POWER & PERMUTATION

10.14.2020

RECAP

* t-tests : can tell you whether the mean of a sample is different from some value (like zero) or from another sample
  * requires the sample to be gaussian-distributed
  * but that’s ~fine because of the central limit theorem

RECOMMENDED READING

* Inferential Thinking Chapter 14
* Prob 140 Textbook Chapter 6.1

POWER OF STATISTICAL TESTS

* **power** is the probability of rejecting the null hypothesis if the alternative hypothesis is actually true

  * *i.e. how often you say "this is significant!" for a real effect*

	Effect is Real	Effect is Not Real
Test Significant	True Positive (TP)	False Positive (FP)
Test Not Significant	False Negative (FN)	True Negative (TN)

```
Power = TP / (TP + FN)
```

POWER OF STATISTICAL TESTS

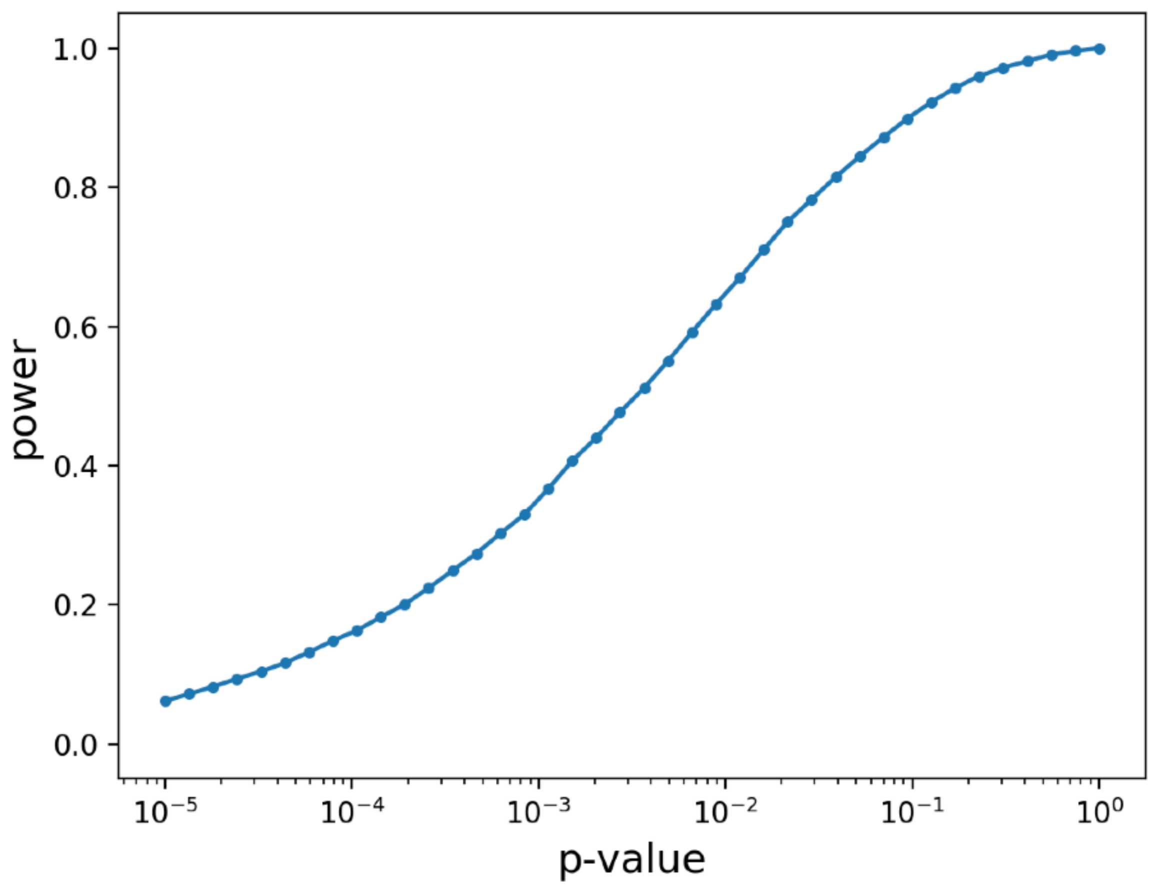
* 80% power means that 20% of the time you get a false negative: the test says “not significant” when the effect is real

POWER OF STATISTICAL TESTS

* power depends upon the p-value threshold that you choose for a test

  * smaller threshold = lower power

  * (but also fewer false positives!)



POWER OF STATISTICAL TESTS

* it’s also related to the **effect size**
  * bigger effects are easier to detect, so bigger effect size = higher power
* and to the **sample size (n)**
  * bigger samples make it easier to detect effects, so have higher power

POWER OF STATISTICAL TESTS

* finally, power is also related to whether the **assumptions of the test are valid**

POWER OF STATISTICAL TESTS

* e.g. if you have paired samples but use an un-paired t-test, then that could reduce your power to find a real effect

PERMUTATION

* many (if not most) data analysis questions boil down to this:

    * *is something about sample A different than sample B?*

* **permutation testing** turns this question into a counterfactual

PERMUTATION

* *suppose sample A and sample B are not different*

    * then it shouldn’t matter if we scramble up (**permute**) samples A and B and then re-divide them into new samples

PERMUTATION

* let’s say sample A & B are each 10 data points
* imagine we throw all 20 data points from samples A & B into a bag
* then we pull out 10 random data points to form a new “sample A*1”
* and use the other 10 to form a new “sample B*1”

PERMUTATION

* let’s compute the difference between the means of sample A*1 and sample B*1, and call this M1

* now let’s suppose we do this 1000 times, each time randomizing which data goes into A*i and B*i, and computing Mi

PERMUTATION

* this gives us a distribution of **permuted mean differences** [M1, M2, M3, …]

* if there really was no difference between sample A and sample B, then the *true* mean difference, M=mean(A-B), should be somewhere in the middle of the distribution of permuted mean differences

PERMUTATION

* recall the definition of the p-value:
  * “if the null hypothesis (there is no difference between A and B) were true, what’s the probability of finding a value at least this extreme?”

PERMUTATION

* the permuted mean differences tell us exactly what the null distribution looks like

* so we can ask directly: how often does randomly dividing the data into A* and B* yield a mean difference as extreme as the one we see?

PERMUTATION

* NB: what I’ve described here is not a “true” permutation test, because we typically don’t test every possible permutation

* instead, we test only a small number (like 1000)

* technically this is known as “Monte Carlo permutation test” because it involves randomization

PERMUTATION

* permutation tests can work with many different statistics, not just the mean

  * for example, you could use permutation to test whether the variances of two samples are the same

PERMUTATION

* permutation tests don’t care what distribution your data comes from

* so they work even when t-tests don’t!

END