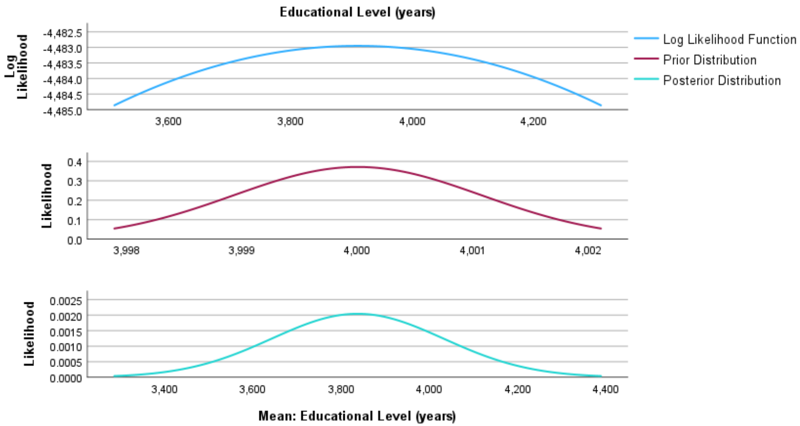
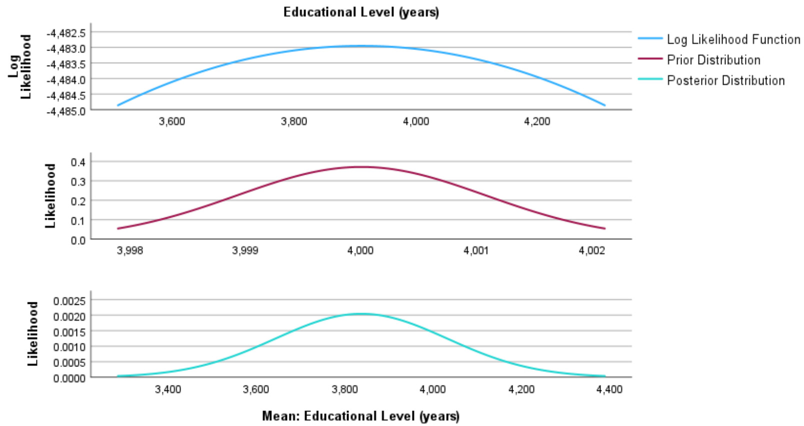


Introduction to Bayesian Analysis with SPSS

Jarlath Quinn – Analytics Consultant

Just waiting for all attendees to join...

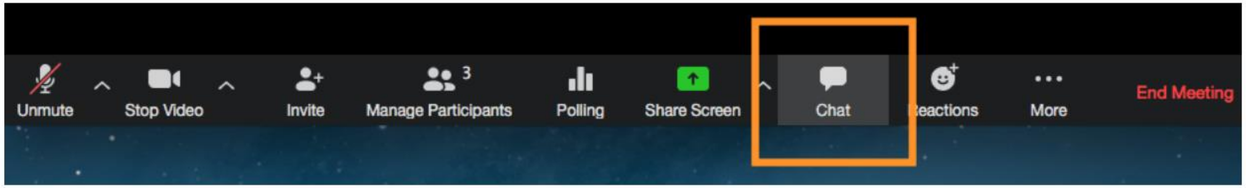


Introduction to Bayesian Analysis with SPSS

Jarlath Quinn – Analytics Consultant

FAQ's

- Is this session being recorded? Yes
- Can I get a copy of the slides? Yes, we'll email links to download materials after the session has ended.
- Can we arrange a re-run for colleagues? Yes, just ask us.
- How can I ask questions? All lines are muted so please use the chat panel – if we run out of time we will follow up with you.



