Q1. Refer to “Question 1” section in the R script.

Estimate for mu:

> sleep.post %>%

+ as.data.frame() %>%

+ median\_hdi(mu)

# A tibble: 1 × 6

mu .lower .upper .width .point .interval

<dbl> <dbl> <dbl> <dbl> <chr> <chr>

1 -1.42 -2.28 -0.480 0.95 median hdi

Estimate for sigma2:

> sleep.post %>%

+ as.data.frame() %>%

+ median\_hdi(sig2)

# A tibble: 1 × 6

sig2 .lower .upper .width .point .interval

<dbl> <dbl> <dbl> <dbl> <chr> <chr>

1 1.72 0.543 4.25 0.95 median hdi

Q2. Refer to “Question 2” section in the R script.

Estimate for mu:

> sleep.post.2 %>%

+ as.data.frame() %>%

+ median\_hdi(mu)

# A tibble: 1 × 6

mu .lower .upper .width .point .interval

<dbl> <dbl> <dbl> <dbl> <chr> <chr>

1 -1.50 -2.35 -0.641 0.95 median hdi

Estimate for sigma2

> sleep.post.2 %>%

+ as.data.frame() %>%

+ median\_hdi(sig2)

# A tibble: 1 × 6

sig2 .lower .upper .width .point .interval

<dbl> <dbl> <dbl> <dbl> <chr> <chr>

1 1.60 0.547 3.93 0.95 median hdi

Here, the estimated difference has moved slightly further away from zero. Also, the range of values in the HDI has increased, indicating a wider range of credible values for the difference. This difference likely arose because using ultrawide priors – relative to the narrower default priors – allocates additional probability to more extreme values. Because of the additional weight afforded to values at either end of the distribution, the range of values that can be considered credible has widened and the median estimate has taken on a slightly more extreme value.

Q3. Refer to “Question 3” section in the R script.

> sleep.post.a %>%

+ as.data.frame() %>%

+ mean\_qi(mu)

# A tibble: 1 × 6

mu .lower .upper .width .point .interval

<dbl> <dbl> <dbl> <dbl> <chr> <chr>

1 -1.07 -1.44 -0.646 0.95 mean qi

> sleep.post.b %>%

+ as.data.frame() %>%

+ mean\_qi(mu)

# A tibble: 1 × 6

mu .lower .upper .width .point .interval

<dbl> <dbl> <dbl> <dbl> <chr> <chr>

1 -1.81 -1.83 -1.80 0.95 mean qi

> sleep.post.c %>%

+ as.data.frame() %>%

+ mean\_qi(mu)

# A tibble: 1 × 6

mu .lower .upper .width .point .interval

<dbl> <dbl> <dbl> <dbl> <chr> <chr>

1 -1.33 -1.68 -0.879 0.95 mean qi

Each three-sample model produced vastly different estimates of mu and vastly different credible intervals. For example, if we based our conclusions on model B, we would conclude that credible values range from -1.83 to -1.80, whereas model C suggests a range from -1.44 to -0.65. On the other hand, our 10000-sample model ranges from -2.29 to -0.49. Accordingly, the latter model provides a much wider range of estimates. Because of the iterative nature of Bayesian analysis, the three-sample models provide a very poor representation what the posterior distribution should look like. By chance, model B produced an incredibly narrow range of credible values, while the 10000-sample model shows that the range of credible values should be much larger in either direction. This is the nature of sampling from a distribution – we may only draw values from a narrow region, leading to poor estimates. However, this issue can be mitigated by drawing many samples from the posterior, vastly increasing the chances that we draw samples truly representative of the distribution; this allows us to derive a much better representation of the possible range of values and more precise estimates. Accordingly, there is no question that we should use the 10000-sample model to draw conclusions about the data.

Q4. Refer to “Question 4” section in the R script (histograms there).

For the unstandardized difference, values are most densely clustered around ~ -1.5 to -1. Thin tails suggest that values in the posterior can be as low as ~ -3 or as high as ~ 0.6, although little probability is afforded to values this extreme. Tails are fairly thick for values as low as ~ -2 or as high as ~ -0.4, indicating some probability for values in this range; this aligns well with the credible interval calculated in Q1. Overall, this histogram suggests that values range from around -3 to 0.6, but that more *probable* values range from around -2 to 0.4, with the highest probability concentrated in the range of -1.5 to -1.

