450A Political Methodology I
TA Section 3

Vincent Bauer and William Marble

October 13, 2016

Road Map

1. Multinomial Distribution
2. Multinomial Goodness of Fit
3. Paired t-test, blocking, research design (most of the slides)

   R code

   3.1 ggplot
   3.2 reshape

Multinomial Distribution

Most of our hypotheses tests have compared proportions from a binary outcome variable, i.e. support for Hillary Clinton. But the homework introduced a more complicated scenario, a categorical outcome variable, i.e. preferences between four trade deals.

Multinomial Distribution

Most of our hypotheses tests have compared proportions from a binary outcome variable, i.e. support for Hillary Clinton. But the homework introduced a more complicated scenario, a categorical outcome variable, i.e. preferences between four trade deals.

The binary outcome has been implicitly assuming a binomial distribution, which is a special case of the multinomial distibution with $k = 2$ outcomes.

Multinomial Distribution

Most of our hypotheses tests have compared proportions from a binary outcome variable, i.e. support for Hillary Clinton. But the homework introduced a more complicated scenario, a categorical outcome variable, i.e. preferences between four trade deals.

The binary outcome has been implicitly assuming a binomial distribution, which is a special case of the multinomial distibution with $k = 2$ outcomes.

A multinomial trials process is a sequence of independent, identically distributed random variables each taking k possible values.

Multinomial Distribution

The expected number of times the outcome i was observed over n trials is

$$E[X_i] = n \cdot p_i$$

Multinomial Distribution

The expected number of times the outcome i was observed over n trials is

$$E[X_i] = n \cdot p_i$$

Similarly, the variance of the expected number of times we observe outcome i is stored in a diagonal matrix.

$$Var[X_i] = n \cdot p_i(1 - p_i)$$

Multinomial Distribution

The expected number of times the outcome i was observed over n trials is

$$E[X_i] = n \cdot p_i$$

Similarly, the variance of the expected number of times we observe outcome i is stored in a diagonal matrix.

$$Var[X_i] = n \cdot p_i(1 - p_i)$$

The off-diagonal entries are the covariances between outcomes, and are negative because increasing the probability of one outcomes decreases the probability of other outcomes.

$$Cov[X_i, X_j] = -n \cdot p_i \cdot p_j$$

Multinomial Distribution

The expected number of times the outcome i was observed over n trials is

$$E[X_i] = n \cdot p_i$$

Similarly, the variance of the expected number of times we observe outcome i is stored in a diagonal matrix.

$$Var[X_i] = n \cdot p_i(1 - p_i)$$

The off-diagonal entries are the covariances between outcomes, and are negative because increasing the probability of one outcomes decreases the probability of other outcomes.

$$Cov[X_i, X_j] = -n \cdot p_i \cdot p_j$$

If we care about proportions and not the number of times, we can drop the n's.

Multinomial Goodness of Fit

What if we wanted to check whether two sets of statistics came from the same population.

Multinomial Goodness of Fit

What if we wanted to check whether two sets of statistics came from the same population.

Say that the university released the following statistics about smoking rates on campus but we are skeptical so we decide to run our own survey.

Multinomial Goodness of Fit

What if we wanted to check whether two sets of statistics came from the same population.

Say that the university released the following statistics about smoking rates on campus but we are skeptical so we decide to run our own survey.

Table: University Statistics

Heavy	Regular	Occasional	Never
4.5%	7.5%	8.5%	79.5%

The university's numbers are our expected outcomes.

Multinomial Goodness of Fit

```r
library(MASS) #the data lives here
library(dplyr)  #pipe operator
library(xtable)  #make a table

#school statistics
smoke.prob = c(.045, .795, .085, .075)

#survey results
smoke.freq = table(survey$Smoke) %>% prop.table() %>% round(3)*100
xtable(smoke.freq)
```

	V1
Heavy	4.70
Never	80.10
Occas	8.10
Regul	7.20

Multinomial Goodness of Fit

We want to test whether the results of our sample and the statistics presented by the university could have come from the same population.

Multinomial Goodness of Fit

We want to test whether the results of our sample and the statistics presented by the university could have come from the same population.

We use a Chi-squared test for independence (Pearson's)

$$X^2 = \sum_k \frac{(f_k - e_k)^2}{e_k}$$

Multinomial Goodness of Fit

We want to test whether the results of our sample and the statistics presented by the university could have come from the same population.

