UNSW SYDNEY | Australia's Global University

Analysis of Pretest-Posttest Data
It's not as straightforward as you might think!

Nancy Briggs
Stats Central
Mark Wainwright Analytical Centre
February 14, 2019

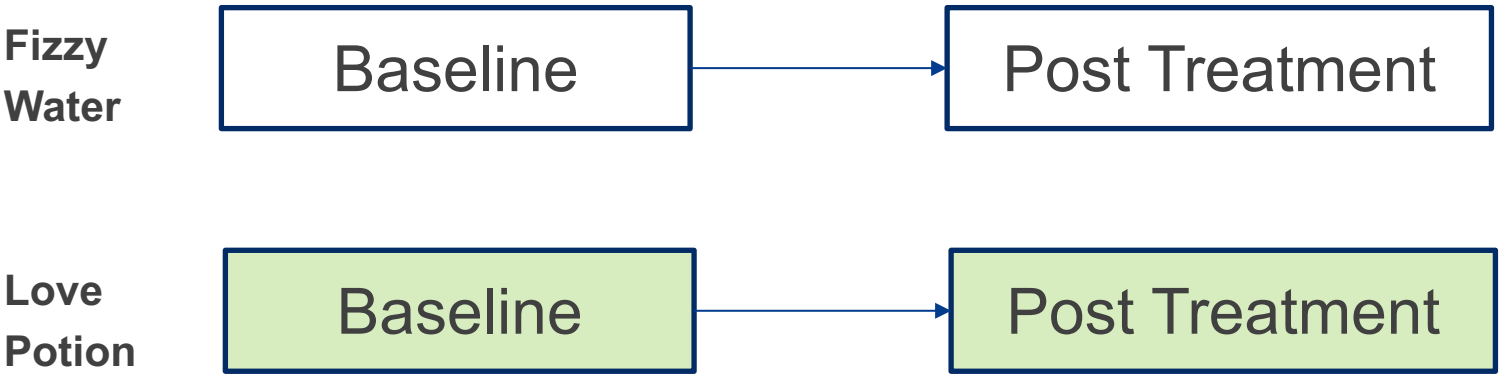
Pre-Post Studies

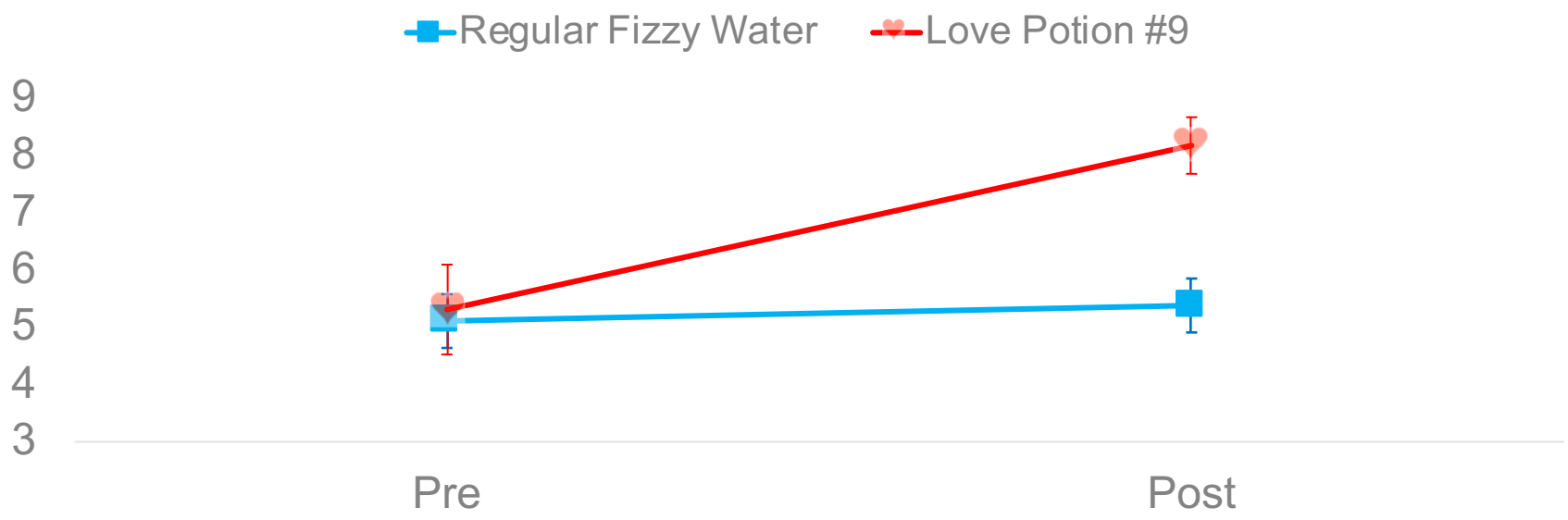
- Research Question: Does an intervention affect an outcome?

