Week 4

Live Discussion Session

Starts at 2.05pm

Assuming 6 in 49 ball lottery: each ticket has approx. 1/(14 million) probability of winning jackpot since there are

$$\binom{49}{6} = \frac{49!}{43! \times 6!} = 13{,}983{,}816$$ combinations of 6 numbers from 49

The probability a player buys a single ticket on two consecutive weeks and wins each the jackpot each time is

$$\left(\frac{1}{13983816}\right)^2 = \frac{1}{195{,}547{,}109{,}921{,}856} \approx \frac{1}{200\ trillion}$$

But on average each player buys 2 tickets per week so in each week the probability of winning is about 1/7million
So the probability of winning in two consecutive weeks is $(1/7\text{million})^2$

But, in a 20 year period there are 1040 weeks. So each person playing has

$$\binom{1040}{2} = \frac{1040!}{1038! \times 2!} = \frac{1040 \times 1039}{2} = 540{,}280$$ combinations of pairs of weeks when they could win

By the Binomial theorem the Probability each player wins exactly 2 times in 1040 weeks is:

$$540280 \times \left(\frac{1}{7\ million}\right)^2 \times \left(1 - \frac{1}{7\ million}\right)^{1038} \approx 1.0245 \times 10^8$$

That’s just over 1 in 100 million. Still very low

However, assume there are 60 million players.
We need to calculate the probability that at least one player wins twice in a 20-year period.
That is 1 minus the probability that NO player wins twice in 20 years.
The probability a player does NOT win twice in 20 years is:

$$1-(1.0245\times10^{8}) = 0.9999999889755$$

By the Binomial Theorem the probability 60 million players all fail to win twice in 20 years is

