p-problems

What's a (statistical) model?

frequentist

likelihood

$$P(D \mid \theta)$$

Bayes

likelihood

$$P(D \mid \theta)$$

&

prior

$$P(\theta)$$

p-value

$$p_{x_{\text{obs}}} = P_{H_0}(T(X) \geq T(x_{\text{obs}}))$$

probability, under the assumption that H_0 is true, of observing a value of the test statistic that is at least as extreme as that of the observed data

"Bayes sucks because it relies on flimsy subjective priors"

A completely new example

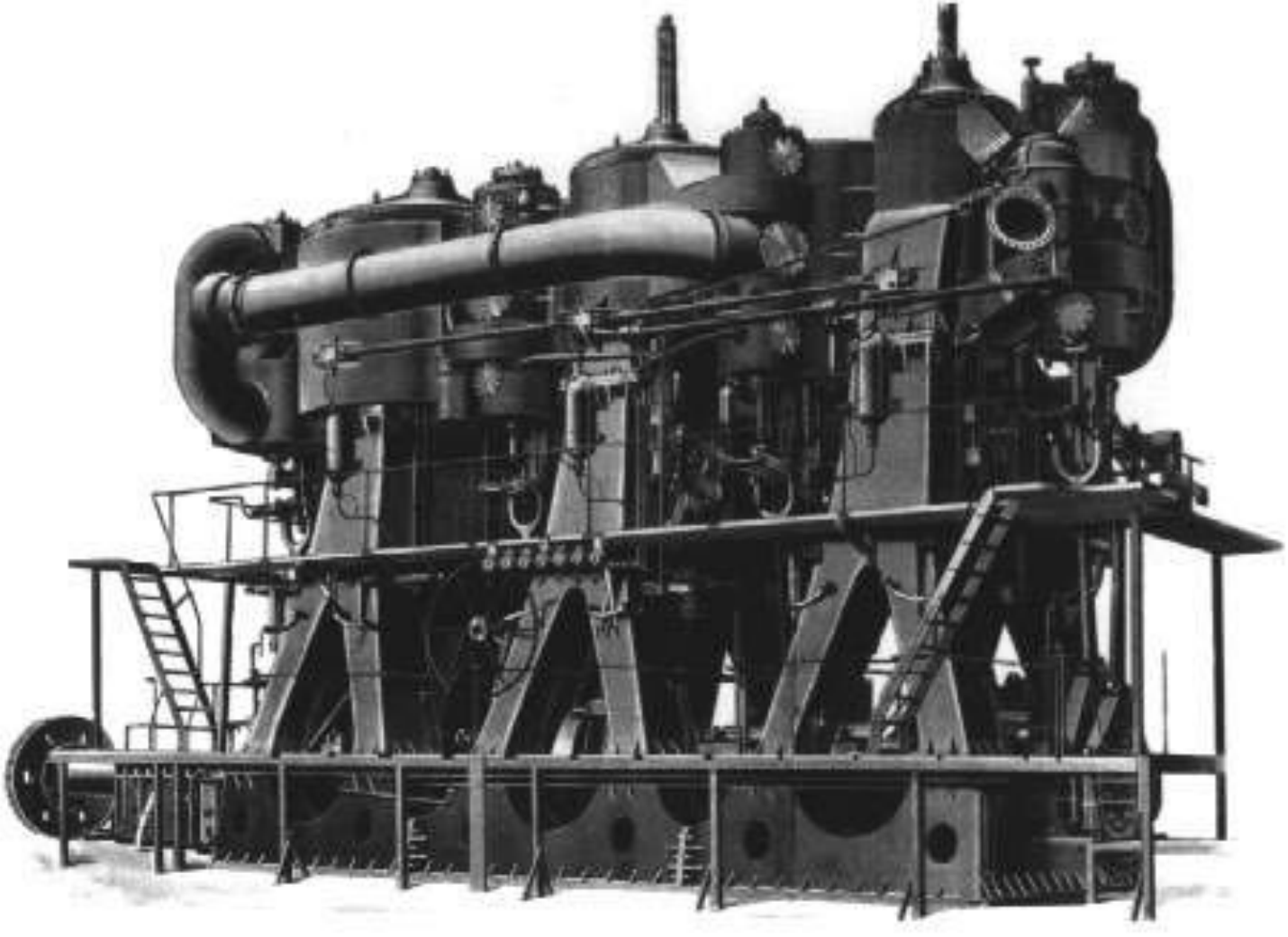
```
binom.test(x = 7, n = 25, p = 0.5)
```

```
##
##  Exact binomial test
##
## data:  7 and 25
## number of successes = 7, number of trials = 25, p-value = 0.04329
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.1207167 0.4938768
## sample estimates:
## probability of success
##                   0.28
```

