

Randomization and Permutation Lab: Infant Handling in Female Baboons

Statistical Software: R

Step 1) Paper Review: Read the paper by Bentley-Condit, Moore and Smith, “Analysis of Infant Handling and the Effects of Female Rank Among Tana River Adult Female Yellow Baboons Using Permutation/Randomization Tests”, *American Journal of Primatology*, Vol. 55, pp. 117-130, 2001.

1. If there are any words or terminology that you do not understand, look them up and provide a short definition for each.
2. Identify or answer the following:
 - a) Objective of the study
 - b) Any relevant background (from journals that were referenced)
 - c) Response variable(s)
 - d) Explanatory variables and levels that were tested
 - e) Variables that were held constant during the study
 - f) Nuisance factors (i.e. factors that are not of interest, but may influence the results)
 - g) Were any interactions between variables tested, and if so, what was observed?
 - h) What type of design was used in the study?
 - i) How (if at all) would you modify this study if you were going to test the same hypotheses?

Be ready to submit your answers as well as discuss this material in class.

Step 2) Developing a Research Hypothesis

Adult female primates interact with the infants of other females. Such behavior has spawned several descriptive terms to reflect the range of behavior: aunting, babysitting, play-mothering, allomothering, and kidnapping. The term “infant handling” is a neutral and inclusive term for all such interactions, which include, but are not limited to, pulling, hitting, holding, and carrying infants. The research of such behavior spans four decades and the social, functional, and evolutionary understanding of infant handling continue to be the subject of study and debate.

Primatologist Vicki Bentley-Condit of Grinnell College studied interactions between female and infant yellow baboons (*Papio cynocephalus cynocephalus*) at the Tana River National Primate Reserve, Kenya. She collected the data for her study by observing baboons in twenty-minute focal samples over an 11-month period in 1991-92. Her subjects included 23 female baboons, 11 of which were mothers with infants (no mother with more than one offspring). Bentley-Condit observed and recorded interactions between females and infants, excluding interactions between a mother and her own offspring. One objective of her study was to see if female rank (described in the next paragraph) impacted the pattern or success rate of these infant-handling interactions.

Separately from the infant-handler interactions, Bentley-Condit computed a “dominance hierarchy score” for each female using a calculation based on aggressive and submissive interactions between female dyads in the troop. These scores use a standard method and exhibit natural break points that made it possible to translate the scores into High, Mid, and Low ranks, which we will code respectively as 1, 2, and 3 throughout the paper (see Figure 1.)

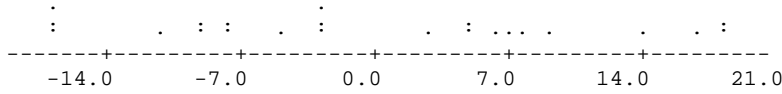


Figure 1: Dominance hierarchy scores for 23 female handlers. The data exhibit 3 natural clusters: High ranks (code=1) are above 10; Mid ranks (code=2) are between 0 and 10; Low ranks (code=3) are below 0. Scores can theoretically range between -22 and +22.

From other published studies, Bentley-Condit expected that patterns of infant handling would be influenced by established female relationships, and she used these dominance ranks as a measure of established relationships within the troop. Her research hypothesis was that female handlers would be disinclined to handle above their own status, and she established the following research hypothesis for investigation¹:

Research Hypothesis 1 (RH1): Females will tend to handle the infants of females who are ranked the same as or lower than themselves.

Table 1 gives the total number of interactions for each female-infant pair *over the entire study period*. For example, the value of 13 in cell (2,1) represents 13 interactions by Handler KM of Infant HZ over the observational period of the study. Handler KM was ranked 1 (High) and the infant HZ had a mother, HQ, which was ranked as a 2 (Mid). The basic summary of the data that directly addresses the research hypothesis is the 3-by-3 table in Table 2A, which categorizes each of the 678 interactions observed over the study period into a cell based upon the rank of the infant (meaning the infant’s mother) and the rank of the female handler. For example, the 5 count in the (1,2) cell is the sum total of all interactions of a high-ranked infant by a mid-ranked female, that is $5 = 1 + 0 + 0 + 0 + 3 + 1 + 0$, reading across the counts corresponding to mid-ranked handlers and the one high-ranked infant in Table 1. If the research hypothesis were true, we would expect counts to be relatively high in cells corresponding to the rank of the handler being at least as high as the infant and low relatively low in the other cells.

Note: We order the dominance rankings as High > Mid > Low, which numerically translates to rank 1 > rank 2 > rank 3 in Table 1.

