Lord’s paradox

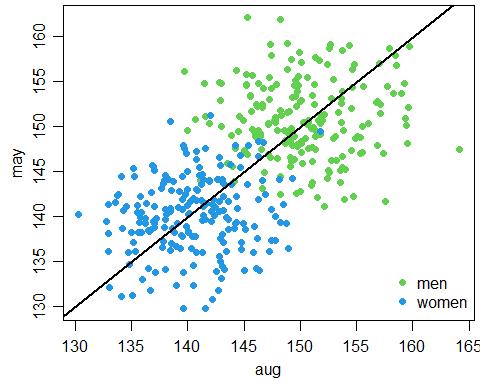
Philip Dixon

2023-03-22

### “Study” description.

Students in a university dining hall. Want to evaluate effect of the dining hall diet on men and women. Is it the same? Collect weight in August (start of the year) and again in May (end of the year). 200 men and 200 women

### Plot of all data, with line of no change pre -> post



### Analysis of Differences

## Analysis of Variance Table  
##   
## Response: diff  
## Df Sum Sq Mean Sq F value Pr(>F)  
## sex 1 67.2 67.229 2.0137 0.1567  
## Residuals 398 13287.5 33.386

## sex emmean SE df lower.CL upper.CL  
## men 0.531 0.409 398 -0.272 1.335  
## women -0.289 0.409 398 -1.092 0.515  
##   
## Confidence level used: 0.95

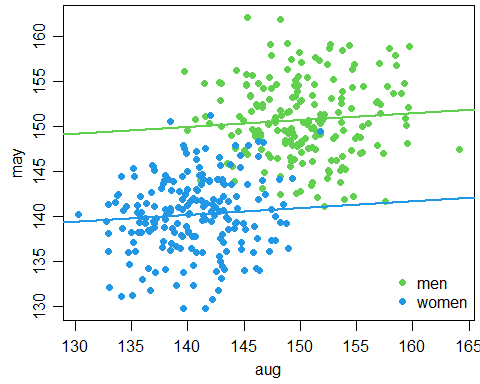
## contrast estimate SE df t.ratio p.value  
## men - women 0.82 0.578 398 1.419 0.1567

### ANCOVA, using aug value as covariate

## sex emmean SE df lower.CL upper.CL  
## men 150 0.388 397 150 151  
## women 141 0.388 397 140 141  
##   
## Confidence level used: 0.95

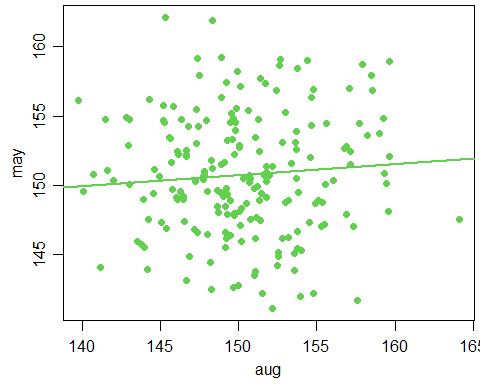
##   
## Call:  
## lm(formula = may ~ aug + sex, data = food)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.5187 -2.8354 -0.0872 3.0011 11.7404   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 139.15907 7.57481 18.371 <2e-16 \*\*\*  
## aug 0.07727 0.05038 1.534 0.126   
## sexwomen -9.78746 0.64894 -15.082 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.259 on 397 degrees of freedom  
## Multiple R-squared: 0.6075, Adjusted R-squared: 0.6056   
## F-statistic: 307.3 on 2 and 397 DF, p-value: < 2.2e-16

### Plot of all data with ANCOVA fitted lines



### Why are the within group slopes close to zero?

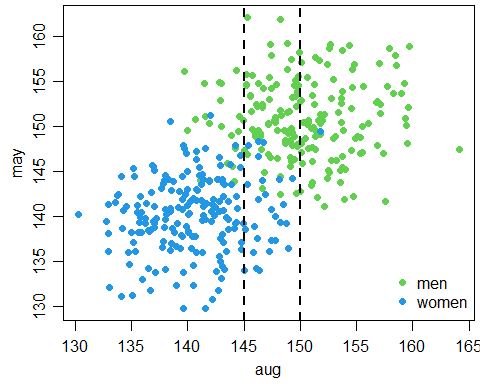
#### Look only at the men and their fitted line



