Factors of Happiness: An Analysis On Culture Vs. Money

Andrew Casas

Abstract

The report uses Denmark's World Happiness score to explore culture and economy as influences of happiness. The World Happiness report of 2021 was used to present a comparison between countries' ladder scores and logged GDP's, versus the ladder scores among different cultures. The methods were conducted through a series of analyses including bootstrap confidence interval and Bayesian credible interval. This will demonstrate that the countries with a higher logged GDP have a higher mean ladder score. This report concludes that although both logged GDP and culture contribute to ladder scores, logged GDP showed to have a slightly stronger influence than culture.

Introduction

Living in a pandemic has made me question many things, one of them being people's happiness. Throughout my time in isolation, I read 'The Little Book of Lykke,' a danish book written by the CEO of The Happiness Research Institute of Copenhagen (Meik). The book addresses how people in Denmark have adopted a very communal way of living which is a contributing factor to Denmark ranking second in World Happiness. This made me question if their happiness was wholly due to their communal way of living or if income played any role, given that Denmark is one of the most economically secure countries in the world. In the data used for this analysis, the term 'ladder score' will be used to denote the ranking of each country/global region's happiness score out of 10. Additionally, each country's economic status will be evaluated by logged GDP, because logged GDP – in comparison to general GDP – is often used for analyzing data in order to better visualize growth rates. Overall, this report will be considering whether money or culture is more important for peoples happiness. Having lived through a year in a pandemic, I hypothesize that culture will be more important to people's happiness than money.

Data

The data used in this report was found on Kaggle (Singh). From this data set I was able to find the global regions which had the largest ladder score mean and ladder score median. North America and Western Europe both had the highest mean and median out of all of the global regions. However, since North America only had 4 countries and a negligible difference in ladder score mean, I decided to use Western Europe as the representative of culture since it has a larger spread in logged GDP as well as other factors. Since Western Europe contains 21 countries, I then used the top 21 countries of the world – who had the highest logged GDP – to compare.

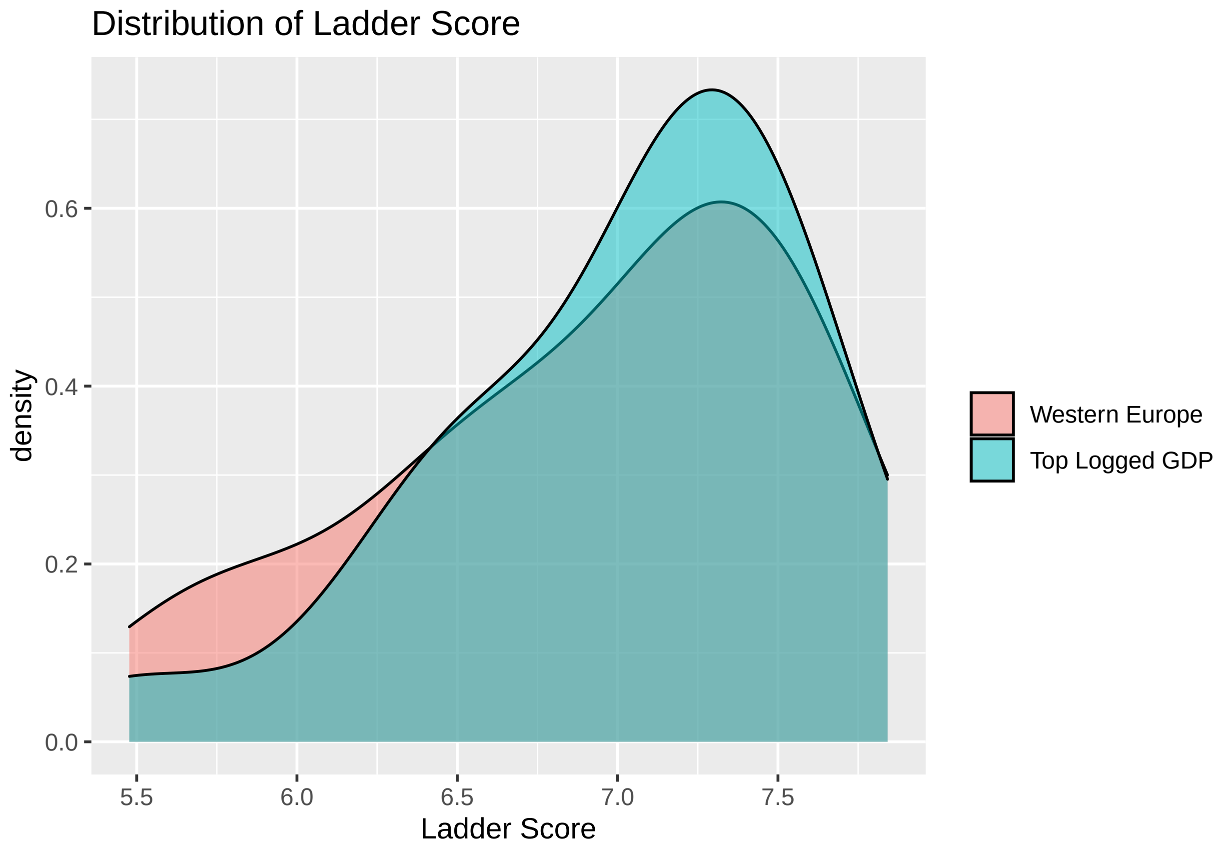
Regional.indicator	Ladder.score	Logged.GDP.per.capita
Western Europe	7.324	11.647
Southeast Asia	6.377	11.488
Western Europe	7.085	11.342
Western Europe	7.571	11.117
Middle East and North Africa	6.561	11.085
Western Europe	7.392	11.053