Infants/
Mothers

Handler’s Names and Ranks

¹ We call this Research Hypothesis 1, as we will introduce the possibility of other research hypotheses later.

Names & Mother's Ranks	KM	KN	NQ	PO	HQ	LL	NY	PS	SK	ST	WK	AL	CO	DD	LS	LY	MH	ML	MM	PA	PH	PT	RS
	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3
KG/KM 1	0	0	4	1	1	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	2	1
HZ/HQ 2	13	23	7	5	0	2	1	1	5	6	18	1	6	3	0	1	4	1	0	9	0	10	1
LC/LL 2	4	0	1	4	3	0	2	1	1	5	3	1	0	0	1	0	2	1	1	1	0	1	6
NK/NY 2	12	4	10	5	9	1	0	2	3	11	7	8	6	3	1	0	2	1	1	5	3	2	3
PZ/PS 2	1	3	4	1	0	0	0	0	0	0	2	0	2	0	0	0	3	0	1	1	0	3	0
CY/CO 3	2	2	7	3	1	1	2	0	3	12	16	3	0	2	0	0	2	0	0	1	0	0	2
LZ/LS 3	1	0	3	2	1	1	0	0	2	0	5	2	2	2	0	1	9	2	0	0	0	3	2
MQ/ML 3	0	1	5	2	2	4	2	2	2	4	5	7	5	2	1	1	7	0	4	4	1	0	2
MW/MH 3	3	0	7	4	2	3	0	5	2	8	13	7	14	2	0	0	0	4	0	8	0	13	6
MX/MM 3	2	3	4	5	0	0	0	0	0	5	2	9	3	1	0	0	2	0	0	1	2	2	3
PK/PH 3	2	0	6	4	3	4	1	0	0	15	10	8	5	1	0	3	1	1	6	3	0	7	5

Table 1: Data matrix giving the total number of interactions by female handler (columns) and infants (rows) over the course of the observational period. Boldface numbers give ranks of handlers or infants, with 1 being High ranking, 2 being Mid ranking, and 3 being Low ranking. Each handler and infant have a two-letter ID, infant ID's are separated by their mother's ID with a /. Horizontal and vertical lines separate the rank categories.

Table 2 provides a summary of the infant-handler interactions by infant (mother) rank and handler rank.

	Handler's Rank					Handler's Rank			
		Hi	Mid	Low			Hi	Mid	Low
Infant	Hi	5	5	3	Infant	Hi	2.9%	2.2%	1.1%
Mother's	Mi	97	83	95	Mother's	Mi	57.1%	36.7%	33.9%
Rank	Lo	68	138	184	Rank	Lo	40.0%	61.1%	65.0%
(A)					(B)				
	Handler's Rank					Handler's Rank			
		Hi	Mid	Low			Hi	Mid	Low
Infant	Hi	3.26	4.33	5.41	Infant	Hi	0.96	0.32	-1.04
Mother's	Mi	68.95	91.67	114.38	Mother's	Mi	3.38	-0.90	-1.81
Rank	Lo	97.79	130.00	162.21	Rank	Lo	-3.01	0.70	1.71
(C)					(D)				

Table 2. Interactions by Infant Rank and Handler Rank (A); column percentages in (B). Table (C) gives the expected values from a chi-squared test of independence and table (D) gives the adjusted, standardized Pearson residuals: $(\text{observed-expected}) \div \sqrt{\text{expected}}$.

Note that there are a few instances where Mid or Low handlers interact with the High ranked infant. This disobeys the research hypothesis, so the question before us is whether the data exhibit a statistical tendency toward the research hypothesis. That is, allowing for variability, is the general pattern in favor of the research hypothesis (RH1) strong enough to rule out chance alone as a plausible explanation for the pattern?

This last question is the question that Professor of Anthropology Bentley-Condit brought to Professor of Statistics Moore after she herself had observed patterns, in a descriptive fashion, which suggested support for the research hypothesis.

In this module, we will work through these steps to investigate the research hypothesis in light of the data:

- We will interpret Table 2 descriptively in light of RH1.
- We will then apply the concept of permutation tests to the data to confirm or deny our descriptive analysis.

In the process of pairing the descriptive with the inferential analysis, we will learn these lessons:

- Permutation tests give an inferential tool in situations where standard methods do not exist.
- Permutation tests are easy to compute in these new situations, given the availability of inexpensive and fast computing power (through R).
- Through the choice of alternative test statistics, permutation tests provide more flexible significance tests than standard methods do.

