

Analysis of Variance

While t-tests allow comparisons of pairs of conditions, it is often more useful to compare all conditions in a single test. I'll talk about Analysis of Variance (ANOVA) in a later lecture. For now, here is the R code and output to run this on the ToothGrowth dataset.

```
> attach(ToothGrowth)
```

```
> oneway.test(len~supp)
```

One-way analysis of means (not assuming equal variances)

data: len and supp

F = 3.6683, num df = 1.000, denom df = 55.309, p-value = 0.06063

You will notice that this only compares supplement, and is thus a pairwise test (and produces the same result as the t-test). What we need to do is include both independent variables.

```
> res.aov2 <- aov(len~supp+dose, data=ToothGrowth)
```

```
> summary(res.aov2)
```

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
supp	1	205.4	205.4	11.45	0.0013	**
dose	1	2224.3	2224.3	123.99	6.31e-16	***
Residuals	57	1022.6	17.9			

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

This tells us that we have a significant main effect for 'supp' [$F(1,57) = 11.45$, $p < 0.005$] and for 'dose' [$F(1,57) = 123.99$, $p < 0.0001$]. What it doesn't tell us is whether there is a relationship between supplement and dose (as our boxplot suggests). For this, we need to add a further variable which describes the interaction between supplement and dose.

```
> res.aov3 <- aov(len~supp+dose+supp:dose, data=ToothGrowth)
```

```
> summary(res.aov3)
```

		Df	Sum Sq	Mean Sq	F value	Pr(>F)
supp	1	205.4	205.4	12.317	0.000894	***
dose	1	2224.3	2224.3	133.415	< 2e-16	***
supp:dose	1	88.9	88.9	5.333	0.024631	*
Residuals	56	933.6	16.7			

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Dealing with Non-Normal Data and Repeated Measures ANOVA

ii. Repeated Measures ANOVA

Conducting a Repeated Measures ANOVA in R is not the same as the other ANOVA tests we've looked at (and there are several approaches that can be used – in this lab we'll look at two of these). Whereas ANOVA assumes that data are drawn from independent populations (i.e., three separate groups of participants), in Repeated Measures ANOVA we use data from the same population (i.e., the same participants in three different conditions).

The assumptions for Repeated Measures ANOVA are:

1. Each participant has a value in each condition (if you are missing a value from a participant in one condition, you need to exclude that participant's data from all conditions);
2. Normality of distribution (check with Shapiro Wilke)
3. Sphericity (i.e., equality of variance across conditions)
4. Data for each condition should not contain outliers (extremes should be checked for and, if necessary, the data winsorized - and refer back to 1. If this means that too many participants are removed, then you should consider a non-parametric alternative such as a Friedman test).

Import the DronesMetrics dataset.

Create a data frame for the beacon data.

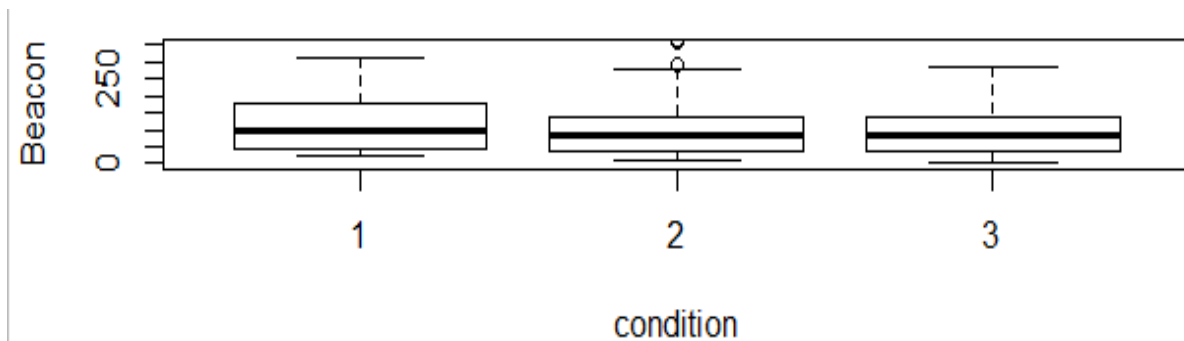
```
> Beacon<-data.frame(auto_beacon,both_beacon,self_beacon)
> Beacon
  auto_beacon both_beacon self_beacon
1         40         66         78
2        122        108         87
3         41         52         88
4         56         44         42
5         40         36         36
6         36         24         32
7         50         20         12
```

