Psychology 205: Research Methods
Analysis of Roediger and McDermott False Memory study

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

October 28, 2013

Outline

Data entry and checking

- Perhaps the most tedious part of research is data entry.
  - You need to make sure that the data are entered correctly.
  - Check and double check.
  - Enter the data twice and automatically compare files.
  - Keep the original data sheets so that you can go back and check once again.
- Also need to look for unusual participants (outliers)
  - People who did not cooperate.
  - People who did not understand the instructions.
  - Might also be due to subject fatigue, distraction, confusion.
  - Most likely due to data entry error!
- After data entry, preliminary graphical and descriptive checks are necessary.
  - In fact, last week we found data entry errors by doing this analysis.
  - Also important to make sure that the coding variables are correct.

Preliminary Analyses

- Two kinds of data
  - Recall: Write down the words you saw/heard
  - Recognition: Have you seen/heard the word before?
- Analysis of recall data
  - Was there a serial position effect for recall? (More of a test of whether Participants were following instructions).
  - Did modality of presentation affect recall?
- Preliminary analysis of recognition data
  - Did recognition differ for words that were presented versus not presented?
  - Were there more false recognitions of words that were similar to the words presented than for control words?
  - Does this depend upon modality of presentation?
  - Is there a tendency to false recognize?
    - Do people who falsely recognize Target words also falsely recognize non-presented, non-primed words?

Data entry

```
> recall <- read.clipboard.tab()  #read in from an excel file that was copied to the clipboard
> dim(recall)  #how many rows (subjects) and columns (variables)
[1]21 241
> colnames(recall)  #what are they called

  colnames(recall)
   [1] "V1"     "V2"     "V3"     "V4"     "V5"     "V6"     "V7"     "V8"     "V9"     "V10"    "V11"    "
  [13] "V13"    "V14"    "V15"    "V1.1"   "V2.1"   "V3.1"   "V4.1"   "V5.1"   "V6.1"   "V7.1"   "V8.1"   "
  [25] "V10.1"  "V11.1"  "V12.1"  "V13.1"  "V14.1"  "V15.1"  "V1.2"   "V2.2"   "V3.2"   "V4.2"   "V5.2"   "
  [37] "V7.2"   "V8.2"   "V9.2"   "V10.2"  "V11.2"  "V12.2"  "V13.2"  "V14.2"  "V15.2"  "V1.3"   "V2.3"   "
  [49] "V4.3"   "V5.3"   "V6.3"   "V7.3"   "V8.3"   "V9.3"   "V10.3"  "V11.3"  "V12.3"  "V13.3"  "V14.3"  "
  [61] "V1.4"   "V2.4"   "V3.4"   "V4.4"   "V5.4"   "V6.4"   "V7.4"   "V8.4"   "V9.4"   "V10.4"  "V11.4"  "
  [73] "V13.4"  "V14.4"  "V15.4"  "V1.5"   "V2.5"   "V3.5"   "V4.5"   "V5.5"   "V6.5"   "V7.5"   "V8.5"   "
  [85] "V10.5"  "V11.5"  "V12.5"  "V13.5"  "V14.5"  "V15.5"  "V1.6"   "V2.6"   "V3.6"   "V4.6"   "V5.6"   "
  [97] "V7.6"   "V8.6"   "V9.6"   "V10.6"  "V11.6"  "V12.6"  "V13.6"  "V14.6"  "V15.6"  "V1.7"   "V2.7"   "
 [109] "V4.7"   "V5.7"   "V6.7"   "V7.7"   "V8.7"   "V9.7"   "V10.7"  "V11.7"  "V12.7"  "V13.7"  "V14.7"  "
 [121] "V1.8"   "V2.8"   "V3.8"   "V4.8"   "V5.8"   "V6.8"   "V7.8"   "V8.8"   "V9.8"   "V10.8"  "V11.8"  "
 [133] "V13.8"  "V14.8"  "V15.8"  "V1.9"   "V2.9"   "V3.9"   "V4.9"   "V5.9"   "V6.9"   "V7.9"   "V8.9"   "
 [145] "V10.9"  "V11.9"  "V12.9"  "V13.9"  "V14.9"  "V15.9"  "V1.10"  "V2.10"  "V3.10"  "V4.10"  "V5.10"  "
 [157] "V7.10"  "V8.10"  "V9.10"  "V10.10" "V11.10" "V12.10" "V13.10" "V14.10" "V15.10" "V1.11"  "V2.11"  "
 [169] "V4.11"  "V5.11"  "V6.11"  "V7.11"  "V8.11"  "V9.11"  "V10.11" "V11.11" "V12.11" "V13.11" "V14.11" "
 [181] "V1.12"  "V2.12"  "V3.12"  "V4.12"  "V5.12"  "V6.12"  "V7.12"  "V8.12"  "V9.12"  "V10.12" "V11.12" "
 [193] "V13.12" "V14.12" "V15.12" "V1.13"  "V2.13"  "V3.13"  "V4.13"  "V5.13"  "V6.13"  "V7.13"  "V8.13"  "
 [205] "V10.13" "V11.13" "V12.13" "V13.13" "V14.13" "V15.13" "V1.14"  "V2.14"  "V3.14"  "V4.14"  "V5.14"  "
 [217] "V7.14"  "V8.14"  "V9.14"  "V10.14" "V11.14" "V12.14" "V13.14" "V14.14" "V15.14" "V1.15"  "V2.15"  "
 [229] "V4.15"  "V5.15"  "V6.15"  "V7.15"  "V8.15"  "V9.15"  "V10.15" "V11.15" "V12.15" "V13.15" "V14.15" "
 [241] "Group"
```

