

Psychology 205: Research Methods in Psychology

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA



NORTHWESTERN
UNIVERSITY

September, 2015

Outline

Preliminaries

Data = Model + Residual (error)

1. Data = Model + Residual
2. Observed data may be represented by a model of the data.
What is left over is residual (or error)
3. The process of research is to reduce the residual
4. We do this by a progression of models, ranging from the very simple to the complex
5. We want to know how each model fits the data

Consider the Recall and Recognition data

1. How to describe it
 - Raw data
 - Summary statistics
 - Graphically
2. All tables and graphs are prepared by using the R computer package. For details on using R, consult the tutorials, particularly the short tutorial, listed in the syllabus
 - First, install R from <http://r-project.org> (just do this once)
 - Then, install the *psych* (just do this once)
 - `install.packages("psych")`
 - `library(psych) # everytime you start R`

The Very RAW data as entered into Excel – but just showing some of it

This is clearly not very useful. We need to think of ways to organize it.

Condition L1P1 L1P2 L1P3 L1P4 L1P5 L1P6 L1P7 L1P8 L1P9 L1P10 L1P11 L1P12 L1P13 L1P14 L1P15 L1Tot
 L2P1 L2P2 L2P3 L2P4 L2P5 L2P6 L2P7 L2P8 L2P9 L2P10 L2P11 L2P12 L2P13 L2P14 L2P15 L2Tot L3P1 L3P2
 L3P3 L3P4 L3P5 L3P6 L3P7 L3P8 L3P9 L3P10 L3P11 L3P12 L3P13 L3P14 L3P15 L3Tot

```

2 1 0 1 1 1 0 0 0 0 0 0 0 0 1 1 6 1 1 1 1 1 0 0 0 1 0 0 0 0 0 1 7 1 1 0 1 1 1 0 0 0 1 0 1 0 0 1 0 7 1 1 1 0 1 0 1 1 1 0
1 0 0 0 1 1 1 9 2 1 1 1 1 1 1 0 0 1 1 0 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 0 1 1 2 1 1 1 1 1 1 1 1 1 0 0 1 0 1 1 1
1 2 2 1 1 0 0 1 1 1 0 1 0 1 1 0 1 1 1 0 1 0 1 1 0 1 1 1 1 0 0 0 0 1 1 9 2 1 0 1 0 1 1 1 1 1 0 1 1 1 1 1 1 1 2 1 1 0 1 0 1
1 0 1 0 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 0 1 1 1 1 1 0 0 1 2 1 1 1 1 1 1 0 0 1 1 1 0 1 1 1 1 2 1 0 0 0 1 0 1 1 1 1 0 1 1 1
0 0 8 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 1 1 1 1 1 1 1 1 1 0 0 0 1 0 0 1 1 1 0 1 1 1 1 1 1 0 1 0 1 0 1 1 1 1 1 1 2 1 0 1
0 1 0 0 1 0 1 0 1 1 1 1 0 8 2 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 4 1 0 1 0 1 1 1 0 1 0 0 1 1 1 1 1 0 2 1 1 1 1 1 1 1 0 1 0 1
1 1 1 1 1 1 3 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 4 1 1 0 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 2 1 1 0 0 0 0 0 0 0 0 0 0 0 2 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 3 2 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 0 1 2 1 0 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 4 1 1 1 1 1 1 1 0 1 0 1 0 1 0 1 0 1 0 0 9
  
```

The raw data as read into R replacing blanks with NA

```
> recall <- read.clipboard.tab() > recall <- recall[-1] > recall
Condition L1P1 L1P2 L1P3 L1P4 L1P5 L1P6 L1P7
L1P8 L1P9 L1P10 L1P11 L1P12 L1P13 L1P14 L1P15 L1Tot L2P1 L2P2 L2P3 L2P4 L2P5 L2P6 L2P7 L2P8 L2P9
L2P10 L2P11 L2P12 L2P13 L2P14 L2P15 L2Tot L3P1 L3P2 L3P3 L3P4 L3P5 L3P6 L3P7 L3P8 L3P9 L3P10
L3P11 L3P12 L3P13 L3P14 1 1 1 0 1 1 1 0 0 0 1 0 1 0 0 1 0 7 NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA 2 1 1 1 0 1 0 1 1 0 1 0 0 0 1 1 1 9 NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA 3
1 1 1 1 1 1 1 1 1 0 0 1 0 1 1 1 1 2 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA 5 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1
13 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA 6 1 1 1 1 1 1 1 0 0 0 1 0 0 1 1 10 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA 7 1 1 1 1 1 1 0 1 0 1 0 1 1 1 1 1 2 NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA 8 1 0 1 0 1 0
0 1 0 1 0 1 1 1 1 0 8 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA 9 1 1 0 1 1 1 1 1 1 0 0 0 1 1 1 1 NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
```