I am a researcher interested in the effect of type of drink on a validated, well-researched Valentine's Spirit Score. I think that drinking Love Potion #9 results in different scores than drinking Fizzy Water.

- What's a good research design to answer the question?
- Pre-post design!

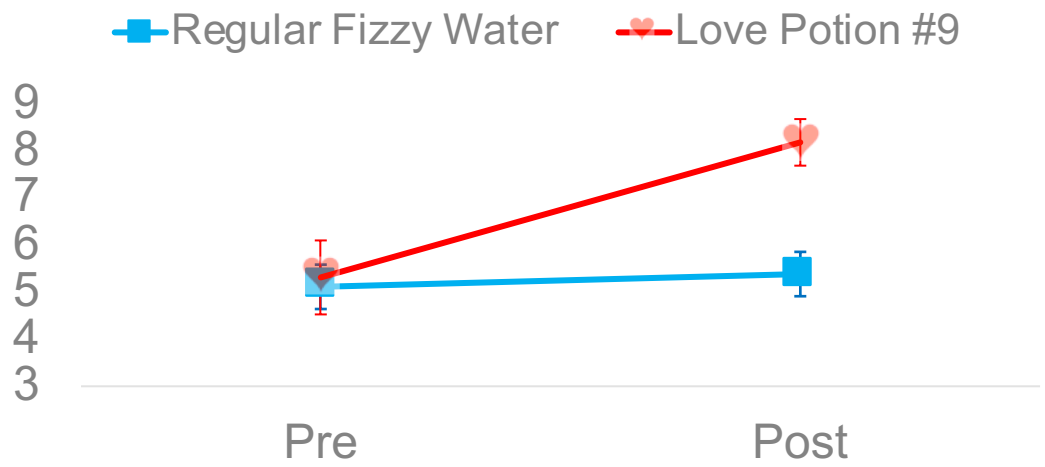
Pre-Post, 2 Groups Design





Options for Analysis

- Repeated Measures Analysis of Variance (RM-ANOVA)
  - 2 factors (1 between, 1 within)
  - Raw data



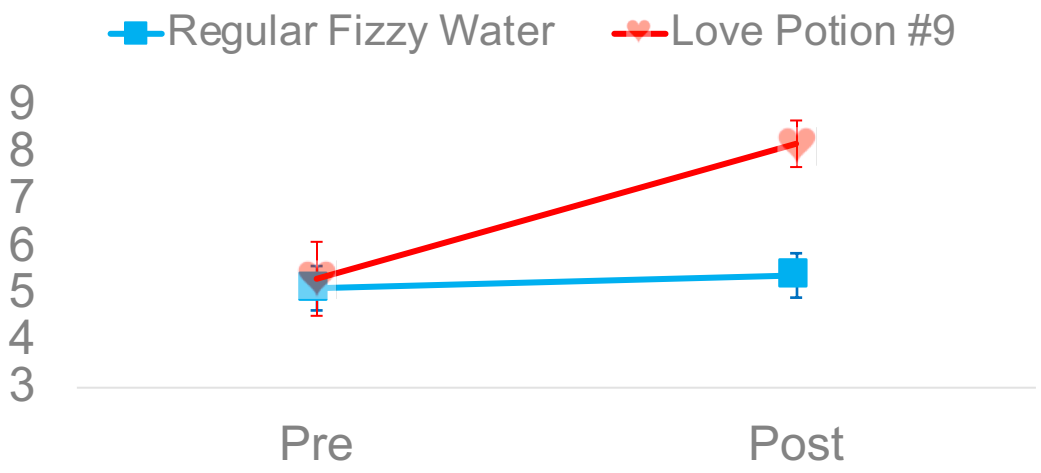
- Change score

$$diff = post - pre$$

  - T-test on the difference
- Analysis of Covariance (ANCOVA)
  - Analyse Post score, using Pre score and treatment as predictors

RM-ANOVA

- Raw data
- The focus here would be the interaction term between the group factor and time.
- Are the lines parallel?