Step 3) Descriptive analysis of the data

Table 2B uses simple percentages to describe the pattern in the data. You should verify with a calculator that the counts from Table 2A are consistent with the percentages of Table 2B. In describing a pattern in a two-way table, it is almost always helpful to first identify the explanatory and response variables. In observational studies, as we have here, this distinction can be a bit arbitrary, but it is still a useful distinction to make. Here we treat Handler Rank as the explanatory variable (X) and Infant Rank as the response variable (Y).

Notice how Table 2B appears to support the research hypothesis: In the top row, we see that the highest probability of handling a high-ranked infant is from the high-ranked handler (2.9% versus 2.2% and 1.1%). Similarly, mid-ranked infants are more likely to be handled by high- or mid-ranked females than by low-ranked females.

Tables 2C and 2D (which should be viewed and interpreted together) provide an alternative path to the same descriptive conclusion. You may skip this part of the discussion if you have not studied the chi-square test for two-way tables. In the chi-square test, the null hypothesis being tested is whether the explanatory and response variables are independent. This significance test is sometimes called an “omnibus test,” meaning that we use it to test the null hypothesis (that X and Y are independent) without having any specific alternative in mind; when one rejects the null, there is further work needed to ascertain the nature of the dependency between X and Y.

The first step in the test is computing expected counts assuming the null to be true. Table 2C gives these expected counts, for which we conventionally retain a decimal place or two, to

distinguish their theoretical nature from the concrete “observed counts,” which must, of course, be whole numbers. Notice that the counts in Table 2C adhere perfectly to the null hypothesis: if one computes column percentages in Table 2C, all columns are the same: X and Y are perfectly independent. Notice also that the column and row totals are the same in Table 2C as in the original data (Table 2A).

Question 1: Verify the calculation that shows the expected counts in Table 2C adhere perfectly to the null hypothesis.

The extent to which the counts in Table 2C differ from the observed counts in Table 2A, as well as the direction of these differences, will give clues to the strength of support in the data for the research hypothesis. Table 2D highlights these differences in a cool way. Notice that cells in Table 2D are positive or negative depending on whether the corresponding observed count is greater or lesser than the expected count. Moreover, the magnitude of the value in Table 2D reflects how different the observed and expected counts are and (this is the cool part) the values in Table 2D can be interpreted just as we interpret z-scores. So, for example, the 1.71 in the (3,3) cell says that the observed number of interactions of low-ranked handlers on low-ranked infants is much higher than one would expect if handler rank and infant rank were independent, to the tune of almost 2 standard deviations higher. This value gives support to the research hypothesis, as do other entries in Table 2D.

Question 2: Tables 2C and 2D lend support to the research hypothesis and, moreover, they suggest a somewhat nuanced view of the research hypothesis, giving evidence for the propensity for certain kinds of interactions over others. Give a succinct interpretation of Tables 2C and 2D in terms of infant handling in the context of handler vs. mother status.

Question 3: Read in the baboon data into R.

The file `baboons.txt` contains the data that we will need to carry out the analysis. Save this txt file with the baboon data on your computer. Open R and use the File menu to change your working directory to the same directory where you just saved this data file [In R, select File, then Change dir...]. Then, read the data file into R and “attach” it for use as follows:

```
baboons.df = read.table("baboon.txt", header=T)
attach(baboons.df)
names(baboons.df)
```

Each row of the data frame is a record of the interaction attempts for one infant-handler combination. Attempts by mothers to handle their own infant are not recorded. There are 23 mothers, 11 infants, and thus, $(23 \times 11) - 11 = 242$ rows in the data frame. Here is information about a few of the columns (variables) recorded:

- `Infant` = infant id number
- `MotherRnk` = dominance rank of the infant's mother (1, 2, or 3)
- `HandlRnk` = handler rank (1, 2, or 3)
- `Hand.Name` = handler name (a two-letter code)
- `Hand.ID` = handler id number (essentially a numeric version of Hand.Name)

- `Pass.Act` = the total number of interaction attempts by the handler

Question 4: Re-create Table 1 using R.

The following code creates a table that places `Infant` in rows and `Hand.ID` in columns and then sums `Pass.Act`, the total number of interactions for each combination:

```
tapply(Pass.Act, list(Infant, Hand.ID), sum)
```

11 entries read `NA` = "missing value" since mothers are not considered a "handler" of their own infant. We make note of which of the handlers are mothers (i.e., which columns of Table 1 correspond to the mothers) and we will replace all `NA`'s with zeros as in Table 1:

```
mothers = c(1, 5, 6, 7, 8, 13, 15, 18, 17, 19, 21)
mothers
Pass.Act.Table1 = tapply(Pass.Act, list(Infant, Hand.ID), sum)
Pass.Act.Table1[is.na(Pass.Act.Table1)] = 0
```

Finally, Table 1 re-created in R:

```
Pass.Act.Table1
```

Question 5: Recreate Table 2A using R.

