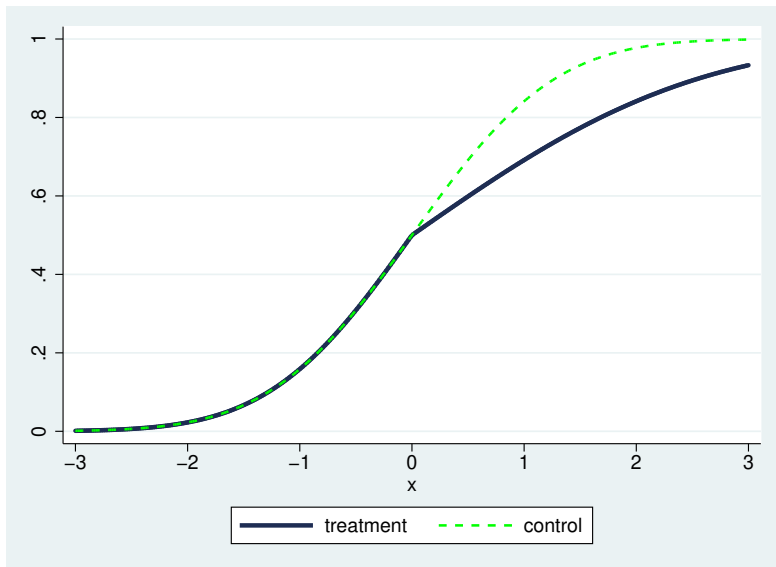


`distcomp`
Comparing Distributions

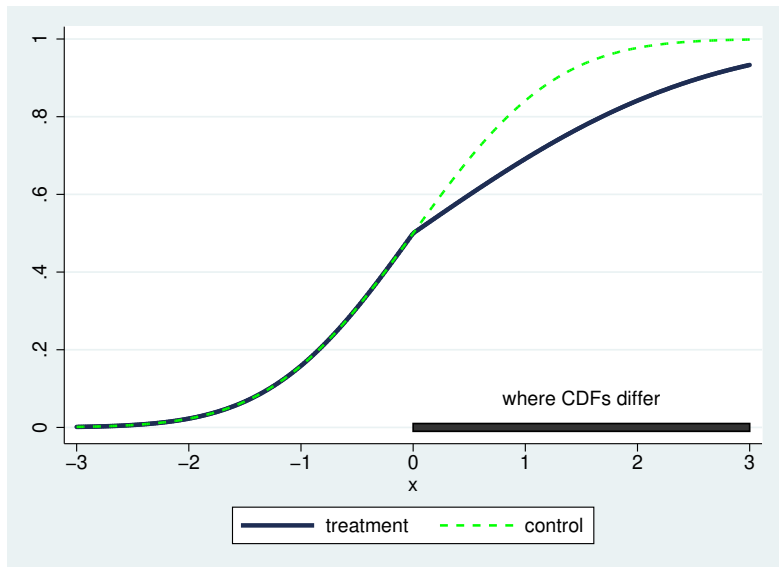
David M. Kaplan
University of Missouri

2021 Stata Conference

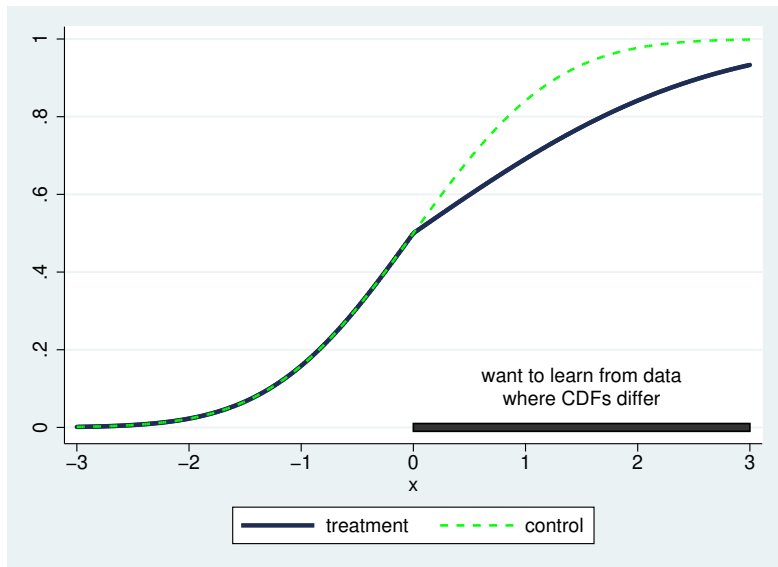
Goal