Regional.indicator	Ladder.score	Logged.GDP.per.capita
North America and ANZ	6.951	11.023
East Asia	5.477	11.000
Western Europe	7.620	10.933
Western Europe	7.464	10.932
Western Europe	7.268	10.906
Western Europe	7.554	10.878
Western Europe	7.155	10.873
East Asia	6.584	10.871
Western Europe	7.363	10.867
Western Europe	6.834	10.823
Middle East and North Africa	6.106	10.817
North America and ANZ	7.183	10.796
North America and ANZ	7.103	10.776
Western Europe	7.842	10.775
Middle East and North Africa	6.494	10.743

Above is a graph demonstrating the countries with the top GDP, with this data including regions from all around the world. Now we will see a chart of the countries in Western Europe along side their ladder score (Logged GDP will not be shown as it is not relevant for this table).

Regional.indicator	Ladder.score
Western Europe	7.842
Western Europe	7.620
Western Europe	7.571
Western Europe	7.554
Western Europe	7.464
Western Europe	7.392
Western Europe	7.363
Western Europe	7.324
Western Europe	7.268
Western Europe	7.155
Western Europe	7.085
Western Europe	7.064
Western Europe	6.834
Western Europe	6.690
Western Europe	6.602
Western Europe	6.491
Western Europe	6.483
Western Europe	6.223
Western Europe	5.929
Western Europe	5.723
Western Europe	5.536