We use a Chi-squared test for independence (Pearson's)

$$X^2 = \sum_k \frac{(f_k - e_k)^2}{e_k}$$

	Heavy	Regular	Occassional	Never
Observed	4.7	7.2	8.1	80.1
Expected	4.5	7.5	8.5	79.5
	=	=	=	=
0.044 $\leftarrow$	0.009 +	0.012 +	0.019 +	0.005

Multinomial Goodness of Fit

```
suppressWarnings(chisq.test(smoke.freq, p=smoke.prob))

##
##  Chi-squared test for given probabilities
##
## data:  smoke.freq
## X-squared = 0.044097, df = 3, p-value = 0.9976
```

The p-value is greater than the .05 significance level so we cannot not reject the null hypothesis that the survey results and the campus-wide smoking statistics came from the same population.

Paired t-test, blocking, research design

Example: Boys' Shoes

Say you wanted to test the effectiveness of two materials for making shoes. Your outcome variable is the amount of wear on shoes made of material A and material B.

Example: Boys' Shoes

Say you wanted to test the effectiveness of two materials for making shoes. Your outcome variable is the amount of wear on shoes made of material A and material B. We therefore have two population parameters of interest.

$$\mu_A = E[Y_A]$$
$$\mu_B = E[Y_B]$$

Example: Boys' Shoes

Say you wanted to test the effectiveness of two materials for making shoes. Your outcome variable is the amount of wear on shoes made of material A and material B. We therefore have two population parameters of interest.

$$\mu_A = E[Y_A]$$
$$\mu_B = E[Y_B]$$

Our null hypothesis is that the wear on the shoes is the same, i.e. $H_0 : \mu_A = \mu_B$.

Example: Boys' Shoes

Say you wanted to test the effectiveness of two materials for making shoes. Your outcome variable is the amount of wear on shoes made of material A and material B. We therefore have two population parameters of interest.

$$\mu_A = E[Y_A]$$
$$\mu_B = E[Y_B]$$

Our null hypothesis is that the wear on the shoes is the same, i.e. $H_0 : \mu_A = \mu_B$. We have no prior beliefs about the strength of these materials so we test the two-sided our alternative hypothesis that the wears are not the same, i.e. $H_A : \mu_A \neq \mu_B$.

Complete Random Assignment

We only have 10 shoes of each type and we are concerned that if the difference between the quality of the shoes is not very large, then our comparisons will be polluted by noise from other differences and we will fail to find a difference where one actually exists (Type-2 Error).

Complete Random Assignment

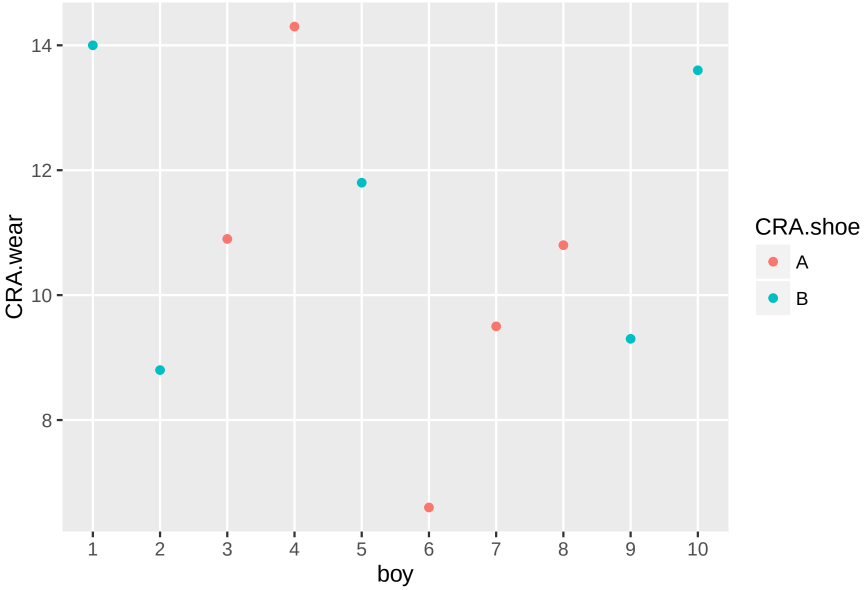
We only have 10 shoes of each type and we are concerned that if the difference between the quality of the shoes is not very large, then our comparisons will be polluted by noise from other differences and we will fail to find a difference where one actually exists (Type-2 Error).