A bit of strange code (can be appreciated or ignored)

```
recall <- read.clipboard.tab()
dim(recall)
[1] 21 443
```

```
W <- seq(2, 257, 16)
W
[1] 2 18 34 50 66 82 98 114 130 146 162
    178 194 210 226 242
```

```
w <- outer(W,0:15,"+")
```

```
w
[1] 2 18 34 50 66 82 98 114 130 146 162 178 194 210 226 242
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14] [,15] [,16]
[1,] 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
[2,] 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33
[3,] 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49
...
[16,] 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257
```

1. First copy the data to the clipboard and then read the clipboard into the recall data.frame
2. How big is this data frame? (What are the dimensions?)
3. Create a vector to show where each list is
4. Then create a vector to show how to add up the items

Find means for each person for each position

```
rec <- matrix(NA,nrow=21,ncol=15)
for (i in 1:15) {rec[,i] <- rowMeans(recall[w[,i]],na.rm=TRUE)}
colnames(rec) <- paste0("P",1:15,"")
rownames(rec) <- paste0("S",1:21,"")
rec
```

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
S1	1.000	0.750	1.000	0.750	0.625	0.625	0.625	0.750	0.750	0.625	0.875	0.750	1.000	1.000	0.750
S2	0.875	0.625	0.875	0.625	0.500	0.500	0.875	0.625	0.625	0.625	0.375	0.625	0.875	0.500	0.750
S3	0.875	1.000	0.750	1.000	0.750	0.625	0.750	0.875	0.500	0.375	0.625	0.500	1.000	0.750	0.875
S4	0.750	0.375	0.625	0.750	0.625	0.500	0.375	0.500	0.500	0.250	0.375	0.375	0.875	0.750	0.875
S5	1.000	0.750	1.000	0.750	0.750	0.750	0.875	0.750	1.000	0.875	0.875	0.625	1.000	0.750	0.750
S6	0.875	0.875	1.000	1.000	1.000	1.000	1.000	0.625	0.500	0.625	0.750	0.875	0.875	0.625	1.000
S7	1.000	0.875	0.625	0.625	0.875	0.375	0.750	0.875	0.500	0.375	0.750	0.875	0.750	0.750	0.875
S8	0.375	0.875	0.375	0.625	0.500	0.375	0.625	0.375	0.625	0.500	0.750	0.500	0.625	0.875	0.875
S9	1.000	0.750	0.875	0.750	0.750	0.500	0.875	0.625	0.500	0.125	0.500	0.500	0.375	0.625	0.875
S10	0.875	0.625	0.750	0.875	0.500	0.500	0.625	0.750	0.875	0.625	0.375	0.625	0.500	0.375	0.750
S11	0.875	0.750	0.875	0.750	0.750	0.625	0.500	0.750	0.500	0.375	0.750	0.125	1.000	0.750	0.875
S12	0.875	0.750	1.000	0.875	0.750	0.750	0.500	0.500	0.625	0.625	0.375	0.750	0.750	0.625	1.000
S13	1.000	1.000	0.875	0.750	0.625	0.750	0.750	0.500	0.750	1.000	0.375	0.750	0.875	0.750	0.750
S14	0.875	0.750	0.750	0.625	0.500	0.750	0.750	0.625	0.500	0.500	0.375	0.500	0.375	0.750	0.750
S15	1.000	0.750	0.750	0.875	0.500	1.000	0.750	0.750	1.000	0.625	0.750	1.000	0.875	0.750	0.625
S16	1.000	1.000	1.000	1.000	0.750	0.625	0.750	0.750	0.750	0.500	0.750	0.625	0.625	0.625	0.750
S17	1.000	0.875	1.000	0.500	0.875	0.750	0.875	0.750	1.000	0.750	0.750	1.000	1.000	1.000	1.000
S18	1.000	0.875	1.000	0.875	1.000	0.750	0.875	1.000	0.750	1.000	1.000	1.000	1.000	0.875	1.000
S19	0.500	0.500	0.250	0.000	0.000	0.000	0.125	0.000	0.000	0.125	0.125	0.125	0.125	0.000	0.000
S20	1.000	0.750	0.875	0.750	0.875	0.750	0.875	0.875	0.625	0.875	0.750	0.500	0.875	0.875	0.875
S21	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Find the totals for each list