::: significant test result :::

::: reject null hypothesis :::

::: behave as if the coin was not fair :::



Testing whether subjects have clairvoyance

```
binom.test(x = 7, n = 25, p = 0.5)
```

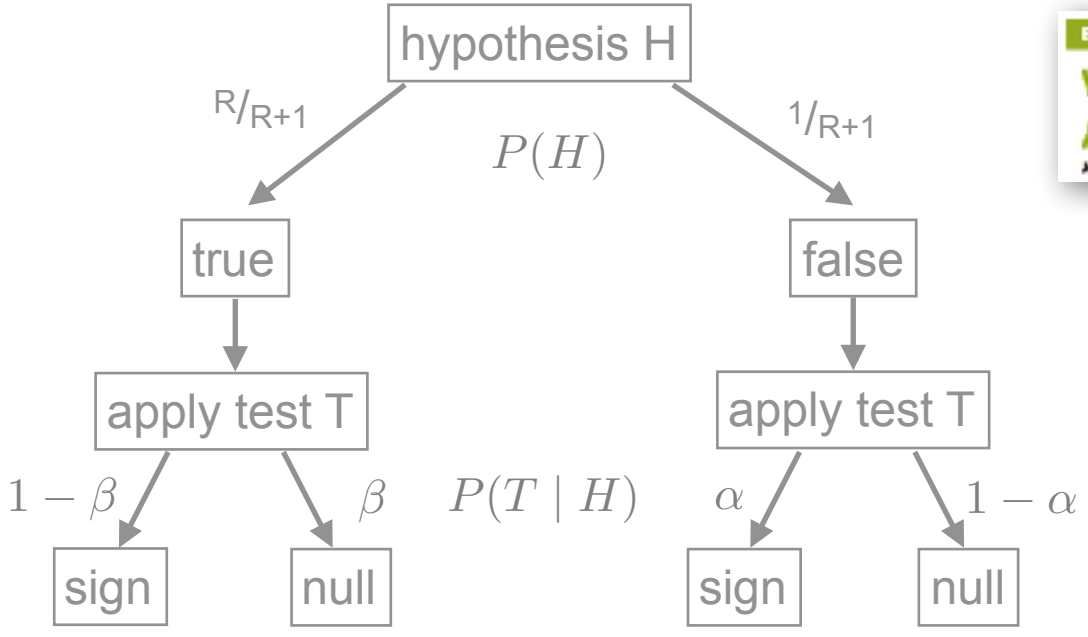
```
##
##  Exact binomial test
##
## data:  7 and 25
## number of successes = 7, number of trials = 25, p-value = 0.04329
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.1207167 0.4938768
## sample estimates:
## probability of success
##                   0.28
```

::: significant test result :::

::: reject null hypothesis :::

::: behave as if clairvoyance exists :::

"*P*-values quantify evidence against the null."




Essay

Why Most Published Research Findings Are False

John P.A. Ioannidis

Positive predictive value

$$P(H = t \mid T = s) = \frac{P(T = s \mid H = t)P(H = t)}{P(T = s)}$$

$$= \frac{R(1-\beta)}{R(1-\beta)+\alpha}$$

"probability that the hypothesis is true, given a significant test result"

Positive predictive value

$$P(H = t \mid T = s) = \frac{P(T = s \mid H = t)P(H = t)}{P(T = s)}$$

$$= \frac{R(1-\beta)}{R(1-\beta)+\alpha}$$

"probability that the hypothesis is true, given a significant test result"

example (coin flip):