To create the 3-by-3 table of infant-handler interaction counts by mother/infant rank and handler rank. Here is Table 2A the "easy" way in R:

```
tapply(Pass.Act, list(MotherRnk, HandlRnk), sum)
```

Later, we will need a slightly more complex function to create a 3-by-3 table from a generalized Table 1, so we develop the tools to do that right now. First, make an ordered list of the handler (column) and infant (row) dominance ratings to coincide with Table 1:

```
col.ranks = c(rep(1,4), rep(2,7), rep(3,12))
col.ranks
row.ranks = col.ranks[mothers]
row.ranks
```

Next, collapse Table 1 into Table 2A by summing across rows (infants) with the same dominance level and then summing across columns (handlers) with the same dominance level. Here are the sums across rows:

```
rowsum(Pass.Act.Table1, row.ranks)
```

While R has a built in function to sum rows, `rowsum`, we will need to write a function to sum across columns. Copy the following code in R to create the new function `colsum()`.

```
colsum = function(x, group) { t(rowsum(t(x), group)) }
```

Although `colsum` may look complicated, this function really just uses the `t()` function for matrix transposition (swapping rows and columns) twice so that it can borrow the existing `rowsum` function to do all of the hard work. The following code will sum across rows and then sum across columns:

```
colsum(rowsum(Pass.Act.Table1, row.ranks), col.ranks)
```

Finally, Save this matrix "compression" operation in an R function by copying and pasting in the following R code:

```
compress = function(Table1, c.ranks, m) {  
  colsum(rowsum(Table1, c.ranks[m]), c.ranks)  
}
```

Now we can use the `compress` function to collapse any Table 1 to into a smaller table, (i.e. Table 2).

```
Pass.Act.Table2 = compress(Table1=Pass.Act.Table1, c.ranks=col.ranks,  
m=mothers)  
Pass.Act.Table2
```

Step 4) Statistical Significance and Null Models

We have now seen a pattern that to some extent supports the research hypothesis. The pattern is not perfect and the strength of support is not overwhelming and so we are in the perfectly balanced situation where researchers usually ask for some confirmation through a statistical significance test. In all situations significance tests help answer the question: could the observed pattern be plausibly attributed to random chance alone. Recall from our introductory module on randomization and permutation tests that in the context of an observational study, statistical significance tests test an "as if by random chance alone" type of null hypothesis, but rejecting the null cannot lead to a firm causal interpretation. Still, the rejection of the null in this context is considered by many scientists to be a useful part of the analysis and a worthwhile next step beyond an observed pattern from the descriptive analysis. Gotelli and Graves [2] call such null hypotheses "null models."

More specifically, Gotelli and Graves define their concept this way: "[A] null model is a pattern-generating model that is based on randomization of ecological data or random sampling from a known or imagined distribution. The null model is designed with respect to some ecological or evolutionary process of interest. Certain elements of the data are held constant, and others allowed to vary stochastically to create assemblage patterns. The randomization is designed to produce a pattern that would be expected in the absence of a particular ecological mechanism." ([2] pp. 3-4)

In practice, a null model is a process of generating data sets that would be equally-likely under the null hypothesis. Gotelli and Graves (p. 6) motivate the scientist's interest in null models, stating that if the data are consistent with a properly constructed null model we can infer that the biological mechanism is not operating, but if the data are inconsistent with the null model, "... this provides some positive evidence in favor of the mechanism."

We can use RH1 to define a null model and use a significance test and the data to compare the null model to RH1. This null model test outlined in Gotelli and Graves's definition above will lead to what we have called heretofore a permutation test. This test will show just how flexible the concept of permutation can be, given the non-standard structure of our data described below.

Before proceeding, we must say a word about the use of a classical significance test for two-by-two tables, namely the *chi-squared test of independence*. The chi-squared test presumes a certain sampling model that is impossible to justify in this case. The chi-squared test would require that the observed 678 data entered the table one at a time. For example, consider the 97 high-on-mid interactions. The chi-squared test assumes that if Handler Rank was independent of Infant Rank, then these 97 interactions could just as likely been distributed into other cells of the two-by-two table. But this ignores the "clustering" of the counts by individual baboons. For example, part of those 97 interactions are the 13 interactions of KM on HZ/HQ, interactions that could, under the null hypothesis, ended up in other cells of the two-by-two table, but they would have entered those other cells as a cluster: all 13 in the cell or not. In a sense, we have two levels of observation: the handler/infant pairs and the individual instances of infant handling. While the expected values (and corresponding residuals in Table 2D) are fine for informal descriptive analysis, the classical chi-squared test is problematic because the clustering of observations from handler/infant pairs occurring in each cell violates key assumptions of the sampling model.