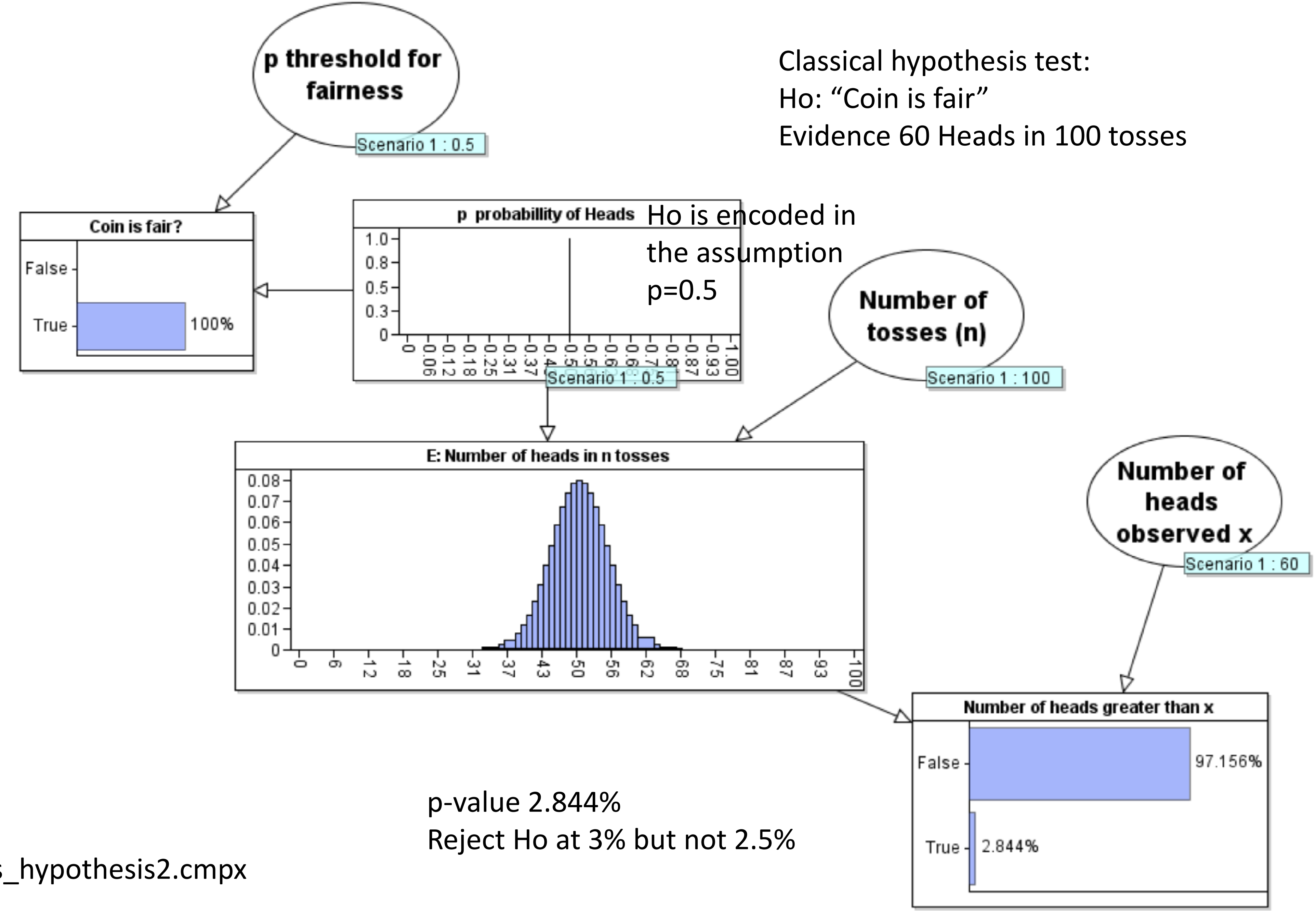
$$0.9999999889755^{60000000} = 0.516$$

So the probability at least one player wins twice in 20 years is

$$1-0.516=0.484$$

In other words there is almost a 50% chance.
Much better then the chance of rolling a 6 on a fair die