- We think that boys and girls may have different wear patterns, so we control for gender and only allow boys into our study.
- But we're still worried about variation within boys, so we decide to try out this random assignment thing that seems to be all the rage.
- We call this experiment the "Complete Random Assignment" Design.

```
library(frt)
data(shoes)

set.seed(123)
#set up data for CRD
#the data is set up for the block design
shoeA <- sample(1:10,5, replace=FALSE) #randomly assign each boy to a treat
shoes$CRA.wear <- shoes$CRA.shoe <- NA
shoes$CRA.shoe[shoeA] <- "A"
shoes$CRA.shoe[-shoeA] <- "B"
shoes$CRA.wear <- ifelse(shoes$CRA.shoe=="A", shoes$matA, shoes$matB)  #get

shoes[1:5, c("boy", "CRA.shoe", "CRA.wear")]

##   boy CRA.shoe CRA.wear
## 1   1        B     14.0
## 2   2        B      8.8
## 3   3        A     10.9
## 4   4        A     14.3
## 5   5        B     11.8
```

```
library(ggplot2)
ggplot(shoes, aes(x=boy, y=CRA.wear, colour=CRA.shoe)) +
 geom_point() +
  scale_x_continuous(breaks = 1:10) +
    theme(panel.grid.minor = element_blank())
```

We run our experiment and obtain the following amount of wear per boy.



We run our two-sample t-test and its disappointing so we sit at our desk dejectedly. Notice that shoe B has higher wear on average but we cannot be sure that this is not due to chance.

```
t.test(shoes$CRA.wear[shoes$CRA.shoe=="A"],
       shoes$CRA.wear[shoes$CRA.shoe=="B"])

##
##  Welch Two Sample t-test
##
## data:  shoes$CRA.wear[shoes$CRA.shoe == "A"] and shoes$CRA.wear[shoes$CR
## t = -0.65881, df = 7.8271, p-value = 0.5289
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.874883  2.714883
## sample estimates:
## mean of x mean of y
##     10.42     11.50
```

Random Block design

Staring at the floor, we realize that people have two feet and that we can give each boy one of each kind of shoe.

Random Block design

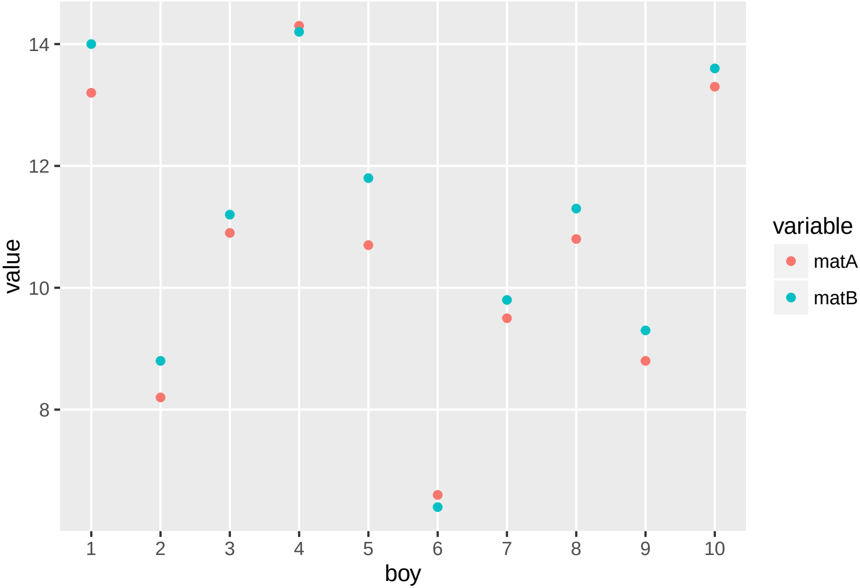
Staring at the floor, we realize that people have two feet and that we can give each boy one of each kind of shoe.

- We decide to give this randomization thing another try and randomize which shoe goes on which foot.
- We promise our boss that we'll find a difference this time and they make twenty more shoes for us and we give them to the same boys.
- We call this the "Random Block Design".

