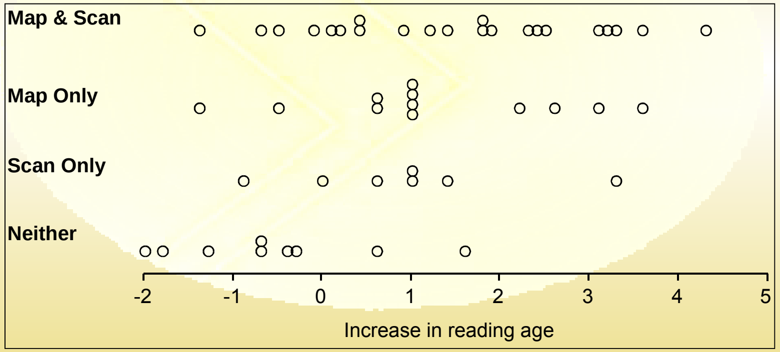
Visual Differences:

Inferential mark-up of plots to give approximate inferences for pair-wise differences

Chris Wild & James Curran

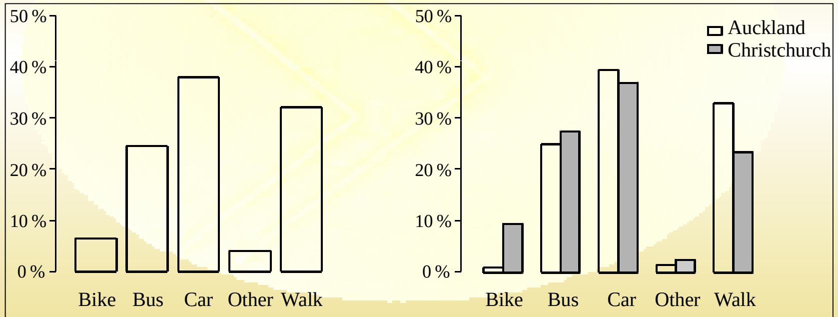
Pair-wise differences

- Many situations where useful to look at pair-wise differences between parameters $\theta_1, \ldots, \theta_I$, i.e. make inferences about $\{\theta_i - \theta_j\}$



Pair-wise differences

- Many situations where useful to look at pair-wise differences between parameters $\theta_1, \ldots, \theta_I$, i.e. make inferences about $\{\theta_i - \theta_j\}$



(a) Comparisons within a sample (b) Comparing two samples

Pair-wise differences

- Many situations where useful to look at pair-wise differences between parameters $\theta_1, \ldots, \theta_I$, i.e. make inferences about $\{\theta_i - \theta_j\}$

```
Call:
svyglm(log(stay + 1) ~ age + aey + adm_typ, design = NZqhsSurv)
Survey design:
svydesign(ids=~cluster_id,strata=~stratum_id,data=NZqhs, weights=~wt)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.010078   0.066113  30.404   0.0209 *
age          0.0091
aey         -0.4095
adm_typAC   -0.214357   0.033520  -6.395   0.0987 .
adm_typWN   -0.394599   0.029617 -13.324   0.0477 *
adm_typZA    0.020351   0.370453   0.055   0.9651
adm_typZC   -0.158209   0.027711  -5.709   0.1104
adm_typZW   -0.659201   0.061893 -10.651   0.0596 .
```

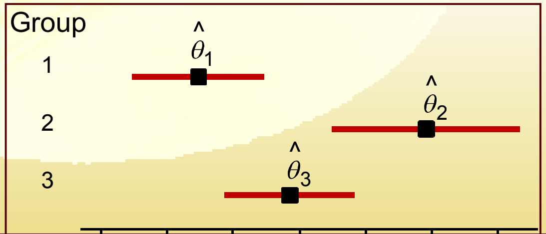
Chisq = 24.18593 on 2 df: p= 5.5988e-06

6-level factor **adm_typ :** AA, AC, WN, ZA, ZC, ZW

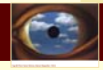
CIs of the form $\hat{\theta}_i - \hat{\theta}_j \pm M_{ij}$

- where M_{ij} is an appropriate margin of error for the (i,j)-comparison
  - may include multiple comparisons adjustments …
- Will approximate $M_{ij} \approx \tilde{M}_{ij} = h_i + h_j$, for suitably chosen $\{h_i\}$ and plot $\{\hat{\theta}_i \pm h_i : i = 1, \dots, I\}$