Do some mild changes to enhance readability

```
item <- seq(1,226,15)
itemi <- 0:14 %+% t(item)
rownames(itemi) <- paste("P",1:15,sep="")
colnames(itemi) <- paste("L",1:16,sep="")
itemi  #show these
keys <- apply(item,1,function(x) as.list(x))  #convert the array to a list
recall.keys <- make.keys(241,keys) #make up the scoring keys
 recall.scores <- score.items(recall.keys,recall,missing=TRUE,impute=FALSE) #score recall
```

```
> itemi   #score items this way to get position effects
    L1 L2 L3 L4 L5 L6  L7  L8  L9 L10 L11 L12 L13 L14 L15 L16
P1   1 16 31 46 61 76  91 106 121 136 151 166 181 196 211 226
P2   2 17 32 47 62 77  92 107 122 137 152 167 182 197 212 227
P3   3 18 33 48 63 78  93 108 123 138 153 168 183 198 213 228
P4   4 19 34 49 64 79  94 109 124 139 154 169 184 199 214 229
P5   5 20 35 50 65 80  95 110 125 140 155 170 185 200 215 230
P6   6 21 36 51 66 81  96 111 126 141 156 171 186 201 216 231
P7   7 22 37 52 67 82  97 112 127 142 157 172 187 202 217 232
P8   8 23 38 53 68 83  98 113 128 143 158 173 188 203 218 233
P9   9 24 39 54 69 84  99 114 129 144 159 174 189 204 219 234
P10 10 25 40 55 70 85 100 115 130 145 160 175 190 205 220 235
P11 11 26 41 56 71 86 101 116 131 146 161 176 191 206 221 236
P12 12 27 42 57 72 87 102 117 132 147 162 177 192 207 222 237
P13 13 28 43 58 73 88 103 118 133 148 163 178 193 208 223 238
P14 14 29 44 59 74 89 104 119 134 149 164 179 194 209 224 239
P15 15 30 45 60 75 90 105 120 135 150 165 180 195 210 225 240
```

Recall analysis

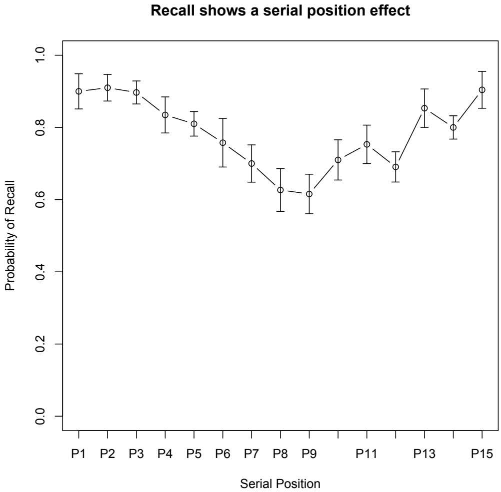
Read in the data, look for serial position effects, look for modality effects. First describe the data.

```
describe(recall.scores$scores)  #show the statistics
  #basic descriptive statistics for just the recall data
> error.bars(recall.scores$scores,ylim=c(0,1),ylab="Probability of Recall",
    xlab="Serial Position",main="Recall shows a serial position effect",typ="b")
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
P1	1	20	0.90	0.10	0.88	0.91	0.19	0.62	1.00	0.38	-0.87	0.20	0.02
P2	2	20	0.91	0.08	0.93	0.92	0.10	0.73	1.00	0.27	-0.43	-0.82	0.02
P3	3	20	0.90	0.07	0.94	0.90	0.09	0.75	1.00	0.25	-0.45	-0.86	0.02
P4	4	20	0.83	0.11	0.85	0.85	0.11	0.62	1.00	0.38	-0.65	-0.58	0.02
P5	5	20	0.81	0.07	0.80	0.81	0.10	0.67	0.93	0.27	-0.05	-0.99	0.02
P6	6	20	0.76	0.14	0.77	0.77	0.23	0.46	0.92	0.46	-0.39	-1.18	0.03
P7	7	20	0.70	0.11	0.69	0.70	0.09	0.44	0.88	0.44	-0.45	-0.27	0.02
P8	8	20	0.63	0.13	0.63	0.63	0.15	0.33	0.87	0.53	-0.29	-0.36	0.03
P9	9	20	0.62	0.12	0.62	0.62	0.09	0.38	0.81	0.44	-0.34	-0.82	0.03
P10	10	20	0.71	0.12	0.67	0.71	0.10	0.53	0.93	0.40	0.24	-1.24	0.03
P11	11	20	0.75	0.11	0.78	0.77	0.14	0.50	0.88	0.38	-0.67	-0.79	0.03
P12	12	20	0.69	0.09	0.69	0.69	0.09	0.56	0.88	0.31	0.22	-1.00	0.02
P13	13	20	0.85	0.11	0.87	0.86	0.10	0.60	1.00	0.40	-0.43	-0.67	0.03
P14	14	20	0.80	0.07	0.81	0.80	0.09	0.62	0.94	0.31	-0.29	0.22	0.02
P15	15	20	0.90	0.11	0.92	0.92	0.12	0.67	1.00	0.33	-0.68	-0.98	0.02

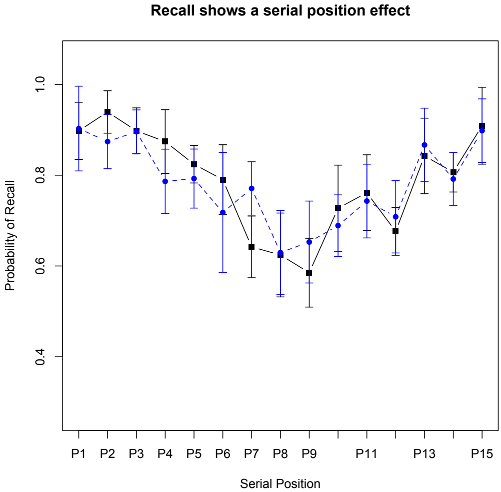
Graph the Serial Position Effects