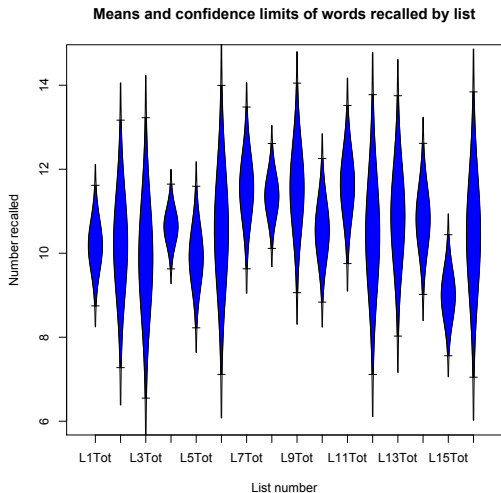
1. The total number recalled for each list was entered as the 16th element for each list
2. We have these data in the spread sheet
3. We can recover them by addressing every 16th position starting at position 17

```
tot<- seq (17,257,16)
recall[,tot]
```

	L1Tot	L2Tot	L3Tot	L4Tot	L5Tot	L6Tot	L7Tot	L8Tot	L9Tot	L10Tot	L11Tot	L12Tot	L13Tot	L14Tot	L15Tot	L16Tot
1	7	NA	NA	11	9	NA	NA	14	NA	14	14	NA	NA	13	11	NA
2	9	NA	NA	11	6	NA	NA	12	NA	10	11	NA	NA	12	8	NA
3	12	NA	NA	12	11	NA	NA	11	NA	10	14	NA	NA	10	10	NA
4	8	NA	NA	10	9	NA	NA	12	NA	9	9	NA	NA	6	5	NA
5	13	NA	NA	11	13	NA	NA	12	NA	13	15	NA	NA	14	9	NA
6	10	NA	NA	12	13	NA	NA	13	NA	14	13	NA	NA	13	13	NA
7	12	NA	NA	13	12	NA	NA	12	NA	11	8	NA	NA	12	7	NA
8	8	NA	NA	11	7	NA	NA	12	NA	6	7	NA	NA	12	8	NA
9	11	NA	NA	8	10	NA	NA	10	NA	9	11	NA	NA	8	10	NA
10	13	NA	NA	9	12	NA	NA	7	NA	8	11	NA	NA	7	10	NA
11	9	NA	NA	9	7	NA	NA	10	NA	12	15	NA	NA	12	8	NA
12	NA	6	7	NA	NA	12	10	NA	12	NA	NA	14	13	NA	NA	12
13	NA	11	12	NA	NA	9	13	NA	12	NA	NA	12	12	NA	NA	11
14	NA	10	9	NA	NA	9	10	NA	9	NA	NA	11	9	NA	NA	8
15	NA	12	11	NA	NA	14	12	NA	14	NA	NA	11	10	NA	NA	12
16	NA	12	12	NA	NA	10	10	NA	13	NA	NA	11	12	NA	NA	12
17	NA	14	10	NA	NA	14	13	NA	12	NA	NA	13	14	NA	NA	15
18	NA	13	14	NA	NA	14	15	NA	14	NA	NA	14	14	NA	NA	14
19	NA	2	0	NA	NA	0	7	NA	4	NA	NA	0	2	NA	NA	0
20	NA	12	14	NA	NA	13	14	NA	14	NA	NA	8	12	NA	NA	9 / 18

