

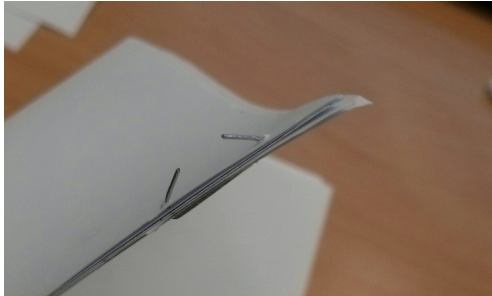
MAS2317/3317

Assignment 2 feedback

March 24, 2014

Stapling

Several assignments were poorly stapled



Staples like this are dangerous and can cause pages to go missing. Many of your assignments had to be re-stapled. It is *your responsibility* to make sure your work is held together securely and safely, by pushing the stapler down *firmly*.

Assignments with unsafe staples will not be marked!

Question 1

We are told that

Henry Cavendish made 23 experimental determinations of the Earth's density. You are provided with data.

Cavendish asserts that the **error standard deviation** of these measurements is 0.2g/cm^3 .

Question 1

We are told that

Henry Cavendish made 23 experimental determinations of the Earth's density. You are provided with data.

Cavendish asserts that the **error standard deviation** of these measurements is 0.2g/cm^3 .

Common problem

You should **check his assertion** that $\sigma = 0.2$ is correct before continuing!

We actually find from the data that $s = 0.1924$, which is close so it should be fine to use.

(-1 marks) if you failed to check $\sigma \approx 0.2$.

Question 14

Part (e)

A football fan tells us it is quite unlikely that her team will get relegated this year. Let $\theta = \Pr(\text{her team will be relegated})$. We are able to elicit from her that there is less than a 1 in 100,000 chance that $\theta > 0.5$.

Pick the most suitable of the 4 distributions to model this situation.

Question 14

Part (e)

A football fan tells us it is quite unlikely that her team will get relegated this year. Let $\theta = \Pr(\text{her team will be relegated})$. We are able to elicit from her that there is less than a 1 in 100,000 chance that $\theta > 0.5$.

Pick the most suitable of the 4 distributions to model this situation.

The correct answer:

We need to have that $\Pr(\theta > 0.5) < 0.00001$:

$$\Pr(\theta > 0.5) = 2.028018e - 05 (= 0.00002) \quad \text{for Be}(10, 37)$$

$$\Pr(\theta > 0.5) = 9.310761e - 10 (= 0.000000001) \quad \text{for Be}(20, 77)$$

meaning that $\theta \sim \text{Be}(20, 77)$ from part (b) is most suitable.

Question 14

Part (e)

A football fan tells us it is quite unlikely that her team will get relegated this year. Let $\theta = \Pr(\text{her team will be relegated})$. We are able to elicit from her that there is less than a 1 in 100,000 chance that $\theta > 0.5$.

Pick the most suitable of the 4 distributions to model this situation.

The correct answer:

We need to have that $\Pr(\theta > 0.5) < 0.00001$:

$$\Pr(\theta > 0.5) = 2.028018e - 05 (= 0.00002) \quad \text{for Be}(10, 37)$$

$$\Pr(\theta > 0.5) = 9.310761e - 10 (= 0.000000001) \quad \text{for Be}(20, 77)$$

meaning that $\theta \sim \text{Be}(20, 77)$ from part (b) is most suitable.

Don't pick $\text{Be}(10, 37)$ for the reason "it's the *nearest value*".

Question 15

Brief summary of task

Data from the case study about traffic accident blackspots in South Tyneside. Perform a regression on the data.

Question 15

Brief summary of task

Data from the case study about traffic accident blackspots in South Tyneside. Perform a regression on the data.

General advice

- Always **include relevant R output**: In this case, the R output that shows the parameters for your regression equation. (-1 marks)
- **Comment on all plots**: Explain what you can learn about the data from the plot. (-2 marks)

Question 15

Write the regression equation

Most students wrote:

$$y = -4.2881 + 0.157x$$

This is missing crucial information a regression equation contains.