$R = 1, \ \beta = 0.2, \ \alpha = 0.05$

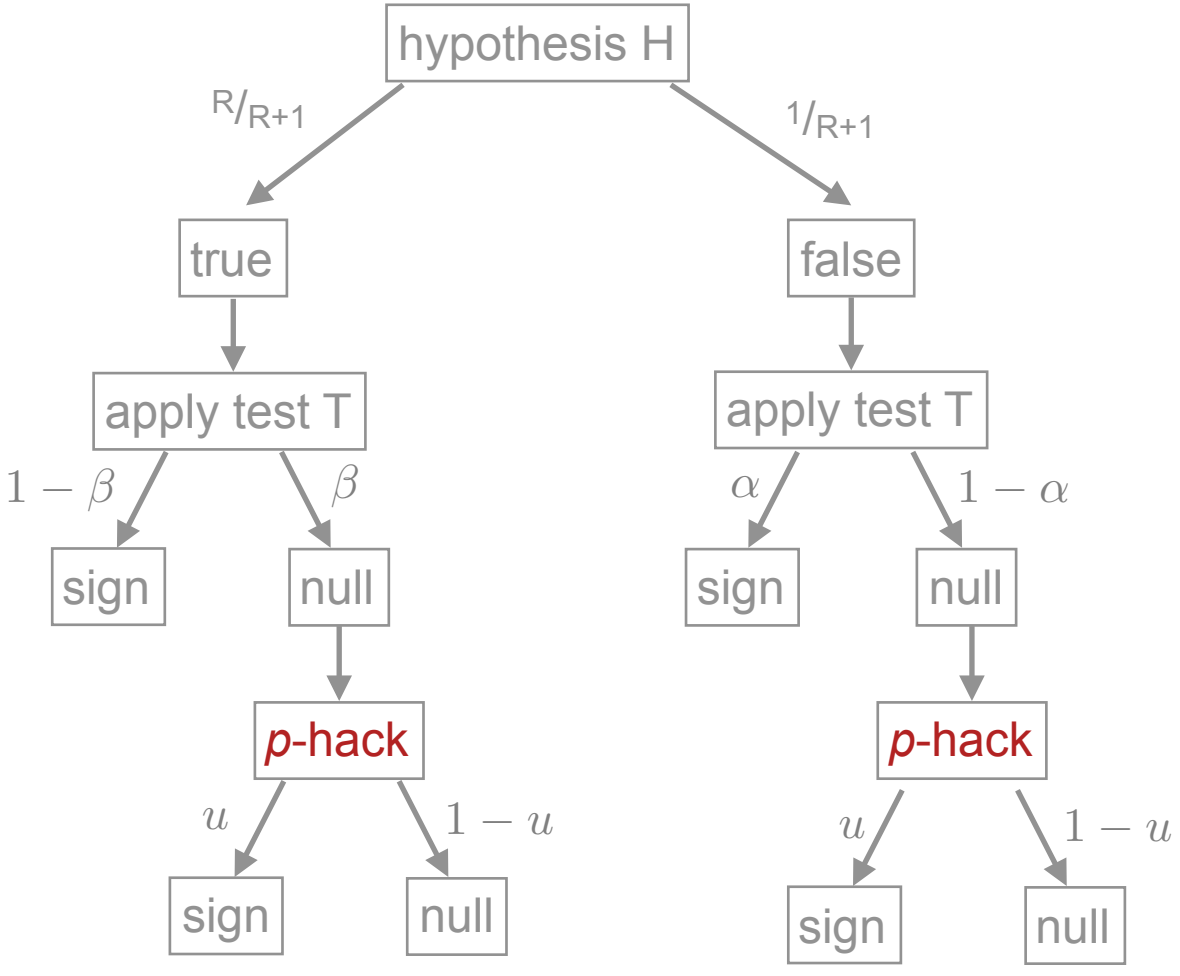
$$P(H = t \mid T = s) = \frac{0.8}{0.85} \approx 0.94$$

example (clairvoyance):

$R = 0.0001, \ \beta = 0.2, \ \alpha = 0.05$

$$P(H = t \mid T = s) = \frac{0.00008}{0.05008} \approx 0.02$$

"NHST gives you rigorous error control."



p-hacking ::: combination of design/presentation/analysis factors that favor a significant test result beyond the normal alpha level