Classical hypothesis test:
Ho: “Coin is fair”
Evidence 60 Heads in 100 tosses

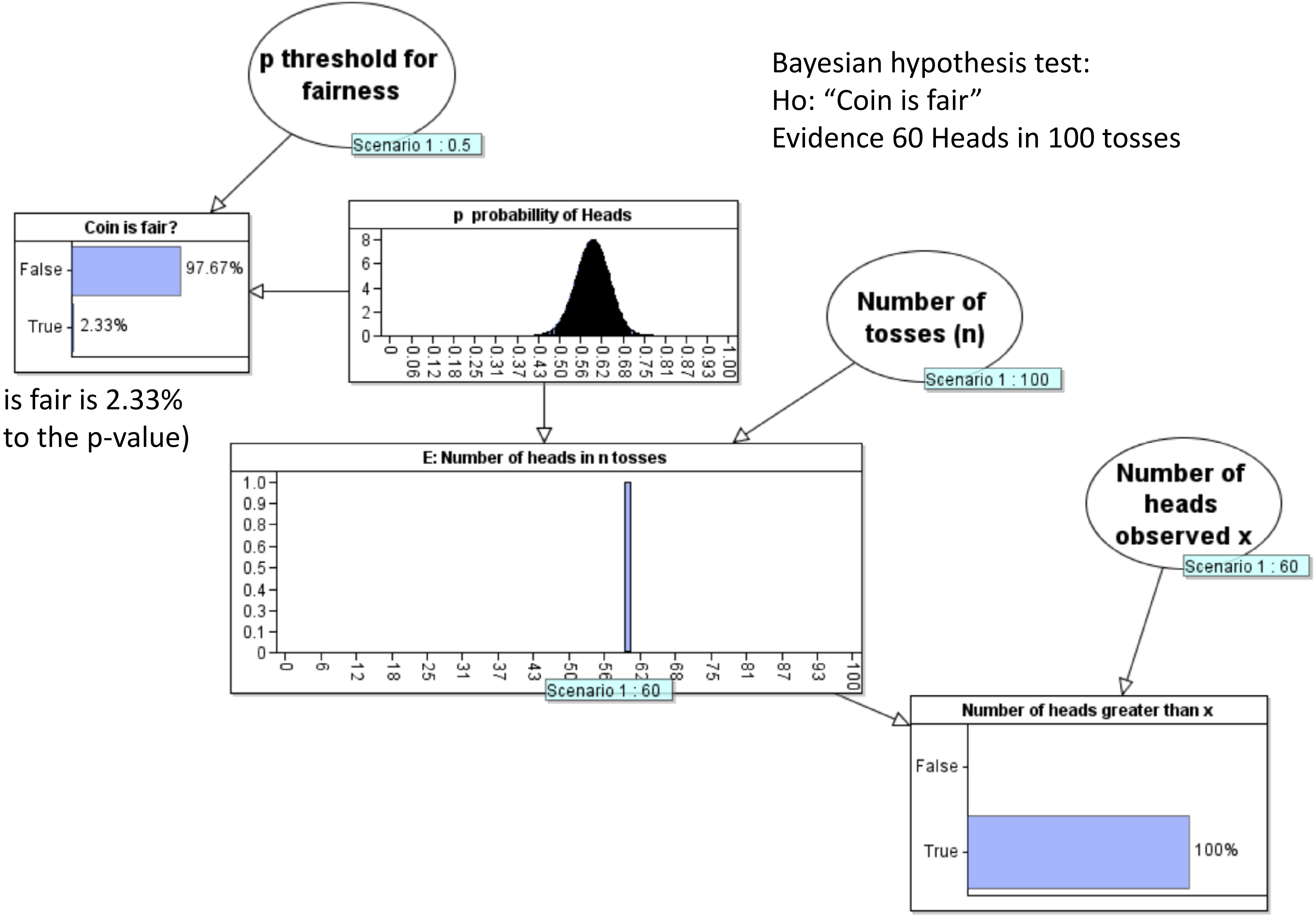