```
error.bars(recall.scores$scores,ylim=c(0,1),
 ylab="Probability of Recall",
  xlab="Serial Position",main="Recall shows a serial position effect",typ="b")
```



Does this differ by A/B groups? – Not visibly

```
error.bars.by(recall.df[,1:15],group=recall.df[,16],
ylab="Probability of Recall",xlab="Serial Position",
main="Recall shows a serial position effect")
```



Score by list

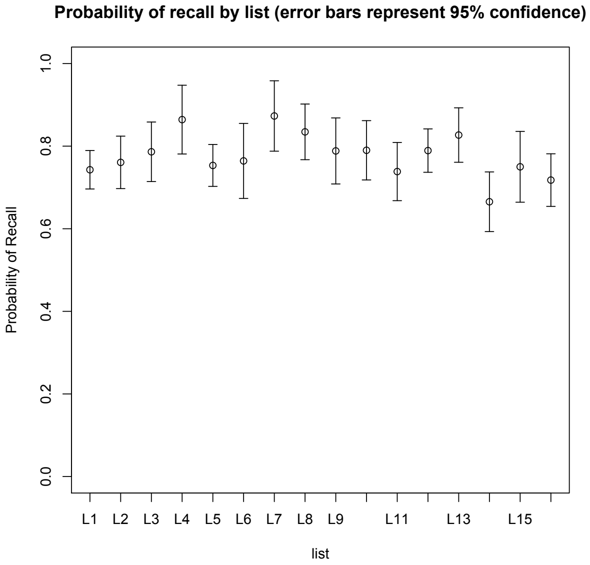
```
itemlist <- item \%+\% t(0:14)
> colnames(itemlist) <- paste("P",1:15,sep="")
> rownames(itemlist) <- paste("L",1:16,sep="")
> itemlist
> keysL <- apply(itemlist,1,function(x) as.list(x))
> recall.keysL <- make.keys(241,keysL)
> colnames(recall.keysL) <- paste("L",1:16,sep="")
>recall.list.df <- score.items(recall.keysL,recall,impute=FALSE,missing=TRUE)
```

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
L1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
L2	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
L3	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
L4	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
L5	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
L6	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
L7	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
L8	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
L9	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
L10	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
L11	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165
L12	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
L13	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195
L14	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
L15	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
L16	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240

Do the lists differ in difficulty – somewhat