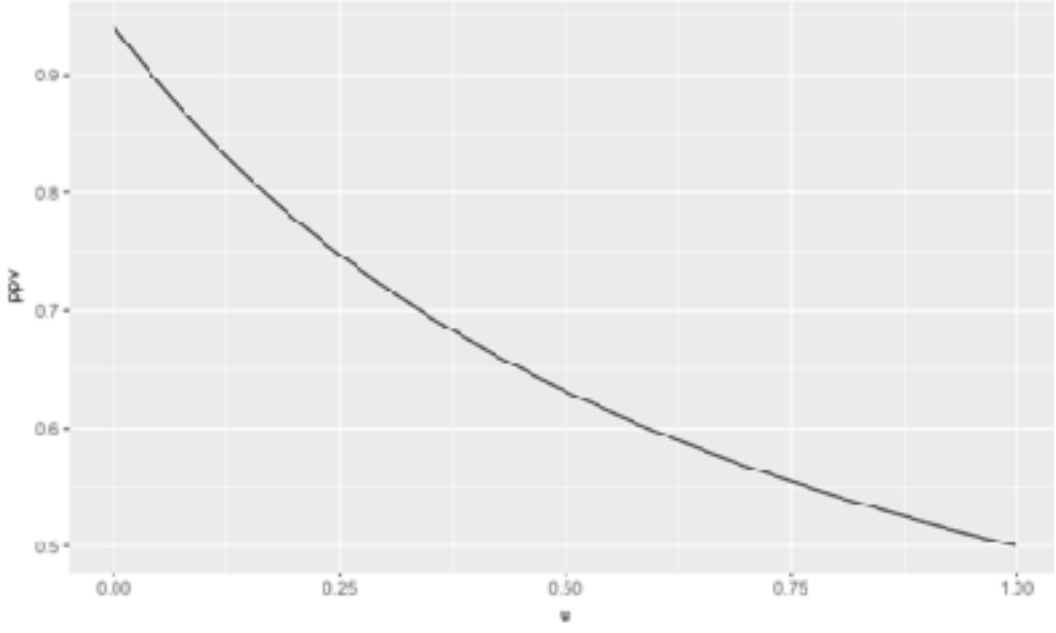
Positive predictive value

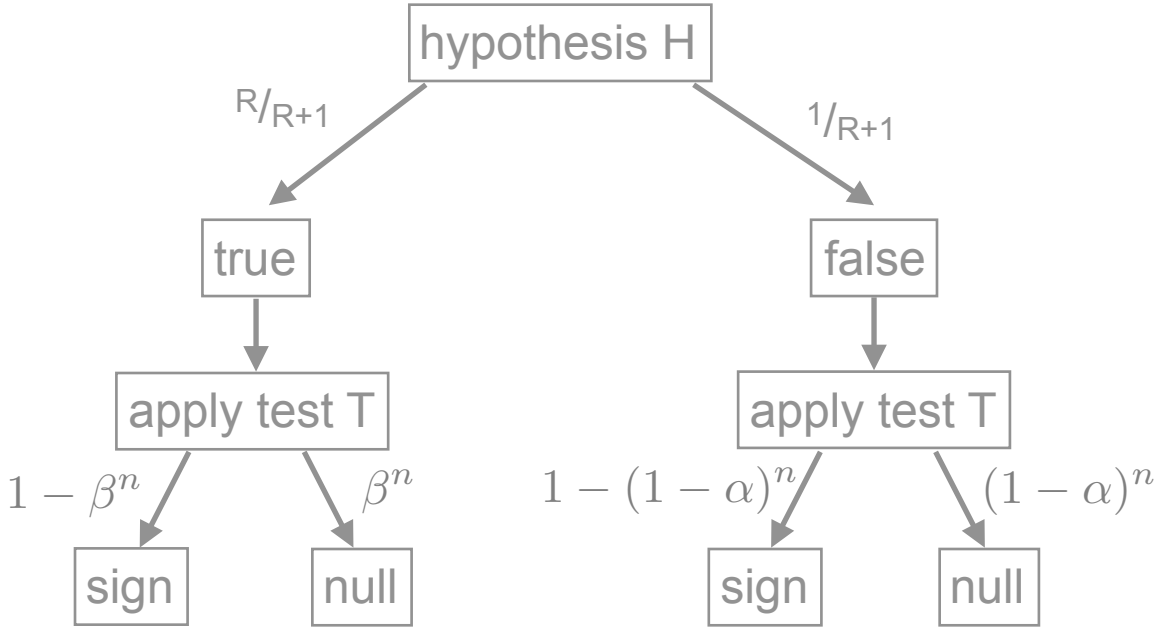
$$P(H = t \mid T = s) = \frac{R(1-\beta) + u\beta R}{R(1-\beta) + u\beta R + \alpha + u(1-\alpha)}$$

p-hacking ::: combination of design/presentation/analysis factors that favor a significant test result beyond the normal alpha level

example:

