**Winter, 2015 Biometry Quiz #3**

1. **[30 pts]** Welbergen (2010; Journal of Mammalogy 91:38-47) examined the relationship between body mass (grams) and forearm length (mm) for three groups of gray-headed flying foxes (*Pteropus poliocephalus*) – adult males, adult females, and juveniles (labeled as “young”). Welbergen’s primary interest with these data was to see if body mass could be predicted from forearm length and whether those predictions were dependent on the group (variable called “sex”) of flying fox. Welbergen’s data was entered into R and analyzed with the commands shown in ***Results #1***. Use this information to answer the following questions.

**[2 pts]**

* 1. Construct the ultimate full model for these data, with an explicit definition of each variable. [*Variables should be entered into the model according to the order discussed in class.*]
  2. Two linear models are presented in the results – lm1 and lm2. Which model do you choose to analyze and why? Be as specific as possible with evidence to support your choice.

*Check your model choice with me before continuing!!*

Use the model you identified above to answer the following questions. Note that I have put the variable names in quotes below because you may be referring to a transformed version of the variable (and you should do so in your answer).

* 1. What can you say about the relationship between “body mass” and “forearm length” for these flying foxes? Be as explicit as you possibly can -- make sure to comment on any differences between groups (i.e., sexes) and specifically state what the relationship is for each group. Provide evidence for your conclusions.
  2. What can you say, if anything, about differences in “body mass” among groups (i.e., sexes) after having factored out the effect of “forearm length” on “body mass”? Again be as explicit as you possibly can. Provide evidence for your conclusions.
  3. At the end of the results for your chosen model are results from predict(). Interpret each of these values and comment on whether these values are statistically significantly different or not (with supporting evidence).

**[2 pts]**

1. **[20 pts]** A graduate student examined the question of whether roads had an impact on the density of Moose. As an initial investigation of this question, the graduate student created a large number of “plots” that were up to 6 miles from the nearest road. For each plot, the student recorded evidence for the presence of Moose (e.g., tracks, droppings) and the straight-line distance from the nearest paved road (determined via GIS analysis). The data are read and analyzed with the R code in ***Results #2***. Use these results to answer the following questions.
   1. Interpret the slope.
   2. Interpret the back-transformed slope.
   3. Predict the log odds for the presence of Moose in a “plot” that is 3 miles from a road.
   4. Predict the odds for the presence of Moose in a “plot” that is 3 miles from a road. Specifically interpret this result.
   5. Predict the probability of the presence of Moose in a “plot” that is 3 miles from a road.
   6. Compute the distance from a road at which 50% of all plots would show a presence of Moose.
   7. Compute the distance from a road at which 90% of all plots would show a presence of Moose.
   8. How much do you expect the odds to change if the distance from a road changes from X to X+2 miles? Provide evidence that does not include computing the odds at both distances.
2. **[15 pts total]** Answer ***two*** of the following three questions.
3. **[worth 10 pts]** Suppose that you have collected a quantitative response variable. In addition you may have collected (i) a quantitative explanatory variable, (ii) a categorical explanatory variable, (iii) both a quantitative and categorical explanatory variable, or (iv) two categorical explanatory variables. Identify which statistical procedure/test you would use in each of these four situations and describe the type(s) of question(s) or hypothesis(es) that you would be able to answer with each.
4. **[worth 5 pts]** Kleiber's law is the observation that, for the vast majority of animals, an animal's metabolic rate scales to the ¾ power of the animal's mass. Thus, if *q* is the animal's metabolic rate and *M* is the animal's mass, then Kleiber's law states that *q* = a*M*3/4, where a is a proportionality constant[[1]](#footnote-1). Suppose that you have metabolism and mass data for a large number and wide variety of animals. What would you do with these data to test Kleiber’s Law (i.e., is the power constant in the relationship equal to ¾). Show each step of your work and be specific with your answer.
5. **[worth 5 pts]** Suppose that you have conducted a one-way ANOVA and received a particular result (either the groups have different means or they do not). Now suppose that you introduce a quantitative covariate to the analysis that is significantly related to the response variable and the relationship between the covariate and the response does not differ by group. The introduction of this covariate to the analysis could lead you to reverse your original conclusion – i.e., changing from concluding that the groups differed to concluding that they do not differ (or vice versa). Explain how each of these reversals could happen (you likely will want to refer to variabilities “explained” and “not explained” and may want to draw some illustrative graphics).

Biometry Quiz #3 R Handout Results #1

**> ff <- read.table("FlyingFox1.txt",header=TRUE)**

**> str(ff)**

'data.frame': 223 obs. of 3 variables:

$ forearm.len: num 152 154 157 157 158 ...

$ body.mass : num 779 788 839 761 763 ...

$ sex : Factor w/ 3 levels "female","male",..: 2 2 2 2 2 2 2 2 2 2 ...

**> ff$sex <- factor(ff$sex,levels=c("young","male","female"))**

**> xtab(~sex,data=ff)**

young male female

79 75 69

**> lm1 <- lm(body.mass~forearm.len\*sex,data=ff)**

**> transChooser(lm1,show.stats=TRUE)**

** **

**> anova(lm1)**

Df Sum Sq Mean Sq F value Pr(>F)

forearm.len 1 9051340 9051340 1792.2167 < 2e-16 \*\*\*

sex 2 987212 493606 97.7368 < 2e-16 \*\*\*

forearm.len:sex 2 30210 15105 2.9908 0.05232 .