For the standardized difference, values are most densely clustered around ~ -1. Thin tails suggest that values in the posterior can be as low as ~ -2.8 or as high as ~ 0.6, although little probability is afforded to values this extreme. Tails are fairly thick for values as low as ~ -1.8 or as high as ~ -0, indicating some probability for values in this range. Overall, this histogram suggests that values range from around -2.8 to 0.6, but that more *probable* values range from around -1.5 to 0.5, with the highest probability concentrated in the range of -1.

Q5. Refer to “Question 5” section in the R script.

Based on our data, the first theorist is almost certainly wrong, with the posterior probability for the hypothesis that Group 1 should have greater sleep gain than Group 2 estimated at only 0.3%. The second theorist did a lot better, with the hypothesis that Group 2 should have greater sleep gain than Group 1 having a posterior probability of 99.7%. The third theorist also seemed to have a good idea about what we would find, with the posterior probability of Group 2 gaining at least 30 mins of extra sleep estimated at 97.3%. The fourth theorist’s hypothesis was pretty good too, with a posterior probability of 74.3% that Group 2 gained between 1 and 2 hours of sleep. Finally, the fifth theorist was way off base, with only 0.8% posterior probability that the difference between groups was no more than 15 minutes in either direction. Because the fifth theorist was particularly aggressive, I decided to put him in his place by calculating an evidence ratio showing that the fourth theorist’s hypothesis was 96.6 times more likely than his.

Q6. Refer to “Question 6” section in the R script.

> toothAOV

$ANOVA

Effect DFn DFd SSn SSd F p p<.05 ges

1 supp 1 36 66.82225 517.505 4.648459 3.783927e-02 \* 0.1143576

2 dose 1 36 2400.95025 517.505 167.021012 4.349447e-15 \* 0.8226785

3 supp:dose 1 36 71.02225 517.505 4.940631 3.260150e-02 \* 0.1206779

$`Levene's Test for Homogeneity of Variance`

DFn DFd SSn SSd F p p<.05

1 3 36 26.19275 161.955 1.940743 0.1404507

Q7. Refer to “Question 7” section in the R script.

> summary(mt.1)

Family: gaussian

Links: mu = identity; sigma = identity

Formula: len ~ supp \* dose

Data: TG\_noMid (Number of observations: 40)

Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;

total post-warmup draws = 4000

Population-Level Effects:

Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

Intercept 13.26 1.25 10.77 15.74 1.00 2087 2925

suppVC -5.28 1.76 -8.76 -1.80 1.00 1939 2271

dose2 12.81 1.77 9.37 16.33 1.00 1865 2278

suppVC:dose2 5.35 2.48 0.52 10.22 1.00 1764 2063

Family Specific Parameters:

Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

sigma 3.93 0.48 3.11 4.99 1.00 3597 3013

Draws were sampled using sampling(NUTS). For each parameter, Bulk\_ESS

and Tail\_ESS are effective sample size measures, and Rhat is the potential

scale reduction factor on split chains (at convergence, Rhat = 1).

For the Frequentist ANOVA, each term (i.e., the effect of supp, the effect of dose, and the interaction between the two) was significant. For the Bayesian model, we see that the intercept (representing the reference level, i.e., tooth length in the supp = OJ and dose = 0.5) was credibly higher than zero. Further, all slopes (representing the difference in tooth length between each term specified by the coefficients and the reference level) were credible, as the credible intervals for each coefficient do not contain zero. This model indicates that relative to the reference level, tooth growth decreases in the supp = VC and dose = 0.5 condition and increases in the supp = OJ and dose = 2 and the supp = VC and dose = 2 conditions. Accordingly, the conclusions we would draw from this model are the same, given that the effects of supp, dose, and the interaction between them are credible at the 95% level, analogous to statistical significance.

> prior\_summary(mt.1)

prior class coef group resp dpar nlpar lb ub source

(flat) b default

(flat) b dose2 (vectorized)

(flat) b suppVC (vectorized)

(flat) b suppVC:dose2 (vectorized)

student\_t(3, 20, 11.7) Intercept default

student\_t(3, 0, 11.7) sigma 0 default

For all beta coefficients, the default priors are flat. Accordingly, we are affording equal probability to *all* possible values of tooth length. That means that we consider the possibility of tusk-sized or near-nonexistent teeth to be just as likely as reasonable values for guinea pig tooth length. These priors are highly uninformative and the approach is similar to that taken by Frequentist analysis – surely we can do better than that. For the intercept and sigma, the priors are T distributions with three degrees of freedom. These distributions have thicker tails than normal distributions, which affords a little bit more probability to extreme values; however, they are still largely uninformative. The prior for sigma does specify that sigma must be positive, however, which provides some valuable information for the model.