$R = 1, \quad \beta = 0.2, \quad \alpha = 0.05$





p-fishing ::: reporting at least one significant test results from *n* (equally powered) studies

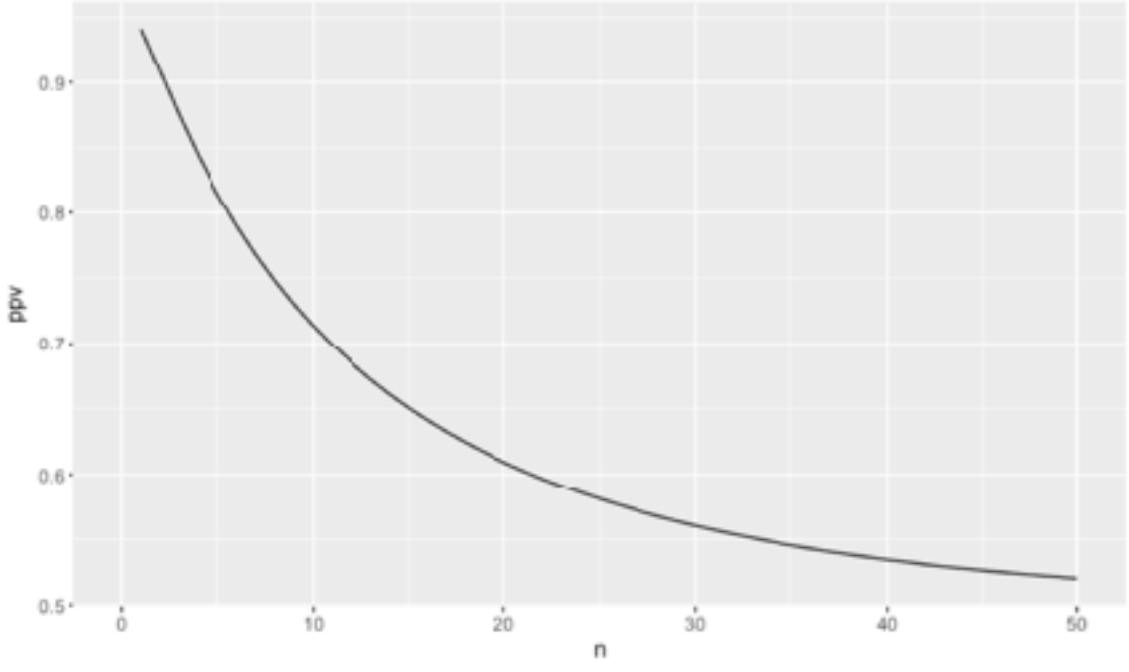
Positive predictive value

$$P(H = t \mid T = s) = \frac{R(1-\beta^n)}{R + 1 - (1-\alpha)^n - R\beta^n}$$

p-fishing ::: reporting at least one significant test results from *n* (equally powered) studies

example:

$R = 1, \quad \beta = 0.2, \quad \alpha = 0.05$



"The method is not to blame for the abuse."


GUNS DONT KILL PEOPLE
PEOPLE KILL PEOPLE

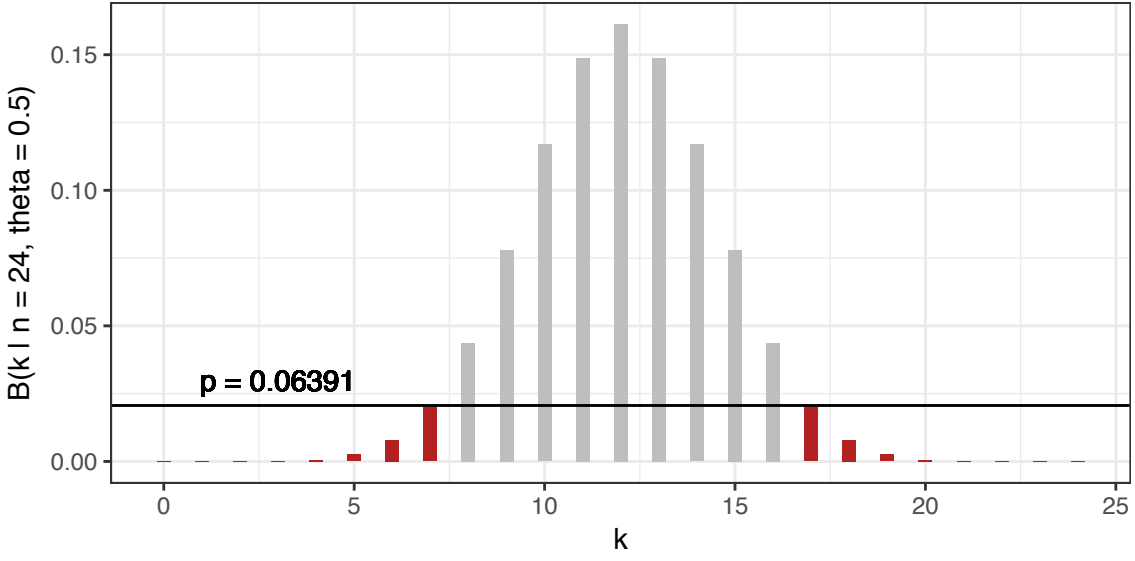
“But the priors! So subjective, so sad! Real shame!”

