1. Suppose you are performing k hypothesis tests at exact level α, and assume all hypotheses are true. In other words, if R_i denotes the event that the ith hypothesis is rejected, *i.e.*, of making an error, then $P(R_i) = \alpha$. However, when considering the tests as a group, there is a higher probability of making an error, as the individual error rates compound.It can be shown from the inclusion-exclusion principle that:

$$P\left(\bigcup_{i=1}^{k} R_i\right) \geq \sum_{i=1}^{k} P(R_i) - \sum_{1\leq i<j\leq k} P(R_i \cap R_j)$$

   i. Use Boole's inequality to find an upper bound for the probability of making at least one error.
   ii. Use the fact above to derive a lower bound for the probability of making at least one error, assuming that the tests are independent.
   iii. Compute the bound for performing 20 tests at $\alpha = 0.05$.
   iv. Find the smallest number of tests for which the lower bound in (ii) exceeds $\frac{1}{2}$. (Hint: write 0.05 as $\frac{5}{100}$; you may find it useful to know that $41^2 = 1681$.)

Remark: this last part is maybe a little more involved in terms of calculation than what I'd put on an exam.

i. at least one error: $\cup_{i=1}^{k} R_i$, so $P[\cup_{i=1}^{k} R_i] \leq \sum_{i=1}^{k} P(R_i) = \sum_{i=1}^{k} \alpha = k\alpha$

ii. since tests are independent $P(R_i \cap R_j) = P(R_i)P(R_j) = \alpha^2$, so $\sum_{1\leq i<j\leq k} P(R_i \cap R_j) = \sum_{1\leq i<j\leq k} \alpha^2 = \binom{k}{2}\alpha^2 = \frac{k(k-1)}{2}\alpha^2$

then by fact above,

$$P(\cup_{i=1}^{k} R_i) \geq \sum_{i=1}^{k} P(R_i) - \sum_{1\leq i<j\leq k} P(R_i \cap R_j) = k\alpha - \tfrac{1}{2}k(k-1)\alpha^2$$

* note the exact probability is

$$P(\cup_{i=1}^{k} R_i) = 1 - P\left(\bigcap_{i=1}^{k} R_i^c\right) = 1 - \prod_{i=1}^{k} P(R_i^c) = 1 - \prod_{i=1}^{k}(1-\alpha) = 1-(1-\alpha)^k$$

iii. if $k = 20$, $\alpha = 5/100$, $P(\cup_{i=1}^{k} R_i) \geq 20 \cdot 5/100 - \frac{1}{2}\cdot 20 \cdot 19 \cdot 25/10000 = 1 - 19\cdot 25/1000 = 1 - 475/1000 = 525/1000$

$= 0.525$

iv. want to solve for k:

$$\frac{1}{2} = k\alpha - \tfrac{1}{2}k(k-1)\alpha^2 \iff 1 = 2k \cdot 5/100 - (k^2-k)(5/100)^2 = k\cdot 1/10 - 25/10000\, k^2 + 25/10000\, k$$

multiply through by 10:

$$10 = k - 25/1000\, k^2 + 25/1000\, k = -25/1000\, k^2 + 1025/1000\, k = -1/40\, k^2 + 41/40\, k \iff 0 = -k^2 + 41k - 400$$

so:

$$k = \frac{-41 \pm \sqrt{41^2 - 1600}}{-2} = \frac{41 \pm \sqrt{81}}{2} = \frac{41 \pm 9}{2} \quad \text{so } k = 16, 25$$

so $k = 16$

* note the exact k is given by $\frac{1}{2} = 1 - (1-\alpha)^k$ so

$$(1-\alpha)^k = \tfrac{1}{2} \Rightarrow k\log(1-\alpha) = -\log 2 \Rightarrow k = -\frac{\log 2}{\log(1-\alpha)} \approx 13.51$$

so $k = 14 \Rightarrow P(\cup_i R_i) > 1/2$

2. Suppose you are reading an article on admission rates at a selective college for different demographic groups reporting that 20% of applicants from underrepresented groups were admitted in 2022-2023 compared with 10% of other applicants.

   i. What is the probability that a randomly selected incoming student is from an underreppresented group if there were twice as many applicants from non-underrepresented groups?
   ii. What is the probability that a randomly selected incoming student is from an underrepresented group if there were ten times as many applicants from non-underrepresented groups?
   iii. Under the scenario in (i), find the probability that a randomly selected applicant was admitted.

let A : {admit applicant}, U : {applicant is from underrepresented group}

$P(A \mid U) = 0.2$

$P(A \mid U^c) = 0.1$

i. if there are 1 : 2 underrepresented to non under-represented,

$\frac{P(U)}{P(U^c)} = \frac{1}{2} \Rightarrow P(U) = \frac{1}{3}$ (since $P(U) + P(U^c) = 1$)

so

$$P(U \mid A) = \frac{P(A \mid U)P(U)}{P(A \mid U)P(U) + P(A \mid U^c)P(U^c)} = \frac{2/10 \cdot 1/3}{2/10 \cdot 1/3 + 1/10 \cdot 2/3} = \frac{2}{4} = \frac{1}{2}$$

ii. if 1 : 10, $P(U) = \frac{1}{11}$, so $P(U \mid A) = \frac{2/10 \cdot 1/11}{2/10 \cdot 1/11 + 1/10 \cdot 10/11} = \frac{2}{12} = \frac{1}{6}$

iii. $P(A) = P(A \mid U)P(U) + P(A \mid U^c)P(U^c)$

$= 2/10 \cdot 1/3 + 1/10 \cdot 2/3$

$= 4/30$

$= 2/15$

3. Imagine you're out trick-or-treating with a friend and you arrive at a house handing out only two treats: Kit Kats and Snickers. The homeowner has mixed them up in a large pillowcase. You and your friend can each draw n pieces of candy; assume either treat is equally likely to be selected on each draw. Supposing that when you arrive, there are $N > 2n$ Snickers and $M > 2n$ Kit Kats, and you draw first, find an expression for the probability that you draw k_1 Snickers and your friend draws k_2 Snickers.

Extra practice: find an expression for the probability that your friend draws k_2 Snickers.

let A_{k_1} : you draw k_1

B_{k_2} : friend draws k_2

if you draw first,

$$P(A_{k_1}) = \frac{\binom{N}{k_1}\binom{M}{n-k_1}}{\binom{N+M}{n}}$$

and then there are $N-k_1$ snickers, $M-(n-k_1)$ kit kats, $N+M-n$ pieces.

so

$$P(B_{k_2} \mid A_{k_1}) = \frac{\binom{N-k_1}{k_2}\binom{M-(n-k_1)}{n-k_2}}{\binom{N+M-n}{n}}$$

so:

$$P(A_{k_1} \cap B_{k_2}) = P(B_{k_2} \mid A_{k_1})\, P(A_{k_1}) = \frac{\binom{N}{k_1}\binom{M}{n-k_1}}{\binom{N+M}{n}} \cdot \frac{\binom{N-k_1}{k_2}\binom{M-(n-k_1)}{n-k_2}}{\binom{N+M-n}{n}}$$