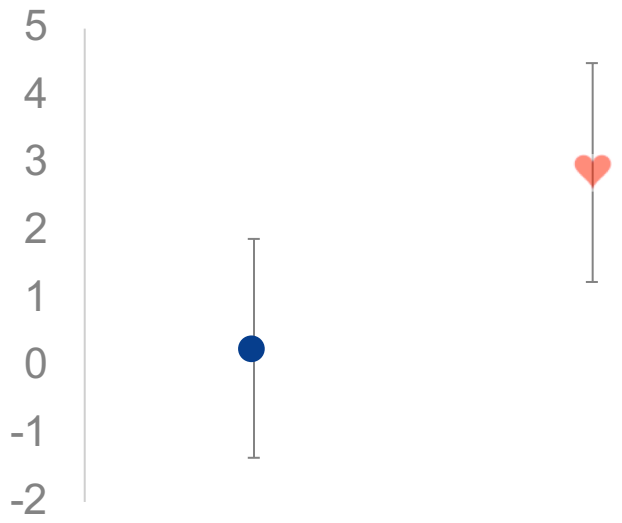
RM-ANOVA

```
rmaov.valentines <- aov(out~group + time + group*time +
                        Error(id), data = data)
summary(rmaov.valentines)

##
## Error: id
##           Df Sum Sq Mean Sq F value   Pr(>F)
## group      1  45.00   45.00   35.97 5.74e-07 ***
## Residuals 38  47.54    1.25
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##             Df Sum Sq Mean Sq F value   Pr(>F)
## time         1  57.36   57.36  151.09 8.21e-15 ***
## group:time   1  33.80   33.80   89.03 1.67e-11 ***
## Residuals   38  14.43    0.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Change Score

- Also called Gain Score
- Independent Samples T-test for the difference in mean change between groups.

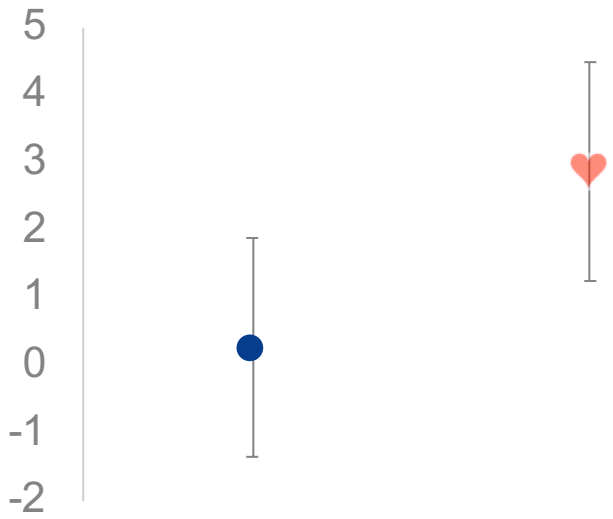
$$diff = post - pre$$



Change Scores

```
##        pre       group     post       diff
## 1 3.797986 Fizzy Water 5.116964  1.3189774
## 2 4.567972 Fizzy Water 5.750108  1.1821363
## 3 4.039042 Fizzy Water 4.513607  0.4745650
## 4 7.176230 Fizzy Water 5.586878 -1.5893517
## 5 4.775449 Fizzy Water 5.795275  1.0198257
## 6 4.117518 Fizzy Water 4.461389  0.3438713
```

Change Score




```
t.test(diff$diff~diff$group,var.equal=T)

##
##  Two Sample t-test
##
## data:  diff$diff by diff$group
## t = -9.4357, df = 38, p-value = 1.673e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.157822 -2.042178
## sample estimates:
## mean in group Fizzy Water mean in group Love Potion
##                 0.3935285                 2.9935285
```

Change Score

- So, what's the problem here?
- This t-test is actually mathematically equivalent to the F-statistic from the interaction term in the RM–ANOVA.
  - You are essentially doing the SAME ANALYSIS.

