# comparing two paired means using Bayes factors

Dr. Merlise Clyde

Duke University

# zinc in drinking water

| location | Surface | Bottom | Difference |
|---|---|---|---|
| 1 | 0.43 | 0.415 | 0.015 |
| 2 | 0.266 | 0.238 | 0.028 |
| 3 | 0.567 | 0.39 | 0.177 |
| 4 | 0.531 | 0.41 | 0.121 |
| 5 | 0.707 | 0.605 | 0.102 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 10 | 0.723 | 0.612 | 0.111 |

# model and hypotheses

- random sample of n = 10 differences $Y_1, \dots Y_n$

# model and hypotheses

- random sample of n = 10 differences $Y_1, \ldots Y_n$
- normal population with mean $\mu \equiv \mu_B - \mu_S$

# model and hypotheses

- random sample of n $=$ 10 differences $Y_1, \dots Y_n$
- normal population with mean $\mu \equiv \mu_B - \mu_S$
- variance $\sigma^2$ unknown

# model and hypotheses

- random sample of n = 10 differences $Y_1, \ldots Y_n$
- normal population with mean $\mu \equiv \mu_B - \mu_S$
- variance $\sigma^2$ unknown

no differences $H_1 : \mu_B = \mu_S \Leftrightarrow \mu = 0$

# model and hypotheses

- random sample of n = 10 differences $Y_1, \ldots Y_n$
- normal population with mean $\mu \equiv \mu_B - \mu_S$
- variance $\sigma^2$ unknown

no differences $H_1 : \mu_B = \mu_S \Leftrightarrow \mu = 0$

means are different $H_2 : \mu_B \neq \mu_S \Leftrightarrow \mu \neq 0$

# model and hypotheses

- random sample of n = 10 differences $Y_1, \ldots Y_n$
- normal population with mean $\mu \equiv \mu_B - \mu_S$
- variance $\sigma^2$ unknown

no differences $H_1 : \mu_B = \mu_S \Leftrightarrow \mu = 0$

means are different $H_2 : \mu_B \neq \mu_S \Leftrightarrow \mu \neq 0$

Bayes factor: $BF[H_1 : H_2] = \frac{p(\text{data}|H_1)}{p(\text{data}|H_2)}$

# Bayes factor

Bayes factor:

$$BF[H_1 : H_2] = \frac{\int p(\mathrm{data}|\mu{=}0,\sigma^2)p(\sigma^2|H_1)\ d\sigma^2}{\int\int p(\mathrm{data}|\mu,\sigma^2)p(\mu|\sigma^2,H_2)p(\sigma^2|H_2)\, d\mu\, d\sigma^2}$$

# Bayes factor

Bayes factor:

$$BF[H_1 : H_2] = \frac{\int p(\text{data}|\mu{=}0,\sigma^2)p(\sigma^2|H_1)\ d\sigma^2}{\int\int p(\text{data}|\mu,\sigma^2)p(\mu|\sigma^2,H_2)p(\sigma^2|H_2)\, d\mu\, d\sigma^2}$$

priors:

# Bayes factor

Bayes factor:

$$BF[H_1 : H_2] = \frac{\int p(\text{data}|\mu{=}0,\sigma^2)p(\sigma^2|H_1)\ \ d\sigma^2}{\int\int p(\text{data}|\mu,\sigma^2)p(\mu|\sigma^2,H_2)p(\sigma^2|H_2)\ d\mu\ d\sigma^2}$$

priors:

- $\mu \mid \sigma^2, H_2 \sim \mathsf{N}(0, \sigma^2/n_0)$

# Bayes factor

Bayes factor:

$$BF[H_1 : H_2] = \frac{\int p(\text{data}|\mu{=}0,\sigma^2)p(\sigma^2|H_1)\ d\sigma^2}{\int\int p(\text{data}|\mu,\sigma^2)p(\mu|\sigma^2,H_2)p(\sigma^2|H_2)\,d\mu\,d\sigma^2}$$

priors:

- $\mu \mid \sigma^2, H_2 \sim \mathsf{N}(0, \sigma^2/n_0)$
- $p(\sigma^2) \propto 1/\sigma^2$ for both $H_1$ and $H_2$

# Bayes factor with reference prior $\sigma^2$

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2+\nu}\right)^{\frac{\nu+1}{2}}$$

# Bayes factor with reference prior $\sigma^2$

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2+\nu}\right)^{\frac{\nu+1}{2}}$$

- t-statistic $\quad t = \dfrac{|\bar{Y}|}{s/\sqrt{n}}$

# Bayes factor with reference prior $\sigma^2$

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2+\nu}\right)^{\frac{\nu+1}{2}}$$

