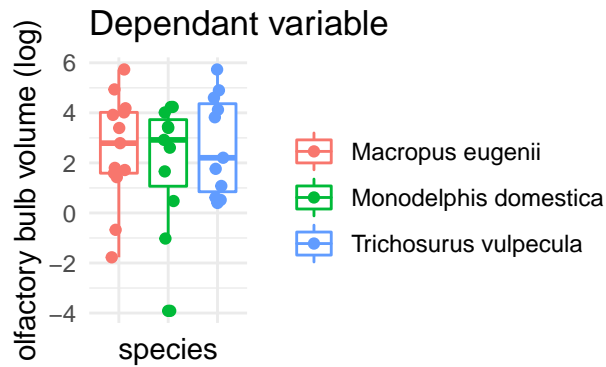
## Covariate(s)

```r
tab.cv = array(NA, dim=c(nlevels(data$Species.name),3))
tab.cv[,1] = levels(data$Species.name)
tab.cv[,2] = summary(data$Species.name)
tab.cv[,3] = tapply(data$Whole_Brain_Volume, data$Species.name,
                    function (x) sprintf("%0.2f (%0.2f)", mean(x), sd(x)))
colnames(tab.cv) = c("group", "n", "mean (SD)")
print(tab.cv)
```

```
##       group                    n    mean (SD)
## [1,] "Macropus eugenii"      "13" "2160.14 (3971.33)"
## [2,] "Monodelphis domestica" "11" "329.58 (289.08)"
## [3,] "Trichosurus vulpecula" "11" "2130.30 (3587.35)"
```

```r
ggplot(data, aes(y=log(Whole_Brain_Volume), x=Species.name, color=Species.name)) +
  geom_boxplot() +
  geom_point(position = position_jitter(width = 0.15, height = 0)) +
  theme_minimal() +
  theme(axis.text.x = element_blank(), legend.title = element_blank()) +
  xlab("group") + ylab("Whole brain volume (log)") + ggtitle("Covarate")
```
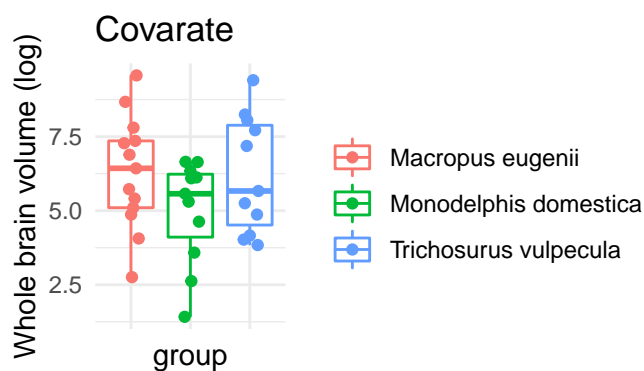


## Main analysis ANCOVA

```r
# Orthogonal contrasts
contrasts(data$Species.name) = contr.helmert(3)
```

```r
# The author did not run a classic ANCOVA with IV + COV, but instead an interaction
# COV*IV similar to testing the homogeneity of slope assumption
fit.ancova = aov(log(Olfactory_Bulb) ~ log(Whole_Brain_Volume-Olfactory_Bulb)*Species, data = data)
result = Anova(fit.ancova, type=3) # Type III
print(result)
```

```
## Anova Table (Type III tests)
##
## Response: log(Olfactory_Bulb)
##                                                Sum Sq Df  F value     Pr(>F)
## (Intercept)                                    20.349  1 190.5605 2.805e-14
## log(Whole_Brain_Volume - Olfactory_Bulb)       54.073  1 506.3676 < 2.2e-16
## Species                                         1.553  2   7.2699   0.00276
## log(Whole_Brain_Volume - Olfactory_Bulb):Species  3.912  2  18.3183 7.181e-06
## Residuals                                       3.097 29
##
## (Intercept)                                    ***
## log(Whole_Brain_Volume - Olfactory_Bulb)       ***
## Species                                        **
## log(Whole_Brain_Volume - Olfactory_Bulb):Species ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Comparing ANCOVA in original study with reanalysis

## Independant variable