```
describe(recall.list.df$scores)
    var  n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
L1    1 20 0.74 0.10   0.79    0.76 0.00 0.43 0.86  0.43 -1.81     2.65 0.02
L2    2 20 0.76 0.14   0.82    0.77 0.05 0.50 1.00  0.50 -0.40    -1.09 0.03
L3    3 20 0.79 0.15   0.86    0.81 0.07 0.45 0.91  0.45 -0.89    -0.51 0.03
L4    4 20 0.86 0.18   1.00    0.88 0.00 0.50 1.00  0.50 -0.61    -1.38 0.04
L5    5 20 0.75 0.11   0.80    0.76 0.00 0.53 1.00  0.47 -0.39     0.30 0.02
L6    6 20 0.76 0.19   0.86    0.79 0.05 0.29 1.00  0.71 -1.33     1.00 0.04
L7    7 20 0.87 0.18   0.96    0.91 0.06 0.31 1.00  0.69 -1.62     2.16 0.04
L8    8 20 0.83 0.14   0.92    0.87 0.00 0.38 0.92  0.54 -1.67     2.23 0.03
L9    9 20 0.79 0.17   0.85    0.81 0.06 0.23 1.00  0.77 -1.72     3.22 0.04
L10  10 20 0.79 0.15   0.87    0.82 0.00 0.27 0.93  0.67 -2.07     4.09 0.03
L11  11 20 0.74 0.15   0.77    0.76 0.00 0.23 0.92  0.69 -1.94     4.01 0.03
L12  12 20 0.79 0.11   0.86    0.81 0.00 0.50 0.93  0.43 -1.16     0.28 0.03
L13  13 20 0.83 0.14   0.92    0.85 0.06 0.54 1.00  0.46 -0.87    -0.68 0.03
L14  14 20 0.67 0.15   0.69    0.69 0.00 0.23 0.92  0.69 -1.43     2.09 0.03
L15  15 20 0.75 0.18   0.86    0.79 0.00 0.29 0.86  0.57 -1.44     0.54 0.04
L16  16 20 0.72 0.14   0.79    0.73 0.00 0.43 0.93  0.50 -0.81    -0.61 0.03
```

Probability of Recall



Does it differ by modality of presentation? The paired t-test.

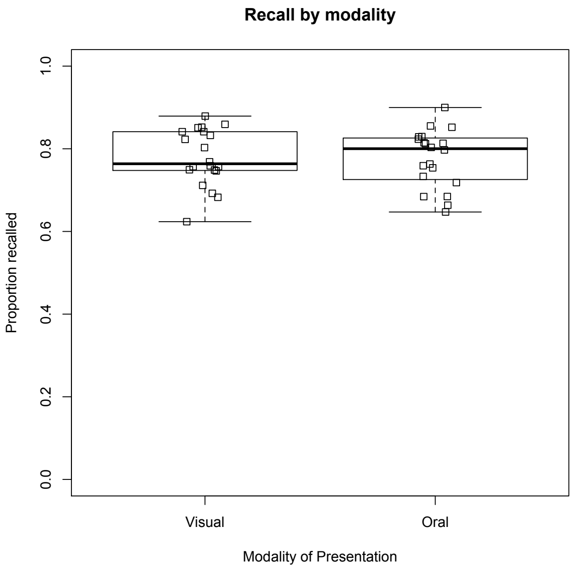
Words were presented two different ways, Visually and Orally. This was within subjects. We first find the total recalled for the two conditions. The means are identical!

```
recall.scores <- recall.list.df$scores
> Visual <- rowMeans(recall.scores[,c(1,2,7,8,11:14)])
> Oral <- rowMeans(recall.scores[,c(3:6,9,10,15:16)])
> recall.df <- data.frame(Visual,Oral)
> describe(recall.df)
>  t.test(Visual,Oral,paired=TRUE)
```

```
   var  n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
Visual   1 20 0.78 0.07   0.76    0.78 0.09 0.62 0.88  0.26 -0.41    -0.80 0.02
Oral     2 20 0.78 0.07   0.80    0.78 0.06 0.65 0.90  0.25 -0.32    -1.04 0.02
Paired t-test
data:  Visual and Oral
t = 0.1533, df = 19, p-value = 0.8797
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.02645649  0.03063939
sample estimates:
mean of the differences
             0.00209145
```

Mean recall does not differ as a function of presentation modality.

Show it graphically – Still no difference



The t-test

- Developed by "student" (William Gosset)
  - A small sample extension of the z-test for comparing two groups
  - Like most statistics, what is the size of the effect versus the error of the effect?
  - Standard error of a mean is $s.e. = \sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{N}}$
- Two cases
  - Independent groups
    - $t = \frac{\bar{X_1} - \bar{X_2}}{\sqrt{\sigma_1^2/N_1 + \sigma_2^2/N_2}}$
    - degrees of freedom $df = N_1 - 1 + N_2 - 1$
  - Paired differences (correlated groups)
    - let $d = X_1 - X_2$ then $t = \frac{\bar{d}}{\sigma_{\bar{d}}}$
    - d.f. = N - 1
- done in R with `t.test` function

Recall results

- As predicted, there was a serial position effect.
  - This suggested that the participants followed instructions.
- There was no difference between the serial position effects for the two experimental groups.
- There was no effect of modality of presentation on immediate recall.
- Now need to see if there is a false memory effect on subsequent recognition.

Preliminary scoring of recognition data

- 5 kinds of words
  - Strong associates
  - Weaker associates
  - Weak associates
  - Targets (Foils): words that were associates with all the words on a list
  - Other words (not presented)
- Simple totals for all 5 word types, expressed as probabilities
  - Data were the "old" versus "new" judgments
  - Scores were averaged for each word type.
  - Scoring was done using the `score.items` function in R
- Look at the data
  - Descriptive statistics using `describe`
  - exploratory graphics using `pairs.panels`

Recognition scoring – using R to process the data