```
t.test(diff$diff~diff$group,var.equal=T)
```

```
## t = -9.4357, df = 38, p-value = 1.673e-11
```

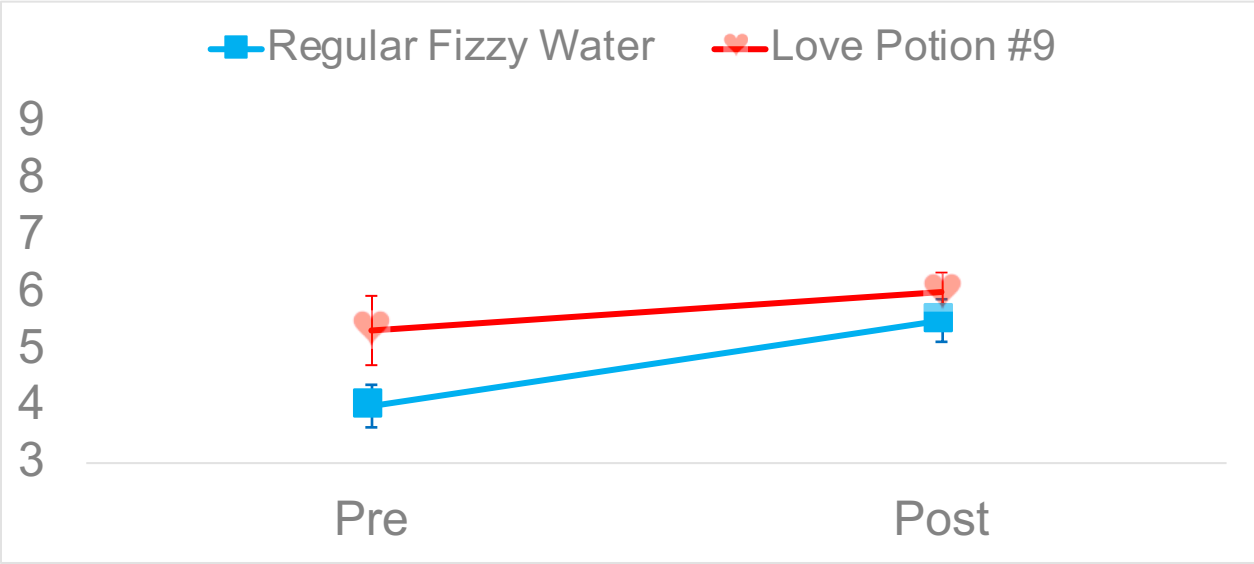
```
rmaov.valentines <- aov(out~group + time + group*time +
                        Error(id), data = data)
summary(rmaov.valentines)
               Df Sum Sq Mean Sq F value   Pr(>F)
## group:time  1  33.80   33.80   89.03 1.67e-11 ***
```

Change Score

- Gain scores may not be reliable
- Estimate of the treatment effect can be biased.
