Errors in 2018 2nd Edition

Last updated 25 April 2022

Page 12, Sidebar 2.2 "less than 1 in 30,000" should say "less than 1 in 300,000" (thanks to Scott McLachlan for noticing that)

Page 19, Box 2.3 "When the correlation coefficient is close to zero there is little evidence of any relationship." should be replaced by "…there is little evidence of a linear relationship". [as pointed out by Mike Nikolou who adds *"But there may well be a strong nonlinear relationship. Example: Air-to-fuel ratio, X, and flame temperature, T, in a burner; T peaks for a certain X and goes down for higher or lower X, leading to almost zero corr. coef., yet the effect of X on Y is strong. I would even venture to say that in your example "a person's age against the number of toys they possess" the corr. coef. could be close to zero, if you count as "toys" all the gadgets etc. that senior people possess as aids for age-reduced capacity (sigh). Again, X = Age and Y = Number of Toys are strongly related, but not linearly."*

Page 20 To be consistent with the full example in Chapter 12 (which we also need to update – see below) we need to change the text of the two bullet points near the bottom of the page as follows:

- For drug Precision the mean weight loss for 100 subjects is 0.5 lbs with sample standard deviation 2
- For drug Oomph the mean weight loss for 100 subjects is 1.5 lbs with sample standard deviation 6

Page 21 Figure 2.15 there should be no arrow connecting Height to Intelligence

Page 34 The numbers in the example relating to absolute versus relative risk for mouth cancer are wrong. They should be as follows:

Assume 500,000 deaths per year in the UK of which 84 are from mouth cancer. Assume 20% of those who died were 'wine drinkers' and that the mouth cancer deaths in each of the two categories were:

	Mouth cancer deaths	% deaths from mouth cancer
Wine drinkers (100,000)	36	0.036%
Non-wine drinkers (400,000)	48	0.012%

So, the relative risk is tripled (from 0.012% to 0.036%) but the absolute risk only increases 0.024%

Page 40 Figure 1.27 should be Figure 2.27 (thanks to Omar Verduga for pointing that out)

Pag 68, Box 4.1 4th paragraph:

"… our probabilities represent **casual** mechanisms..."

should be

> "… our probabilities represent **causal** mechanisms..."

Page 78, Second "Frequentist definition of chance…is 1/n": Mike Nikolaou points out that his is not a definition. Rather it is a model for assessment of probability. He says "*The frequentist definition of probability remains the fraction in the limit of infinite repetitions. It just so happens that in the case of dice the model is exceptionally simple. In other cases (e.g. distribution of molecular velocity in a gas) building such a model is more complicated (for gases it took the genius of Maxwell to sort it out).*"

Page 153 Table 6.4 $P(data) = 0.168756$ and so the denominator used in the calculation of the posterior is equal to 0.168756 rather than 0.08257. The Posterior result shown in the table is correct despite this error.

Page 157 Figure (7.1) lower case "p" should be replaced with upper case "P" i.e.

$$P(T = True \,|N = True) = \frac{P(N = True \,|T = True)P(T = True)}{P(N = True)}$$

Page 160 Box 7.1 Part 2. Refers to Theorem 4.1 when it should refer to Theorem 5.1.

Page 209 Figure 8.13(b). The table should be replaced with this one (thanks to Pascal Bercker for noting this)

Disease	False	True
False	0.95	0.01
True	0.05	0.99

Page 259: Figure 9.10: Bottom left node label should be "CPU" not "PCI" (thanks to Pascal Bercker for noting this).

Page 385-386. Example 12.2 This entire example should be replaced with the following.

