HYPOTHESIS TESTING REGARDING THE MEAN(S)

* We often use hypothesis testing to determine whether a process or treatment actually has an effect on the population of interests
* Or whether two groups are different from one another in the mean of some variable
* We often use hypothesis testing to determine whether a process or treatment actually has an effect on the population of interests

HYPOTHESIS TESTING STOCK AND WATSON EXAMPLE

* We hypothesize that the population mean of $20 hourly wage
* But our random sample gave us an mean of $22.24
  + So our sample mean does not exactly equal the hypothesized value
  + Is this because the true population mean hourly wage isn’t really $20?
  + Or is that the true population mean IS $20 hourly wage but our sample mean just differs from that because of random sampling variation?
* SOLUTION; we can rule out

T-Tests

In this way, T and P are inextricably linked. Consider them simply different ways to quantify the "extremeness" of your results under the null hypothesis. You can’t change the value of one without changing the other.

The larger the absolute value of the t-value, the smaller the p-value, and the greater the evidence against the null hypothesis

Scatterplot

Covariance

Making sense of the formula

* An obs has a \high" value of X if Xi is greater than \_ X
* An obs has a \low" value of X if Xi is less than \_ X
* Do obs with \high" values of X usually \have" high values
* of Y?

Correlation

Why divide by (sX \_ sY )?

* After dividing by std dev, distance between obs is unitless
  + Think \z-scores; e.g., two standard deviations away is \far"
  + Z-score is in standard deviations
* After dividing by sX \_ sY , we have a unitless measure of extent to which X and Y move together
* Can compare whether rXY stronger than rXZ