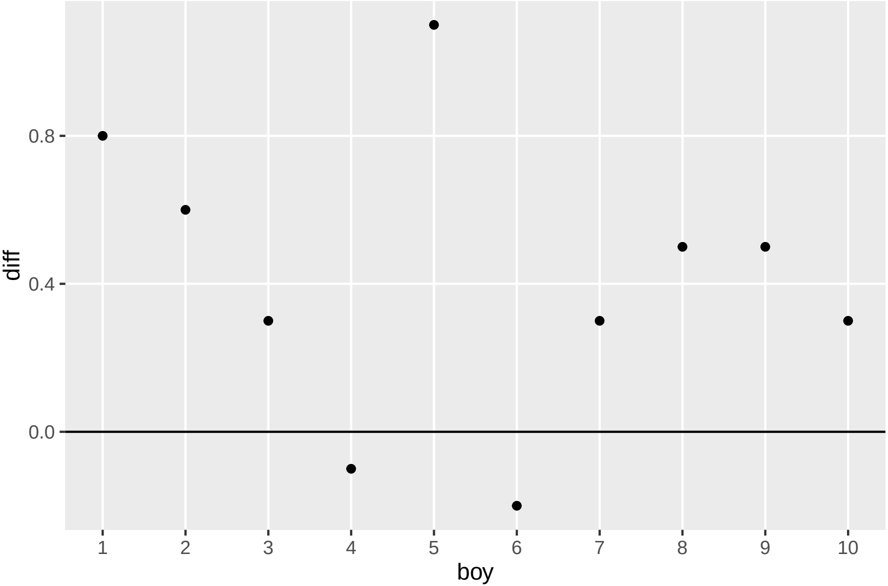
```
library(reshape)
shoes2 <- melt(shoes, id.vars=c("boy"),
               measure.vars=c("matA", "matB"))
shoes2[c(1:3,50, 17:20),]
```

```
##    boy variable value
## 1    1     matA  13.2
## 2    2     matA   8.2
## 3    3     matA  10.9
## NA  NA     <NA>    NA
## 17   7     matB   9.8
## 18   8     matB  11.3
## 19   9     matB   9.3
## 20  10     matB  13.6
```

This time a different pattern emerges. Notice that that differences between the boys is much greater than the differences between the shoes.



And we plot the differences themselves. Most boys show more wear on shoe B.



Then we run a two-sample t-test on these newly randomized shoes, and are disappointed again.

```
t.test(shoes$matB,shoes$matA)

##
##  Welch Two Sample t-test
##
## data:  shoes$matB and shoes$matA
## t = 0.36891, df = 17.987, p-value = 0.7165
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.925046  2.745046
## sample estimates:
## mean of x mean of y
##     11.04     10.63
```

Finally, we decide to pair these observations together.

```
t.test(shoes$matB,shoes$matA, paired=TRUE)

##
##  Paired t-test
##
## data:  shoes$matB and shoes$matA
## t = 3.3489, df = 9, p-value = 0.008539
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1330461 0.6869539
## sample estimates:
## mean of the differences
##                     0.41
```

We're curious what this test is doing, so we run a one-sample t-test on just the differences.

```
t.test(shoes$diff)


##
##  One Sample t-test
##
## data:  shoes$diff
## t = 3.3489, df = 9, p-value = 0.008539
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1330461 0.6869539
## sample estimates:
## mean of x
##      0.41
```

Analysis

Q: Why did we observe a statistically significant difference in the second test but not the first?

Analysis

Q: Why did we observe a statistically significant difference in the second test but not the first?

Intuitive answer:

- The children in our study varied a lot on how they used their shoes, while the differences between the materials was relatively small.
- Our first test did pick up higher wear on shoe B but there was too much variance. If we had had many more children then we could have made a conclusion about this difference.

Analysis

Q: Why did we observe a statistically significant difference in the second test but not the first?

Intuitive answer:

- The children in our study varied a lot on how they used their shoes, while the differences between the materials was relatively small.
- Our first test did pick up higher wear on shoe B but there was too much variance. If we had had many more children then we could have made a conclusion about this difference.

The math to figure out how many children we would need to detect this magnitude of change is called a 'power calculation'.

Let's think about this intuition more formally.

Let's think about this intuition more formally.