"Visual Comparison Intervals"

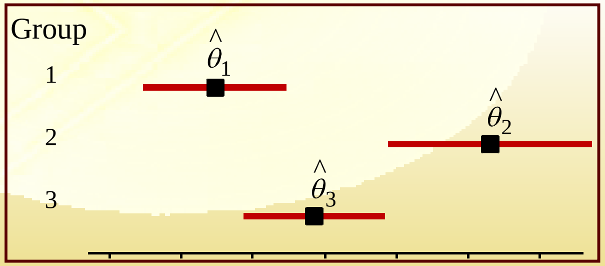


CIs of the form $\hat{\theta}_i - \hat{\theta}_j \pm M_{ij}$

What does the  **see?**

What we can see ***easily*** is:

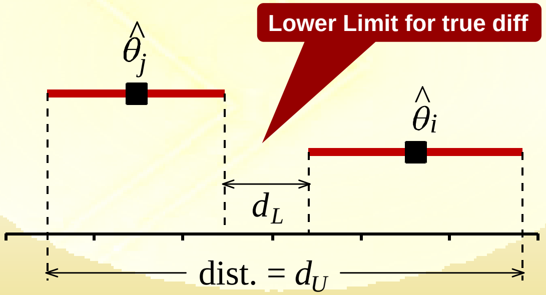
- clear separation
- clear overlap
- "marginal"



CIs of the form $\hat{\theta}_i - \hat{\theta}_j \pm M_{ij}$

What does the  **see?**

Non-overlapping

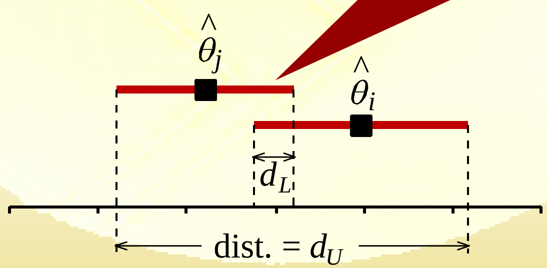


CIs of the form $\hat{\theta}_i - \hat{\theta}_j \pm M_{ij}$

What does the  **see?**

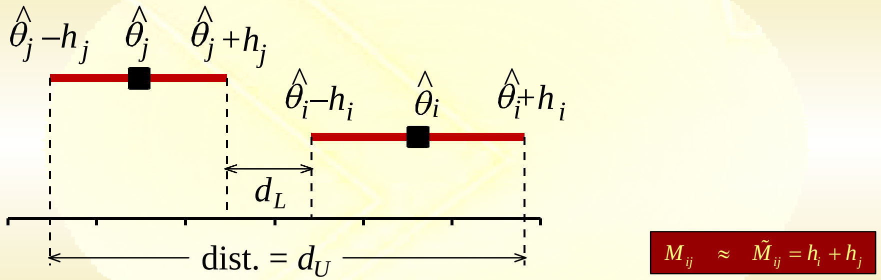
Overlapping

Lower Limit for true diff
- ***Negative***
- CI contains zero *("nonsignificant")*



CIs of the form $\hat{\theta}_i - \hat{\theta}_j \pm M_{ij}$

Which is just as it should be …



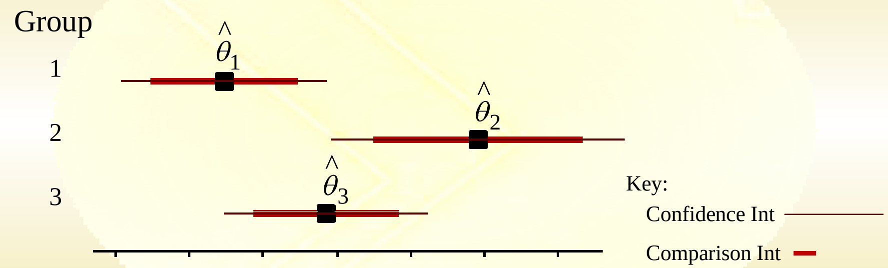
$M_{ij} \approx \tilde{M}_{ij} = h_i + h_j$

$$d_L = \hat{\theta}_i - h_i - \left(\hat{\theta}_j + h_j\right) = \hat{\theta}_i - \hat{\theta}_j - \tilde{M}_{ij} = \tilde{CI}_L$$

$$d_U = \hat{\theta}_i + h_i - \left(\hat{\theta}_j - h_j\right) = \hat{\theta}_i - \hat{\theta}_j + \tilde{M}_{ij} = \tilde{CI}_U$$