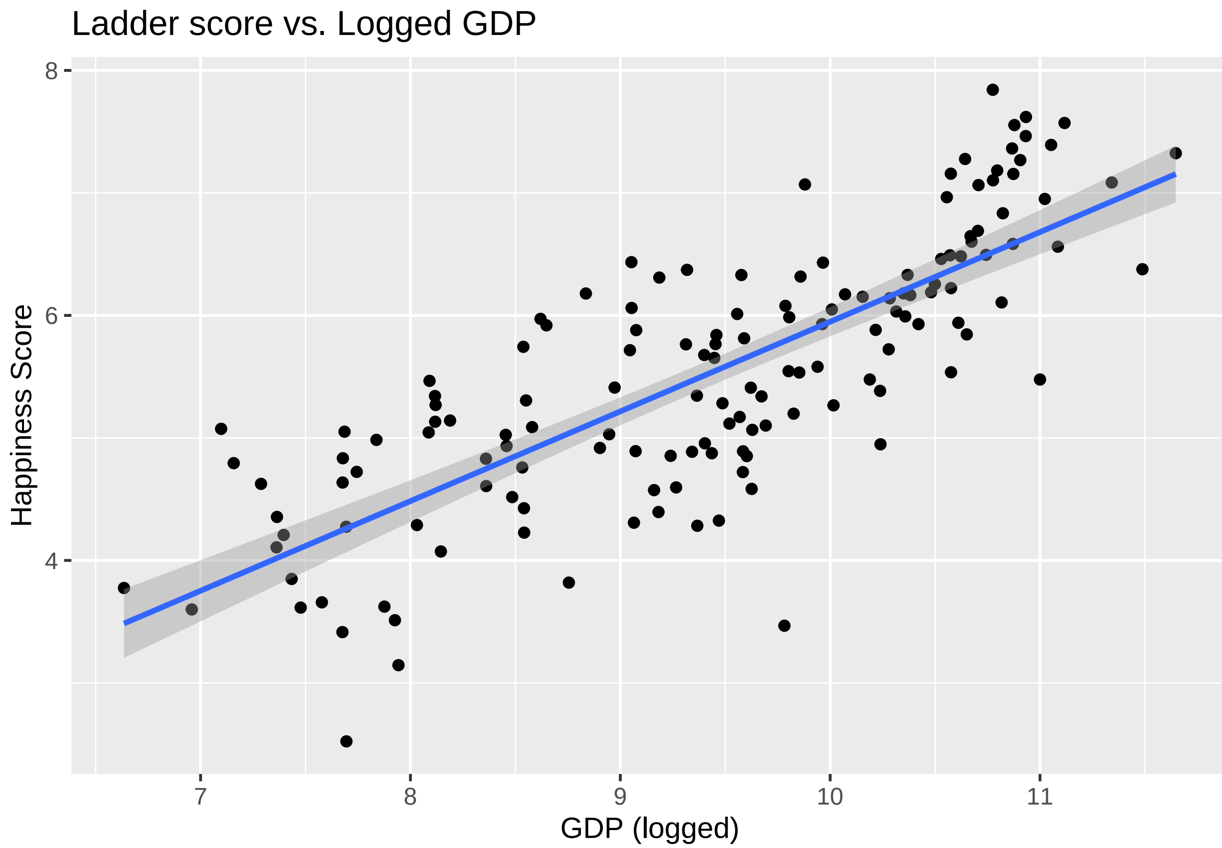
Finally I will see the spread of the two variables to get an understanding of their distribution.



The countries with the top logged GDP tend to be denser further to the right, which means that it is denser where countries are happier. However, this does not yet conclude our question. This report will be bootstrapping the difference of means and finding a confidence interval to see if the true mean of difference of means shows us that countries with a larger logged GDP tend to be happier.

Methodology

Before I address my hypothesis, we first have to answer some questions. Firstly, does logged GDP affect happiness? And are cultures happier than others? Without knowing the answer to these questions, we cannot know whether one factor is more important to happiness than another. In order to figure this out, I will determine if logged GDP affects happiness using a line of regression. I will see if the linear increase of the independent variable, which is Logged GDP, increases the ladder score which is the dependent variable. For this line of regression, I will use all 149 countries (Singh).



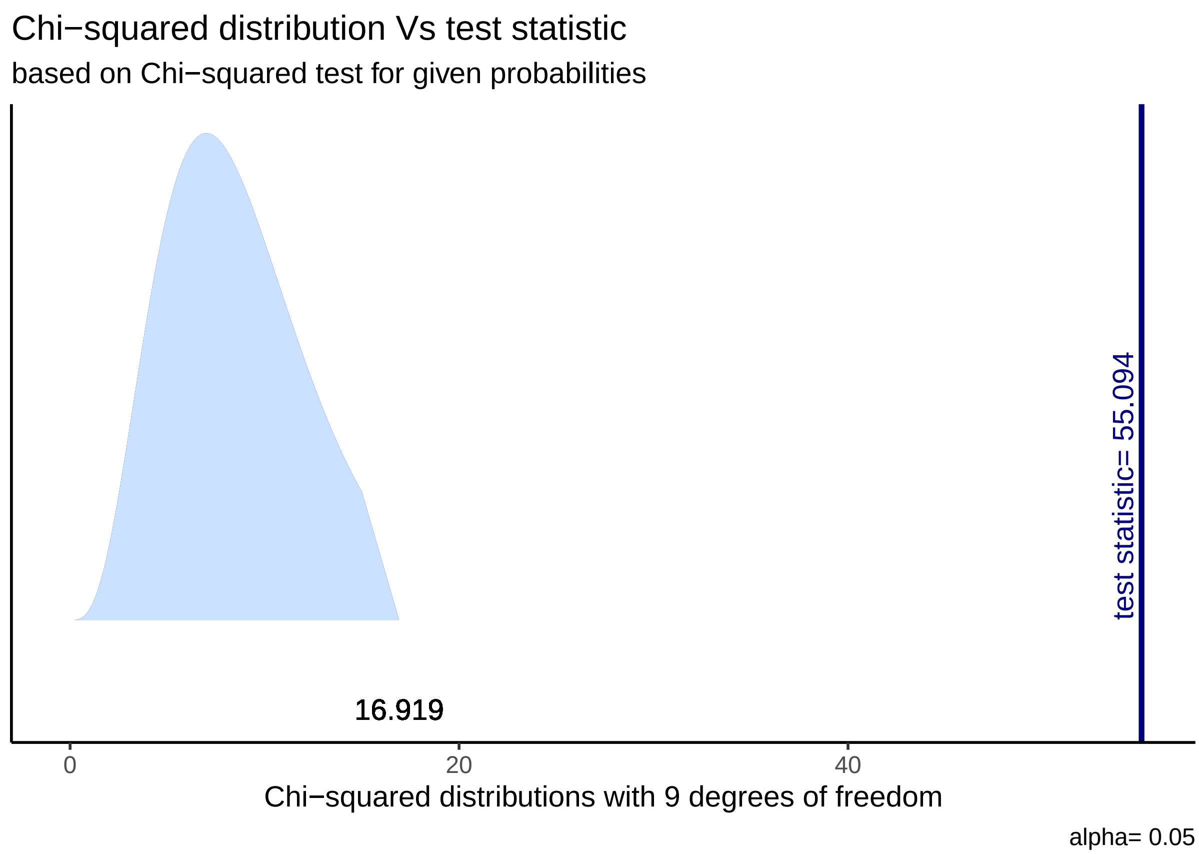
We can see that as logged GDP increases, happiness increases as well.
Now I will explore whether there is a difference in happiness between cultures and countries using a goodness of fit test. For my goodness of fit test, I will distribute the world's countries – based on their corresponding world percentage – within the world's top 25 happiest countries. My alternative hypothesis is that the distribution will not be the same as the null hypothesis.