Wear (Y_A) is determined by 1) shoe material, 2) shoe usage, 3) other random factors

$$Y_A = \mu_A + \beta_i + \epsilon_{i,A}$$
$$Y_B = \mu_B + \beta_i + \epsilon_{i,B}$$

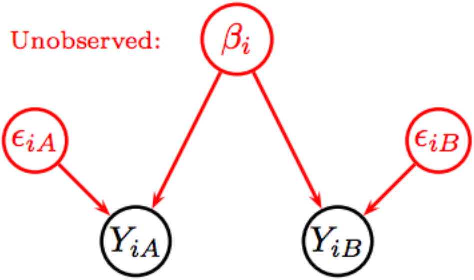
Let's think about this intuition more formally.

Wear (Y_A) is determined by 1) shoe material, 2) shoe usage, 3) other random factors

$$Y_A = \mu_A + \beta_i + \epsilon_{i,A}$$
$$Y_B = \mu_B + \beta_i + \epsilon_{i,B}$$

Once we put both shoes on the same boys, β_i is the same in both equations.

Graphically,



Lets show formally why the Complete Random design is failing to detect statistically significant results. There is no bias in this design, its just inefficient for our small sample size.

Lets show formally why the Complete Random design is failing to detect statistically significant results. There is no bias in this design, its just inefficient for our small sample size.

For simplicity we assume that the material is exactly the same in each shoe, i.e. $Var[\mu] \approx 0$, we assume this in most regressions.

But there is still variation between boys and other random variation.

$$\beta_i \sim N(0, \sigma_\beta^2)$$

$$\epsilon_{i,A/B} \sim N(0, \sigma_\epsilon^2/2)$$

Lets show formally why the Complete Random design is failing to detect statistically significant results. There is no bias in this design, its just inefficient for our small sample size.

For simplicity we assume that the material is exactly the same in each shoe, i.e. $Var[\mu] \approx 0$, we assume this in most regressions.

But there is still variation between boys and other random variation.

$$\beta_i \sim N(0, \sigma_\beta^2)$$

$$\epsilon_{i,A/B} \sim N(0, \sigma_\epsilon^2/2)$$

The $\frac{1}{2}$ is for convenience later and makes no difference to the proof.

What is the variance of our estimate Y_A?

What is the variance of our estimate Y_A?

$$Y_A = \mu_A + \beta_i + \epsilon_{i,A}$$

What is the variance of our estimate Y_A?

$$Y_A = \mu_A + \beta_i + \epsilon_{i,A}$$
$$Var[Y_A] = Var[\mu_A + \beta_i + \epsilon_{i,A}]$$

What is the variance of our estimate Y_A?

$$Y_A = \mu_A + \beta_i + \epsilon_{i,A}$$
$$Var[Y_A] = Var[\mu_A + \beta_i + \epsilon_{i,A}]$$

By assumption, $\beta \perp \epsilon$, and by design $\perp \mu$

What is the variance of our estimate Y_A?

$$Y_A = \mu_A + \beta_i + \epsilon_{i,A}$$
$$Var[Y_A] = Var[\mu_A + \beta_i + \epsilon_{i,A}]$$

By assumption, $\beta \perp \epsilon$, and by design $\perp \mu$

$$Var[Y_A] = 0 + Var[\beta_i] + Var[\epsilon_{i,A}]$$

What is the variance of our estimate Y_A?

$$Y_A = \mu_A + \beta_i + \epsilon_{i,A}$$
$$Var[Y_A] = Var[\mu_A + \beta_i + \epsilon_{i,A}]$$

By assumption, $\beta \perp \epsilon$, and by design $\perp \mu$

$$Var[Y_A] = 0 + Var[\beta_i] + Var[\epsilon_{i,A}]$$
$$Var[Y_A] = \sigma^2_\beta + \sigma^2_\epsilon/2$$

What is the variance of our estimate Y_A?

$$Y_A = \mu_A + \beta_i + \epsilon_{i,A}$$
$$Var[Y_A] = Var[\mu_A + \beta_i + \epsilon_{i,A}]$$

By assumption, $\beta \perp \epsilon$, and by design $\perp \mu$

$$Var[Y_A] = 0 + Var[\beta_i] + Var[\epsilon_{i,A}]$$
$$Var[Y_A] = \sigma^2_\beta + \sigma^2_\epsilon/2$$

So the SE in our two-sample t-test is actually

$$\sqrt{\frac{\sigma^2_\beta + \sigma^2_{\epsilon_A}/2}{N_A} + \frac{\sigma^2_\beta + \sigma^2_{\epsilon_B}/2}{N_B}}$$

What is the variance of our estimate Y_A?