Show the data by person and by list: Is there an pattern?

```
error.bars(recall[,tot],main="Means and confidence  
limits of words recalled by list",xlab="List  
number",ylab="Number recalled")
```



Describe the Position data

```
> describe(rec)
```

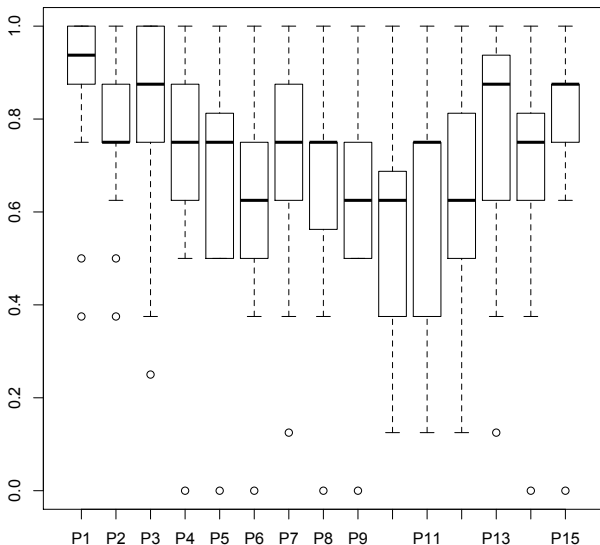
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
P1	1	20	0.89	0.17	0.94	0.93	0.09	0.38	1	0.62	-1.79	2.39	0.04
P2	2	20	0.78	0.16	0.75	0.79	0.19	0.38	1	0.62	-0.64	0.08	0.04
P3	3	20	0.81	0.21	0.88	0.85	0.19	0.25	1	0.75	-1.19	0.62	0.05
P4	4	20	0.74	0.22	0.75	0.77	0.19	0.00	1	1.00	-1.63	3.52	0.05
P5	5	20	0.68	0.23	0.75	0.69	0.19	0.00	1	1.00	-1.02	1.49	0.05
P6	6	20	0.62	0.23	0.62	0.63	0.19	0.00	1	1.00	-0.71	0.90	0.05
P7	7	20	0.71	0.21	0.75	0.73	0.19	0.12	1	0.88	-1.10	0.81	0.05
P8	8	20	0.66	0.22	0.75	0.69	0.19	0.00	1	1.00	-1.22	1.83	0.05
P9	9	20	0.64	0.23	0.62	0.65	0.19	0.00	1	1.00	-0.53	0.78	0.05
P10	10	20	0.57	0.25	0.62	0.57	0.28	0.12	1	0.88	-0.02	-0.86	0.06
P11	11	20	0.61	0.23	0.75	0.62	0.19	0.12	1	0.88	-0.33	-1.11	0.05
P12	12	20	0.63	0.25	0.62	0.65	0.19	0.12	1	0.88	-0.31	-0.62	0.06
P13	13	20	0.76	0.25	0.88	0.79	0.19	0.12	1	0.88	-0.98	-0.04	0.06
P14	14	20	0.70	0.22	0.75	0.73	0.19	0.00	1	1.00	-1.41	2.44	0.05
P15	15	20	0.80	0.22	0.88	0.84	0.19	0.00	1	1.00	-2.39	6.56	0.05

Oops, there was something wrong with the data

1. Note on the previous slide that the minimum for some positions was zero.
2. This does not look right.
3. Lets explore the data graphically to see what is happening.
4. It turns out that one person systematically had very poor recall.
5. How should we treat such an outlier?
6. To be contnued!