```
recog.key <- make.keys(96,list(
high=c(11,83,82,45,81,46,44,25,15,31,59,57,14,49,58,95),
med=c(27,77,2,26,30,36,3,10,78,22,62,96,73,43,16,69),
low=c(66,65,17,60,24,21,47,33,5,74,88,4,85,18,6,94),
target =c(12,19,61,80,39,79,56,42,53,70,1,28,76,91,87,86),
control=c(90,48,51,32,7,35,50,23,13,68,8,89,71,52,84,38,29,63,54,9,64,72,40,
          41,92,75,55,37,67,20,34,93),
nTarget=c(90,48,51,32,7,35,50,23),
nHigh=c(13,68,8,89,71,52,84,38),
nMed=c(29,63,54,9,64,72,40,41),
nLow=c(92,75,55,37,67,20,34,93)))
recognition <- read.clipboard.tab()  #get the data
recog <- recognition #just make a copy
recog[recog>1] <- 1  #ignore the know remember distinction for now
recog.scores <- score.items(recog.key,recog,impute=FALSE)
```

R code for organizing the data for Visual Oral scores

```
recog.list <- list(
high=c(11,83,82,45,81,46,44,25,15,31,59,57,14,49,58,95),
med=c(27,77,2,26,30,36,3,10,78,22,62,96,73,43,16,69),
low=c(66,65,17,60,24,21,47,33,5,74,88,4,85,18,6,94),
target =c(12,19,61,80,39,79,56,42,53,70,1,28,76,91,87,86))
rec.order <- matrix(unlist(rec.list),nrow=4,byrow=TRUE)

visual.recog <- rec.order[,c(1,2,7,8,11:14)]
oral.recog  <- rec.order[,c(3:6,9,10,15:16)]
recog.df <- rbind(visual.recog,oral.recog)
recog.list <- apply(recog.df,1,function(x) as.list(x))
recog.keysVO <- make.keys(96,recog.list)
recog.keys.control <- make.keys(96,list(control))
recog.keys.VOC <- cbind(recog.keysVO,recog.keys.control)
colnames(recog.keys.VOC) <- c(rownames(recog.df),"Control")
```

More R to combine the various keys and score them

```
recog.list <- list(
high=c(11,83,82,45,81,46,44,25,15,31,59,57,14,49,58,95),
med=c(27,77,2,26,30,36,3,10,78,22,62,96,73,43,16,69),
low=c(66,65,17,60,24,21,47,33,5,74,88,4,85,18,6,94),
target =c(12,19,61,80,39,79,56,42,53,70,1,28,76,91,87,86))

rec.order <- matrix(unlist(recog.list),nrow=4,byrow=TRUE)

visual.recog <- rec.order[,c(1,2,7,8,11:14)]
oral.recog  <- rec.order[,c(3:6,9,10,15:16)]
recog.df <- rbind(visual.recog,oral.recog)
recog.list <- apply(recog.df,1,function(x) as.list(x))
recog.keysVO <- make.keys(96,recog.list)
recog.keys.control <- make.keys(96,list(control))
recog.keys.VOC <- cbind(recog.keysVO,recog.keys.control)
colnames(recog.keys.VOC) <- c(rownames(recog.df),"Control")
recognition.scores <- score.items(recog.keys.VOC,recognition,impute=FALSE)
recog.VO.df <- data.frame(correctV=rowMeans(recognition.scores.VO$scores[,1:3]),
 correctO=rowMeans(recognition.scores.VO$scores[,5:7]),TV=recognition.scores.VO$
 TO=recognition.scores.VO$scores[,8],Control=recognition.scores.VO$scores[,9])
 describe(recog.VO.df)
```

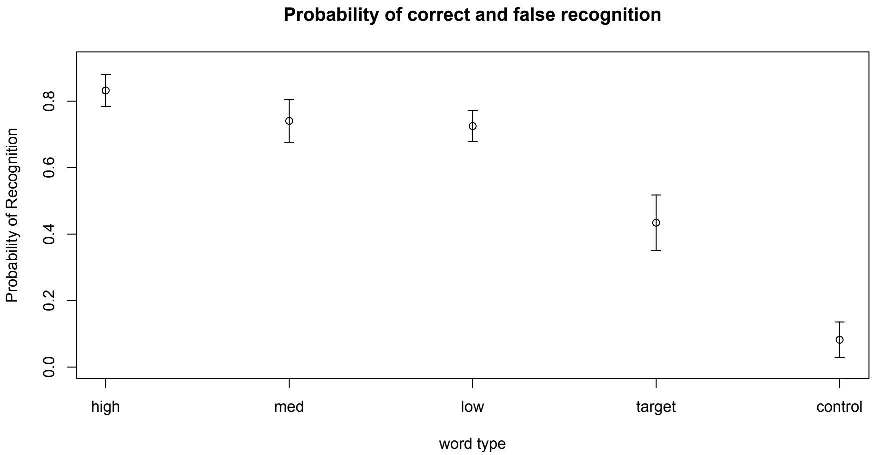
Descriptive Statistics of Recognition data