$$Y_A = \mu_A + \beta_i + \epsilon_{i,A}$$
$$Var[Y_A] = Var[\mu_A + \beta_i + \epsilon_{i,A}]$$

By assumption, $\beta \perp \epsilon$, and by design $\perp \mu$

$$Var[Y_A] = 0 + Var[\beta_i] + Var[\epsilon_{i,A}]$$
$$Var[Y_A] = \sigma_\beta^2 + \sigma_\epsilon^2/2$$

So the SE in our two-sample t-test is actually

$$\sqrt{\frac{\sigma_\beta^2 + \sigma_{\epsilon_A}^2/2}{N_A} + \frac{\sigma_\beta^2 + \sigma_{\epsilon_B}^2/2}{N_B}}$$

We can tell graphically that $\sigma_\beta^2 \gg \sigma_\epsilon^2/2$

There is nothing we can do about $\sigma_\epsilon^2/2$ but ideally we would like to minimize σ_β^2 because it is unrelated to the shoes.

Comparing variation within boys controls for σ_β^2.

$$D_i = Y_A - Y_B$$

There is nothing we can do about $\sigma_\epsilon^2/2$ but ideally we would like to minimize σ_β^2 because it is unrelated to the shoes.

Comparing variation within boys controls for σ_β^2.

$$
\begin{aligned}
D_i &= Y_A - Y_B \\
&= (\mu_A + \beta_i + \epsilon_{i,A}) - (\mu_B + \beta_i + \epsilon_{i,B})
\end{aligned}
$$

There is nothing we can do about $\sigma_\epsilon^2/2$ but ideally we would like to minimize σ_β^2 because it is unrelated to the shoes.

Comparing variation within boys controls for σ_β^2.

$$
\begin{aligned}
D_i &= Y_A - Y_B \\
&= (\mu_A + \beta_i + \epsilon_{i,A}) - (\mu_B + \beta_i + \epsilon_{i,B}) \\
&= (\mu_A + \mu_B) + (\beta_i - \beta_i) + (\epsilon_{i,A} - \epsilon_{i,B})
\end{aligned}
$$

There is nothing we can do about $\sigma_\epsilon^2/2$ but ideally we would like to minimize σ_β^2 because it is unrelated to the shoes.

Comparing variation within boys controls for σ_β^2.

$$\begin{aligned} D_i &= Y_A - Y_B \\ &= (\mu_A + \beta_i + \epsilon_{i,A}) - (\mu_B + \beta_i + \epsilon_{i,B}) \\ &= (\mu_A + \mu_B) + (\beta_i - \beta_i) + (\epsilon_{i,A} - \epsilon_{i,B}) \\ Var[D_i] &= 0 + 0 + Var[\epsilon_{i,A} - \epsilon_{i,B}] \end{aligned}$$

There is nothing we can do about $\sigma_\epsilon^2/2$ but ideally we would like to minimize σ_β^2 because it is unrelated to the shoes.

Comparing variation within boys controls for σ_β^2.

$$\begin{aligned}
D_i &= Y_A - Y_B \\
&= (\mu_A + \beta_i + \epsilon_{i,A}) - (\mu_B + \beta_i + \epsilon_{i,B}) \\
&= (\mu_A + \mu_B) + (\beta_i - \beta_i) + (\epsilon_{i,A} - \epsilon_{i,B}) \\
Var[D_i] &= 0 + 0 + Var[\epsilon_{i,A} - \epsilon_{i,B}] \\
&= \sigma_\epsilon^2/2 + \sigma_\epsilon^2/2 = \sigma_\epsilon^2
\end{aligned}$$

Q: What does this toy example tell us about performing research?

- We will learn how to control for the effects of confounding variables in in multivariate regression.
- Analygous to fixed effects or first differences. All cancel out the effects of shared confounders.
- But we would not be able to estimate the boy specific effects had we not given them both kinds of shoes.
- Thinking carefully about design can make a huge difference.
- Paired comparison design is a form of randomized block design.
    1. Create blocks of relatively homogenous subunits
    2. Randomize treatment within blocks
    3. Analyze appropriately

George Box: "Block what you can; Randomize what you cannot!"

1. Block to control for large, known, sources of variation
2. Randomize to eliminate bias from unknown sources of variation

Comparative politics,

- You could randomize treatment assignment between villages
- But probably better to match villages on observable characteristics, creating blocks or pairs, and then randomize within.