- Less powerful than ANCOVA for assessing treatment effect

RM-ANOVA & Change Score

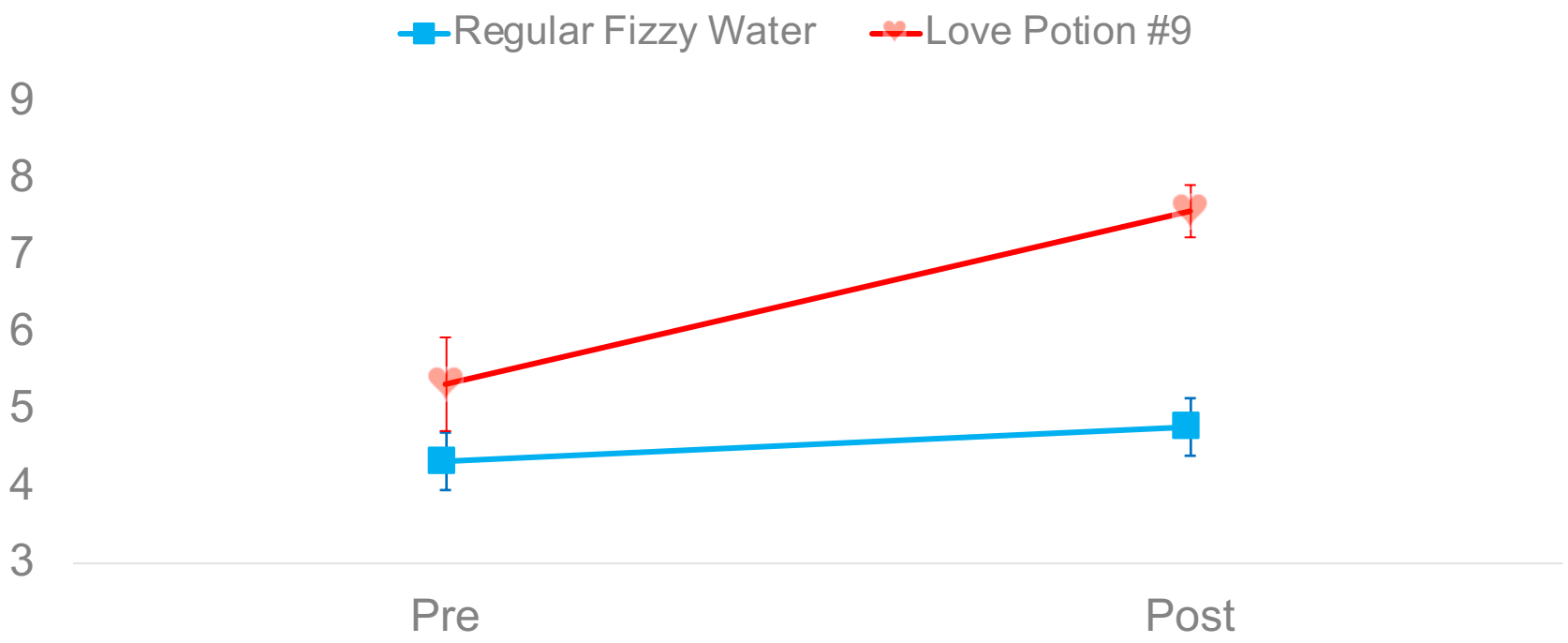
- BUT:
- Assumptions of RM-ANOVA
  - Randomisation of factors
    - Time is not randomised here
    - What if you don't have a random allocation of people to groups?



ANCOVA

- Outcome is the post measure
- Pre-test measure is a predictor (covariate)
- Term of interest is the group effect.

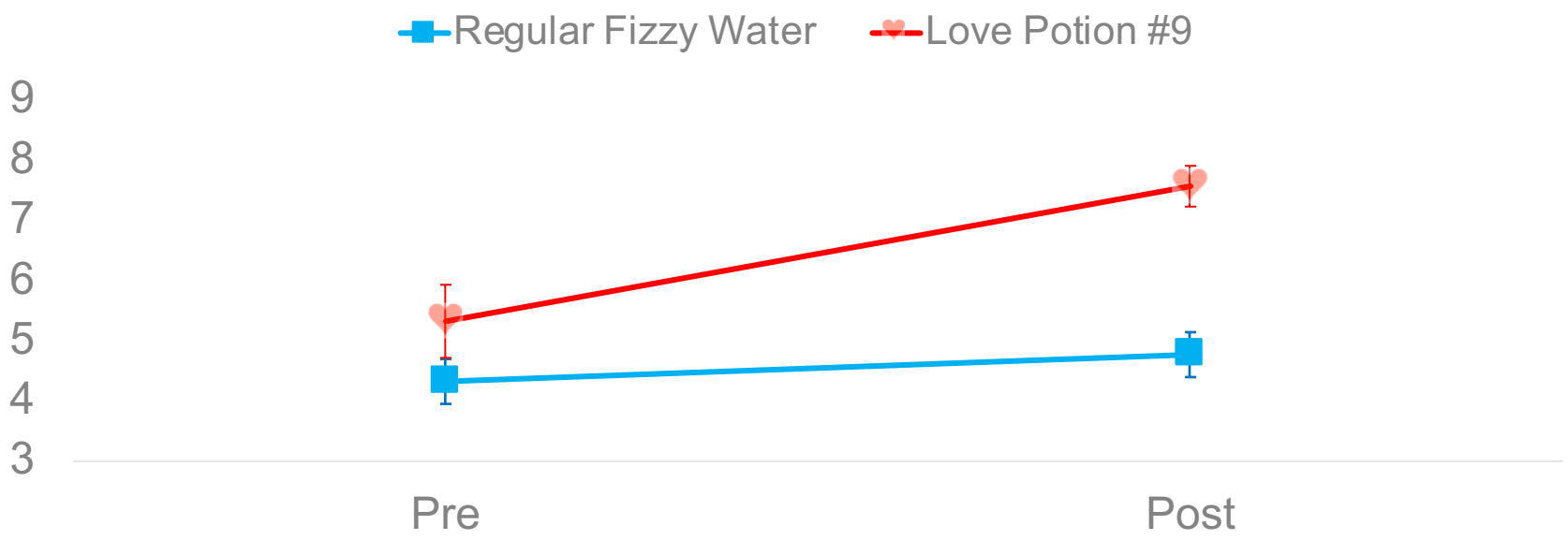
$$post = pre + group$$



ANCOVA results