- Premier accredited partner to IBM, Predictive Solutions and DataRobot specialising in advanced analytics & big data technologies
- Work with open source technologies (R, Python, Spark etc.)
- Team each has 15 to 30 years of experience working in the advanced and predictive analytics industry

- Deep experience of applied advanced analytics applications across sectors
  - Retail
  - Gaming
  - Utilities
  - Insurance
  - Telecommunications
  - Media
  - FMCG

The Two Paradigms

Frequentist vs Bayesian

R.A. Fisher

Karl Pearson

Egon Pearson

Jerzy Neyman

Interest in Bayesian Approaches

- An increasing use of Bayesian techniques in research articles over the last few years
- Advances in computing technology
- Ability to deal with smaller samples
- ‘The Replication Crisis’
- Flexibility with experimental designs
- More intuitive evidence-based approach to hypothesis testing
- Incorporated in many Machine Learning algorithms

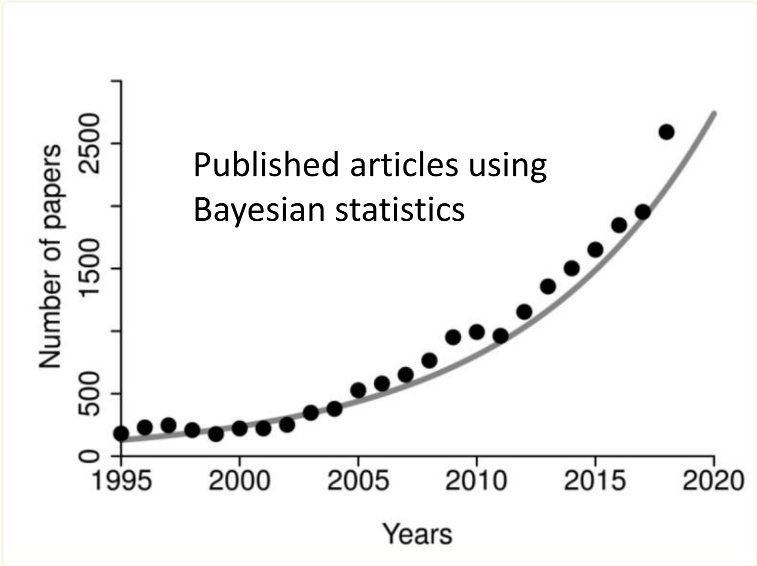


Figure 1

The number of published medical articles using Bayesian statistics in the period from 1995 to 2018 (sciencedirect.com, February 2019).

Frequentist vs Bayesian

Frequentist

- Probabilities are related to how frequently something occurs
- Data is King!
- The <u>parameter</u> is a fixed value
- The truth is fixed, the estimation is uncertain
- Only relates to the Null Hypothesis.
- *Only* uses the collected data, nothing else

Bayesian

- Probabilities are related to our level of certainty or uncertainty of something
- Data is really important but so is context
- The <u>parameter</u> is a <u>random variable</u>
- The truth is uncertain, the estimation is fixed
- Relates to both the Null and the Alternative Hypothesis
- Uses the collected data and an estimate of the truth in absence of any data

Frequentist vs Bayesian

Frequentist

- Probabilities are related to how frequently something occurs
- Data is King!
- The parameter is a fixed value
- The truth is fixed, the esti… uncertain
- Only relat… Hy…
- …cted data, …se

Bayesian

- Probabilities ar… …vel of certaint… …hing
- … …but so is context
- … is a random variable
- …uth is uncertain, the estimation is fixed
- Relates to both the Null and the Alternative Hypothesis
- Uses the collected data and an estimate of the truth in absence of any data

In most simple analyses, both approaches yield very similar results

For example, a different way of thinking:

Frequentist Confidence Intervals vs Bayesian Credible Intervals

Frequentist

"If we were to repeat the analysis many times, **95% of the computed confidence intervals** would contain the **true value**"

Bayesian

" Based on our data, there is a **95% *probability* that the parameter** lies within **the credible region**"

For example, a different way of thinking:

Frequentist Confidence Intervals vs Bayesian Credible Intervals

Frequentist

Variable

"If we were to repeat the analysis many times, **95% of the computed confidence intervals** would contain the **true value**"

Fixed

Bayesian

Fixed

" Based on our data, there is a **95% *probability* that the parameter** lies within **the credible region**"

Variable

Prior & Posterior Probabilities

Frequentist vs Bayesian

Frequentist

Does not use PRIOR probabilities

- Probabilities are related to how frequently something occurs
- Data is King!
- The parameter is a fixed value
- The truth is fixed, the estimation is uncertain
- Only relates to the Null Hypothesis
- *Only* uses the collected data, nothing else

Bayesian

Includes the calculation of PRIOR probabilities: the probability of an event occurring based on established knowledge or beliefs

- Probabilities are related to our level of certainty or uncertainty of something
- Data is really important but so is context
- The parameter is a random variable
- The truth is uncertain, the estimation is fixed
- Relates to both the Null and the Alternative Hypothesis
- Uses the collected data and an estimate of the truth in absence of any data

What's the probability of the coin landing 'Heads' ?