Regional.indicator	%
Central and Eastern Europe	11
Commonwealth of Independent States	8
East Asia	4
Latin America and Caribbean	13
Middle East and North Africa	11
North America and ANZ	3
South Asia	5
Southeast Asia	6
Sub-Saharan Africa	24
Western Europe	14

Our observed observations were much different. You can see them shown below.

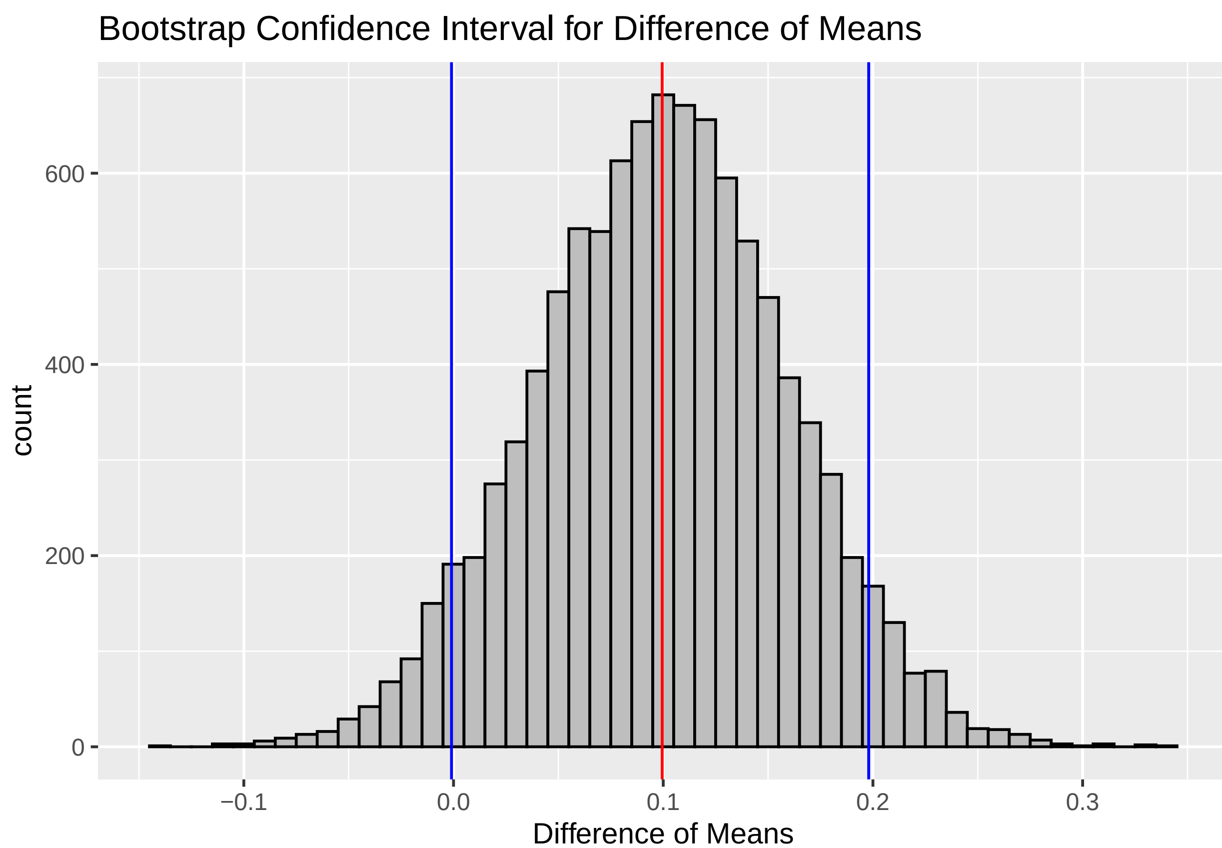
Regional.indicator	%
Central and Eastern Europe	4
Commonwealth of Independent States	0
East Asia	4
Latin America and Caribbean	4
Middle East and North Africa	12
North America and ANZ	16
South Asia	0
Southeast Asia	0
Sub-Saharan Africa	0
Western Europe	60