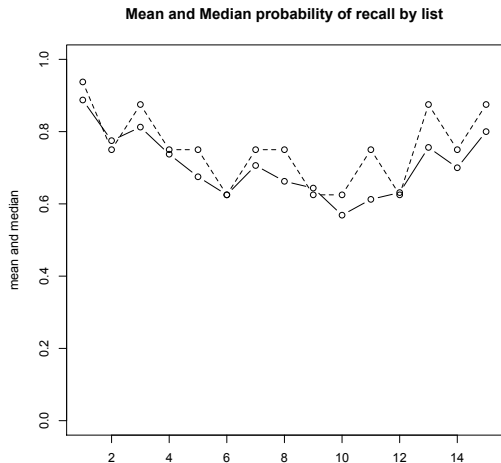
Boxplot the data to try to figure out what is happening

show the distributions of the data



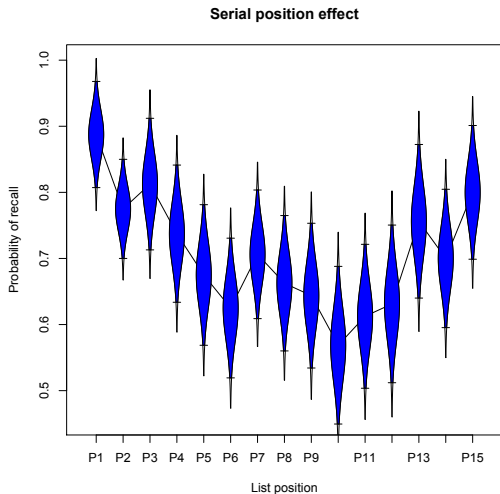
Plot some summary estimates of central tendency; Is there a pattern?

```
plot(stats[, "mean"], typ="b", xlab="Position", ylab="Probability of recall", main="Serial position (means)", ylim=c(0,1))
```



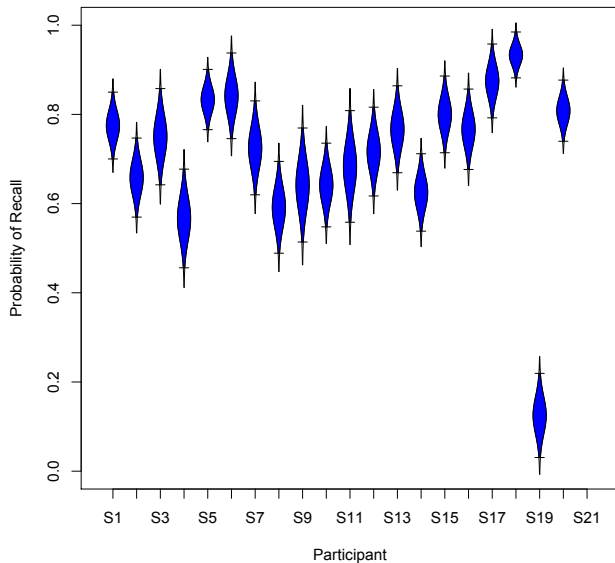
The recall data show a serial position– add in the standard errors

```
error.bars(rec,ylab="Probability of recall",xlab="List position",main="Serial position effect",typ="b")
```



People are a major source of difference

People differ in their probability of endorsement



R code for person graphs — supplementary information

R is a syntax driven language, but each line of syntax is pretty straightforward. This is shown here for demonstration purposes on how to draw some graphs using some of the built in functions.

```
#first, plot the means and medians
plot(colMeans(rec[,1:15],na.rm=TRUE),ylim=c(0,1),ylab="mean and median",xlab="Position",
      main="Mean and Median probability of recall by list",type="b")
points(apply(rec[,1:15],2,median,na.rm=TRUE)
       ,type="b",lty="dashed")
#now show the error bars
error.bars(t(rec[,1:15]),ylab="Probability of Recall",ylim=c(0,1),
           xlab="Participant",main="People differ in their probability of endorsement")
#show them by group
error.bars.by(rec[,1:15],group=recall$Condition,ylim=c(0,1),
             ylab="Probability of Recall",xlab="Serial Position")

#plot by person rather than by item (this is plotting the matrix transpose)
error.bars(t(rec[,1:15]),ylab="Probability of Recall",ylim=c(0,1),
           xlab="Participant",main="People differ in their probability of endorsement")

#find the individual total scores
tot <- rowSums(rec[,1:15])
ord <- order(tot)
#plot subjects ordered by total score
error.bars(t(rec[ord,1:15]),ylab="Probability of Recall",ylim=c(0,1),xlab=
           "Participant",main="People differ in their probability of endorsement")
```

Preliminary analysis of the recall data

Coming soon.