```
describe(recog.scores$scores)
 pairs.panels(recog.scores$scores)

      var  n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
high       1 20 0.83 0.18   0.89    0.86 0.11 0.29 1.00  0.71 -1.45     1.48 0.04
med        2 20 0.74 0.19   0.81    0.75 0.19 0.38 1.00  0.62 -0.41    -1.29 0.04
low        3 20 0.72 0.21   0.72    0.73 0.23 0.38 1.00  0.62 -0.40    -1.23 0.05
target     4 20 0.43 0.19   0.44    0.45 0.23 0.00 0.69  0.69 -0.53    -0.73 0.04
control    5 20 0.08 0.12   0.04    0.06 0.05 0.00 0.43  0.43  1.47     0.99 0.03
nTarget    6 20 0.11 0.17   0.00    0.08 0.00 0.00 0.60  0.60  1.45     1.50 0.04
nHigh      7 20 0.09 0.16   0.00    0.05 0.00 0.00 0.50  0.50  1.44     0.58 0.04
nMed       8 20 0.08 0.16   0.00    0.05 0.00 0.00 0.50  0.50  1.52     0.73 0.04
nLow       9 20 0.06 0.10   0.00    0.04 0.00 0.00 0.29  0.29  1.29     0.26 0.02

#now by V and O condition
        var  n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
correctV    1 20 0.77 0.17   0.79    0.78 0.19 0.29 1.00  0.71 -0.92     0.42 0.04
correctO    2 20 0.76 0.19   0.79    0.79 0.22 0.33 0.96  0.62 -0.71    -0.72 0.04
TV          3 20 0.39 0.23   0.38    0.38 0.19 0.00 0.88  0.88  0.52    -0.52 0.05
TO          4 20 0.48 0.25   0.50    0.48 0.19 0.00 1.00  1.00  0.16    -0.52 0.06
Control     5 20 0.08 0.12   0.04    0.06 0.05 0.00 0.43  0.43  1.47     0.99 0.03
```

Data entry for recall data

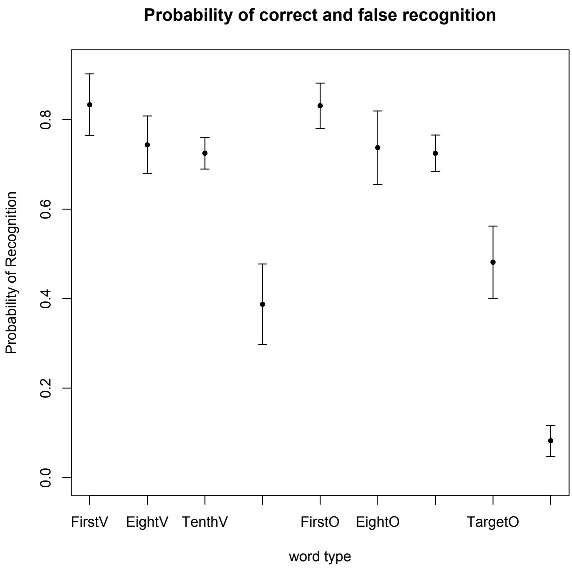
- Recall by list order
- Recognition by prior recall
  - Were words that were recalled more likely to be later recognized?
  - Were studied words more likely to be later recognized?
  - Were “foil” words from lists that were not recalled more likely to be recognized
- Data sheet for scoring data
- Data entry

Recognition by condition