> Consider the hypothesis *H*:
>
> *H:* "People taking the drug lose weight over a 6-month period"
>
> ***Null hypothesis (not H)***: "People taking the drug lose no weight over a 6-month period"
>
> ***D***: We observe 100 people using the drug. The average (mean) weight loss is 0.5 lb and the sample standard deviation 2.05. The *standard deviation of the mean* is then calculated as
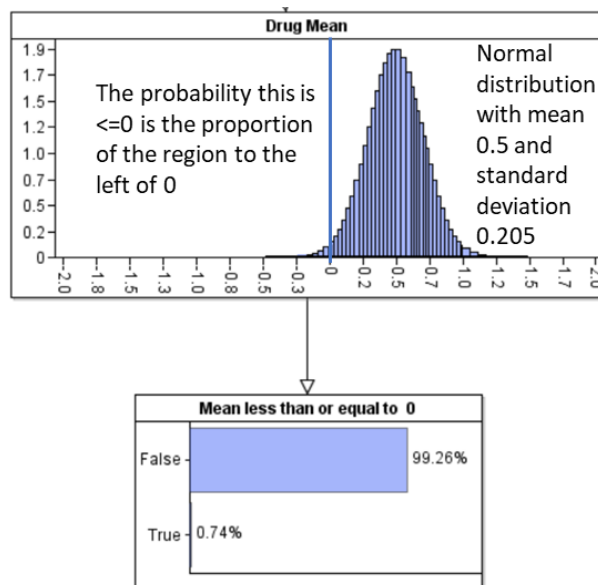>
> $$\frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}} = \frac{2.05}{10} = 0.205$$
>
> Providing (as in the case) that the sample size is at least 30, the 'classic' way to estimate P(D | not H) is to:

1. Assume the (true) *mean weight loss* has a Normal distribution whose mean is 0.5 and standard deviation 0.205

2. Then calculate the probability that this distribution is less than or equal to zero

Using standard tables, excel or AgenaRisk you can see in this case that the probability (which is also called the p-value) is 0.0074, i.e. 0.74%

As the p-value < 1% we reject the null hypothesis at the 1% p-value level – and hence 'accept' that there is 'significant' support for H.
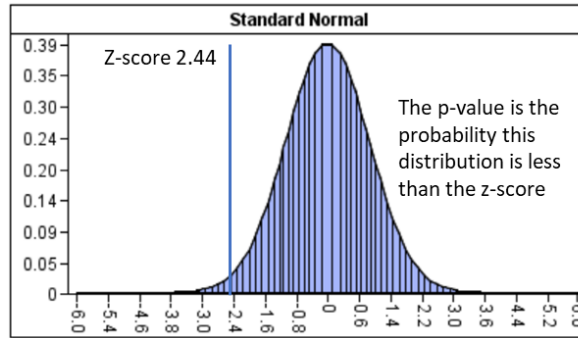

In the old days it was necessary to 'transform' the particular Normal distribution into a 'standard Normal distribution' (one with mean 0 and standard deviation 1) in order to calculate the p-value because – in the absence of computers – people relied on tables that had the standard normal distribution.

To do the transformation we calculate what is called the z-score:

$$z = \frac{\text{sample mean } - \text{ null hypothesis mean}}{\text{standard deviation of mean}} \qquad = \frac{0.5-0}{0.205} = 2.44$$

This z-score is the distance from the mean of the 'standard Normal distribution' (one with mean 0 and standard deviation 1). The p-value – which is exactly equivalent to the p-value 0.0074 we previously calculated is equal to the probability that the distribution is less than the z-score.

Tables of standardized Z-scores show that any value above 2.326 has a probability less than 1% - so we can reject the null hypothesis at the 1% level.
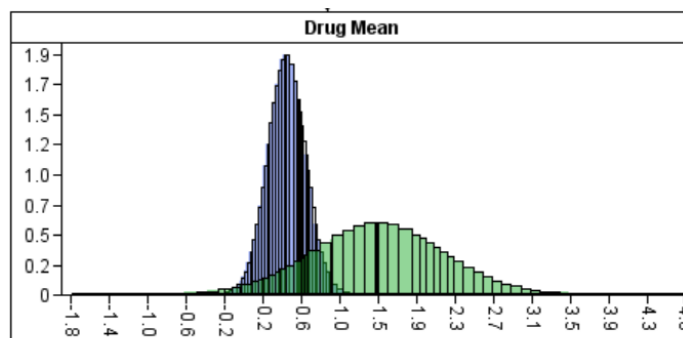
Standard Normal

Z-score 2.44

The p-value is the probability this distribution is less than the z-score

So now we ask: which of the following two weight loss drugs is best?