The problem is that our data are in a 'wide format' (three columns that represent each independent variable). This implies that the data in each column is produced by a separate activity. We need to convert our data into a 'long format' (so that we can label data by participant). There are several ways to do this, but that does not involve calling a new library is the command 'reshape'.

```
> LongBeacon<-reshape(Beacon,varying =
list(c("auto_beacon","both_beacon","self_beacon")),direction="long",timevar="condition",idvar="part",
v.names="Beacon")
> LongBeacon
  condition Beacon part
1.1      1    40    1
2.1      1   122    2
3.1      1    41    3
4.1      1    56    4
5.1      1    40    5
6.1      1    36    6
7.1      1    50    7
8.1      1   180    8
9.1      1    63    9
10.1     1   120   10
11.1     1    79   11
```

This defines the Dependent Variable, 'Beacon', that varies three levels of Independent Variable, condition.

```
> boxplot(Beacon~condition, data=LongBeacon, ylim=c(0,350))
```



Analysis

We can apply the one-way test (as we are looking at a single independent variable, Beacon, with three levels) using the AOV function from previously.

```
> analysis<-aov(Beacon~condition+Error(part/condition), data=LongBeacon)
```

Notice that we have defined the ANOVA model differently; here, we have indicated an 'error' term (which assumes that the results could be confounding by within subject factors, i.e., individual differences). The result of the ANOVA is shown on the bottom lines: $F(1,200) = 0.892$, $p=0.346$ (so, not a significant main effect).

```
> summary(analysis)
```

```
Error: part
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  1 234881   234881

Error: part:condition
      Df Sum Sq Mean Sq
condition  1  14493   14493

Error: within
      Df Sum Sq Mean Sq F value Pr(>F)
condition  1  14453   14453   0.892  0.346
Residuals 200 3241262   16206
```

(iii.) Non-Parametric ANOVA

If your data fail the Shapiro-Wilk test (and this means if any single condition produces a value >0.05) then you should apply a non-parametric test (or transform the data so then the distribution is normal). Let's assume that our original TPT data could not be transformed to normality. The Friedman test supports Repeated Measures ANOVA and is called with the command:

```
> friedman.test(Beacon~part|condition, data=LongBeacon)
```

Friedman rank sum test

data: Beacon and part and condition

Friedman chi-squared = 91.559, df = 67, p-value = 0.02484

We can test for pairwise comparisons using a non-parametric test:

```
> wilcox.test(auto_beacon,both_beacon, paired=TRUE, data=Beacon)
```

wilcoxon signed rank test with continuity correction

```
data: auto_beacon and both_beacon
v = 1435.5, p-value = 0.1094
alternative hypothesis: true location shift is not equal to 0
```

MIXED ANOVA

Taken from <https://cran.r-project.org/doc/contrib/Baron-rpsych.pdf>

Suppose you have 1 between and 2 within Independent Variables.

Test of drug treatment effect by one between-subject: group (2 x 8Pp) and two within-subject factors (drug (2) and dose(3)).

