

What is p-hacking?

Excellent example found here:

<http://fivethirtyeight.com/features/science-isnt-broken/>

- Remember that there is (somewhat unconscious) bias present in many experiments:
 - Collect data until a statistically significant result is obtained.
 - If initial results aren't statistically significant, keep collecting the data. ← Ad Hoc Sequential Sample Size Determination (SSSD)

Data Dredging:

*When you use multiple tests on a data set, the probability of making **at least one** type I error, α , is larger than the significance level suggests - each hypothesis test has some probability of error and these errors compound as more tests are conducted*

$P(\text{No type I errors}) = (1 - \alpha)^N$, where N = independent tests

$$\begin{aligned} P(\geq 1 \text{ type I error}) &= 1 - P(\text{No Type I Errors}) \\ &= 1 - (1 - \alpha)^N \end{aligned}$$

Bonferroni

- If your goal is to determine which variable really did respond to treatment rather than just preliminary data exploration
- Only reject H_0 if the **P** value is less than the bonferroni-adjusted α^*
- $P(\geq 1 \text{ Type I error}) = \alpha$
- lose power \leftarrow penalized for ‘asking too many questions of the data’

$$\alpha^* = \frac{\alpha}{Num_tests}$$

Data Dredging:

α	Test Number	P(No type I errors)	α^*
0.05	10	0.60	0.005
	100	0.006	0.0005
0.01	10	0.90	0.001
	100	0.37	0.0001

Clinical test can have 10+ tests

Gene location can have 100+ tests

False discovery Rate

- Carry out multiple tests at fixed significance level
- Gather all tests that yield statistically significant result (discoveries)
- Estimate the false discovery proportion from the total discoveries

$$P(\text{Reject}|H_0 \text{ true})/\{P(\text{reject}| H_0 \text{ true})+ P(\text{reject}| H_A \text{ true})\}$$