Precision: the mean weight loss for 100 subjects is 0.5 lbs with sample standard deviation 2.05 (so standard deviation of mean is 0.205)

Z-score 2.44, p-value essentially 0.0074 (i.e. 0.74%). **Null hypothesis rejected at 1%**

Oomph: the mean weight loss for 100 subjects is 1.5 lbs with sample standard deviation 6.97 (so standard deviation of mean is 0.697)
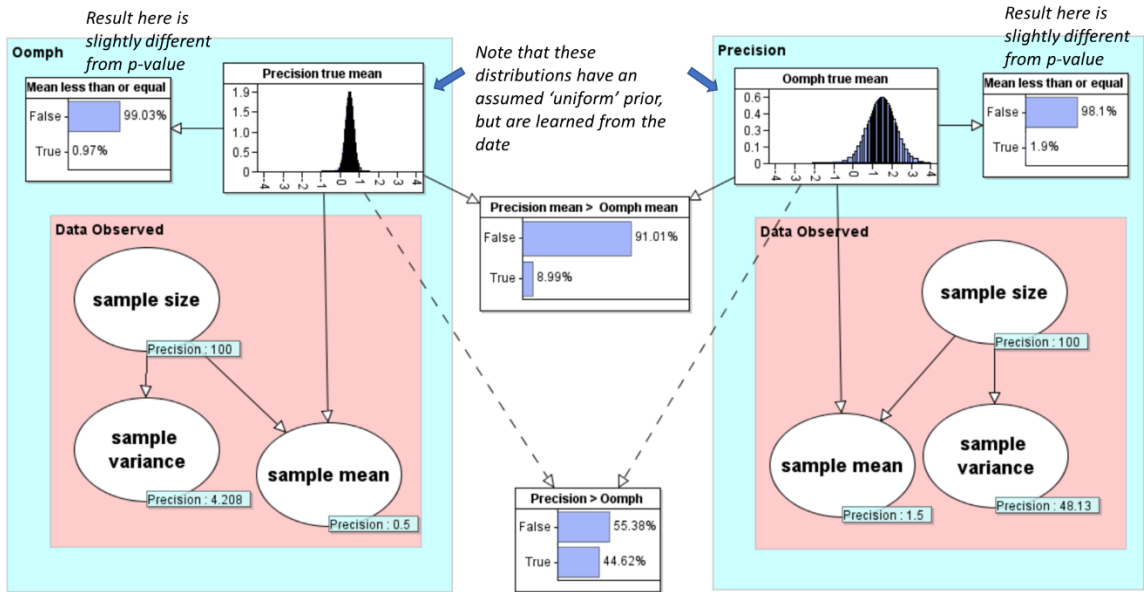
Z-score 2.15, p-value 0.016 (i.e. 1.6%). **Null hypothesis NOT rejected at 1%**
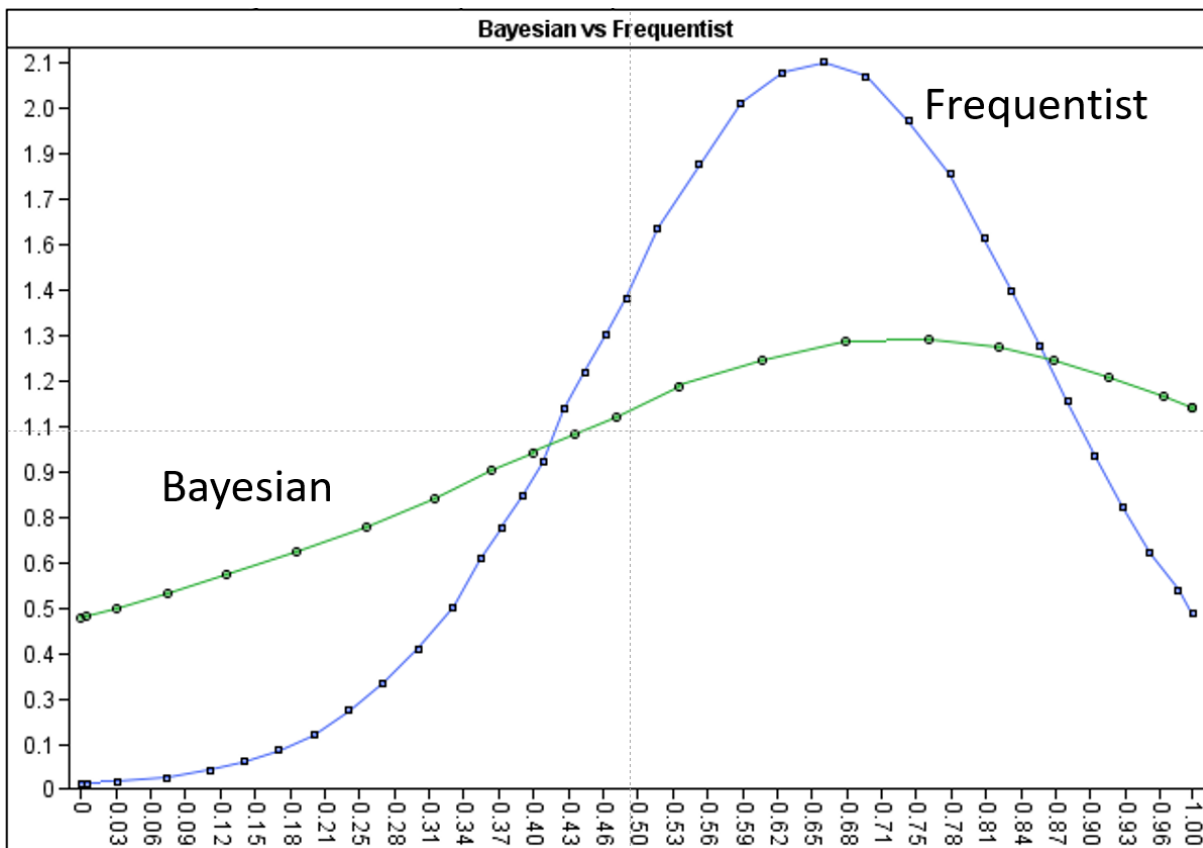