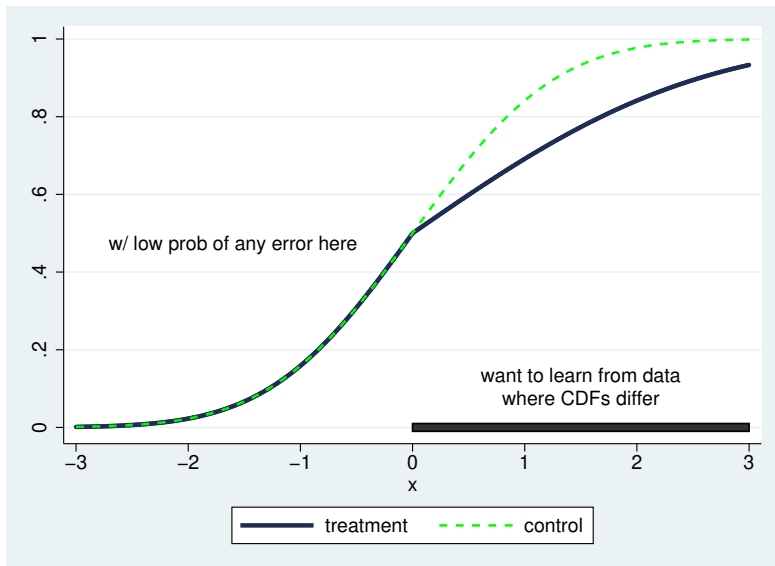
Goal



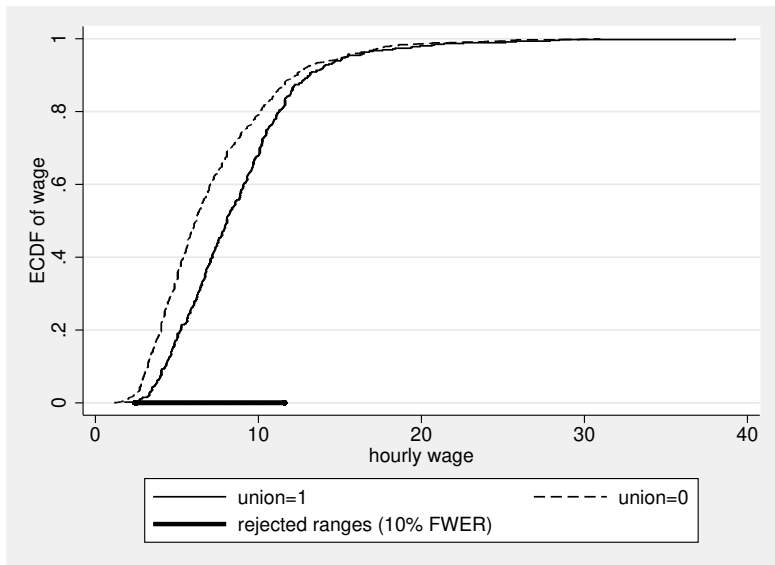
Goal



Goal



Wages: union vs. not



Wages: code

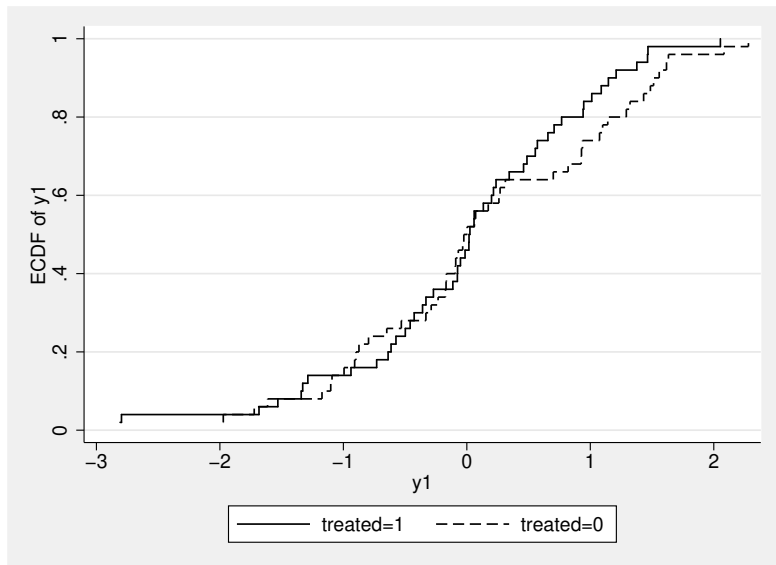
```
. sysuse nlsw88 , clear  
. distcomp wage , by(union) alpha(0.10)
```