Question 15

Write the regression equation

Most students wrote:

$$y = -4.2881 + 0.157x$$

This is missing crucial information a regression equation contains.

A good regression equation

$$y_j = -4.2881 + 0.157x_j + \varepsilon_j, \quad \varepsilon_j \sim N(0, \sigma^2),$$

where x_j is the average observed vehicle speed and y_j is the number of casualties, for each reference site $j = 1, \dots, 67$.

Question 15

Write the regression equation

Most students wrote:

$$y = -4.2881 + 0.157x$$

This is missing crucial information a regression equation contains.

A good regression equation

$$y_j = -4.2881 + 0.157x_j + \varepsilon_j, \quad \varepsilon_j \sim N(0, \sigma^2),$$

where x_j is the average observed vehicle speed and y_j is the number of casualties, for each reference site $j = 1, \dots, 67$.

About 25 assignments contained the ε_j term. **None** of the assignments contained a *fully specified* regression equation. **(-1 mark)** for failing to acknowledge there is an error term in a regression equation.

Question 16

Part (f)

Calculate the posterior distribution for your accident blackspot site

Question 16

Part (f)

Calculate the posterior distribution for your accident blackspot site

Several people gave answers like this ...

$$\theta \mid \mathbf{x} \sim \text{Ga}(2.5 + 9, \frac{2.5}{0.678} + 1)$$

... and just left it like this for their final answer.

Question 16

Part (f)

Calculate the posterior distribution for your accident blackspot site

Several people gave answers like this ...

$$\theta \mid \mathbf{x} \sim \text{Ga}(2.5 + 9, \frac{2.5}{0.678} + 1)$$

... and just left it like this for their final answer.

You should be able to finish calculating simple sums like this to give a final answer, where parameters are represented as a decimal or single fraction.

A note about plotting

- Make sure you plot probability densities over a sufficient amount of points.
- Plotting over too few points creates jagged lines in your densities and make your plot look completely wrong in extreme cases.

A note about plotting

- Make sure you plot probability densities over a sufficient amount of points.
- Plotting over too few points creates jagged lines in your densities and make your plot look completely wrong in extreme cases.

The following is an example taken from one student's assignment that was submitted last year . . .

A note about plotting

This code only plots over 10 points

```
x = seq(0,10,1) # Only 10 points to plot over
prior = dgamma(x,2.5,2500/579)
posterior = dgamma(x,9.5,3079/579)
plot(x,posterior,type='l',lty=2)
lines(x,prior)
```

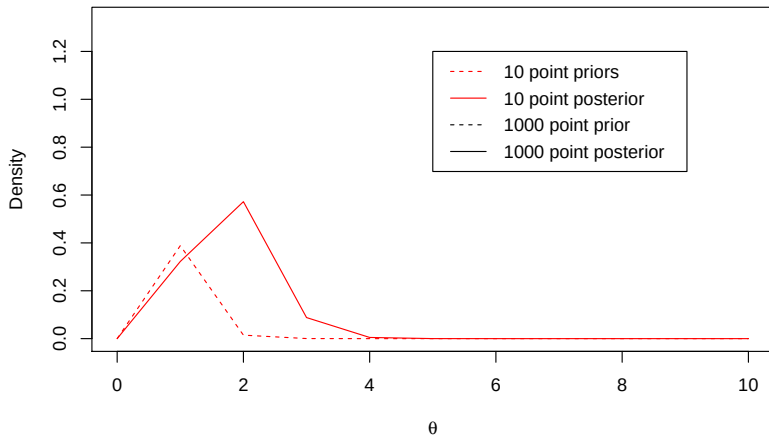
While this code has only one change in the seq() command

```
x = seq(0,10,0.01) # <= Now plots over 1000 points.
prior = dgamma(x,2.5,2500/579)
posterior = dgamma(x,9.5,3079/579)
plot(x,posterior,type='l',lty=2)
lines(x,prior)
```

A note about plotting

This produces very different plots!

Prior and posterior distributions



A note about plotting

This produces very different plots!

Prior and posterior distributions