Now we will conduct a goodness of fit test with a 95% to see if the distribution is suitable.



This shows that the test statistic is 55.094. However, our test statistic with our observed statistic is much larger, 181.01, therefore the distribution does not fit because the observed test statistic is in the rejection region. This demonstrates that there is a difference in happiness between the different world cultures. This is because, although Western Europe accounts for only 14% of the countries in the world, they take up 60% when distributed in top 25 happiest countries.

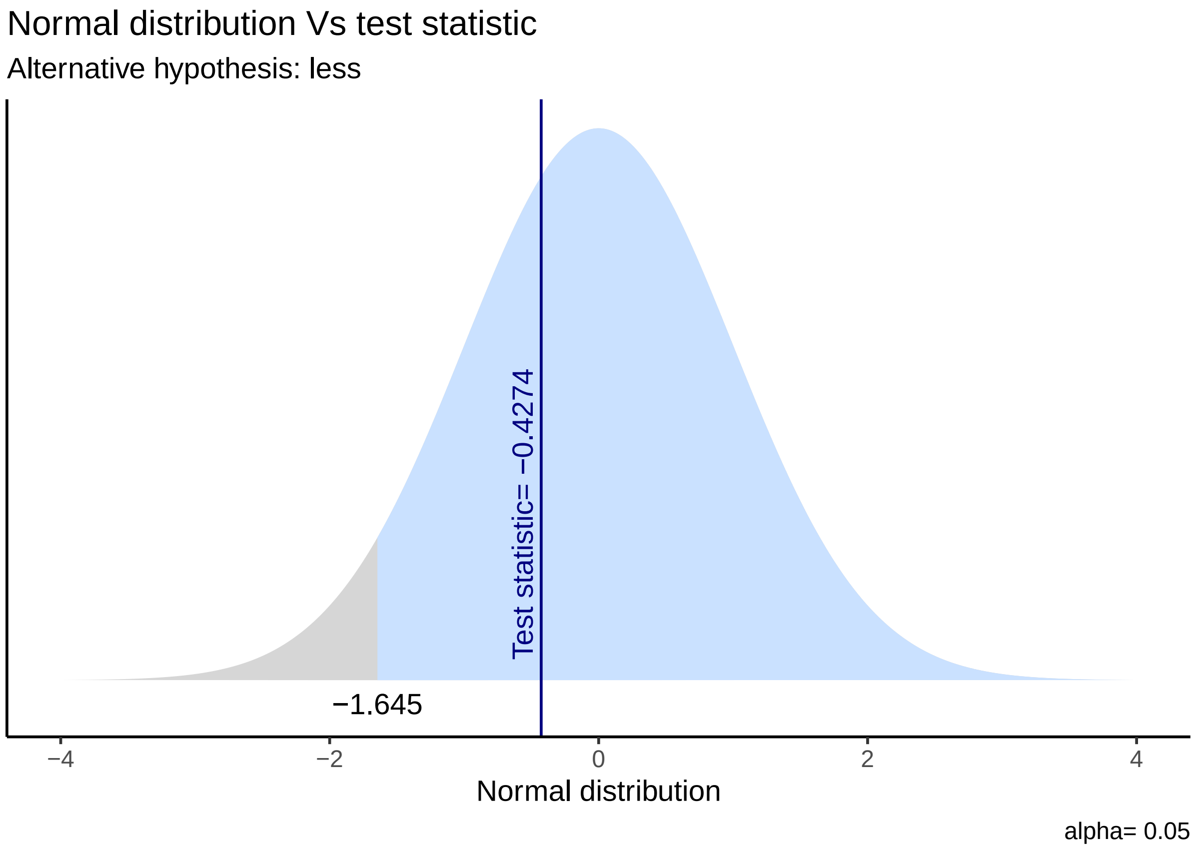
Now that we know that there is a correlation between logged GDP and happiness, as well as a correlation between culture and happiness, we can finally answer which factor affects happiness more. I will simulate a bootstrap 90% confidence interval for difference of means to evaluate the true difference of means between the sample of the countries with the top logged GDP as well as the sample of the countries in Western Europe.