The beauty of the permutation test is that it is adaptable enough to respect the clustering of the counts by handler-infant pairs.

Question 6: Perform a chi-squared test of independence on Table 2A. State the p-value. (R will give a pretty precise p-value.) Here is some help with the `chisq.test` function in R:

```
chitest = chisq.test(Pass.Act.Table2)
names(chitest)
chitest$observed      # This is Table 2A (observed counts)
chitest$expected      # This is Table 2C (expected counts)
chitest$residuals     # This is Table 2D (standardized residuals)
chitest$p.value
```

For our permutation test, the null hypothesis and corresponding null model are as follows:

H_0 : Handler rank and infant rank are independent

Null Model: For the data in Table 1, the dominance ranks can be viewed as meaningless labels attached at random to handlers and infants. Thus data sets produced by permuting these ranks in all possible ways that respect infant-mother pairs are equally likely.

In the null model, the elements to be held constant are the counts. The elements that are allowed to vary stochastically (randomly) to create "assemblage patterns" are the female and infant ranks. The assemblage patterns generated become new, hypothetical data sets from which we can construct the sampling distribution of a permutation test statistic. To complete the picture, we need some test statistic that reflects the level of agreement between the data

and the null model. For now, denote the test statistic by the letter C. Here then is the formal process that becomes our test:

Steps for randomization test:

1. Assign ranks at random to infants and handlers using the rank distributions of the data set. That is, assign ranks at random so that the 11 infants are assigned, in this case, 1 High, 4 Mid, and 6 Low and so that the 23 handlers are assigned 4 High's, 7 Mid's, and 12 Low's (with the restriction that the 11 handlers that are the mothers must be assigned to the same rank as their infants) . This assignment leads to the original data table but with permuted ranks.
2. Re-form the 3-by-3 table of interaction counts by handler and mother rank (as in Table 2A).
3. Compute the value of C for this table.

This 3-step process defines a sampling distribution for C under H₀. The p-value of a particular data set is defined to be P(C ≥ C_D) where C_D represents the value of the test statistic observed in the original data.

To compute the sampling distribution exactly would require a complete enumeration of all possible permutations. There are 42,688,800 such permutations, a prohibitively large number to enumerate, so we instead approximate the sampling distribution by randomly generating permutations of the row and column ranks according to the 3-step process outlined above. *We can then determine an empirical p-value as the proportion of permutations that result in C ≥ C_D.*

For what follows, we will use a test statistic that reflects RH1. We denote this test statistic by **LTE** (“Less Than or Equal to”); so replace C with LTE in the above 3-step process. Here is the definition of LTE:

$$\begin{array}{ccccccc}
 1 & -1 & -1 & & a & b & c \\
 1 & 1 & -1 & * & d & e & f \\
 1 & 1 & 1 & & g & h & i
 \end{array} = a + d + g - b + e + h - c - f + i$$

The * symbol is meant to represent the familiar dot product from vector algebra. For the 3-by-3 table in Table 2A the calculation of LTE is thus:

$$\begin{array}{ccccccc}
 1 & -1 & -1 & & 5 & 5 & 3 \\
 1 & 1 & -1 & * & 97 & 83 & 95 \\
 1 & 1 & 1 & & 68 & 138 & 184
 \end{array} = 5+97+68-5+83+138-3-95+184 = 472$$

Notice that the LTE statistic adds counts that support RH1 and subtracts counts that run counter to RH1. Hence large LTE values lend evidence for RH1, while small values lend evidence against RH1. The question is, is 472 a large value? Is it large enough to be considered “statistically significant?”

The permutation test (steps 1-3 above), permutes the infant-mother ranks at random many times, re-computing the 3-by-3 table and the LTE statistic many times, to ascertain the answer to this question.

Question 7: Calculate the LTE statistic for the observed data using the following R code . This code multiplies the observed Table 2 by the appropriate “1”s and “-1”s..

```
Pass.Act.Table2
as.vector(Pass.Act.Table2)
LTEvector = c(1,1,1,-1,1,1,-1,-1,1)
LTEvector
sum(LTEvector * as.vector(Pass.Act.Table2))
```

Again, save this code into and R:

```
teststat = function(statvector, Table2vector) {
sum(statvector * Table2vector)
}
```

Note that this function is general enough to work for a variety of test statistics. Here we specify that we are interested in the LTE statistic by using LTEvector.
teststat(statvector=LTEvector, Table2vector=as.vector(Pass.Act.Table2))

Question 8: Table 3 gives the original data set, but with one random re-arrangement (permutation) of the infant and handler ranks. Answer the questions that follow.