Stop at n=24

flip coin n=24 times; observe k=7 heads

likelihood function: **binomial distribution**

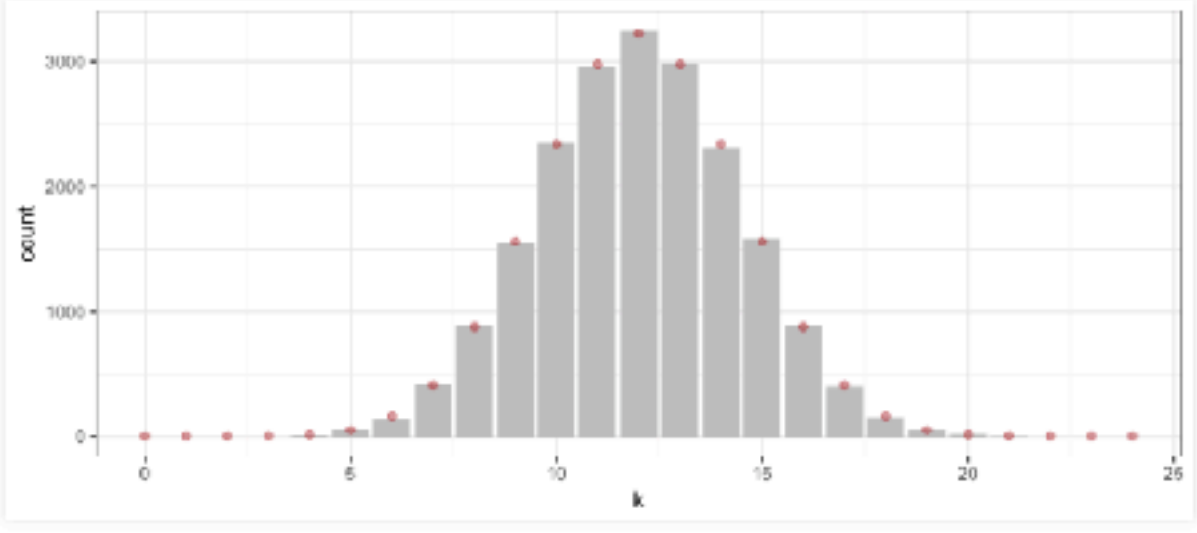
$$B(k; n = 24, \theta = 0.5) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$



Monte Carlo sampling [excursion]

use a large number of samples to approximate a random variable

```
# repeat 24 flips of a fair coin 20,000 times
n.samples = 20000
x.reps = map_int(1:n.samples, function(i) sum(sample(x = 0:1, size = 24, replace = T, prob = c(0.5, 0.5))))
ggplot(data.frame(k = x.reps), aes(x = k)) + geom_bar(fill = "gray") +
  geom_point(data = tibble(k = 0:24, exp = dbinom(0:24, size = 24, prob = 0.5)),
             aes(x = k, y = exp * n.samples), color = "firebrick", alpha = 0.5)
```