Ho is encoded in the assumption p=0.5

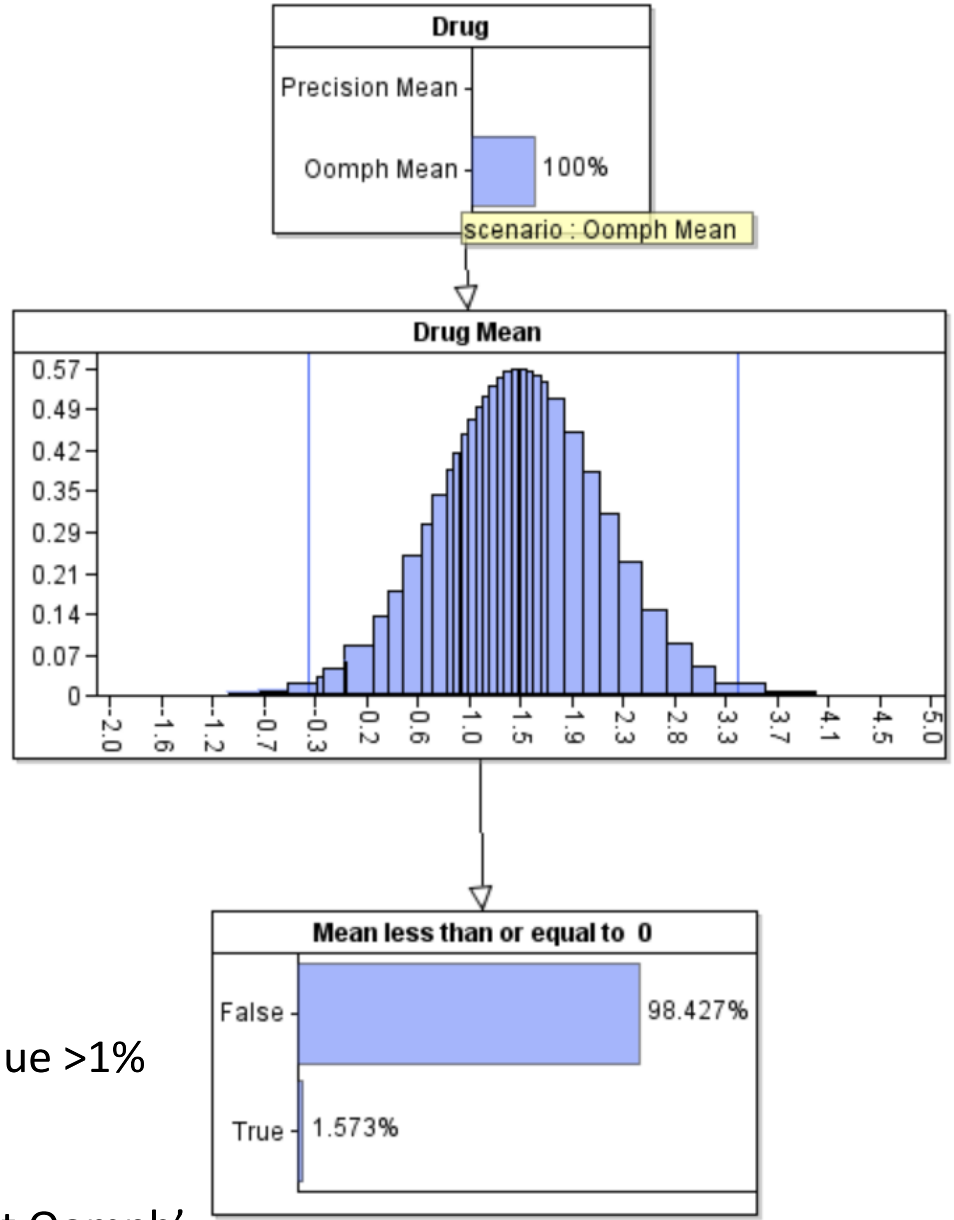
p-value 2.844%
Reject Ho at 3% but not 2.5%