Q8. Refer to “Question 8” section in the R script.

> summary(mt.2)

Family: gaussian

Links: mu = identity; sigma = identity

Formula: len ~ supp \* dose

Data: TG\_noMid (Number of observations: 40)

Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;

total post-warmup draws = 4000

Population-Level Effects:

Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

Intercept 13.20 1.27 10.64 15.65 1.00 2549 2592

suppVC -5.18 1.76 -8.54 -1.65 1.00 2090 2301

dose2 12.83 1.77 9.43 16.31 1.00 2235 2880

suppVC:dose2 5.22 2.49 0.30 10.15 1.00 1938 2345

Family Specific Parameters:

Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

sigma 3.93 0.50 3.11 5.05 1.00 2764 2346

Draws were sampled using sampling(NUTS). For each parameter, Bulk\_ESS

and Tail\_ESS are effective sample size measures, and Rhat is the potential

scale reduction factor on split chains (at convergence, Rhat = 1).

This model was nearly identical to the model fit with default priors, again showing an intercept that was credibly different from zero and credible slopes for each coefficient (i.e., supp, dose, and the interaction between the two). The estimates for each model term and the intervals surrounding each estimate are very similar to those in the previous model, and all inferences that can be drawn from this model are again identical.

Q9. Refer to “Question 9” section of the R script.

**Table 1**

*Mean Tooth Length and 95% HPD Intervals as a Function of Supp and Dose*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Supp | Dose | Mean | HPD95% LB | HPD95% UB |
| OJ | 0.5 | 13.20 | 10.63 | 15.60 |
| 2.0 | 26.00 | 23.40 | 28.30 |
| VC | 0.5 | 8.00 | 5.66 | 10.40 |
| 2.0 | 26.10 | 23.52 | 28.50 |

For the guinea pigs receiving vitamin C via OJ, average tooth length was credibly higher for doses of 2.0 relative to doses of 0.5, with the difference between doses estimated at 12.80 (HPD95% = 9.35 – 16.20). The same pattern was observed in the VC group, with the difference in tooth length between dose levels estimated at 18.10 (HPD95% = 14.43 – 21.50). Overall, this suggests that tooth length was greater for doses of 2.0 regardless of the vehicle through which vitamin C was delivered. Further contrasts revealed a slightly more nuanced pattern of results: For dose levels of 2.0, there was no credible difference in tooth length between the OJ and VC groups, estimate = 0.01 (HPD95 = -3.40 – 3.48). For dose levels of 0.5, however, tooth growth was credibly higher in the OJ relative to VC group, estimate = 5.21 (HPD95% = 1.85 – 8.66). Considered in aggregate, our investigation suggests that at low doses, OJ stimulates tooth growth better than VC. At higher doses, on the other hand, these supplements produce equal gains.

A white grid with red and blue dots

Description automatically generated

Q10. Refer to “Question 10” section in the R script.

Intercept:

> hypothesis(mt.2, 'Intercept = 0')

Hypothesis Tests for class b:

Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star

1 (Intercept) = 0 13.2 1.27 10.64 15.65 NA NA \*

---

'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.

'\*': For one-sided hypotheses, the posterior probability exceeds 95%;

for two-sided hypotheses, the value tested against lies outside the 95%-CI.

Posterior probabilities of point hypotheses assume equal prior probabilities.

suppVC:

> hypothesis(mt.2, 'suppVC = 0')

Hypothesis Tests for class b:

Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star

1 (suppVC) = 0 -5.18 1.76 -8.54 -1.65 0.28 0.22 \*

---

'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.

'\*': For one-sided hypotheses, the posterior probability exceeds 95%;

for two-sided hypotheses, the value tested against lies outside the 95%-CI.

Posterior probabilities of point hypotheses assume equal prior probabilities.

dose2:

> hypothesis(mt.2, 'dose2 = 0')

Hypothesis Tests for class b:

Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star

1 (dose2) = 0 12.83 1.77 9.43 16.31 0 0 \*

---

'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.