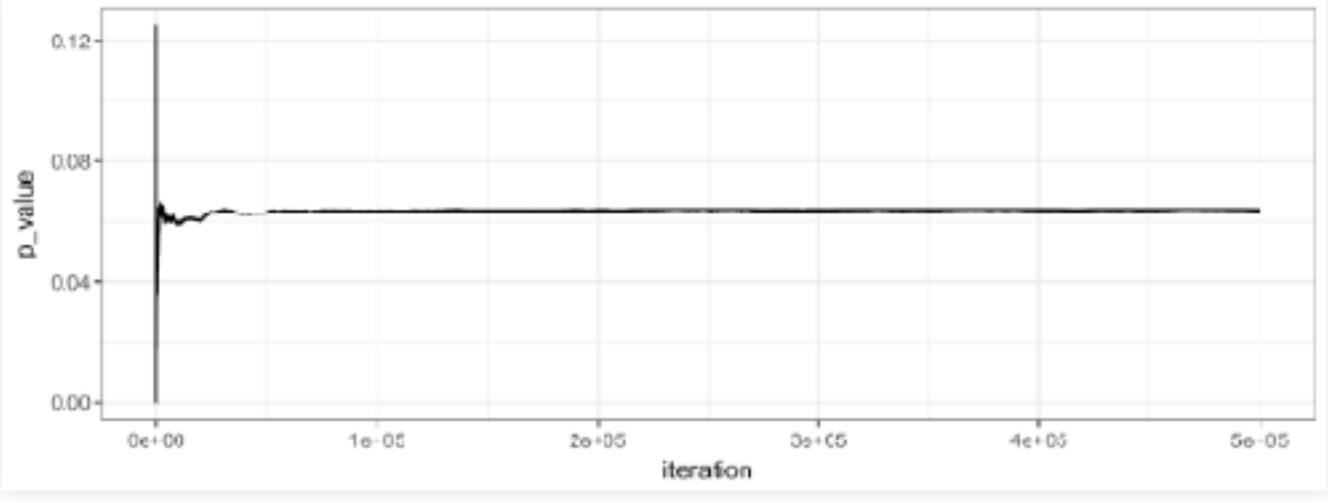


Approximate *p*-value by simulation [excursion]

```
k_obs = 7
n_obs = 24
x_reps = 500000

lhs = map_dbl(1:x_reps, function(i) {
  k_hyp = rbinom(1, size = n_obs, prob = 0.5)
  dbinom(k_hyp, size = n_obs, prob = 0.5)
})

lh_obs = dbinom(k_obs, size = n_obs, prob = 0.5)

mean(lhs <= lh_obs) %>% show()
```

```
## [1] 0.06389
```

```
tibble(iteration = 1:x_reps,
       p_value = cumsum(lhs <= lh_obs) / 1:x_reps) %>%
  ggplot(aes(x = iteration, y = p_value)) + geom_line()
```