```
error.bars(recog.scores$scores[,1:5],ylab="Probability of Recognition",
 xlab="word type",main="Probability of correct and false recognition",
  within=TRUE)
```

Recognition by Visual and Oral condition and word type



Descriptive versus inferential

- Descriptive statistics is also known as Exploratory Data Analysis; Analogous to a detective trying to solve a crime
  - We are acting as a detective, trying to understand what is going on
  - Looking for strange behaviors
  - Developing hypotheses (ideally hypotheses are developed before collecting data, but it is important for future studies to examine the current data to develop hypotheses)
- This is in contrast to Inferential Statistics; analogous to a court proceeding with the presumption of innocence.
  - Are the results different enough from what is expected by a simpler hypothesis to reject that simpler hypothesis.
  - A typical "simple" hypothesis is the "Null Hypothesis" (aka "Nill" hypothesis).
  - What is the likelihood of observing our data given the Null hypothesis versus our alternative hypothesis.
- What do the data show? How certain are we that they show it?

Inferential tests: t, F, r

- The basic inferential test is the t-test.
  - Is the difference between two group means larger than expected by chance?
  - "Nill hypothesis" is that two groups are both sampled from the same population.
  - Alternative hypothesis is that the two groups come from different populations with different means.
- The basic test was developed by William Gosset (publishing under the pseudonym of "student")
- Two cases
  - Independent groups
    - $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/N_1 + \sigma_2^2/N_2}}$
    - degrees of freedom $df = N_1 - 1 + N_2 - 1$
  - Paired differences (correlated groups)
    - let $d = X_1 - X_2$ then $t = \frac{\bar{d}}{\sigma_{\bar{d}}}$
    - d.f. = N - 1

Hypothesis testing using inferential statistics

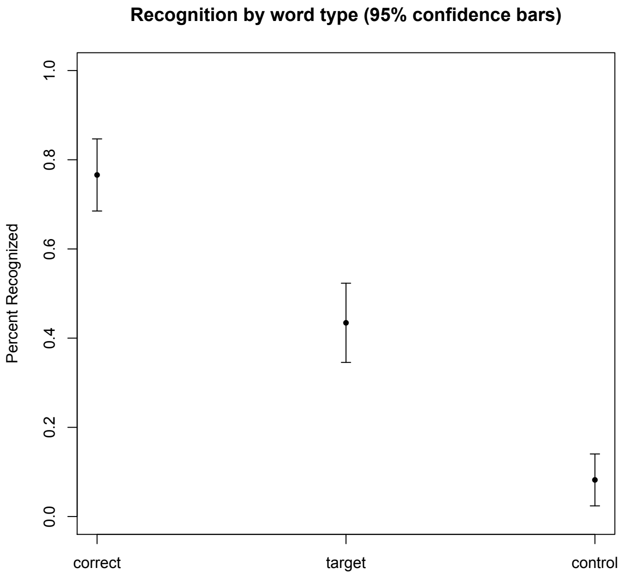
- How likely are the observed data given the hypothesis that an Independent Variable has no effect.
- Bayesian statistics compare the likelihood of the data given the hypothesis of no differences as contrasted to the likelihood of the data given competing hypotheses.
  - This takes into account our prior willingness to believe that the IV could have an effect.
  - Also takes into account our strength of belief in the hypothesis of no effect
- Conventional tests report the probability of the data given the "Null" hypothesis of no difference.
- The less likely the data are to be observed given the Null, the more we tend to discount the Null.
  - Three kinds of inferential errors: Type I, Type II and Type III
  - Type I is rejecting the Null when in fact it is true
  - Type II is failing to reject the Null when it is in fact not true
  - Type III is asking the wrong question

Descriptive statistics of the recognition data

```
recog.df <- data.frame(recog.scores$scores,correct = rowMeans(recog.scores$scores[,1:3]))
describe(recog.df)
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
high	1	20	0.83	0.18	0.89	0.86	0.11	0.29	1.00	0.71	-1.45	1.48	0.04
med	2	20	0.74	0.19	0.81	0.75	0.19	0.38	1.00	0.62	-0.41	-1.29	0.04
low	3	20	0.72	0.21	0.72	0.73	0.23	0.38	1.00	0.62	-0.40	-1.23	0.05
target	4	20	0.43	0.19	0.44	0.45	0.23	0.00	0.69	0.69	-0.53	-0.73	0.04
control	5	20	0.08	0.12	0.04	0.06	0.05	0.00	0.43	0.43	1.47	0.99	0.03
nTarget	6	20	0.11	0.17	0.00	0.08	0.00	0.00	0.60	0.60	1.45	1.50	0.04
nHigh	7	20	0.09	0.16	0.00	0.05	0.00	0.00	0.50	0.50	1.44	0.58	0.04
nMed	8	20	0.08	0.16	0.00	0.05	0.00	0.00	0.50	0.50	1.52	0.73	0.04
nLow	9	20	0.06	0.10	0.00	0.04	0.00	0.00	0.29	0.29	1.29	0.26	0.02
correct	10	20	0.77	0.17	0.78	0.78	0.22	0.39	0.98	0.59	-0.64	-0.78	0.04

Percent of words recognized



Two correlated t-tests

with(recog.df,t.test (correct,target,paired=TRUE))

```
with(recog.df,t.test (correct,target,paired=TRUE))
Paired t-test
data:  correct and target
t = 6.3952, df = 19, p-value = 3.919e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2230393 0.4400559
sample estimates:
mean of the differences
              0.3315476
> with(recog.df,t.test (control,target,paired=TRUE))
Paired t-test
data:  control and target
t = -8.0851, df = 19, p-value = 1.432e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4434155 -0.2610488
sample estimates:
mean of the differences
             -0.3522321
```