Frequentist vs Bayesian

- Could be anything. Could be 0.0001, 0.2 0.5 or 0.99.
- Only way to find out is collect the data.

- Given there are only two outcomes, and the coin looks 'fair', might be sensible to assume it's 0.5
- We can 'update' this prior with the probability from the data we collect.

Prior Probabilities

- In Bayesian statistics, the Prior Probability is the probability distribution that would describes the likelihood of a given value before some new evidence is taken into account
- As such, those working in the Frequentist tradition have often regarded this idea as the *most controversial* aspect of the Bayesian approach
- The prior could be the expected proportion of people likely to vote in a referendum. Likewise, it could refer to the expected mean and standard deviation of an age distribution describing when people get married
- These prior probabilities could be based on past research, subject matter expertise or reasonable assumptions based on the context of the study

Prior Probabilities

- In situations where no information is available as to the likely prior probability, an **uninformative prior** is often employed where it has a minimal effect on the inference. Examples include probabilities where every outcome is equally likely
- If the prior and posterior probabilities come from the *same* statistical distribution family, they are called **conjugate distributions** and the prior is called a **conjugate prior**
- A major criticism of Bayesian approaches is that there is no single, prescribed and well-defined method for choosing a prior. Using different priors for the same problem can mean that the analysts reach different conclusions

Posterior Probabilities

- The Posterior probability is the probability that results from updating the prior probability with new evidence
- This new evidence usually takes the form of collected data and is often referred to as the Likelihood Function
- Posterior probability = prior probability x new evidence (called the likelihood)

Bayes Theorem

Bayes Rule

H **A hypothesis**

E **Some new evidence**

$$P(H|E) = \frac{P(H)\,P(E|H)}{P(E)}$$

Bayes Rule

$$\overset{(1)}{P(H|E)} = \frac{\overset{(2)}{P(H)}\ \overset{(3)}{P(E|H)}}{\underset{(4)}{P(E)}}$$

1 **P(H|E)** The probability of the Hypothesis being true based on the Evidence

2 **P(H)** The *prior* probability of the Hypothesis being true

3 **P(E|H)** The probability of the Evidence being true based on the Hypothesis

4 **P(E)** The *prior* probability of the Evidence being true

Berni thinks his detector has found a horde of Saxon gold

The signal is so strong!

What’s the probability this is true?

Berni thinks his detector has found a horde of Saxon gold

P(H) = Prior probability of finding gold = 0.0001

P(E|H) = Probability of getting that signal *if it was* gold = 0.75

P(E) = Prior Probability of getting a strong signal = 0.002

Berni thinks his detector has found a horde of Saxon gold

$$P(H|E) = \frac{0.0001 \times 0.75}{0.002} = 0.0375$$

This is a Posterior Probability

Parameter Estimation and Credible Intervals

Parameter Estimation and Credible Intervals

- In the previous example we introduced the fundamental principles of Bayesian analysis using Bayes Theorem
- In particular we saw how Bayes Theorem can combine evidence from new data (the likelihood) and prior probabilities to provide a posterior probability
- Population parameters include not just means and proportions, but also things like:
  - The *difference* between two means (such average blood pressure for left handed and right handed people)
  - The *proportional* differences between the voting habits of different ethnicities
  - The *correlation* between travel distance and shopping frequency

Parameter Estimation and Credible Intervals

- In Bayesian analysis we estimate a parameter's posterior distribution by combining a prior distribution with the likelihood obtained from the data (specifying a prior distribution is an important aspect of this process)
- Bayesian analysis will often return a mean estimate of the parameter value, such as the mean difference or mean correlation. But really, as Bayesian analysis sees the parameter as a distribution of values reflecting uncertainty, analysts focus on the **credible intervals** that encapsulate 95% of this distribution
- **Credible Intervals** are the Bayesian equivalent of the Frequentist **Confidence Intervals**. Except that here, we *can* say that there is a 95% probability that the parameter lies within these bounds.
- If the credible intervals estimated a range of values that related to some sort of *difference* between groups, think about what this would mean if the lower interval was a minus number and the upper interval was positive…

