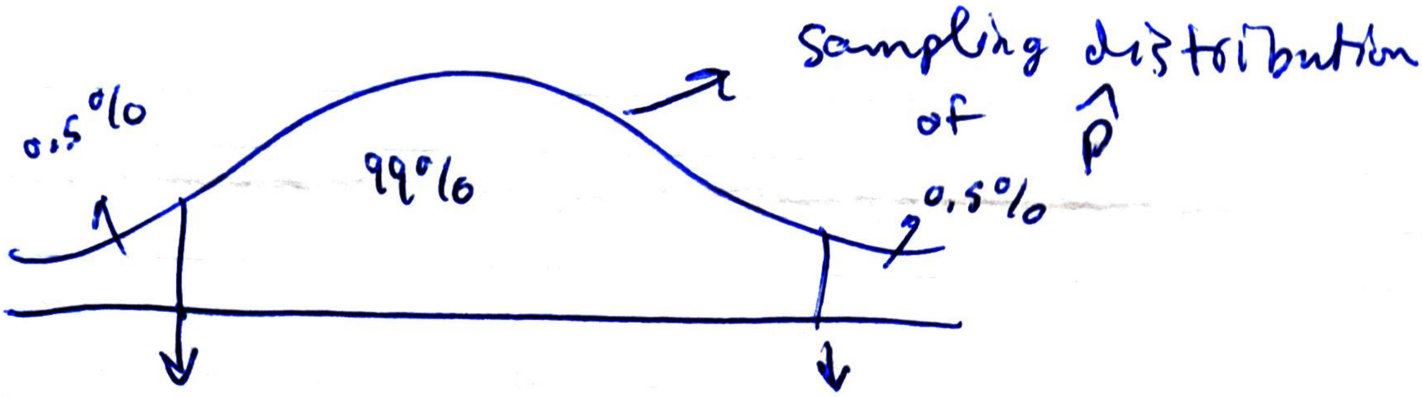
1. Drunken Cyclists

Assumptions:

1. A SRS of bicycle riders in the U.S.
2. we want a representative sample of All bicycle riders in the U.S.
3. 10% conditions → you haven't sampled more than 10% of all bicycle riders in the U.S.
4. Success / Failure condition → $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$
5. sample size is fixed at 1711
6. Binary event → +/- for alcohol
7. CLT has kicked in, i.e. sample size is large enough to approximate the binomial distribution with a Normal.

99% CI for p => ?

$$\text{qnorm}\left(p = c(0.005, 0.995),\ \text{mean} = \overset{\hat{p}}{\frac{542}{1711}},\ \text{sd} = \underbrace{\sqrt{\frac{\left(\frac{542}{1711}\right)\left(1 - \frac{542}{1711}\right)}{1711}}}_{sd(\hat{p})}\right)$$



$\hat{p} + z_{0.005} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\qquad$ $\hat{p} + z_{0.995} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

2) Handling contact lenses.

<u>Assumptions</u>

① SRS of Americans who wear contact lenses

② Representative sample of All Americans who wear contact lenses

③ 10% condition => same probability of not washing hands for each individual.

④ Success / Failure condition:

$n\hat{p} = 139 \qquad n(1-\hat{p}) = 142 \qquad \hat{p} = \frac{139}{281}$

CLT has kicked in

⑤ Binary event and fixed sample size of 281.

"plus 4" 99% CI => adding two events and two not events to your sample.

qnorm(p=c(0.005, 0.995), mean = $\frac{139+2}{281+4}$, sd = $\sqrt{\frac{\left(\frac{139+2}{281+4}\right)\left(1-\frac{139+2}{281+4}\right)}{281+4}}$)

3) Cancer-detecting dogs

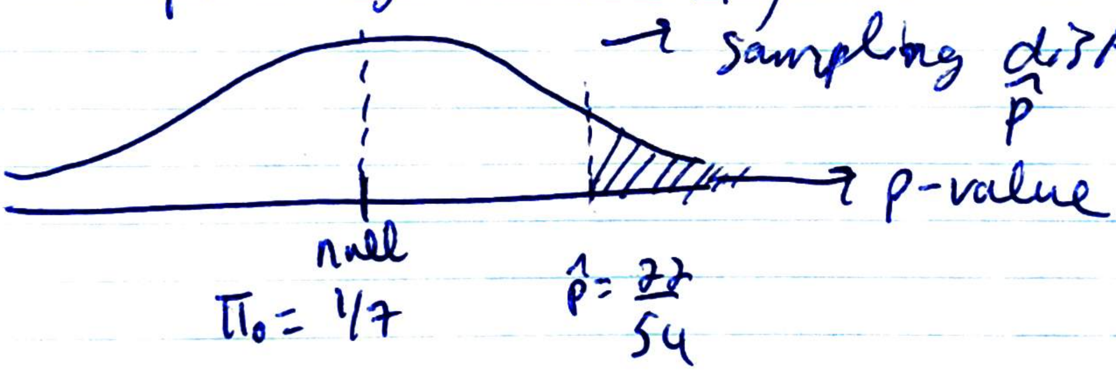
population parameter : π =>

the population proportion of ~~cancer-detec~~ detecting bladder cancer from a urine sample.