Wages: code

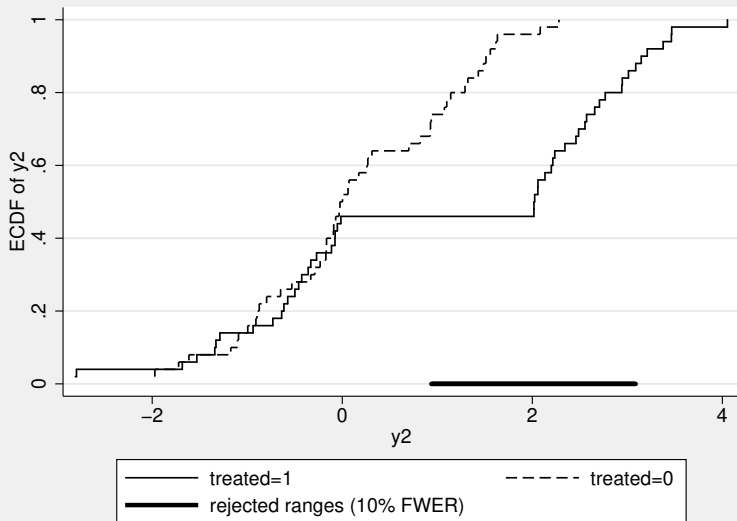
```
. sysuse nlsw88 , clear
. distcomp wage , by(union) alpha(0.10)

. distcomp wage , by(union) alpha(0.10) p
> groptline0(lcolor(cyan))
> groptline1(lcolor(green))
> gropttwoway(title(Dave's picture))
> groptrej(lwidth(vthick) lcolor(blue)
>          lpattern(solid))
```

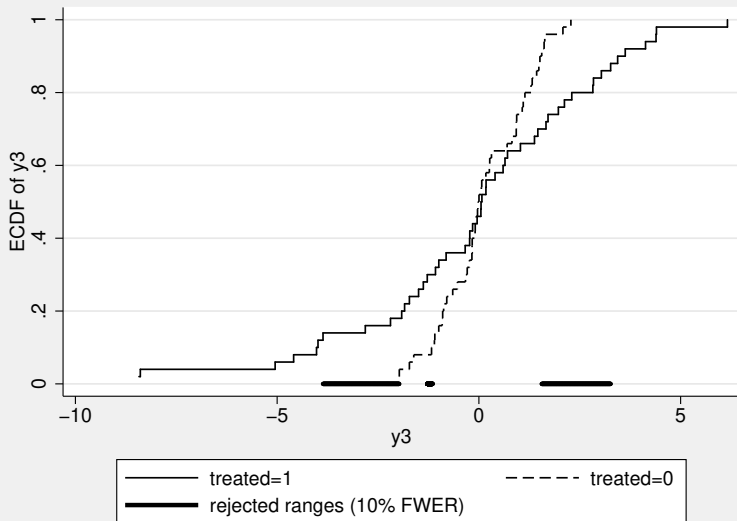

Simulated data: both $N(0, 1)$, no significance



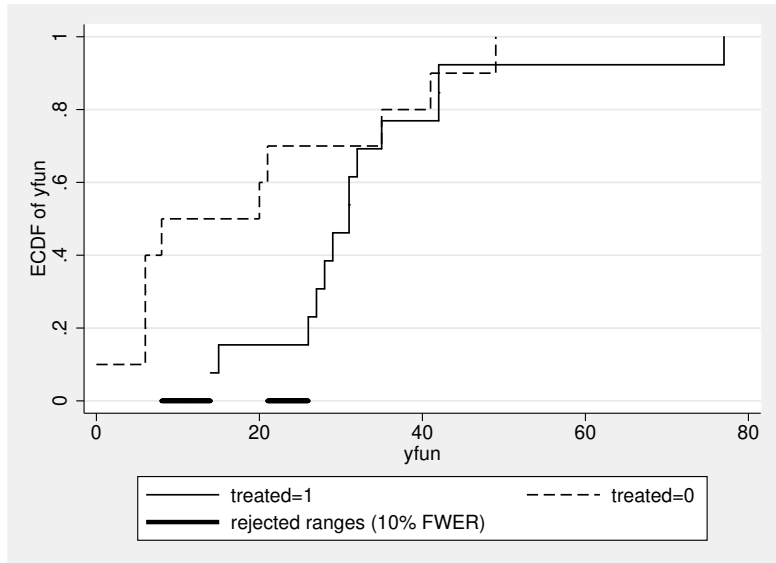
$Y_0 \sim N(0, 1)$, +2 effect above median