Parameter Estimation and Credible Intervals

- In the SPSS output below, we can see the mean difference between the previous experience of clerical workers and managers. Although the overall estimate is -7.42, the Credible Intervals indicate the actual value could range from -26.1 to 11.26
- In other words, there's not much evidence that one group has more previous experience than the other

Bayesian Independent

Group Statistics

	Employment Category	N	Mean	Std. Deviation	Std. Error Mean
Previous Experience (months)	= Clerical	363	85.04	95.275	5.001
	= Manager	84	77.62	73.260	7.993

Posterior Distribution Characterization for Independent Sample Mean[a]

	Posterior			95% Credible Interval	
	Mode	Mean	Variance	Lower Bound	Upper Bound
Previous Experience (months)	-7.42	-7.42	90.616	-26.10	11.26

a. Prior for Variance: Diffuse. Prior for Mean: Diffuse.

Hypotheses and Bayes Factor

Hypotheses and Bayes Factor

- In statistical inference, there is an important difference between estimating a population parameter and testing hypotheses about it: they are separate things
- In the previous example, we looked at how Bayesian estimations attempt to approximate a parameter's posterior distribution. Now, we turn our attention to *testing* hypotheses using Bayes Factor, in same way we might do so with a frequentist P-value
- In the Frequentist approach to inference, much attention is focussed on the *Null Hypothesis*
- When a P value is calculated and found to be lower than a pre-specified alpha level (e.g. 0.05), the convention is to 'Reject the Null Hypothesis'.
- What we *can't* do with this information, is use it to *accept* the Alternative Hypothesis. The test simply doesn't consider this condition.

Hypotheses and Bayes Factor

- So what happens when the result is non-significant? Such as P = 0.38? All that happens is that the Null Hypothesis is not rejected
- As such, the frequentist approach has very little to say about the evidence for one hypothesis vs another
- Whereas a P-value is used to reject or not reject a null hypothesis, Bayes factor compares the evidence for separate hypotheses by dividing one by the other
- In SPSS, the resultant ratio of evidence attempts to quantify much more evidence there is for the null hypothesis vs the alternative hypothesis
- By doing so, it compares the evidence for both hypotheses, which is something that a frequentist P-value does not do

Bayes Factor

- In the Bayesian world, analysts use Bayes Factors (*B*) to indicate the relative strength of evidence for two theories
- These are simple ratios comparing evidence for the alternative and null hypotheses
- For example, a Bayes Factor, comparing a null hypothesis to the alternative hypothesis, would show that the data are *B* times more likely under the null than under the alternative.
- Because the *B* value is a ratio, it can range from 0 to infinity. But a value of 1 would indicate that there is exactly the same evidence for the null hypothesis as there is for the alternative hypothesis
- Values greater than 1 indicate increasing evidence for one theory over the other (e.g., the null over the alternative hypothesis) and values less than 1 the opposite (e.g., increasing evidence for the alternative over the null hypothesis).

BF_{10} vs BF_{01}

- Bayes Factors with the label $\mathbf{BF_{10}}$ refer to situations where evidence for the Alternative Hypothesis divided by evidence for the Null Hypothesis. In such situations, **larger values** indicate **more evidence** for the **Alternative Hypothesis**
- Conversely, Bayes Factors with the label $\mathbf{BF_{01}}$ refer to situations where evidence for the Null Hypothesis divided by evidence for the Alternative Hypothesis. In such situations, **larger values** indicate **more evidence** for the **Null Hypothesis**
- Rather confusingly, SPSS uses both methods depending on the test*

$$BF_{01} = \frac{\text{H0 Evidence}}{\text{H1 Evidence}} \qquad BF_{10} = \frac{\text{H1 Evidence}}{\text{H0 Evidence}}$$