We can either enter the data into R by hand and then construct the data frame for our analysis (which is what the next steps illustrate) or we can copy the data in the Ela.uni table (a couple of pages on from here), paste this in a spreadsheet, save it in tab-separated format and then import the dataset to R (which is much easier to do).

```
> Ela.mat <-matrix(c(
+ 19,22,28,16,26,22,
+ 11,19,30,12,18,28,
+ 20,24,24,24,22,29,
+ 21,25,25,15,10,26,
+ 18,24,29,19,26,28,
+ 17,23,28,15,23,22,
+ 20,23,23,26,21,28,
+ 14,20,29,25,29,29,
+ 16,20,24,30,34,36,
+ 26,26,26,24,30,32,
+ 22,27,23,33,36,45,
+ 16,18,29,27,26,34,
+ 19,21,20,22,22,21,
+ 20,25,25,29,29,33,
+ 21,22,23,27,26,35,
+ 17,20,22,23,26,28), nrow = 16, byrow = T)
```

Create multivariate format

```
> Ela.mul <- cbind.data.frame(subj=1:16, gp=factor(rep(1:2,rep(8,2))), Ela.mat)

> dimnames(Ela.mul)[[2]] <-c("subj","gp","d11","d12","d13","d21","d22","d23")

> Ela.mul
  subj gp d11 d12 d13 d21 d22 d23
1    1  1  19  22  28  16  26  22
2    2  1  11  19  30  12  18  28
3    3  1  20  24  24  24  22  29
4    4  1  21  25  25  15  10  26
5    5  1  18  24  29  19  26  28
6    6  1  17  23  28  15  23  22
7    7  1  20  23  23  26  21  28
```

```

8  8 1 14 20 29 25 29 29
9  9 2 16 20 24 30 34 36
10 10 2 26 26 26 24 30 32
11 11 2 22 27 23 33 36 45
12 12 2 16 18 29 27 26 34
13 13 2 19 21 20 22 22 21
14 14 2 20 25 25 29 29 33
15 15 2 21 22 23 27 26 35
16 16 2 17 20 22 23 26 28

```

```

> Ela.uni <- data.frame(effect = as.vector(Ela.mat),subj = factor(paste("s", rep(1:16, 6), sep="")),gp =
factor(paste("gp", rep(rep(c(1, 2), c(8,8)), 6), sep="")),drug = factor(paste("dr", rep(c(1, 2), c(48, 48)),
sep="")),dose=factor(paste("do", rep(rep(c(1,2,3), rep(16, 3)), 2), sep="")),row.names = NULL)

```

```

> Ela.uni
  effect subj gp drug dose
1    19  s1 gp1 dr1 do1
2    11  s2 gp1 dr1 do1
3    20  s3 gp1 dr1 do1
4    21  s4 gp1 dr1 do1
5    18  s5 gp1 dr1 do1
6    17  s6 gp1 dr1 do1
7    20  s7 gp1 dr1 do1
8    14  s8 gp1 dr1 do1
9    16  s9 gp2 dr1 do1
10   26 s10 gp2 dr1 do1
11   22 s11 gp2 dr1 do1
12   16 s12 gp2 dr1 do1
13   19 s13 gp2 dr1 do1
14   20 s14 gp2 dr1 do1
15   21 s15 gp2 dr1 do1
16   17 s16 gp2 dr1 do1
17   22  s1 gp1 dr1 do2
18   19  s2 gp1 dr1 do2
19   24  s3 gp1 dr1 do2
20   25  s4 gp1 dr1 do2
21   24  s5 gp1 dr1 do2
22   23  s6 gp1 dr1 do2
23   23  s7 gp1 dr1 do2
24   20  s8 gp1 dr1 do2
25   20  s9 gp2 dr1 do2
26   26 s10 gp2 dr1 do2
27   27 s11 gp2 dr1 do2
28   18 s12 gp2 dr1 do2
29   21 s13 gp2 dr1 do2
30   25 s14 gp2 dr1 do2
31   22 s15 gp2 dr1 do2
32   20 s16 gp2 dr1 do2
33   28  s1 gp1 dr1 do3
34   30  s2 gp1 dr1 do3
35   24  s3 gp1 dr1 do3
36   25  s4 gp1 dr1 do3
37   29  s5 gp1 dr1 do3
38   28  s6 gp1 dr1 do3
39   23  s7 gp1 dr1 do3
40   29  s8 gp1 dr1 do3
41   24  s9 gp2 dr1 do3
42   26 s10 gp2 dr1 do3
43   23 s11 gp2 dr1 do3
44   29 s12 gp2 dr1 do3