Drug Mean

Precision is blue, Oomph is green

Mean less than or equal to 0

False	99.26%
	98.427%
True	0.74%
	1.573%

The full BN model to determine which is 'best':

Page 326 Figure 10.25 should be replaced with:

Page 427 Example 13.1 $P(Collision) = 0.01 \times 0.1 \times 0.1 \times 0.5 \times 0.5 = 0.000025$

Page 467 penultimate paragraph replace "….to have even high reliability…." with "….to have even higher reliability…".

1. First we fit the model for each hypothesis. In this example we can assume that the binomial distribution is an appropriate choice, where q is the probability of the death penalty and is the only parameter we need to fit. We can fit the parameters from the data $\{n, r\}$ for the number of cases, n, and the number of death penalties applied, r.

Let's take each hypothesis in turn

$H_1 : S$ independent of V and D

We need to consider the case where V and D are independent of S :
$$P(q \mid r = 36, n = 326, H_1) \propto P(q \mid H_1)P(r = 36 \mid q_{ww}, n = 326, H_1)$$

Here we are effectively pooling the data. When we calculate the result in AgenaRisk, using a uniform prior distribution, $q \sim U[0,1]$, and a Binomial likelihood, $P(q \mid r = 36, n = 326, H_1) \sim Bin(r = 36, n = 326)$, we get the full marginal distribution, with a mean of 0.11.

$H_2 : S$ dependent on D

Now, here we have two parameters for the death penalty each dependent on different values of the parent variable, $D = \{D - White, D - Black\}$. We will label these $q_{D-White}$ and $q_{D-Black}$ and apply uniform priors as before.

$$P(q_{D-White} \mid r = 19, n = 160, H_2) \propto P(q_{D-White} \mid H_2)P(r = 19 \mid q_{D-White}, n = 160, H_2)$$
$$P(q_{D-Black} \mid r = 17, n = 166, H_2) \propto P(q_{D-Black} \mid H_2)P(r = 17 \mid q_{D-Black}, n = 166, H_2)$$

The mean values for the death penalty probability are 0.12 and 0.107 respectively.