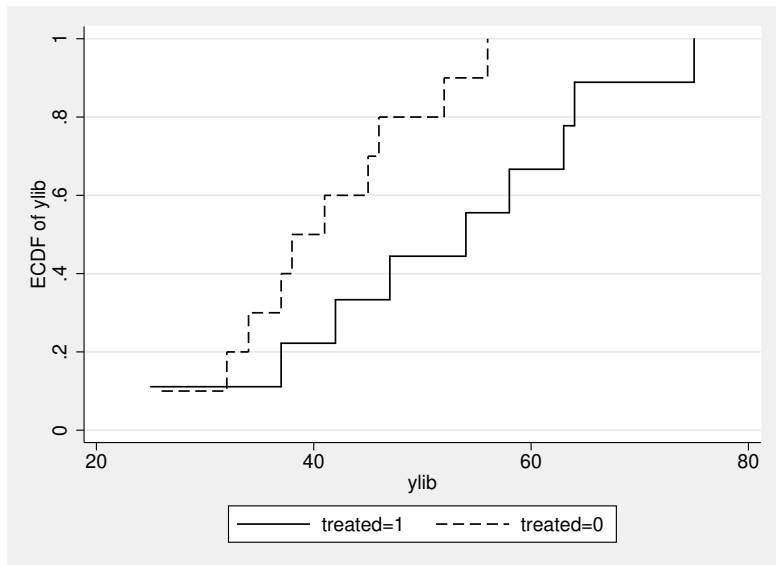
$N(0, \sigma^2)$, $\sigma = 1$ vs. $\sigma = 3$



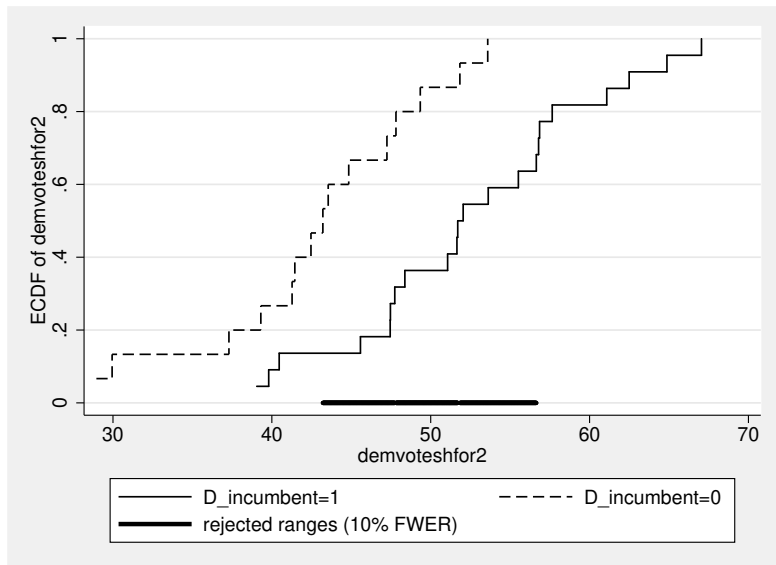
Gneezy and List (2006): fundraising



Gneezy and List (2006): data entry



Regression discontinuity: incumbent effect



Bonus: “power” vs. ksmirnov

```
. set obs 69
. gen grp = (_n>49)
. gen y = (_n<=49)*(_n/50) + (_n>49)*(_n-49)/21
. replace y = 10^6+_n if _n>63
. ksmirnov y , by(grp) exact
. distcomp y , by(grp) p
```

Bonus: “power” vs. ksmirnov

```
. set obs 69
. gen grp = (_n>49)
. gen y = (_n<=49)*(_n/50) + (_n>49)*(_n-49)/21
. replace y = 10^6+_n if _n>63
. ksmirnov y , by(grp) exact
. distcomp y , by(grp) p
```

ksmirnov: $p = 0.121(!!)$

distcomp: $p = 0.0056$

Conclusion

`distcomp` (Kaplan, 2019; Goldman and Kaplan, 2018)

- ▶ `net` from <https://kaplandm.github.io/stata>
- ▶ (also `sivqr` for IV quantile regression!)
- ▶ beyond “if” to *where* do distributions differ?
- ▶ better than KS (power more evenly distributed)

Thank you / questions & comments appreciated!

References I

- Gneezy, Uri and John A. List. 2006. “Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments.” *Econometrica* 74 (5):1365–1384. URL <https://doi.org/10.1111/j.1468-0262.2006.00707.x>.
- Goldman, Matt and David M. Kaplan. 2018. “Comparing distributions by multiple testing across quantiles or CDF values.” *Journal of Econometrics* 206 (1):143–166. URL <https://doi.org/10.1016/j.jeconom.2018.04.003>.
- Kaplan, David M. 2019. “distcomp: Comparing distributions.” *Stata Journal* 19 (4):832–848. URL <https://doi.org/10.1177/1536867x19893626>.