The confidence interval is represented within the quantiles (-0.001, 0.198)

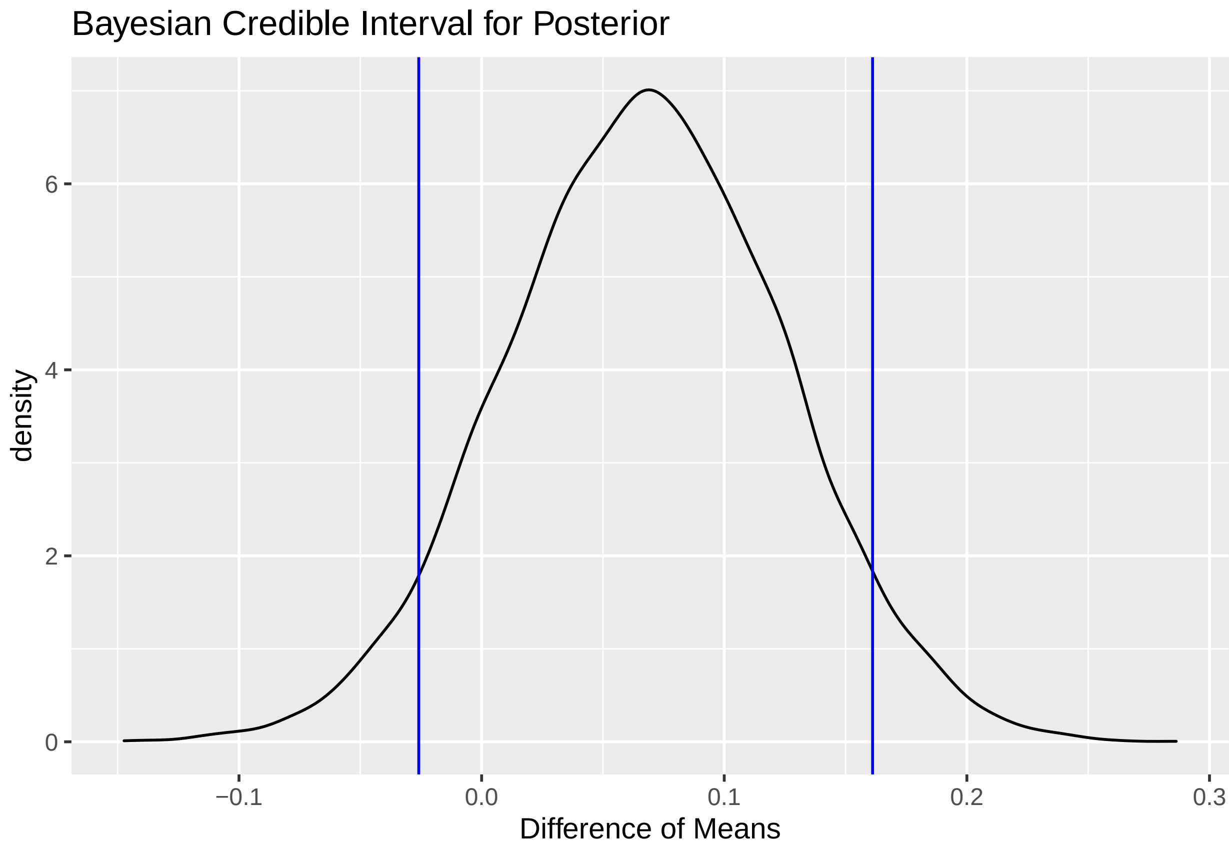
The bootstrap sample was calculated by subtracting the logged GDP ladder score mean from Western Europe's ladder score mean. Based on the confidence interval above, I am 90% confident that the countries with a higher logged GDP will be 0 to 0.2 points happier in terms of ladder score. We can also see that this data is normally distributed, where it is most likely that countries with a higher logged GDP will be happier by 0.1 points in terms of ladder score. This was calculated by finding the Maximum Likelihood Estimator (MLE). As we can see from the graph above, the normally distributed bootstrap has a mean of 0.0995052 ≈ 0.1. With a normal likelihood function, I then found the loglikelihood function, which I then differentiated with respect to μ and then equated it to 0. We isolated μ which gave us the MLE which was $\frac{\sum x_i}{n} = \bar{x}$, therefore the mean, 0.1, is the most likely difference between the difference of means.