$H_3 : S$ dependent on V

This is similar to the previous case with states $V = \{V - White, V - Black\}$
$$P(q_{V-White} \mid r = 30, n = 214, H_3) \propto P(q_{V-White} \mid H_3)P(r = 30 \mid q_{V-White}, n = 214, H_3)$$
$$P(q_{V-Black} \mid r = 6, n = 112, H_3) \propto P(q_{V-Black} \mid H_3)P(r = 6 \mid q_{V-Black}, n = 112, H_3)$$

The mean values for the death penalty probability are 0.143 and 0.061 respectively.

$H_4 : S$ dependent on V and D

This is similar to the previous case with death penalty variables for each combination of $D = \{D-White, D-Black\}$ and $V = \{V-White, V-Black\}$:

$$\{V-White/D-White, V-White/D-Black, V-Black/D-White, V-Black/D-Black\}$$
thus:

$$P(q_{V-White/D-White} \mid r = 19, n = 151, H_4) \propto P(q_{V-White/D-White} \mid H_4)P(r = 19 \mid q_{V-White/D-White}, n = 151, H_4)$$
$$P(q_{V-White/D-Black} \mid r = 11, n = 63, H_4) \propto P(q_{V-White/D-Black} \mid H_4)P(r = 11 \mid q_{V-White/D-Black}, n = 63, H_4)$$
$$P(q_{V-Black/D-White} \mid r = 0, n = 9, H_4) \propto P(q_{V-Black/D-White} \mid H_4)P(r = 0 \mid q_{V-Black/D-White}, n = 9, H_4)$$
$$P(q_{V-Black/D-Black} \mid r = 6, n = 103, H_4) \propto P(q_{V-Black/D-Black} \mid H_4)P(r = 6 \mid q_{V-Black/D-Black}, n = 103, H_4)$$

The mean values for the death penalty probability are 0.131, 0.184, 0.09 and 0.066 respectively.

2. Next we calculate the Bayesian scores for all hypotheses:

$H_1 : S$ independent of V and D

We need to consider the case where V and D are independent of S. Unlike when data fitting we need to predict all of the cells in the original data table, i.e. every combination of states being considered, in order to assess how whether the model is surprised by the data:

$$P(r = 19, n = 151 \mid H_1) = \sum_q \left[P(q \mid r = 36, n = 326, H_1)P(r = 36, n = 326 \mid q, H_1) \right] = 0.071$$
$$P(r = 11, n = 63 \mid H_1) = \sum_q \left[P(q \mid r = 36, n = 326, H_1)P(r = 36, n = 326 \mid q, H_1) \right] = 0.049$$
$$P(r = 0, n = 9 \mid H_1) = \sum_q \left[P(q \mid r = 36, n = 326, H_1)P(r = 36, n = 326 \mid q, H_1) \right] = 0.034$$
$$P(r = 6, n = 103 \mid H_1) = \sum_q \left[P(q \mid r = 36, n = 326, H_1)P(r = 36, n = 326 \mid q, H_1) \right] = 0.036$$

The joint probability of the data is therefore:

$$P(r = 19, n = 151 \mid H_1)P(r = 11, n = 63 \mid H_1)P(r = 0, n = 9 \mid H_1)P(r = 6, n = 103 \mid H_1)$$
and this equals 4.26E-06.

From the example AgenaRisk model you will see that this is simply the marginal probability of the data nodes (a trick for getting these is to switch off dynamic discretization in these nodes and insert the state for the data value you need to calculate the probability of).