When parameters themselves meaningful

(& not just the differences)



What is going on here? (10 secs)

TABLE 10.3.1 Increase in Reading Age

Both:	0.1	3.2	4.3	-0.5	1.9	3.3	2.5	3.6	0.4	2.3	-1.4	-0.7
	-0.1	0.2	0.4	0.9	1.2	1.4	1.8	1.8	2.4	3.1		
Map Only:	1.0	-0.5	1.0	0.6	0.6	1.0	1.0	-1.4	2.2	3.6	3.1	2.6
Scan Only:	1.0	3.3	1.4	-0.9	1.0	0.0	0.6					
Neither:	-0.3	-1.3	1.6	-0.4	-0.7	0.6	-1.8	-2.0	-0.7			

Kindly provided by Mary Matthews, Carmel College.

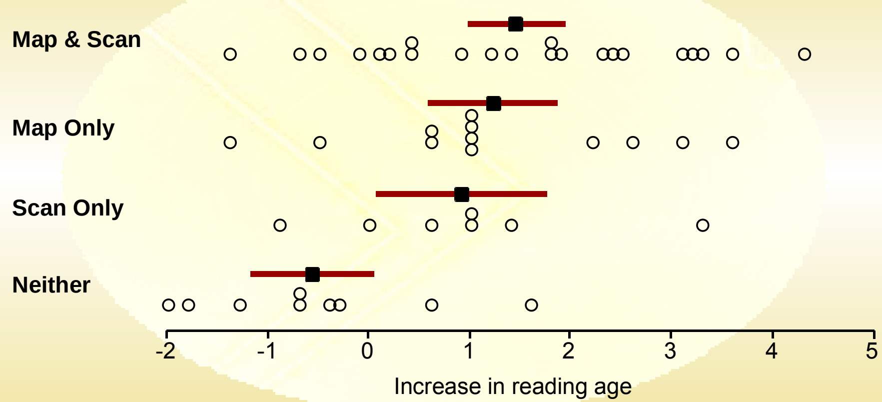
One-way Analysis of Variance

```
Analysis of Variance for Increase      F-statistic  P-value
Source    DF       SS       MS        F        P
Grp        3    27.06     9.02     4.45    0.008      } Anova Table
Error     46    93.35     2.03
Total     49   120.41

Level      N    Mean   StDev
MapScan   22   1.459   1.544
MapOnly   12   1.233   1.441
ScanOnly   7   0.914   1.302
Neither    9  -0.556   1.135

Pooled StDev =    1.425
```

Figure 10.3.2 Minitab analysis of variance output for reading ages

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

vs … What is going on here? (10 secs)



What is going on here?

TABLE 11.2.7 Melanoma Data (reproduces Table 11.2.1)

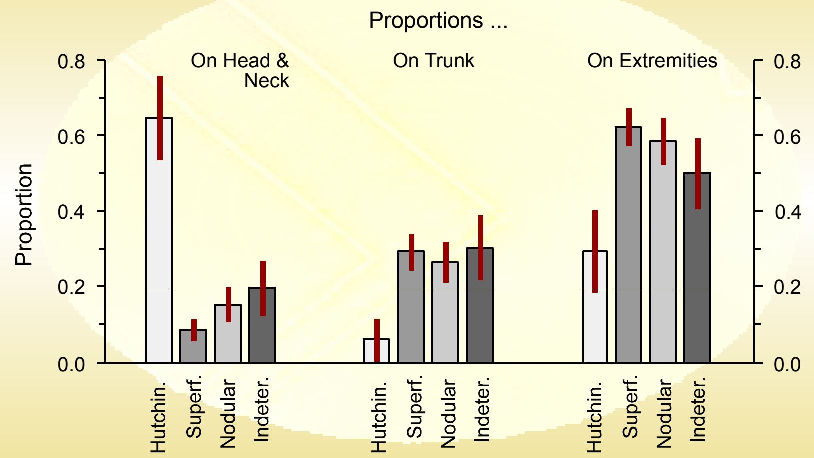
Type	SITE: Head and neck	Trunk	Extremities	Row totals
Hutchinson's	22	2	10	34
Superficial	16	54	115	185
Nodular	19	33	73	125
Indeterminate	11	17	28	56
Column Totals	68	106	226	400