Using the bootstrap sample, I will conduct a hypothesis test with a null hypothesis of $\mu = 0.1$ and with an alternative hypothesis of $\mu < 0$. Our null hypothesis is less than 0 because, if the difference of mean is less than 0, then the happiness mean of Western Europe is larger than that of the countries with the top logged GDP.



From this hypothesis test we can see that the null hypothesis remains true. Since the test statistic is in the 'accepted' region we can assure ourselves with 95% confidence that the countries with higher logged GDP will have higher happiness in ladder score.

To be completely sure of this hypothesis, I will also do a 90% Bayesian credible interval to see if the countries with higher logged GDP will have larger ladder score means. To do this I will derive our posterior distribution, which is found in the appendix.
Our posterior distribution is a normal distribution $\sim$ N(0.067,0.003)



The credible interval is represented between (-0.026, 0.161)

With the Bayesian credible interval of 90% we can see that although there is some area in the negative – being below zero – it is almost certain that the difference of means will almost always be positive, thus the countries with larger logged GDP tend to have a larger ladder score.

Results

Before answering my main question, I first answered some preliminary questions. I proved that there is a correlation between ladder score and logged GDP. I did this using a simple linear regression and noticed that the line was positive and the points around it were very close, meaning that the two variables shared a strong correlation. We then did a goodness of fit test to prove that there was a distinct difference in happiness between cultures. I achieved this by having the null hypothesis have a distribution which was the same as the distribution of countries per region across the world. The observed hypothesis showed us that Western Europe and North America had a large majority of the countries with the largest ladder score.
I then did a bootstrap of the difference of means in order to do a confidence interval of 90%, where I was 90% confident that the mean ladder score of the countries with the largest logged GDP would be greater than the mean of the countries from Western Europe. I then analyzed the MLE for the mean, which was about 0.1, with the results showing that it is most likely that the countries with higher logged GDP will have a larger ladder score by 0.1. This gave predictable results as the difference of means between the two samples is about 0.1. With the same bootstrap data, I was then able to conduct a hypothesis test with the null hypothesis being $\mu = 0.1$ and the alternative hypothesis stating it would be less. This resulted in the hypothesis test indicating that the null hypothesis is correct. Finally, I took a different approach and derived the posterior distribution under the assumption that the likelihood and prior distribution were both normal. This resulted with a normal posterior distribution with mean 0.067, and standard deviation of 0.057. This distribution was very similar to the bootstrap sample which was expected. When calculating the Bayesian

credible interval of 90% I was reassured to find that the countries with higher logged GDP have a larger ladder score mean. Overall, the results among all the tests were very consistent with each finding and seem very reasonable. Although my original hypothesis was wrong, the difference in happiness between the two samples were minuscule, and therefore logged GDP and culture are both strong influences of happiness, however logged GDP showed to be more influential of a countries World Happiness score.

Conclusion

We began our study hypothesizing that the a countries happiness is based on shared culture more than a higher logged GDP. I then created a bootstrap sample of the difference of means, resulting in a normal distribution with the mean at 0.1. This indicated that, with 90% confidence, that the countries with the higher logged GDP will have a higher mean of happiness. This also found that the maximum likelihood of these countries have a larger mean than the Western European countries by 0.1. An apparent drawback of this report was the ability to truly capture a sample representative of culture – as geographic region do not always share a culture altogether. Attempting to find a sample to represent the idea of culture versus logged GDP was challenging and I feel like there could have been a better way of analysing this hypothesis. In future analyses, I will most likely not address such a subjective concept as it is more much more challenging and difficult to compare in heavily objective-based terms.