'\*': For one-sided hypotheses, the posterior probability exceeds 95%;

for two-sided hypotheses, the value tested against lies outside the 95%-CI.

Posterior probabilities of point hypotheses assume equal prior probabilities.

suppVC:dose2:

> hypothesis(mt.2, 'suppVC:dose2 = 0')

Hypothesis Tests for class b:

Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star

1 (suppVC:dose2) = 0 5.22 2.49 0.3 10.15 0.91 0.48 \*

---

'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.

'\*': For one-sided hypotheses, the posterior probability exceeds 95%;

for two-sided hypotheses, the value tested against lies outside the 95%-CI.

Posterior probabilities of point hypotheses assume equal prior probabilities.

Because we tested the hypotheses that each slope was equal to 0, the evidence ratio corresponds to evidence for a null effect (i.e., equal to 0) rather than evidence for an effect that is not equal to 0 (which would typically correspond to the alternative model. Thus, I will interpret the evidence ratio for a null effect the *BF*01 and the inverse of this as the *BF*10.

For the slope of suppVC, the *BF*01 was 0.28, indicating that the slope being equal to zero is 0.28 times more likely than the slope not being equal to zero. The corresponding *BF*10 was 3.57, indicating that an effect not equal to zero is 3.57 times more likely than an effect that is equal to zero. This reflects moderate evidence that the slope for suppVC is not equal to zero. In other words, we are confident enough in this slope to publish the finding, but not that confident in the grand scheme of things.

For the slope of dose2, the *BF*01 calculated by hypothesis was 0, implying very small evidence that a null effect is more likely than an effect not equal to zero. The corresponding *BF*10 cannot be calculated directly using this result, but very strong evidence for a slope not equal to zero is implied. This reflects very strong (perhaps even “extreme”) evidence that the slope for dose2 is not equal to zero. In other words, we are incredibly confident in this slope.

For the slope of suppVC:dose2, the *BF*01 was 0.91, indicating that the slope being equal to zero is 0.91 times more likely than the slope not being equal to zero. The corresponding *BF*10 was 1.10, indicating that an effect not equal to zero is 1.10 times more likely than an effect that is equal to zero. This reflects anecdotal and essentially equivocal evidence that the slope for suppVC is not equal to zero. In other words, we are not at all confident in this slope.

Q11. Refer to the “Question 11” section in the R script.

> hypothesis(mt.2, 'dose2 + Intercept = 0')

Hypothesis Tests for class b:

Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star

1 (dose2+Intercept) = 0 26.03 1.24 23.51 28.43 NA NA \*

---

'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.

'\*': For one-sided hypotheses, the posterior probability exceeds 95%;

for two-sided hypotheses, the value tested against lies outside the 95%-CI.

Posterior probabilities of point hypotheses assume equal prior probabilities.

Q12. Refer to the “Question 12” section in the R script.

> hypothesis(mt.2, 'dose2 > 0')

Hypothesis Tests for class b:

Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star

1 (dose2) > 0 12.83 1.77 9.92 15.71 Inf 1 \*

---

'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.

'\*': For one-sided hypotheses, the posterior probability exceeds 95%;

for two-sided hypotheses, the value tested against lies outside the 95%-CI.

Posterior probabilities of point hypotheses assume equal prior probabilities.

Evidence suggesting that the slope for dose is positive is extreme, with a directional bayes factor of +Inf in favor of the alternative hypothesis (i.e., a positive effect). Because 100% of the posterior probability for this slope is above zero, we are 100% certain that the effect is positive.

> hypothesis(mt.2, 'suppVC > 0')

Hypothesis Tests for class b:

Hypothesis Estimate Est.Error CI.Lower CI.Upper Evid.Ratio Post.Prob Star

1 (suppVC) > 0 -5.18 1.76 -7.99 -2.35 0 0

---

'CI': 90%-CI for one-sided and 95%-CI for two-sided hypotheses.

'\*': For one-sided hypotheses, the posterior probability exceeds 95%;

for two-sided hypotheses, the value tested against lies outside the 95%-CI.

Posterior probabilities of point hypotheses assume equal prior probabilities.

Evidence suggesting the slope for supp is positive is essentially non-existent, with a directional bayes factor of 0 for the alternative hypothesis (i.e., a positive effect). On the other hand, implied evidence suggesting the slope is *negative* is extreme. Because 0% of the posterior probability lies above zero, we are 100% certain that the effect is not positive.