* Bayesian One-way ANOVA uses $\mathbf{BF_{10}}$

SPSS Bayes Factor BF_{01} – Null vs Alternative Hypothesis

Bayes Factor	Evidence Category	Bayes Factor	Evidence Category	Bayes Factor	Evidence Category
>100	Extreme Evidence for H0	1-3	Anecdotal Evidence for H0	1/30-1/10	Strong Evidence for H1
30-100	Very Strong Evidence for H0	1	No Evidence	1/100-1/30	Very Strong Evidence for H1
10-30	Strong Evidence for H0	1/3-1	Anecdotal Evidence for H1	1/100	Extreme Evidence for H1
3-10	Moderate Evidence for H0	1/10-1/3	Moderate Evidence for H1		

- In SPSS, larger Bayes Factor values indicate more evidence for the Null Hypothesis. It should be noted that in other analytical programs, *the opposite interpretation is made*
- The above table, from the SPSS Help system, is provided to give users some guidance as to how to interpret the Bayes Factor and the strength of evidence for the Null (H0) vs the Alternative (H1) hypotheses

SPSS Bayes Factor

- Using the same analytical example as before, we can see that, based on the previous table, the resultant Bayes Factor of 8.441 in provides 'Moderate Evidence' for the Null Hypothesis that there is no difference in the average amount of previous experience for Clerical and Managerial workers

Group Statistics

	Employment Category	N	Mean	Std. Deviation	Std. Error Mean
Previous Experience (months)	= Clerical	363	85.04	95.275	5.001
	= Manager	84	77.62	73.260	7.993

Bayes Factor Independent Sample Test (Method = Rouder)[a]

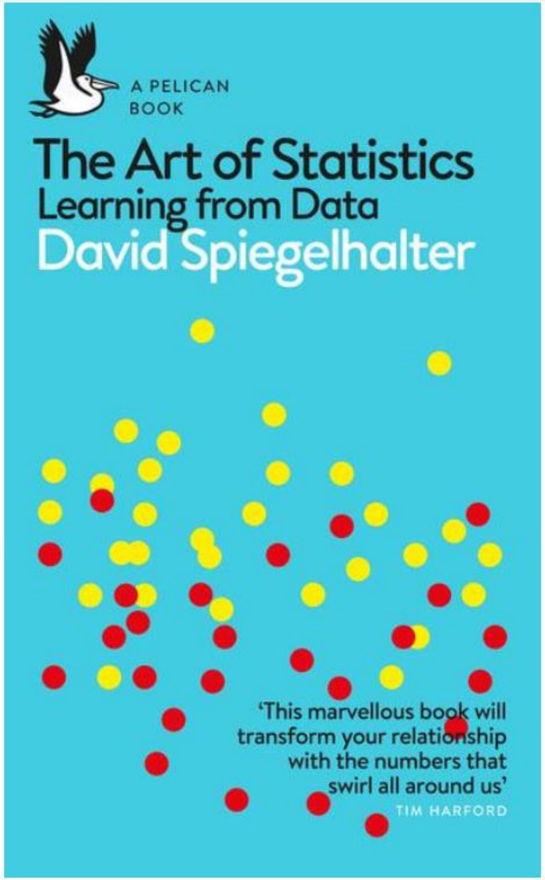
	Mean Difference	Pooled Std. Error Difference	Bayes Factor[b]	t	df	Sig.(2-tailed)
Previous Experience (months)	-7.42	11.087	8.441	-.669	445	.504

a. Assumes unequal variance between groups.

b. Bayes factor: Null versus alternative hypothesis.

Bayesian Analysis in SPSS Statistics

Further Reading: Recommended



The Art of Statistics Learning from Data - Pelican Books

D. J. Spiegelhalter (author)

Working with Smart Vision Europe Ltd.

- **Sourcing Software**
  - You can buy your analytical software from us often with discounts
  - Assist with selection, pilot, implementation & support of analytical tools
  - http://www.sv-europe.com/buy-spss-online/
- **Training and Consulting Services**
  - Guided consulting & training to develop in house skills
  - Delivery of classroom training courses / side by side training support
  - Identification & recruitment of analytical skills into your organisation
- **Advice and Support**
  - offer 'no strings attached' technical and business advice relating to analytical activities
  - Technical support services

Contact us:

+44 (0)207 786 3568
info@sv-europe.com
Twitter: @sveurope
Follow us on Linked In
Sign up for our Newsletter

Thank you