Appendix

A. Deriving the MLE

Let $X_1, ..., X_n \overset{iid}{\sim} Normal(\mu = 0, \sigma^2)$.
Let $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$
Since $\mu = 0$, let μ be denoted as μ_0

Now we will use the likelihood function of a normal distribution to find the loglikelihood function.

$$\begin{aligned} L(\sigma^2|x, \mu_0) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu_0)^2}{\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2} \end{aligned} \tag{1}$$

Now we will apply the ln function to both sides to simplify.

$$ln(L) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2 \tag{2}$$

This function will be denoted as l. This is the loglikelihood function.

Now we will differentiate l with respect to μ and equate it to 0 to find the maximum likelihood estimator with respect to μ.

$$\begin{aligned} \frac{\partial l}{\partial \mu} &= \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu_0)^2 \\ \mu &= \frac{1}{n}\sum_{i=1}^{n}(x_i)^2 \end{aligned} \tag{3}$$

As we can see the maximum likelihood estimator is to the mean of the distribution.

B. Deriving the Posterior

Assumptions: Based on the data, we will assume that the likelihood and the prior will follow a normal distribution. We will assume this because we are finding the difference of means, and in both of the sample sets the data has an even spread and it is most likely that the density of the data will be centered around the mean. We will also assume all of the data points are iid.

We will be deriving the posterior distribution of μ.

Prior ~ N($\mu_0 = 0$, $\sigma_0^2 = 0.1$) = $f(\mu)$
likelihood ~ N($\mu, \sigma^2 = 0.5$) = $f(x_1, x_2...x_n|\mu)$
n = 10000

$$\begin{aligned} f(\mu|x_1, x_2...x_n) \propto f(\mu)f(x_1, x_2...x_n|\mu) &= f(\mu)f(x_1|\mu)f(x_2|\mu)...f(x_n|\mu) \\ &= \frac{1}{\sqrt{2\pi\sigma_0^2}} exp(\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}) \prod_{i=1}^{n} \frac{1}{sqrt2\pi\sigma^2} exp(\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}) \\ &= \frac{1}{2\pi^{\frac{n+1}{2}}\sqrt{\sigma_0^2\sigma^{2n}}} exp(\frac{-\mu^2 + 2\mu\mu_0 - \mu_0^2}{2\sigma_0^2} - \sum_{i=1}^{n}\frac{x_i^2 - 2\mu x_i + \mu^2}{2\sigma^2}) \\ &= exp(-\frac{\mu_0^2\sigma^2 + \sum_{i=1}^{n}\sigma_0^2 x_i^2}{2\sigma_0^2\sigma^2}) \end{aligned} \tag{4}$$

We can see that

$$
\begin{aligned}
\mu_1 &= \frac{\mu_0^2\sigma^{-2} + \sum_{i=1}^{n}\sigma_0^{-2}x_i}{\sigma_0^{-2} + n\sigma^{-2}} \\
&= \sigma_1^2(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\sigma^2/n}) \\
\sigma_1^2 &= \frac{1}{\sigma_0^{-2} + n\sigma^{-2}} \\
&= (\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n})^{-1}
\end{aligned}
\tag{5}
$$

Which finally gives us,

$$
\begin{aligned}
f(\mu|x_1, x_2 ... x_n) &\propto exp(\frac{-(\mu - \mu_1)^2}{2\sigma_1^2}) \\
f(\mu|x_1, x_2 ... x_n) &\sim N(\mu_1 = 0.067, \sigma_1^2 = 0.003)
\end{aligned}
\tag{6}
$$

We can conclude that our posterior has a normal distribution.

C. Test statistic for Goodness of Fit test

$$
\begin{aligned}
C^2 &= \frac{\sum_{i=1}^{n}(O_i - E_i)^2}{E_i} \\
&= 181.01
\end{aligned}
\tag{7}
$$

Bibliography

1. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: January 15, 2021)

2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

3. Grolemund, G. (2014, July 16) *Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: January 15, 2021)

4. Singh, A. (2021, March 22). World happiness REPORT 2021. Retrieved April 05, 2021, from https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021?select=world-happiness-report-2021.csv

5. Wiking, Meik. The Little Book of Lykke the Danish Search for the World's Happiest People. Penguin Life, 2017.