Q13. Refer to “Question 13” section in the R script.

I conducted an experiment to evaluate the strengths of different interventions in increasing tooth growth in guinea pigs. The reason for this is that I want to harvest ivory from these animals in order to construct piano keys and decorative reliefs. If I can figure out how to produce enormously long teeth in guinea pigs, I should be well equipped to produce grandiose reliefs that rival those of the ancient Romans.

In my experiment, I compared two different interventions: Guinea pigs were supplied with either Vodka Crans (VC group) or smoothies from Orange Julius (OJ group) at a dose of either 0.5 or 1. Previous studies have championed the benefits of both Vodka Crans and Orange Julius in increasing guinea pig tooth length. Typically, Vodka Crans demonstrate larger gains. For either supplement, studies have shown that increases are more pronounced at higher doses. Thus, I predicted an increase in tooth length for either supplement, with larger increases associated with higher doses. I also expect to observe symptoms of alcohol intoxication for guinea pigs administered Vodka Crans.

Because the guinea pigs I purchased for this experiment are from a species known for their prominent teeth – and because my guinea pig supplier assured me that all specimens in this shipment had their teeth intact – I expect tooth length in my reference level to be fairly high. Formally, I expect tooth length in the reference level (Orange Julius at dose 0.5) to range from about 5 to 25, centered around 15. Although literature generally shows that both VC and OJ increase tooth length, reported effects have sometimes been very heterogeneous. Accordingly, I want to account for widely varying differences and will place a prior on the model slopes allowing variation to range from -50 to 50, centered on zero. Finally, I expect little residual error in the model and will implement a prior reflecting my belief that expect sigma to range from 0 to 2.

> mt.3

Family: gaussian

Links: mu = identity; sigma = identity

Formula: len ~ supp \* dose

Data: TG\_noHi (Number of observations: 40)

Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;

total post-warmup draws = 4000

Population-Level Effects:

Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

Intercept 13.22 1.10 11.05 15.38 1.00 2302 2459

suppVC -5.25 1.56 -8.25 -2.17 1.00 2012 2287

dose1 9.47 1.52 6.49 12.47 1.00 1892 2271

suppVC:dose1 -0.67 2.18 -4.86 3.62 1.00 1790 2066

Family Specific Parameters:

Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS

sigma 3.47 0.39 2.81 4.32 1.00 2913 2911

Draws were sampled using sampling(NUTS). For each parameter, Bulk\_ESS

and Tail\_ESS are effective sample size measures, and Rhat is the potential

scale reduction factor on split chains (at convergence, Rhat = 1).

**Table 2**

*Mean Tooth Length and 95% HPD Intervals as a Function of Supp and Dose*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Supp | Dose | Mean | HPD95% LB | HPD95% UB |
| Orange Julius | 0.5 | 13.22 | 11.10 | 15.40 |
| 1.0 | 22.70 | 20.70 | 25.00 |
| Vodka Cran | 0.5 | 8.98 | 5.90 | 10.20 |
| 1.0 | 16.75 | 14.70 | 18.90 |

A grid with red and blue dots

Description automatically generated

For the guinea pigs receiving Orange Julius smoothies, average tooth length was credibly higher for doses of 1.0 relative to doses of 0.5, with the difference between doses estimated at 9.46 (HPD95% = 6.43 – 12.40). The same pattern was observed for guinea pigs administered Vodka Crans, with the difference in tooth length between dose levels estimated at 8.82 (HPD95% = 5.70 – 11.90). Overall, this suggests that tooth length was greater for doses of 1.0 regardless of the supplement administered. Further contrasts revealed the superiority of Orange Julius smoothies in improving tooth length. For dose levels of 0.5, tooth length was credibly higher in the OJ group relative to the VC group, estimate = 5.29 (HPD95% = 2.14 – 8.17), which also held true for doses of 1.0, estimate = 5.93 (HPD95% = 2.81 – 8.94).

Considered in aggregate, my investigation suggested that both VC and OJ stimulate tooth growth better at higher doses, confirming one of my predictions. Contrary to my expectations, however, the model also revealed OJ to be superior to Vodka Crans. However, it is important to note that hope for Vodka Crans is not lost, as guinea pigs receiving this supplement reported higher life satisfaction. In terms of tooth length for the purposes of harvesting ivory, however, the model certainly suggests that I should administer OJ at does of 1.0 to future specimens.