THE BASICS OF MULTIPLE TESTING

ELLY LEE

RECAP: HYPOTHESIS TESTING

- Statistical method used to decide whether the data provides significant evidence to reject a hypothesis
- Null hypothesis (H_0): a gene is NOT related to colon cancer
- Alternative hypothesis (H_1): a gene is related to colon cancer
- Type 1 error (false positive): rejecting H_0 when H_0 is true
  - By construction, valid tests satisfy: $P(FP) \leq \alpha$
- Type 2 error (false negative): failing to reject H_0 when H_0 is false

WHAT IS MULTIPLE TESTING?

- Testing many hypotheses on the same dataset
  - Each hypothesis test has its own p-value (p_1, p_2, …)
- Ex: Genome-wide association studies (GWAS) with 278,869 SNPs
  - Each SNP is associated with a separate hypothesis

THE BIG PROBLEM WITH DOING SO MANY TESTS

- If $\alpha = 0.05$ for each test, then:
  - For one test -> 5% chance of a false positive
  - For 20 independence tests -> chance of at least on FP $\approx 1 - (1 - 0.05)^{20} \approx 0.64$
    - 64% chance of getting a wrong "significant" result?! TOO HIGH!
- The more tests you run, the more likely to "discover" something that's just "random noise"
- Back to the GWAS example:
  - With a $p < 0.05$ threshold, 13,943 false positives are expected even if no true associations exist.

FWER OR FDR

- **Family Wise Error Rate** (**FWER**): Probability of making at least one FP among all tests
  - FWER = P(at least one $H_0^{(k)}$ rejected for k = 1, 2, …, m | $H_0^{(k)}$ true for all k = 1, …, m)
- **False Discovery Rate** (**FDR**): expected proportion of FP of the total number of positive findings (rejection of a null)
  - An FDR of 5% = among all discoveries (rejections of a null), 5% are expected to be FP (null is actually true)

FWER OR FDR

- Controlling FWER is best for:
  - When you hope not to have any false positives (having FP too costly)
  - The # of hypotheses is small – controlling FWER is more conservative
- Controlling FDR is best for:
  - When you have a very large # of tests, and controlling FWER is too stringent
  - When some number of false positives is acceptable

CONTROLLING FWER - BONFERRONI

- If given the significance level α and k tests, then…
- Bonferroni-adjusted threshold $\alpha^* = \boldsymbol{\alpha/k}$
- Example: $p_1 = 0.004$, $p_2 = 0.021$, $p_3 = 0.037$, $p_4 = 0.056$, $p_5 = 0.093$, $\alpha = 0.05$
  - $\alpha^* = 0.05/5 = 0.01$
  - Before Bonferroni: reject H_1, H_2, H_3
  - After Bonferroni: reject H_1

CONTROLLING FWER – HOLM'S STEP DOWN

- Given *m* hypotheses with p-values $p_1, p_2, \ldots, p_m$

1. Order them from lowest to highest: $p_{(1)} \leq p_{(2)} \ldots \leq p_{(m)}$ with hypotheses $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$
   - Note: Not always $p_1 \neq p_{(1)}$
     - Ex: If $p_1 = 0.021$, $p_2 = 0.05$, $p_3 = 0.004$, then $p_{(1)} = p_3 \leq p_{(2)} = p_1 \leq p_{(3)} = p_2$
2. For each p-value from lowest to highest, test whether $p_{(k)} \leq \alpha/(\mathbf{m+1-k})$
3. If so, reject $H_{(k)}$, otherwise EXIT
4. Continue to examine the next large p-value until EXIT or last p-value is examined
   - If $p_{(1)} \leq \alpha/m$, reject $H_{(1)}$ and continue, otherwise EXIT
   - If $p_{(2)} \leq \alpha/(m - 1)$, reject $H_{(2)}$ and continue, otherwise EXIT
   - And so on…

CONTROLLING FWER – HOCHBERG'S STEP UP

- Similar procedure as Holm's, but instead of starting from the smallest p-value, we begin from the **largest** p-value
  - If $p_{(m)} \leq \alpha$, reject $H_{(m)}$ and below and EXIT, otherwise continue
  - If $p_{(m-1)} \leq \alpha/2$, reject $H_{(m-1)}$ and below and EXIT, otherwise continue
  - And so on…

CONTROLLING FDR – BENJAMINI-HOCHBERG

- Given *m* hypotheses with p-values $p_1, p_2, \ldots, p_m$

1. Order them from lowest to highest: $p_{(1)} \leq p_{(2)} \ldots \leq p_{(m)}$ with hypotheses $H_{(1)}, H_{(2)}, \ldots, H_{(m)}$
2. Choose a target FDR level/desired FDR threshold (e.g. $\alpha = 0.05$)
3. For each p-value from lowest to highest, test whether $p_{(k)} \leq \mathbf{(k * \alpha)/m}$
4. If so, reject $H_{(k)}$, otherwise EXIT
5. Continue to examine the next large p-value until EXIT or last p-value is examined
   - If $p_{(1)} \leq \alpha/m$, reject $H_{(1)}$ and continue, otherwise EXIT
   - If $p_{(2)} \leq 2\alpha/m$, reject $H_{(2)}$ and continue, otherwise EXIT
   - And so on…

CONCLUSION

- Here, we have discussed a few methods for controlling FWER or FDR:
  - FWER – Bonferroni, Holm's, Hochberg's
  - FDR – Benjamini-Hochberg
- These corrections allow us to make more meaningful statements when we conduct many hypothesis tests.
- There are many, MANY more methods out there!
- Some methods will take into account **dependence** between hypotheses, hypotheses conducted **sequentially**, and many other situations.