```r
tab.IV = rbind(stats.orig.IV, stats.rep.IV)
rownames(tab.IV) = c("original Study", "reanalysis")
print(t(tab.IV))
```

```
##         original Study reanalysis
## Fvalue  NA             "18.32"
## df1     NA             " 2"
## df2     NA             "29"
## pvalue  "< 0.0001"     "< 0.0001"
## MD      NA             NA
## lowerCI NA             NA
## upperCI NA             NA
```

## Covariate

```r
tab.CV = rbind(stats.orig.CV, stats.rep.CV)
rownames(tab.CV) = c("original Study", "reanalysis")
print(t(tab.CV))
```

```
##         original Study reanalysis
## Fvalue  NA             "506.37"
## df1     NA             " 1"
## df2     NA             " 2"
```

```
## pvalue  NA            "< 0.0001"
## MD      NA            NA
## lowerCI NA            NA
## upperCI NA            NA
```

# Assumptions

## 1. Homogeneity of variance

- ANOVA/ANCOVA is fairly robust in terms of the error rate when sample sizes are equal.
- When groups with larger sample sizes have larger variances than the groups with smaller sample sizes, the resulting F-ratio tends to be conservative. That is, it's more likely to produce a non-significant result when a genuine difference does exist in the population.
- Conversely, when the groups with larger sample sizes have smaller variances than the groups with smaller samples sizes, the resulting F-ratio tends to be liberal and can inflate the false positive rate.

```
tapply(log(data$Olfactory_Bulb), data$Species.name, sd)
```

```
##     Macropus eugenii Monodelphis domestica Trichosurus vulpecula
##            2.156197              2.565358              1.977223
```

```
leveneTest(log(Olfactory_Bulb) ~ Species.name, data = data)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  2  0.0086 0.9914
##       32
```

## 2. Independence between covariate and IV

- When the covariate and the experimental effect (independent variable) are not independent the treatment effect is obscured, spurious treatment effects can arise and the interpretation of the ANCOVA is seriously compromised.

- We test whether our groups differ on the CV. If the groups do not significantly differ then is appropriate to use the covariate.

```
fit.cv = aov(log(Whole_Brain_Volume-Olfactory_Bulb) ~ Species.name, data = data)
Anova(fit.cv, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: log(Whole_Brain_Volume - Olfactory_Bulb)
##               Sum Sq Df  F value Pr(>F)
## (Intercept)  1170.15  1 336.5831 <2e-16 ***
## Species.name   12.83  2   1.8458 0.1743
## Residuals     111.25 32
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 3. Homogeneity of regression slopes

- We test the interaction between the IV and the CV

```
fit.hrs = aov(log(Olfactory_Bulb) ~ log(Whole_Brain_Volume-Olfactory_Bulb)*Species.name, data = data)
Anova(fit.hrs, type=3)
```

```
## Anova Table (Type III tests)
##
## Response: log(Olfactory_Bulb)
##                                                     Sum Sq Df    F value
## (Intercept)                                         59.023  1   552.7269
## log(Whole_Brain_Volume - Olfactory_Bulb)           156.487  1  1465.4368
## Species.name                                         1.553  2     7.2699
## log(Whole_Brain_Volume - Olfactory_Bulb):Species.name  3.912  2    18.3183
## Residuals                                            3.097 29
##                                                       Pr(>F)
## (Intercept)                                        < 2.2e-16 ***
## log(Whole_Brain_Volume - Olfactory_Bulb)           < 2.2e-16 ***
## Species.name                                         0.00276 **
## log(Whole_Brain_Volume - Olfactory_Bulb):Species.name 7.181e-06 ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(data, aes(y=log(Olfactory_Bulb), x=log(Whole_Brain_Volume-Olfactory_Bulb), color=Species.name, s
  geom_point() +
  geom_smooth(formula = y ~ x,method=lm, se=FALSE, fullrange=TRUE) +
  theme_minimal() +
  theme(legend.title = element_blank()) +
  xlab("whole brain volume") + ylab("olfactory bulb")
```
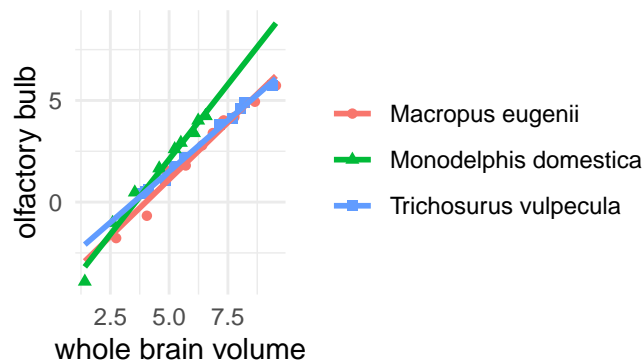