Coin_toss_hypothesis2.cmpx

Bayesian hypothesis test:
Ho: “Coin is fair”
Evidence 60 Heads in 100 tosses

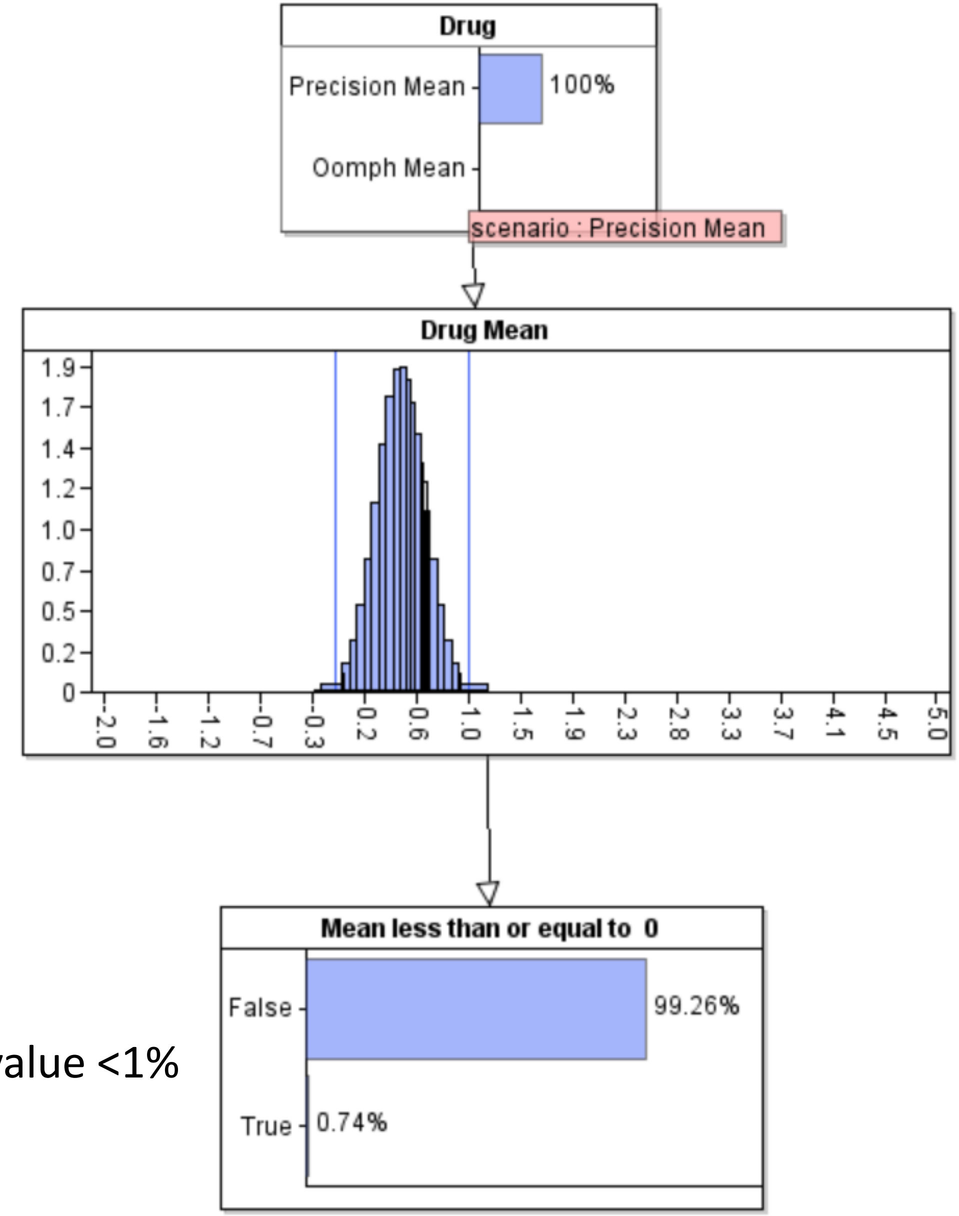


Prob coin is fair is 2.33%
(different to the p-value)