a) $H_0: \pi = \frac{1}{7}$ $\qquad H_a: \pi > 1/7$

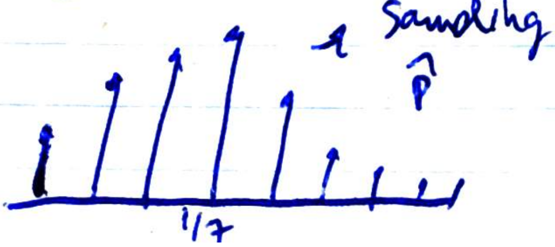
b) 2 ways

(1) Assuming CLT (and all assumptions that go with it)



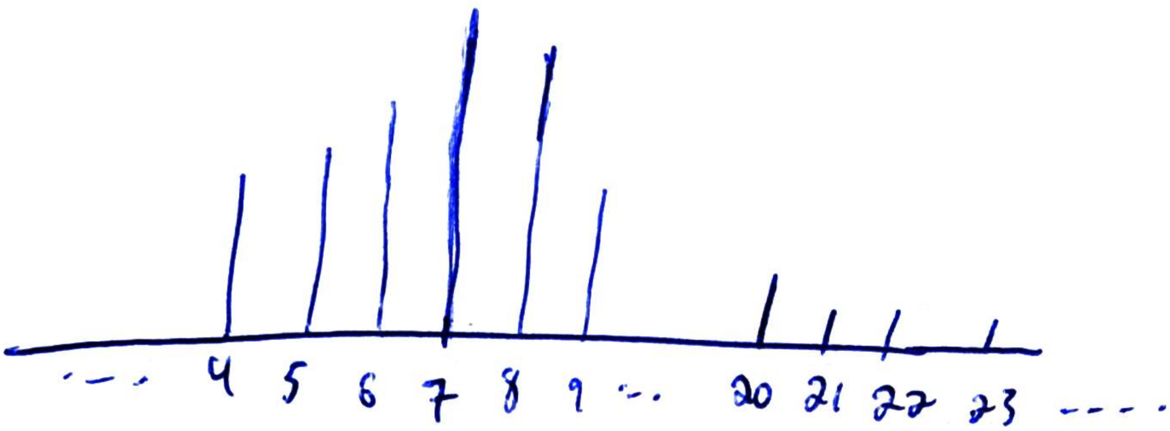
p-value = pnorm ($q = \frac{22}{54}$, mean = $\frac{1}{7}$, sd = $\sqrt{\frac{\frac{22}{54}\left(1-\frac{22}{54}\right)}{54}}$,
lower.tail = FALSE)

(2) Exact p-value using Binomial model.



pbinom(q = 21 , size = 54 , prob = 1/7 ,
lower.tail = F)

Sampling distribution



p-value = $P(Y \geq 22 \mid H_0)$

where Y is the # of successes

$= P(Y=22) + P(Y=23) + P(Y=24) + \cdots + P(\quad)$

= dbinom (22, 54, 1/7) + dbinom(23, 54, 1/7) + ... + ...

= pbinom (21, 54, 1/7, lower.tail = FALSE)

= 1 - pbinom(21, 54, 1/7, lower.tail = TRUE)

pbinom (22, 54, 1/7, lower.tail = F)

$= P(Y=23) + P(Y=24) + \cdots +$

pbinom (22, 54, 1/7, lower.tail = T)

$= P(Y=0) + P(Y=1) + \cdots + P(Y=21) + P(Y=22)$

4) Presidential campaign.

a) No difference between proportions of times a taller candidate wins compared to proportion of times a smaller candidate wins.

b) $H_0 : \pi_{tall} = 0.5 = \pi_{small}$

$H_a : \pi_{tall} > 0.5$

c) ① fixed sample size
② 10% condition → violated
③ Random sample ?

R code to calculate a p-value.

```
pbinom(q=19, size=25, prob=0.5,
      lowertail = FALSE)
```

d) 95% CI from the Nomogram is

~~[0.61, .80]~~ [0.6, 0.9].

Since the 95% CI does not contain the null, there is evidence to suggest that taller candidates have a larger chance of winning at an alpha = 0.05.