$H_2 : S$ dependent on D

This is similar to the previous case except we need to condition each calculation on the particular combination of parent variables we are considering:

$$P(r=19,n=151\,|\,H_2)=\sum_{q_{D-White}}\left[P(q_{D-White}\,|\,r=19,n=160,H_2)P(r=19,n=160\,|\,q_{D-White},H_2)\right]=0.068$$

$$P(r=0,n=9\,|\,H_2)=\sum_{q_{D-White}}\left[P(q_{D-White}\,|\,r=19,n=160,H_2)P(r=19,n=160\,|\,q_{D-White},H_2)\right]=0.315$$

$$P(r=11,n=63\,|\,H_2)=\sum_{q_{D-Black}}\left[P(q_{D-Black}\,|\,r=17,n=166,H_2)P(r=17,n=166\,|\,q_{D-Black},H_2)\right]=0.042$$

$$P(r=6,n=103\,|\,H_2)=\sum_{q_{D-Black}}\left[P(q_{D-Black}\,|\,r=17,n=166,H_2)P(r=17,n=166\,|\,q_{D-Black},H_2)\right]=0.052$$

The joint probability of the data is therefore:

$$P(r=19,n=151\,|\,H_2)P(r=0,n=9\,|\,H_2)P(r=11,n=63\,|\,H_2)P(r=6,n=103\,|\,H_2)$$

and this equals 4.67 E-5.

$\underline{H_3: S \text{ dependent on } V}$

Again we condition appropriately:

$$P(r=19,n=151\,|\,H_3)=\sum_{q_{D-White}}\left[P(q_{D-White}\,|\,r=30,n=214,H_3)P(r=30,n=214\,|\,q_{D-White},H_3)\right]=0.067$$

$$P(r=11,n=63\,|\,H_3)=\sum_{q_{D-Black}}\left[P(q_{D-Black}\,|\,r=30,n=214,H_3)P(r=30,n=214\,|\,q_{D-Black},H_3)\right]=0.042$$

$$P(r=0,n=9\,|\,H_3)=\sum_{q_{D-White}}\left[P(q_{D-White}\,|\,r=6,n=112,H_3)P(r=6,n=112\,|\,q_{D-White},H_3)\right]=0.575$$

$$P(r=6,n=103\,|\,H_3)=\sum_{q_{D-Black}}\left[P(q_{D-Black}\,|\,r=6,n=112,H_3)P(r=6,n=112\,|\,q_{D-Black},H_3)\right]=0.119$$

The joint probability of the data is therefore:

$$P(r=19,n=151\,|\,H_3)P(r=0,n=9\,|\,H_3)P(r=11,n=63\,|\,H_3)P(r=6,n=103\,|\,H_3)$$

and this equals 1.93 E-4.

$\underline{H_4: S \text{ dependent on } V \text{ and } D}$

Again we need to consider all combinations:

$$P(r=19,n=151\,|\,H_4)$$
$$=\sum_{q_{V-White/D-White}}\left[P(q_{V-White/D-White}\,|\,r=19,n=151,H_4)P(r=19,n=151\,|\,q_{V-White/D-White},H_4)\right]$$
$$=0.068$$
$$P(r=11,n=63\,|\,H_4)$$
$$=\sum_{q_{V-Black/D-White}}\left[P(q_{V-Black/D-White}\,|\,r=11,n=63,H_4)P(r=11,n=63\,|\,q_{V-Black/D-White},H_4)\right]$$
$$=0.092$$

$$P(r = 0, n = 9 \mid H_4)$$

$$= \sum_{q_{V-White/D-Black}} \left[P(q_{V-White/D-Black} \mid r = 0, n = 9, H_4) P(r = 0, n = 9 \mid q_{V-White/D-Black}, H_4) \right]$$

$$= 0.525$$

$$P(r = 6, n = 103 \mid H_4)$$

$$= \sum_{q_{V-Black/D-Black}} \left[P(q_{q_{V-Black/D-Black}} \mid r = 6, n = 103, H_4) P(r = 6, n = 103 \mid q_{V-Black/D-Black}, H_4) \right]$$