```

```

45 20 s13 gp2 dr1 do3
46 25 s14 gp2 dr1 do3
47 23 s15 gp2 dr1 do3
48 22 s16 gp2 dr1 do3
49 16 s1 gp1 dr2 do1
50 12 s2 gp1 dr2 do1
51 24 s3 gp1 dr2 do1
52 15 s4 gp1 dr2 do1
53 19 s5 gp1 dr2 do1
54 15 s6 gp1 dr2 do1
55 26 s7 gp1 dr2 do1
56 25 s8 gp1 dr2 do1
57 30 s9 gp2 dr2 do1
58 24 s10 gp2 dr2 do1
59 33 s11 gp2 dr2 do1
60 27 s12 gp2 dr2 do1
61 22 s13 gp2 dr2 do1
62 29 s14 gp2 dr2 do1
63 27 s15 gp2 dr2 do1
64 23 s16 gp2 dr2 do1
65 26 s1 gp1 dr2 do2
66 18 s2 gp1 dr2 do2
67 22 s3 gp1 dr2 do2
68 10 s4 gp1 dr2 do2
69 26 s5 gp1 dr2 do2
70 23 s6 gp1 dr2 do2
71 21 s7 gp1 dr2 do2
72 29 s8 gp1 dr2 do2
73 34 s9 gp2 dr2 do2
74 30 s10 gp2 dr2 do2
75 36 s11 gp2 dr2 do2
76 26 s12 gp2 dr2 do2
77 22 s13 gp2 dr2 do2
78 29 s14 gp2 dr2 do2
79 26 s15 gp2 dr2 do2
80 26 s16 gp2 dr2 do2
81 22 s1 gp1 dr2 do3
82 28 s2 gp1 dr2 do3
83 29 s3 gp1 dr2 do3
84 26 s4 gp1 dr2 do3
85 28 s5 gp1 dr2 do3
86 22 s6 gp1 dr2 do3
87 28 s7 gp1 dr2 do3
88 29 s8 gp1 dr2 do3
89 36 s9 gp2 dr2 do3
90 32 s10 gp2 dr2 do3
91 45 s11 gp2 dr2 do3
92 34 s12 gp2 dr2 do3
93 21 s13 gp2 dr2 do3
94 33 s14 gp2 dr2 do3
95 35 s15 gp2 dr2 do3
96 28 s16 gp2 dr2 do3

```

```

> analysis<-aov(effect ~ gp * drug * dose + Error(subj/(dose+drug)), data=Ela.uni)
> summary(analysis)

```

```

Error: subj
      Df Sum Sq Mean Sq F value Pr(>F)
gp      1   270   270.01   7.092 0.0185 *
Residuals 14   533   38.07

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: subj:dose
Df Sum Sq Mean Sq F value Pr(>F)
dose 2 758.8 379.4 36.510 1.58e-08 ***
gp:dose 2 42.3 21.1 2.034 0.15
Residuals 28 291.0 10.4

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: subj:drug
Df Sum Sq Mean Sq F value Pr(>F)
drug 1 348.8 348.8 13.00 0.00287 **
gp:drug 1 326.3 326.3 12.16 0.00362 **
Residuals 14 375.6 26.8

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: Within
Df Sum Sq Mean Sq F value Pr(>F)
drug:dose 2 12.06 6.031 0.682 0.514
gp:drug:dose 2 14.81 7.406 0.837 0.444
Residuals 28 247.79 8.850