```
ancova<-aov(diff$post ~ diff$pre + diff$group)
summary(ancova)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## diff$pre      1  13.24   13.24    27.5 6.62e-06 ***
## diff$group    1  72.63   72.63   150.9 1.27e-14 ***
## Residuals    37  17.81    0.48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



What Can Happen When Pre Scores Differ?

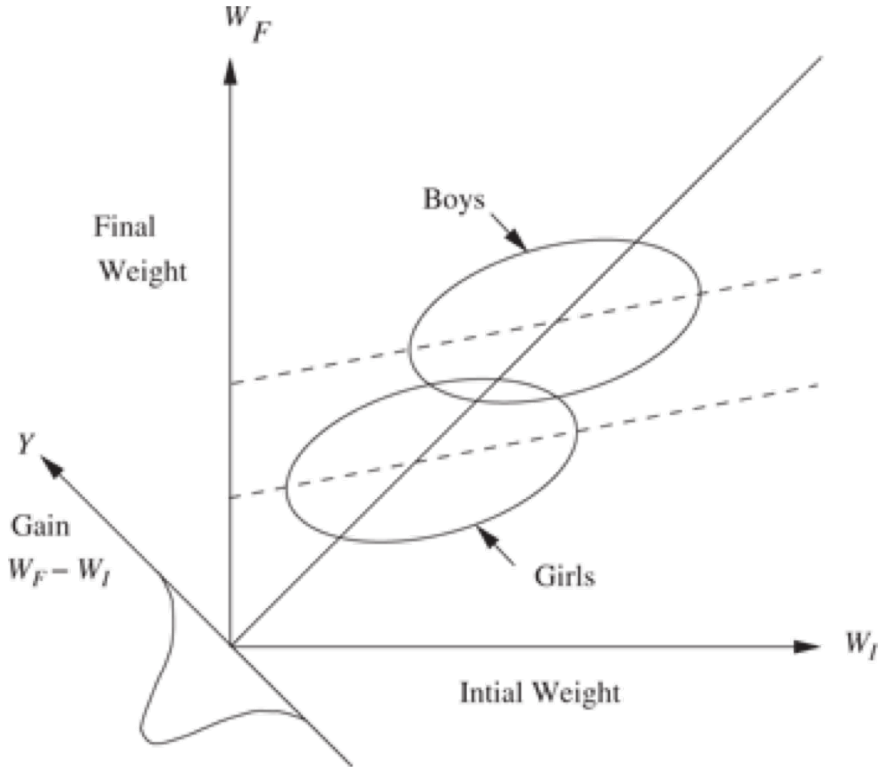
```
mean(data$pre[data$group=="Fizzy Water"])

## [1] 4.168234

mean(data$pre[data$group=="Love Potion"])

## [1] 5.168234

t.test(data$diff~data$group,var.equal=T)

##  Two Sample t-test
##
## data:  data$diff by data$group
## t = 3.6291, df = 38, p-value = 0.0008341

ancova2<-aov(data$post ~ data$pre +  data$group)
summary(ancova2)

##             Df Sum Sq Mean Sq F value  Pr(>F)
## data$pre     1  5.873   5.873  12.198 0.00126 **
## data$group   1  1.601   1.601   3.326 0.07630 .
## Residuals   37 17.813   0.481
```

What's going on?