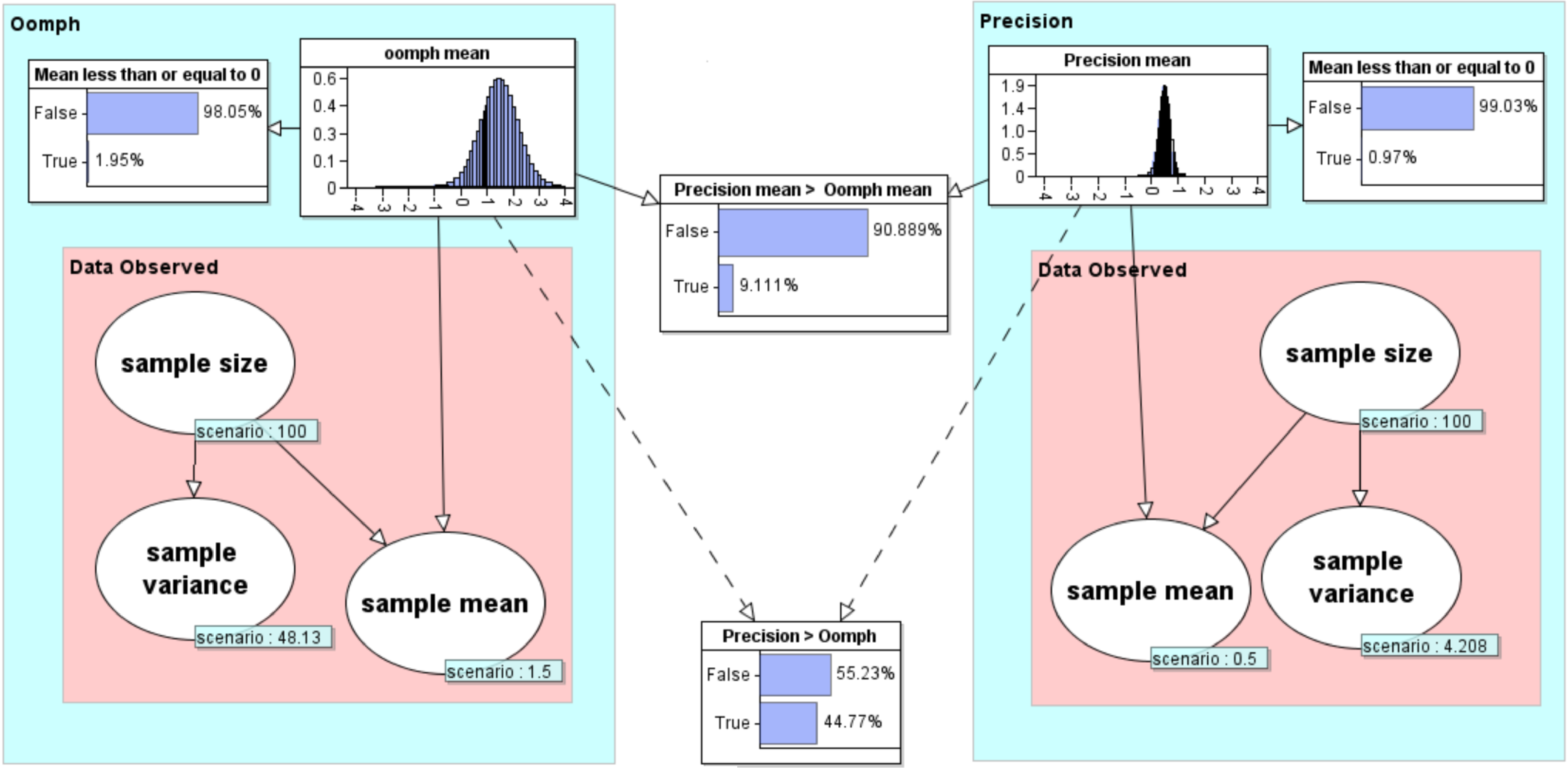
p-value >1%

‘reject Oomph’



p-value <1%

‘accept Precision’



\weight loss hypothesis testing classical v Bayesian_CORRECTED.cmpx

Question 1

A new virus is affecting the population. People who have the virus will normally have specific symptoms such as a cough and the loss of the sense of taste and/or smell.

It is estimated that 1 in 4 of people who suffer these symptoms have the virus and 1 in 1000 people without these symptoms have the virus.

A test for the virus has the following accuracy

- For people with symptoms, the true positive rate is 90% and the false positive rate is 5%
- For people without symptoms, the true positive rate is 80% and the false positive rate is 1%

Answer the following questions:

a) If we know that 10% of the population have symptoms, what percentage of the population has the virus? **[2 marks]**

b) What is the probability that a person with symptoms will test positive? **[2 marks]**

c) What is the probability that a person without symptoms will test positive? **[2 marks]**

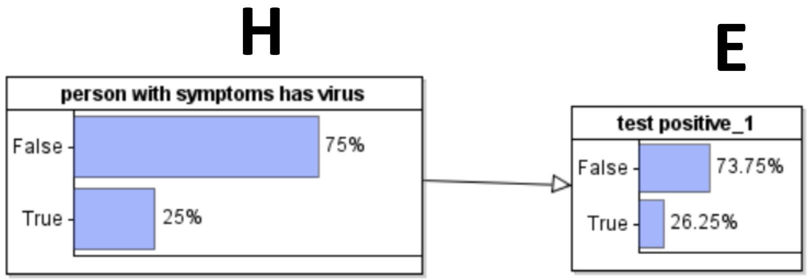
d) A person with symptoms tests positive. What is the probability they have the virus? **[2 marks]**

e) A person with symptoms tests negative. What is the probability they have the virus? **[2 marks]**