$$= 0.115$$

The joint probability of the data is 3.78 E-4..
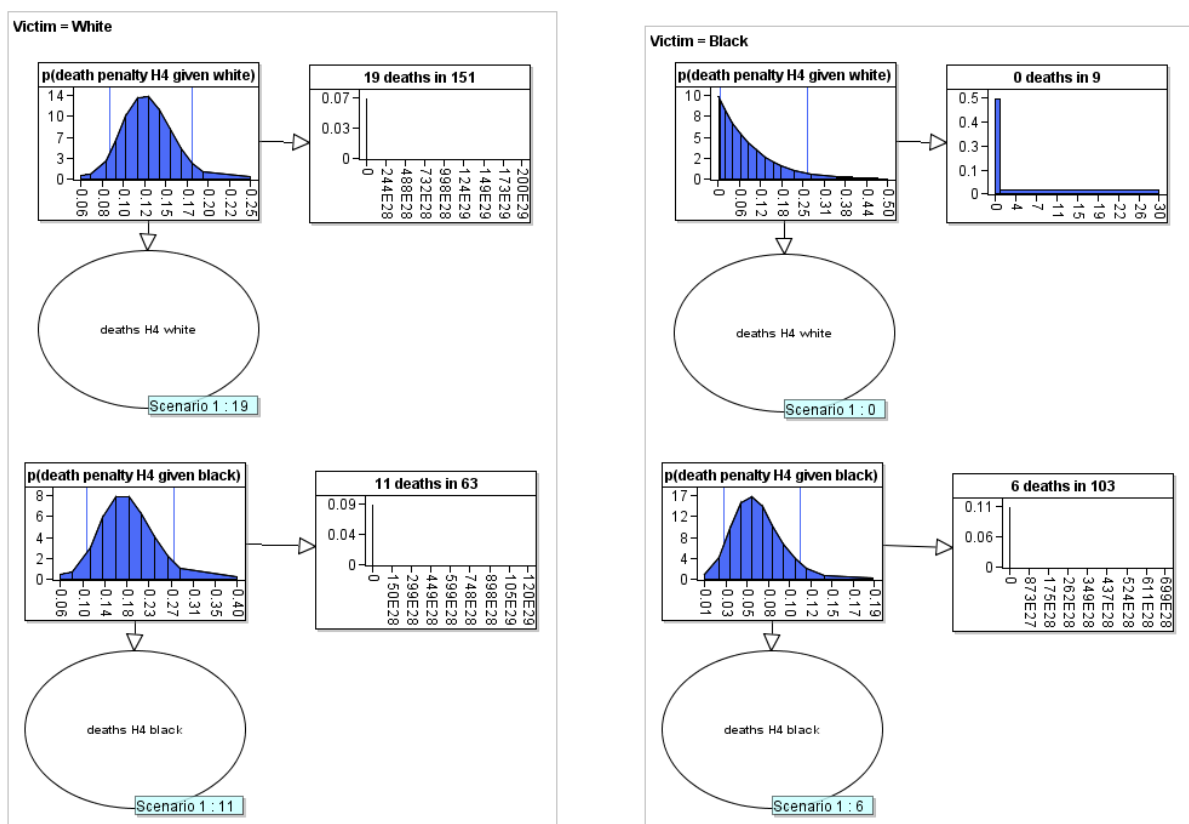
Figure 12.24 replaced with:

Figure 12.24 Data Fitting score calculations for Example 12.9 hypothesis scores.

Page 415 replace penultimate two paragraphs with:

The results for all the hypotheses, after normalization, are:

$$P(D|H_1: S \text{ independent of } V \text{ and } D) = 0.0068$$
$$P(D|H_2: S \text{ dependent on } D) = 0.075$$
$$P(D|H_3: S \text{ dependent on } V) = 0.31$$
$$P(D|H_4: S \text{ dependent on } V \text{ and } D) = 0.61$$

So given that H_4 has the greatest score we might conclude that both the victim's and the defendant's race are significant causal factors in determining whether the death penalty was applied. Of course, there may be other factors that are unaccounted so we cannot conclude that race is the only factor. Neither can we conclude that the race factors are not surrogates for any other unobserved factors.