## Additonal analyses

```
# Reproducing Figure 4
ggplot(data, aes(x = log(Whole_Brain_Volume), y = log(Olfactory_Bulb))) +
  geom_point() + theme_minimal()
```
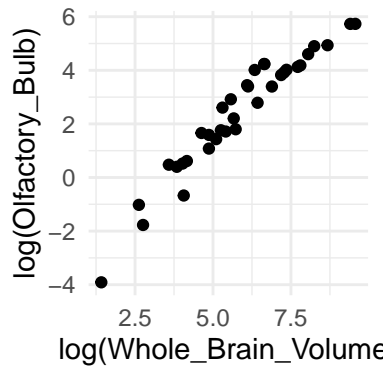
```
# Reproducing Table 1
fit = lm(log(Olfactory_Bulb)~log(Whole_Brain_Volume-Olfactory_Bulb), data=data)
summary(fit)
```

```
##
## Call:
## lm(formula = log(Olfactory_Bulb) ~ log(Whole_Brain_Volume - Olfactory_Bulb),
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47470 -0.31060 -0.05929  0.41685  1.13776
##
## Coefficients:
##                                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                               -3.99436    0.34643  -11.53 4.08e-13
## log(Whole_Brain_Volume - Olfactory_Bulb)   1.10161    0.05657   19.47  < 2e-16
##
## (Intercept)                               ***
## log(Whole_Brain_Volume - Olfactory_Bulb) ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6301 on 33 degrees of freedom
## Multiple R-squared:  0.9199, Adjusted R-squared:  0.9175
## F-statistic: 379.2 on 1 and 33 DF,  p-value: < 2.2e-16
```

```
fit = lm(log(Olfactory_Bulb)~log(Whole_Brain_Volume-Olfactory_Bulb),
         data=subset(data, subset = Species == 1))
summary(fit)
```

```
##
## Call:
## lm(formula = log(Olfactory_Bulb) ~ log(Whole_Brain_Volume - Olfactory_Bulb),
##     data = subset(data, subset = Species == 1))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7233 -0.2690  0.1020  0.2343  0.6669
##
## Coefficients:
##                                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                               -4.49546    0.39378  -11.42 1.94e-07
```

```
## log(Whole_Brain_Volume - Olfactory_Bulb)  1.12144    0.06026    18.61 1.16e-09
##
## (Intercept)                                ***
## log(Whole_Brain_Volume - Olfactory_Bulb) ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3951 on 11 degrees of freedom
## Multiple R-squared:  0.9692, Adjusted R-squared:  0.9664
## F-statistic: 346.3 on 1 and 11 DF,  p-value: 1.156e-09
```

# Notes

- First main result reproduced
- Assumptions hold. Homogeneity of regression slopes was part of the research questions (Table 2).
- ANCOVA used with interactions term for species and whole brain volume, not to (globally) regress covariate
- Log transform of IV and CV was successful
- Helpful R code was supplied
- Model was clearly specified in the methods section
- Explain ANCOVA in the framework of regression (slope, intercept)
- Table 2 is confusing as p-values were associated with the DV instead of the IV
- Lots of models tested, no correction for multiplicity (7 ANCOVAs were run across different brain regions)

Data was analyzed according to recommendations by Field, Miles, & Field (2012).