- The random permutation should not change the number of elements in each group (i.e. the frequency in each rank needs to stay the same). Count the frequency of ranks 1, 2, and 3 for the Infants. Is the distribution the same as in the original data set?
- In similar fashion, is the distribution of ranks for the handlers the same as in the original data set?
- Now, verify that any handler who is also a mother has the same rank as her infant. (This makes this permutation a valid permutation of ranks as described in the first step of the randomization test above.)
- Complete Step 2 for this table by hand. That is re-form the table into a 3-by-3 table.
- Finally, compute the LTE statistic for this table by hand
- Repeat d. and e using R code.

```
ncol.ranks = c(1,3,1,3,3,3,2,2,3,1,3,2,2,3,3,1,3,3,3,2,2,3,2)
nrow.ranks = col.ranks[mothers]
Table2 = compress(Table1=Pass.Act.Table1, c.ranks=ncol.ranks, m=mothers)
Table2
teststat(statvector=LTEvector, Table2vector=as.vector(Table2))
```

Infants/ Mothers Names & Mother's Ranks	Handler's Names and Ranks																						
	KM	KN	NQ	PO	HQ	LL	NY	PS	SK	ST	WK	AL	CO	DD	LS	LY	MH	ML	MM	PA	PH	PT	RS
	1	3	1	3	3	3	2	2	3	1	3	2	2	3	3	1	3	3	3	2	2	3	2
KG/KM 1	0	0	4	1	1	0	0	0	3	1	0	0	0	0	0	0	0	0	0	0	0	2	1
HZ/HQ 3	13	23	7	5	0	2	1	1	5	6	18	1	6	3	0	1	4	1	0	9	0	10	1
LC/LL 3	4	0	1	4	3	0	2	1	1	5	3	1	0	0	1	0	2	1	1	1	0	1	6
NK/NY 2	12	4	10	5	9	1	0	2	3	11	7	8	6	3	1	0	2	1	1	5	3	2	3
PZ/PS 2	1	3	4	1	0	0	0	0	0	0	2	0	2	0	0	0	3	0	1	1	0	3	0
CY/CO 2	2	2	7	3	1	1	2	0	3	12	16	3	0	2	0	0	2	0	0	1	0	0	2
LZ/LS 3	1	0	3	2	1	1	0	0	2	0	5	2	2	2	0	1	9	2	0	0	0	3	2
MQ/ML 3	0	1	5	2	2	4	2	2	2	4	5	7	5	2	1	1	7	0	4	4	1	0	2
MW/MH 3	3	0	7	4	2	3	0	5	2	8	13	7	14	2	0	0	0	4	0	8	0	13	6
MX/MM 3	2	3	4	5	0	0	0	0	0	5	2	9	3	1	0	0	2	0	0	1	2	2	3
PK/PH 2	2	0	6	4	3	4	1	0	0	15	10	8	5	1	0	3	1	1	6	3	0	7	5

Table 3: This is the original data table but with a random permutation of the infant/mother ranks and the handler ranks.

Question 8: Use the following code in R to set up new (randomly permuted) ranks for the mothers and then for the remaining handlers. The results will be stored as the column ranks and row ranks as before.

```
pcol.ranks = col.ranks # Set up new variable to hold permuted column ranks
pcol.ranks[mothers] = sample(col.ranks[mothers]) # Permute mother's ranks
pcol.ranks[-mothers] = sample(col.ranks[-mothers]) # Permute handler's
(non-mothers)
```

This separate permutation for mothers and non-mothers (-mothers) is needed to maintain the distribution of mothers: one mother of High rank, four mothers of Mid rank and six mothers of Low rank. Once a mother's new rank is determined, it's infant also needs the same new rank. Thus rows are determined by the permutation of mothers.

```
pro.w.ranks = pcol.ranks[mothers] # Copy mother's new ranks to infants
```

Compare the original ranks with the permuted ranks to verify that any handler who is also a mother has the same rank as her infant. (This makes this permutation a valid permutation of ranks as described in the first step of the randomization test above.)

```
col.ranks
pcol.ranks
row.ranks
pro.w.ranks
pcol.ranks[mothers]
```