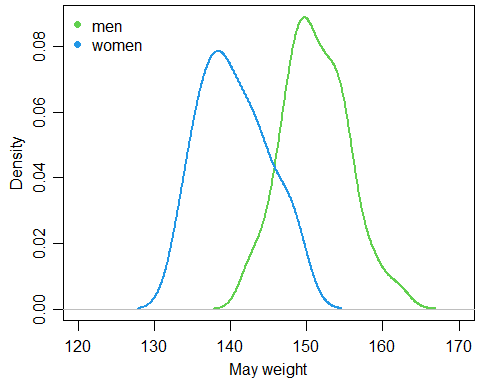
#### The paradox arises because aug is a terrible predictor of may within men (and also within women), slope close to 0

### What happens when you look only at a narrow range of initial values?

#### The vertical fences identify the subset of aug between 145 and 150

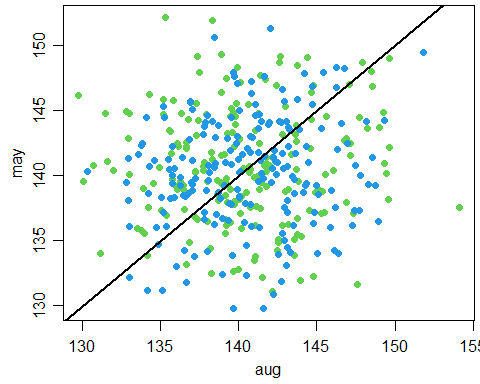


### Smoothed histograms of may values for men and women with aug between 145 and 150



### Does the paradox still exist when men and women are similar at baseline?

#### shift all men values (may and aug) down by 10



### Analysis of differences, when little diff at baseline

#### Average difference for each group

## Analysis of Variance Table  
##   
## Response: diff  
## Df Sum Sq Mean Sq F value Pr(>F)  
## sex 1 67.2 67.229 2.0137 0.1567  
## Residuals 398 13287.5 33.386

## sex emmean SE df lower.CL upper.CL  
## men 0.531 0.409 398 -0.272 1.335  
## women -0.289 0.409 398 -1.092 0.515  
##   
## Confidence level used: 0.95

## contrast estimate SE df t.ratio p.value  
## men - women 0.82 0.578 398 1.419 0.1567

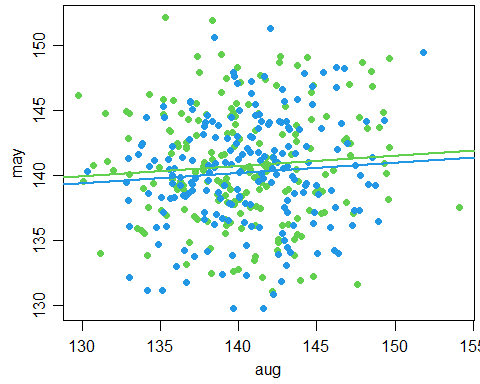
### ANCOVA, when little diff at baseline

#### mean may for individuals at average aug weight

## sex emmean SE df lower.CL upper.CL  
## men 140.8 0.3013 397 140.2 141.4  
## women 140.2 0.3013 397 139.6 140.8  
##   
## Confidence level used: 0.95

## contrast estimate SE df t.ratio p.value  
## men - women 0.56 0.426 397 1.314 0.1895

##   
## Call:  
## lm(formula = may ~ aug + sex, data = food3)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.5187 -2.8354 -0.0872 3.0011 11.7404   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 129.93175 7.07144 18.374 <2e-16 \*\*\*  
## aug 0.07727 0.05038 1.534 0.126   
## sexwomen -0.56013 0.42616 -1.314 0.189   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.259 on 397 degrees of freedom  
## Multiple R-squared: 0.009852, Adjusted R-squared: 0.004863   
## F-statistic: 1.975 on 2 and 397 DF, p-value: 0.1401



### Comparison of difference and ANCOVA analyes

#### Estimated differences are similar (0.82 for difference, 0.56 for ANCOVA)

#### SE of the difference is substantially larger for difference (se = 0.58) than for ANCOVA (se = 0.43)

#### Consequences of the se for design: A study using differences would require 80% more (almost double) the number of individuals than a study using ANCOVA.

#### Best way to improve the study: Find a way to more accurately measure weight (more precise measurement, less subject to daily / weekly variability)