Chi-Square Test

```
Expected counts are printed below observed counts
              Head & N    Trunk   Extremit    Total
Hutchinson's 1     22        2         10        34
                 5.78     9.01      19.21
Superficial  2     16       54        115       185
                31.45    49.03     104.53
Nodular      3     19       33         73       125
                21.25    33.13      70.62
Indeterminate 4    11       17         28        56
                 9.52    14.84      31.64
Total              68      106        226       400
Chi-Sq = 45.517 +  5.454 +  4.416 +
          7.590 +  0.505 +  1.050 +
          0.238 +  0.000 +  0.080 +
          0.230 +  0.314 +  0.419 = 65.813
DF = 6, P-Value = 0.000
```

Figure 11.2.5 Minitab output for the melanoma data.

vs … What is going on here?



Benefits of visual comparison intervals

- Can see what the main stories are almost instantaneously
  - Both significance and effect-size (albeit approximate)
    - Can all happen as annotations on the most obvious plot of the data
      - thus reducing abstraction
  - Can dig for salient details subsequently

But how can we get the h_i's ??

Have $I(I-1)/2$ MoE's to approx …

$\hat{\theta}_i - h_i$ $\quad \hat{\theta}_i \quad$ $\hat{\theta}_i + h_i$

Let's do for I=4 …

$$M_{12} \approx h_1 + h_2$$
$$M_{13} \approx h_1 + h_3$$
$$M_{14} \approx h_1 + h_4$$
$$M_{23} \approx h_2 + h_3$$
$$M_{24} \approx h_2 + h_4$$
$$M_{34} \approx h_3 + h_4$$

$$\rightarrow \begin{bmatrix} M_{12} \\ M_{13} \\ M_{14} \\ M_{23} \\ M_{24} \\ M_{34} \end{bmatrix} \approx \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix}$$

$$\mathbf{Y} \approx \mathbf{X}\beta$$

Simple Least Squares problem !

"But these are just approximations"

- Always exact for $I \leq 3$
  - (includes comparing sets of 3 within subgroups)
- Usually works remarkably well
  - Additionally we ***know the answers***, so …
    - easy to flag the occasional problem comparison:
      - ***"Significance"*** conflict betw. actual and approx.
      - Notable ***length*** misrepresentation betw. actual and approx.

Danny Chang packaged into R library

- **Includes** grabbing the right components of coeff vector & covariance matrix from a model fit, calculating MoE's for diffs and then …
  - **lm()**
  - **glm()**
  - **lm()**
  - **polr()**
  - **coxph()**

Origins: CW's Teaching notes at Auckland early 90s

Same forms of graphics but …

Approximate $M_{ij} = t_{df}\,\text{se}\left(\hat{\theta}_1 - \hat{\theta}_2\right)$

by $\tilde{M}_{ij} = h_i + h_j$ where $h_i = t_{df}\dfrac{\text{se}\left(\hat{\theta}_i\right)}{\sqrt{2}}$

i.e., approximate $\text{se}\left(\hat{\theta}_1 - \hat{\theta}_2\right) = \sqrt{\text{se}\left(\hat{\theta}_1\right)^2 + \text{se}\left(\hat{\theta}_2\right)^2}$

by $\dfrac{1}{\sqrt{2}}\text{se}\left(\hat{\theta}_1\right) + \dfrac{1}{\sqrt{2}}\text{se}\left(\hat{\theta}_2\right)$

So ***independence case only*** but works quite well so long as the std errors not too different (up to a factor of 5 gives err under 10%)

Also related to work on Quasi variances

Menezes (1999), Firth (2000), Firth and Menezes (2003, 2004) – R package qvcalc

Basic idea $v_{ij} \approx q_i + q_j$

where v_{ij} is the variance of the simple contrast $\hat{\beta}_i - \hat{\beta}_j$. The q_i's are estimated by ML using the model

$$\log v_{ij} \sim N(\mu_{ij}, \sigma^2)$$

where

$$\exp(\mu_{ij}) = q_i + q_j$$