Again, copy and paste the following code into R to create and save a function that we can use to repeat this permutation process over and over again later:

```
permute = function(c.ranks, m) {
```

```

pcol.ranks = c.ranks
pcol.ranks[m] = sample(c.ranks[m])
pcol.ranks[-m] = sample(c.ranks[-m])
return(pcol.ranks)
}
permute(c.ranks=col.ranks, m=mothers)

```

We now combine all our previous programming into one function:

```

perm.test = function(Table1, c.ranks, m, statvector, numperm) {
  # create a matrix with numperm rows, each containing the obsvd col.ranks
  col.rank.matrix = matrix(c.ranks, ncol=length(c.ranks), nrow=numperm,
byrow=T)
  # permute each row of col.rank.matrix to produce perm.matrix
  # with each row a valid permutation of the column (handler) ranks
  perm.matrix = t(apply(col.rank.matrix, 1, permute, m=m))
  # calculate a Table 2 for each permutation of ranks within perm.matrix
  # each Table 2 will be in one column
  perm.table2 = apply(perm.matrix, 1, compress, Table1=Table1, m=m)
  # Finish by returning the list of numperm test statistics
  return(apply(perm.table2, 2, teststat, statvector=statvector))
}

```

The following R code will compute the empirical p-value, based on 1000 iterations of perm.test:

```

Pass.Act.LTE.perm.distrn = perm.test(Table1=Pass.Act.Table1,
  c.ranks=col.ranks, m=mothers, statvector=LTEvector, numperm=1000)

Pass.Act.LTE.p.value = sum(Pass.Act.LTE.perm.distrn
>=teststat(statvector=LTEvector,Table2vector=as.vector(Pass.Act.Table2)))/
1000
Pass.Act.LTE.p.value

```

You can also view the distribution of the statistic LTE:

```

hist(Pass.Act.LTE.perm.distrn)
abline(v=472)

```

The precise results will vary with each run, because random samples are being generated by R, but here is an explanation of this code:

Pass.Act.LTE.perm.distrn contains a vector of length 1000, giving the test statistic for each of 1000 iterations of the permutation test steps outlined above. We next asked R for a histogram of this permutation distribution and added a vertical line (with `abline(v=472)`) at the observed LTE statistic value of 472. We notice that only a few of the values of LTE statistic resulting from the permutations are at or beyond 472, so we know that the empirical p-value will be small. The estimated p-value is `Pass.Act.LTE.p.value`. In fact, the author observed an empirical p-value of 0.015, meaning that 15 of the 1000 iterations had LTE statistics exceeded or were equal to the observed LTE statistic value of 472. The small p-

value gives fairly strong evidence that evidence of this magnitude would be hard to explain by chance alone.

Question 9: Does the chi-square test, which we argue above as inappropriate for these data, suggest greater, lesser, or about the same level of significance as the permutation test? Try to reason why this would make sense in light of the way in which the data are structured.

Step 5) A New Research Hypothesis

Within the period of analyzing the data, the authors [1] recognized the possibility of a second research hypothesis, which in the paper is described as “immediately lower”, rather than “lower,” the latter being our RH1. Specifically, we define Research Hypothesis 2 this way:

Research Hypothesis 2 (RH2): Females will tend to handle the infants of females who are ranked immediately below them in the 3-tiered ranking system (or ranked the same, for a 3-ranked handler).

The rationale for RH2 is given in [1].

Question 10: Below we define a new test statistic, LT. Explain why this test statistic “makes sense” when the research hypothesis of interest is RH2. Compute the observed value of LT for Table 1.

$$\begin{array}{ccccccc} -1 & -1 & -1 & & a & b & c \\ 1 & -1 & -1 & * & d & e & f \\ 1 & 1 & 1 & & g & h & i \end{array} = -a+d+g-b-e+h-c-f+i$$

Question 11: Perform a permutation test with LT and state your conclusions.

First recall the `LTEvector`: `LTEvector = c(1,1,1,-1,1,1,-1,-1,1)` corresponding to RH1. Create a new vector, `LTvector` corresponding to RH2.

Second run 1000 iterations of `Pass.ACT.LT.perm.distn` by using `statvector = LTvector`.

Finally, determine the p-value by creating a new function: `Pass.Act.LT.p.value`

Step 6) OPTIONAL Categories of infant-handler interaction

The Bentley-Condit article [1] suggests more nuanced analyses than we have heretofore considered in this lab. First, TABLE 1 of [1] categorizes interactions into three different types:

- (1) *Passive*: movement to within 1m. of the mother-infant pair with no attempt to handle, (the paper calls these interactions Approach);
- (2) *Unsuccessful*: movement to within 1m. of the mother-infant pair with an attempted (but not successful) handle, (the paper calls these Attempted); or
- (3) *Successful*: a successful handle.