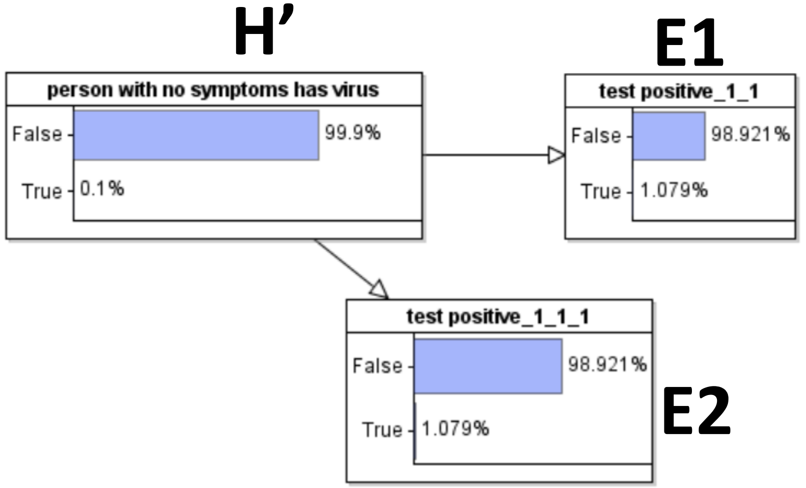
f) A person without symptoms tests positive. What is the probability they have the virus? **[2 marks]**

g) A person without symptoms tests positive and is subject to an additional test. Assuming that a second test is independent of the first, what is the probability they test positive in this second test? **[4 marks]**

h) A person without symptoms tests positive in both the first and second test. What is the probability they have the virus? **[4 marks]**



P(E|H)=0.9
P(E|not H)=0.05



P(E1|H')=0.8
P(E1|not H')=0.01
(same for E2)

a) P(virus) = P(virus with symptoms)*P(symptoms) + P(virus no symptoms)*P(no symptoms)
=P(H)*P(symptoms) + P(H')*P(no symptoms)= **(0.25 x 0.1)+(0.001 x 0.9) = 0.0259 = 2.59%**
as10% of the population have symptoms

b) What is the probability a person with symptoms will test positive?

$P(E) = P(E|H) \times P(H) + P(E|not\ H) \times P(not\ H)$ =0.9 x 0.25 + 0.05 x 0.75 = 0.2625 = 26.25%

c) What is the probability a person without symptoms will test positive?

$P(E1) = P(E1|H') \times P(H') + P(E1|not\ H') \times P(not\ H')$=0.8 x 0.001+ 0.01 x 0.999 = 0.01079 = 1.079%

d) A person with symptoms tests positive. What is the probability they have the virus?

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} = \frac{0.9 \times 0.25}{0.2625} = 0.857 = 85.7\%$$

e) A person with symptoms tests negative. What is the probability they have the virus?

$$P(H|notE) = \frac{P(not\ E|H) \times P(H)}{P(not\ E)} = \frac{0.1 \times 0.25}{0.7375} = 0.0339 = 3.39\%$$

f) A person with no symptoms tests positive. What is the probability they have the virus?

$$P(H'|E1) = \frac{P(E1|H') \times P(H')}{P(E1)} = \frac{0.8 \times 0.001}{0.01079} = 0.07414 = 7.414\%$$

g) A person without symptoms tests positive. Assuming second test is independent of first, what is the probability they test positive in a second test?

We know P(H'|E1) = 0.07414 Let H" = H'|E1
Then P(E2) = P(E2|H") x P(H") + P(E2 | not H") x P(not H") = 0.8 x 0.07414 + 0.01 x 0.92596 = 0.06857 = 6.857%

h) A person without symptoms tests positive in both the first and second test. What is the probability they have the virus?
We know from f) that after the first positive test the revised posterior probability of H' is 0.07414. But from g) we know the probability of testing positive on second if the first was positive is 0.06857. So we now have revised posteriors of P(H')=0.07414 and P(E2)=0.06857

$$P(H'|E2) = \frac{P(E2|H') \times P(H')}{P(E2)} = \frac{0.8 \times 0.07414}{0.06857} = 0.86498 = 86.498\%$$