Test if modality of presentation affects recognition– it doesn't

```
describe(recog.VO.df)

          var  n mean   sd median trimmed  mad  min  max range  skew kurtosis
correctV    1 20 0.77 0.17   0.79    0.78 0.19 0.29 1.00  0.71 -0.92     0.42 0.
correctO    2 20 0.76 0.19   0.79    0.79 0.22 0.33 0.96  0.62 -0.71    -0.72 0.
TV          3 20 0.39 0.23   0.38    0.38 0.19 0.00 0.88  0.88  0.52    -0.52 0.
TO          4 20 0.48 0.25   0.50    0.48 0.19 0.00 1.00  1.00  0.16    -0.52 0.
Control     5 20 0.08 0.12   0.04    0.06 0.05 0.00 0.43  0.43  1.47     0.99 0.

with(recog.VO.df,t.test(TV,TO,paired=TRUE))
Paired t-test
data:  TV and TO
t = -1.4787, df = 19, p-value = 0.1556
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.22644386  0.03894386
sample estimates:
mean of the differences
               -0.09375
```

Analysis of Variance allows for testing multiple hypotheses at once

- The t-test compares group means to the standard error of their differences
- The F-test (developed by R. A. Fisher) compares the variances between group means to the variance within groups.
  - For two groups, F is just t^2, but the theory generalizes to multiple groups.
  - The F is a ratio of variances: $\frac{VarianceBetweenGroups}{Variancewithingroups} = \frac{\sigma^2_{bg}}{\sigma^2_{wg}}$
    - To make this sound more complicated than it really is, variances are called "Mean Squares" and are found by finding the Sums of Squares between Groups and the Sums of Squares within Groups.
    - These Sums of Squares are in turned divided by the "degrees of freedom" or "df" to find MS (Mean Squares) or σ^2
- We now recognize that these variance components can be estimated by linear regression, but some still prefer the ANOVA terminology.

Using ANOVA to compare recognition accuracy

- ANOVA partitions the total variance into various independent parts:
  - Variance between groups
  - Variance within subjects.
  - But if there are more than two groups, the variance between groups can be further partitioned.
  - And, in the case of within subject analyses, the variance within subjects can be partitioned into that which is due to groups and that which is left over (residual variance).
- To do within subject analyses in R is a little tricky for it requires reorganizing the data.
- This is why it took so long to do.
- You should try to understand what the final result is, rather than the specific process of doing the analysis.

Several ways to think about the data

- As a between subjects design
  - This will find how much variance is associated with the differences between conditions
  - And then compares this to a pooled estimate of error within conditions.
  - This ignores the fact that the same subjects were in all conditions
- As a within subjects design
  - The variance between conditions is the same
  - But the variance within conditions is divided into that due to subject differences, and that due to subject within condition differences

What have we found?

1. The only reliable (statistically significant) effect is that Correct words are recognized more than False Words.
    - There is a suggestion (That does not reach conventional levels of statistical significance) that Recall versus Math instructions make a difference.
    - There is no difference between the way we presented the words.
2. Lets look at the means to understand what is happening.
    - We do this by the simple R command of print(model.tables(mod3,“means”))

Do these numbers differ from the base rate for recognition?

1. Multiple ways of doing this
   - Perhaps the most simple is to collapse the various conditions.
   - That is, to consider Correct, False, and Base rate recognition.
   - Base rate of recognition just reflects willingness to say yes, I have seen it.
2. We do this by simply averaging the experimental conditions.
   - This is appropriate because there were no interactions.
3. We can test whether these means differ.
   - The most important test is whether the False Recognition differs from the Base Rate of recognition for words that were never exposed.
   - This is actually just a paired sample t-test!

Summarize what we have found — THESE numbers are CORRECT

- Recall results
  - Recall of the items show a serial position effect.
  - This reflects that the subjects followed directions.
- Recognition results
  - Recognition did not differ as a function of modality nor of the chance to recall the lists.
- Real recognition was greater ($m = .77$) than false recognition ($m = .43$) which was still greater than that of control words ($m = .08$).
  - Real words were recognized more ($m = .77$) than false words ($m = .43$), $t(19) = 6.40$, $p < .00001$
  - False words that were associated with presented words were recognized more ($m = .43$) than were words that were not associated with any list ($m = .08$, $t = 8.09$, $p = .00001$)

Results for the paper

1. What is presented above is enough for the paper
2. Probably include at least two figures -
   - serial position effects
   - Recognition by word type
3. Results should also include the inferential statistics
4. Could also include failure to find effects graphs
   - Lack of effect of modality or recall/math
5. Additional analyses of recognition by strength of associate are not included
   - These attempted to consider recognition as a function of associative strength

Structure of final paper (see detailed instructions from before)

1. Abstract (100-150 words)
   - Why did you do the study, Who were the subjects, What did you find, So what? Write it last.
2. Introduction (2-3 pages)
   - A bit of background (adapt from R & M)
   - Overview of study
3. Method (1-3 pages)
   - With enough detail that someone can carry out the study
   - Can refer to word lists from R & M rather than including the words
4. Results (1-3 pages)
   - Just the most important results
   - Should reference table(s) and figure(s) (to appear at end of paper)
5. Discussion (2-3 pages)
   - Why is this study important
   - What are the most important findings
   - So what? What is next