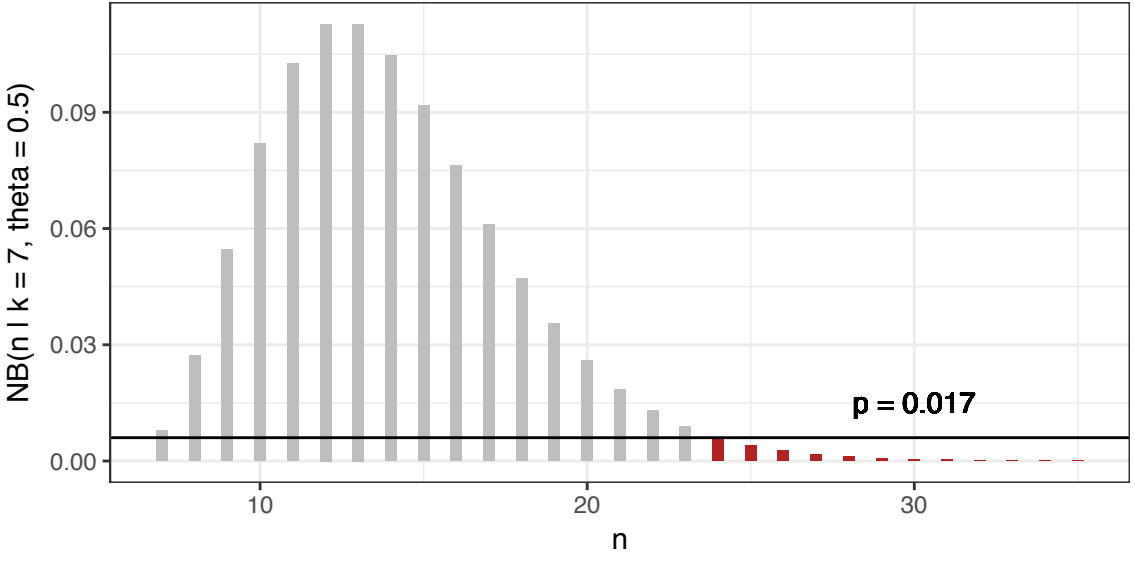


Stop at k=7

flip coin until we had k=7 heads; observe n=24

likelihood function: **negative binomial distribution**

$$NB(n; n = k, \theta = 0.5) = \frac{k}{n}\binom{n}{k}\theta^k(1-\theta)^{n-k}$$

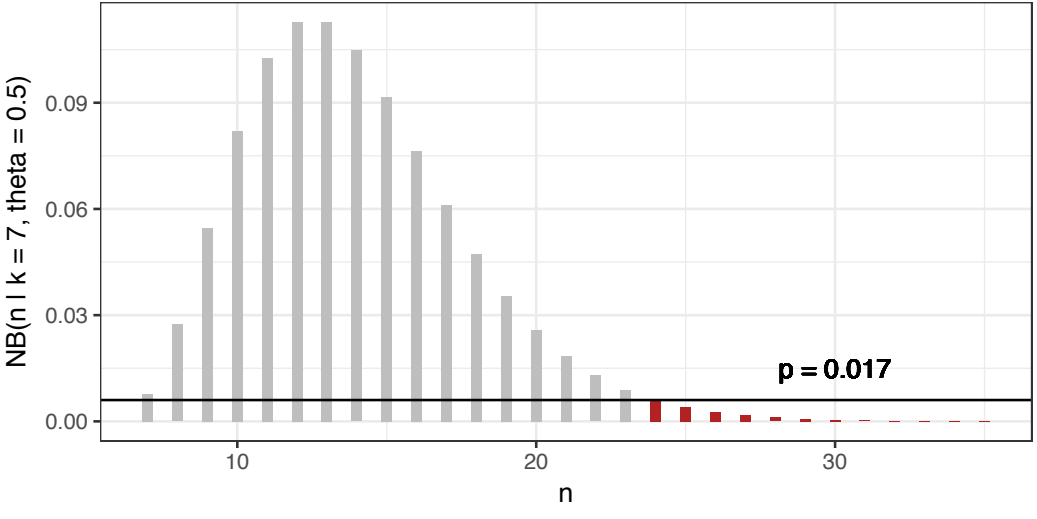


Stop at k=7 [approximation by simulation]

flip coin until we had k=7 heads; observe n=24

likelihood function: **negative binomial distribution**

$$NB(n; n = k, \theta = 0.5) = \frac{k}{n}\binom{n}{k}\theta^k(1-\theta)^{n-k}$$




```
k_obs = 7
n_obs = 24
x_reps = 500000

lhs = map_dbl(1:x_reps, function(i) {
  f_hyp = rnbinom(1, size = k_obs, prob = 0.5)
  dnbinom(f_hyp, size = k_obs, prob = 0.5)
})

lh_obs = dnbinom(n_obs-k_obs, size = k_obs, prob = 0.5)

mean(lhs <= lh_obs) %>% show()
```

```
## [1] 0.017502
```

Stop after a while [approximation by simulation]

collect data for 2 weeks; observe n=24, k=7

likelihood function: Poisson + Binomial process

$n \sim \text{Poisson}(\lambda = 24)$

$k \sim \text{Binomial}(n, \theta = 0.5)$

```
lambda = 24
k_obs = 7
n_obs = 24
n_reps = 500000

lhs = map_dbl(1:n_reps, function(i) {
  n_hyp = rpois(1, lambda = lambda)
  k_hyp = rbinom(1, size = n_hyp, prob = 0.5)
  dbinom(k_hyp, size = n_hyp, prob = 0.5)
})

lh_obs = dbinom(k_obs, size = n_obs, prob = 0.5)

mean(lhs <= lh_obs) %>% show()
```

```
## [1] 0.041226
```

p-values are subjective constructions

researcher intentions determine what it means to “repeat” an experiment

many potential problems:

- exclusion criteria for participants settled after seeing the data
- inability to use data when the way of obtaining it is unknown
    - who can you trust?
    - what's the sampling distribution for large-scale surveys, linguistic corpora...?
- sampling protocol dependent on external circumstance (funding, motivation, ...)
- *ex post* unverifiable researcher reports: was this really what they intended to do?