- t-test and ANCOVA are asking different questions:
- t-test:
  - Is there a difference between the groups in mean change?
- ANCOVA
  - Do post-test means differ between groups, after having adjusted for pre-test scores?

Lord's Paradox

- Sometimes an unadjusted adjusted analysis will totally lead you to the incorrect conclusion.
- For studies with non-randomised groups.



Things to Consider

- Research Design
  - Experimental vs. Observational
    - Lack of random assignment to treatment groups means that a t-test on change scores (or the repeated measures ANOVA) is likely not what you want.
- What is your actual research question?
  - Do you want to assess the difference in the groups on the change itself? This research question is about growth or gains.
  - Do you want to examine post-treatment difference, after accounting for where people are starting out? This research question is about a treatment effect.
- If you have any missing data in the outcome, neither of these approaches is recommended!

Things to Consider

- One statistician even suggests that the only always correct way to analyse these data is graphically:

The only procedure that is always correct in this situation is a scatterplot comparing the scores at time 2 with those at time 1 for the different groups. In most cases you should analyse the data in several ways. If the approaches give different results ... think more carefully about the model implied by each.

Wright (2001)

References

Hand, D. J. (1994). Deconstructing statistical questions. Journal of the Royal Statistical Society: A, 157, 317–356.

Knapp, T & Schafer. W. (2009). From Gain Score t to ANCOVA F (and vice versa). Practical Assessment, Research & Evaluation, 14(6).

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. Psychological Bulletin, 72, 304–305.

Senn, S. (2006). Change from baseline and analysis of covariance revisited. Statistics in Medicine, 25, 4334-4344.

Wainer, H. (1991). Adjusting for differential base rates: Lord's paradox again. Psychological Bulletin, 109, 147–151.

Wright, D. (2003). Making friends with your data: Improving how statistics are conducted and reported. British Journal od Educational Psychology 73, 123-136.

Thank You!


0 0 1 1 1 1 1 0 1 1 1 1 0 1 1 1 0 0 0 0 0
0 1 1 1 1 0 1 0 0 0 1 1 0 1 1 0 0 0 1 0 1
1 0 1 1 1 0 1 1 0 0 0 1 1 0 1 1 0 0 0 1 0
0 0 0 1 0 1 0 0 0 1 0 0 1 1 0 0 0 1 0 0 1
1 1 1 1 1 0 0 0 1 0 0 1 1 0 0 1 0 0 1 1 0
0 1 1 1 1 0 1 0 1 0 1 0 1 0 0 1 0 1 0 1 1
1 1 1 0 0 0 1 1 0 1 0 1 1 1 0 0 0 1 0 0 0
1 0 0 0 0 1 1 1 0 0 1 0 0 0 1 0 1 0 0 0 1
1 1 1 0 0 0 1 1 0 0 1 0 0 0 0 1 0 0 0 0 0
0 0 0 0 1 1 0 0 0 1 0 0 0 1 0 1 0 0 1 0 1
1 1 1 1 1 0 1 0 0 1 0 0 0 0 0 1 1 0 1 1 1
0 0 1 0 1 1 1 1 0 1 1 0 0 0 0 0 1 0 0 1 0
0 0 1 0 1 1 0 1 1 1 0 0 0 0 1 1 0 1 1 1 1
1 1 0 1 1 1 1 1 1 0 1 1 0 0 0 0 0 0 1 0 0
1 0 1 1 0 1 1 0 1 0 0 0 1 0 1 0 0 1 1 0 0
0 1 1 1 0 0 0 1 0 0 0 0 0 1 0 1 0 0 0 0 1
1 1 0 1 1 1 0 1 0 0 0 1 0 1 1 1 0 1 1 1 1
1 1 0 0 0 1 1 0 0 1 1 1 0 1 1 0 0 1 1 1 1
1 1 1 0 1 0 0 1 0 1 1 1 1 0 0 1 0 1 1 0 1
0 0 0 0 0 0 1 1 0 0 1 0 1 1 0 0 0 1 1 1 1
1 1 0 1 1 1 0 1 0 0 1 0 1 1 0 1 1 1 0 1 0
0 1 1 1 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1