Residuals 217 1095928 5050

**> compSlopes(lm1)**

Multiple Slope Comparisons

comparison diff lwr upr raw.p adj.p

1 male-young 5.67494 0.80082 10.54905 0.02270 0.06178

2 female-young -0.08093 -4.76632 4.60445 0.97287 0.97287

3 female-male -5.75587 -11.27945 -0.23228 0.04119 0.06178

Slope Information

level slopes lwr upr raw.p adj.p

3 female 7.2419 3.4534 11.030 0.00021 0.00021

1 young 7.3229 4.5661 10.080 0.00000 0.00000

2 male 12.9978 8.9782 17.017 0.00000 0.00000

**> compIntercepts(lm1)**

Tukey HSD on adjusted means assuming parallel lines.

comparison diff lwr upr p.adj

1 male-young 274.22 247.00 301.44 0

2 female-young 117.43 89.61 145.25 0

3 female-male -156.79 -184.95 -128.63 0

Mean adjusted values at a covariate value of 154.9

young male female

490.56 764.78 607.99

**> fitPlot(lm1,legend="topleft",col="gray")**

****

**> predict(lm1,data.frame(forearm.len=c(155,155,155),sex=c("young","male","female")))**

1 2 3

471.0677 723.9145 619.1563

**> ff$log.bm <- log(ff$body.mass)**

**> ff$log.fl <- log(ff$forearm.len)**

**> lm2 <- lm(log.bm~log.fl\*sex,data=ff)**

**> anova(lm2)**

Df Sum Sq Mean Sq F value Pr(>F)

log.fl 1 31.4080 31.4080 2588.8126 < 2e-16 \*\*\*

sex 2 1.8891 0.9445 77.8534 < 2e-16 \*\*\*

log.fl:sex 2 0.0606 0.0303 2.4973 0.08467 .

Residuals 217 2.6327 0.0121

**> compSlopes(lm2)**

Multiple Slope Comparisons

comparison diff lwr upr raw.p adj.p

1 male-young -0.43715 -1.6134 0.73908 0.46465 0.46465

2 female-young -1.26771 -2.3880 -0.14740 0.02675 0.08025

3 female-male -0.83056 -2.2220 0.56091 0.24070 0.36105

Slope Information

level slopes lwr upr raw.p adj.p

3 female 1.7103 0.75957 2.6610 0.00048 0.00048

2 male 2.5409 1.52483 3.5569 0.00000 0.00000

1 young 2.9780 2.38538 3.5706 0.00000 0.00000

**> compIntercepts(lm2)**

Tukey HSD on adjusted means assuming parallel lines.

comparison diff lwr upr p.adj

1 male-young 0.43712 0.39502 0.47921 0

2 female-young 0.24246 0.19944 0.28548 0

3 female-male -0.19465 -0.23821 -0.15110 0

Mean adjusted values at a covariate value of 5.039

young male female

6.1314 6.5685 6.3739

**> fitPlot(lm2,legend="topleft",col="gray")**

****

**> predict(lm2,data.frame(log.fl=c(5.04,5.04,5.04),sex=c("young","male","female")))**

1 2 3

6.172781 6.575204 6.419336

Biometry Quiz #3 R Handout Results #2

**> df <- read.csv("NewData.csv")**

**> str(df)**

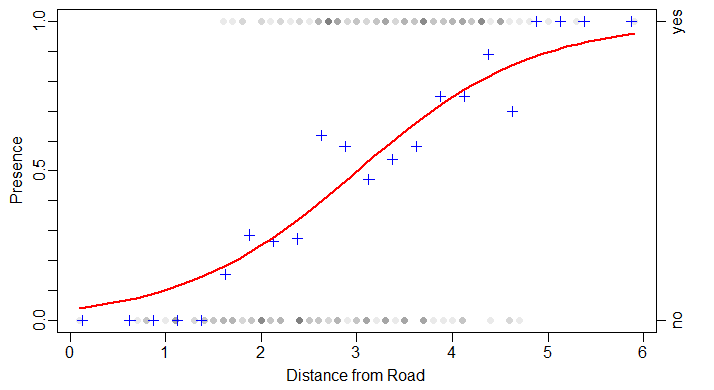
'data.frame': 200 obs. of 2 variables:

$ presence: Factor w/ 2 levels "no","yes": 1 1 1 1 2 1 2 1 1 1 ...

$ distance: num 3.1 2.2 3.7 3.1 3.5 3.1 4.3 1.1 0.8 2 ...

**> glm1 <- glm(presence~distance,data=df,family=binomial)**

**> fitPlot(glm1,breaks=seq(0,7,0.25),xlab="Distance from Road",ylab="Presence")**



**> summary(glm1)**

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -3.2656 0.5693 -5.736 9.71e-09 \*\*\*

distance 1.0874 0.1830 5.942 2.81e-09 \*\*\*

---

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 277.18 on 199 degrees of freedom

Residual deviance: 230.08 on 198 degrees of freedom

AIC: 234.08

Number of Fisher Scoring iterations: 3

1. Thus, for example a cat that has a mass 100 times that of a mouse will have a metabolism roughly 31 times greater than that of a mouse. [↑](#footnote-ref-1)