Our Table 1 above is a sum total of these three types of interactions for each handler-infant pair. For example, of the 13 interactions of handler KM with infant HZ/HQ, 2 were passive, 4 were unsuccessful, and 7 were successful. In the baboons.df the variables Passive, Unsuccess, and Successful refer to these categories of interaction and the variable Pass.Act is the sum of these three.

We can create three tables, one for each category of interaction.

```
inc.mat.Pass <- tapply(Passive, list(Infant, Hand.ID), sum) #create Passive table
inc.mat.Pass[is.na(inc.mat.Pass)] <- 0 #The first matrix contains some NA's #replace NA's

inc.mat.Un <- tapply(Unsuccess, list(Infant, Hand.ID), sum) #create Table 1
inc.mat.Un[is.na(inc.mat.Un)] <- 0 #The first matrix contains some NA's #replace NA's

inc.mat.Succ <- tapply(Successful, list(Infant, Hand.ID), sum) #create Table 1
inc.mat.Succ[is.na(inc.mat.Succ)] <- 0 #The first matrix contains some NA's #replace NA's
```

Analysis 1: The researchers were more interested in the research hypotheses RH1 and RH2 with respect to these categories than they were with the overall table of counts given in Table 1. Consider both RH1 and RH2 for the 3 categories of interactions. Run both descriptive analyses and permutation test analyses for the various combinations of research hypothesis and category of interaction. The goal is to learn about the relationship, if any, between dominance rank and infant-handling behavior and how level of interaction, e.g., whether the interaction is “successful” or not.

Analysis 2: In [1], the authors ended up dropping infant 1, the only high-ranking infant in the data set. The strength of conclusions was compromised by this decision. After the publication of [1], the authors took a suggestion of Lunneborg [3] and performed a sensitivity analysis, which had the virtue of looking more carefully at the role of infant 1 and all individual infants. There are actually several ways to go about such a sensitivity analysis, but one we suggest here is as follows. Consider each infant, one at a time. Set aside that infant from the data set; that infant’s mother remains in the data set, but as a handler only, not a mother. Now, re-compute the p-value for the data set. Consider the set of 11 p-values thus obtained and consider whether it seems as if any one infant has great affect on the original results. For example, with infant 1 removed is the p-value much different from the original p-value or from the p-values obtained with the other 10 one-removed p-values? This determination is subjective, but it does provide a way to ascertain the affect of individuals without just throwing away that infant from the analysis.

The R code in [in the appendix](#) performs this analysis. The vector `p.rem.inf.PA.LTE` contains the 11 p-values, removing one infant at a time. We see that infant #1 does have a decidedly larger p-value:

Deleted: Step 4 of the R code document

```
[1] 0.120 0.014 0.059 0.012 0.061 0.024 0.017 0.021 0.037 0.015 0.035
```

Lunneborg actually suggests looking not at p-values with infants removed but at the test statistic itself. The code does this as well, putting test statistic values for the LTE test statistic in the vector `LTE.rem.inf.PA`:

```
0.7142857 0.7611408 0.7218750 0.7651123 0.7168950 0.6672052 0.6781250  
0.6650407 0.6429809 0.6750789 0.6526138
```

Note: These are not actually LTE values, but rather they are the LTE value divided by the sum of all entries in the 3-by-3 table. The reason for normalizing this way, is that when infants are removed, the total count of the table changes, which would make comparisons across infants impossible. Since the LTE statistic measures the total count in cells consistent with the research hypothesis minus total count in cells inconsistent with the research hypothesis, one should interpret this normalized LTE statistic as a difference in the proportion of consistent events minus the proportion of inconsistent events.

Note, that on this descriptive level, the removal of infant 1 seems to exert less influence on the results. The value for infant one removed (.7142857) is not extreme within the set of 11 values.

Now, go on to investigate the removal of infants with the LT statistic and also with different categories of interaction. Summarize your findings.

Cited Sources

- [1] Bentley-Condit, V.K., Moore, T.L., and Smith, E.O. (2001), "Infant Handling by Tana River Adult Female Yellow Baboons", (*Papio Cynocephalus Cynocephalu*)," *American Journal of Primatology*, 55, 117--130.
- [2] Gotelli, N.J. and Graves, G. R. (1996), *Null Models in Ecology*. Washington, DC: Smithsonian Institution Press.
- [3] Lunneborg, C. E. (2002), "Infant Handling by Female Baboons: A Sensitivity Analysis," *Stats*, 33, 13--15.