- t-statistic $$t = \frac{|\bar{Y}|}{s/\sqrt{n}}$$
- sample standard deviation s

# Bayes factor with reference prior $\sigma^2$

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2 + \nu}\right)^{\frac{\nu+1}{2}}$$

- t-statistic $\quad t = \dfrac{|\bar{Y}|}{s/\sqrt{n}}$
- sample standard deviation s
- degrees of freedom $\quad \nu = n - 1$

# Bayes factor with reference prior $\sigma^2$

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2 + \nu}\right)^{\frac{\nu+1}{2}}$$

- t-statistic $\quad t = \frac{|\bar{Y}|}{s/\sqrt{n}}$
- sample standard deviation s
- degrees of freedom $\quad \nu = n - 1$

$$n_0 \to 0,\ BF[H1 : H_2] \to \infty$$

# Bayes factor with reference prior $\sigma^2$

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2+\nu}\right)^{\frac{\nu+1}{2}}$$

- t-statistic $\quad t = \dfrac{|\bar{Y}|}{s/\sqrt{n}}$
- sample standard deviation s
- degrees of freedom $\quad \nu = n - 1$

  $n_0 \to 0, \ BF[H1 : H_2] \to \infty$
- no improper or vague prior on $\mu$

# information paradox

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2 + \nu}\right)^{\frac{\nu+1}{2}}$$

# information paradox

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2 + \nu}\right)^{\frac{\nu+1}{2}}$$

- fixed $n$ and $n_0$

# information paradox

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2+\nu}\right)^{\frac{\nu+1}{2}}$$

- fixed $n$ and $n_0$
- t-statistic $|t| \to \infty$

# information paradox

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2 + \nu}\right)^{\frac{\nu+1}{2}}$$

- fixed $n$ and $n_0$
- t-statistic $|t| \to \infty$
- bounded Bayes factor

$$BF[H_1 : H_2] \to \left(\frac{n_0}{n_0+n}\right)^{\frac{n-1}{2}}$$

# information paradox

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2 + \nu}\right)^{\frac{\nu+1}{2}}$$

- fixed $n$ and $n_0$
- t-statistic $|t| \to \infty$
- bounded Bayes factor

$$BF[H_1 : H_2] \to \left(\frac{n_0}{n_0+n}\right)^{\frac{n-1}{2}}$$

- Jeffrey's recommended a Cauchy prior $\mathsf{C}(0, r^2\sigma^2)$

# information paradox

$$BF[H_1 : H_2] = \left(\frac{n+n_0}{n_0}\right)^{1/2} \left(\frac{t^2 \frac{n_0}{n+n_0} + \nu}{t^2+\nu}\right)^{\frac{\nu+1}{2}}$$

- fixed $n$ and $n_0$
- t-statistic $|t| \to \infty$
- bounded Bayes factor

$$BF[H_1 : H_2] \to \left(\frac{n_0}{n_0+n}\right)^{\frac{n-1}{2}}$$

- Jeffrey's recommended a Cauchy prior $\mathsf{C}(0, r^2\sigma^2)$
- Jeffrey-Zellner-Siow or 'JZS' prior

# zinc concentration example

```r
bayes_inference(difference, data=zinc, statistic="mean", type="ht",
                prior="JZS", mu_0=0, method="theo", alt="twosided")

## Single numerical variable
## n = 10, y-bar = 0.0804, s = 0.0523
## (Using Zellner-Siow Cauchy prior:  mu ~ C(0, 1*sigma)
## (Using Jeffreys prior: p(sigma^2) = 1/sigma^2
##
## Hypotheses:
## H1: mu = 0 versus H2: mu != 0
## Priors:
## P(H1) = 0.5 , P(H2) = 0.5
## Results:
## BF[H2:H1] = 50.7757
## P(H1|data) = 0.0193  P(H2|data) = 0.9807
##
```

# summary

- Cauchy prior for $\mu$ and hypothesis testing

# summary

- Cauchy prior for $\mu$ and hypothesis testing
- requires numerical integration

# summary

- Cauchy prior for $\mu$ and hypothesis testing
- requires numerical integration
- modify scale $r < 1$ if expect smaller effect

# summary

- Cauchy prior for $\mu$ and hypothesis testing
- requires numerical integration
- modify scale $r < 1$ if expect smaller effect
- use `bayes_inference` for credible intervals

# summary

- Cauchy prior for $\mu$ and hypothesis testing
- requires numerical integration
- modify scale $r < 1$ if expect smaller effect
- use `bayes_inference` for credible intervals

next: comparing two means