comparing two independent means: hypothesis testing

Dr. Merlise Clyde

Duke University

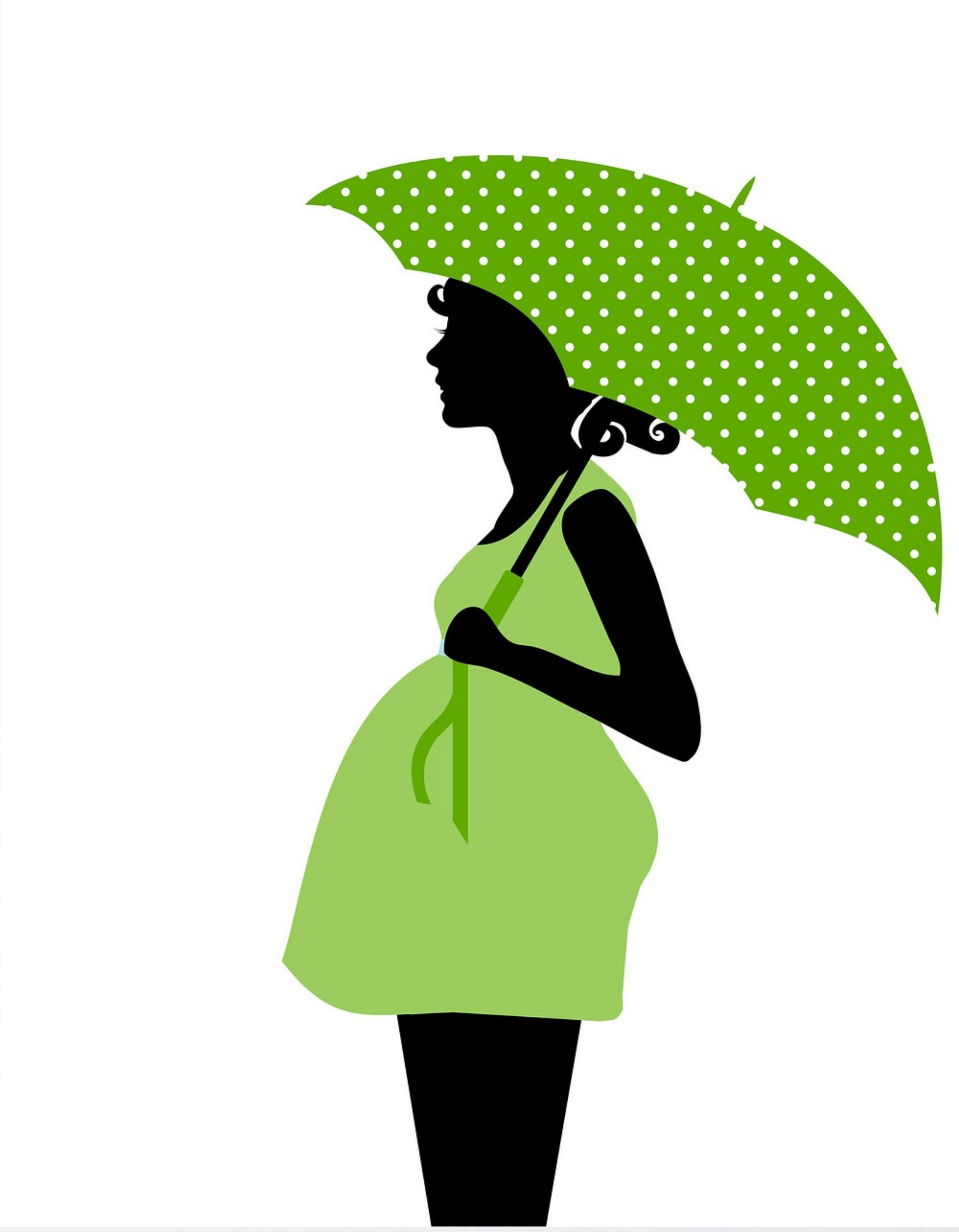
weight gain during pregnancy and mother's age

2004 survey of births on North Carolina

weight gain during pregnancy and mother's age

2004 survey of births on North Carolina

gained weight gain during pregnancy in pounds

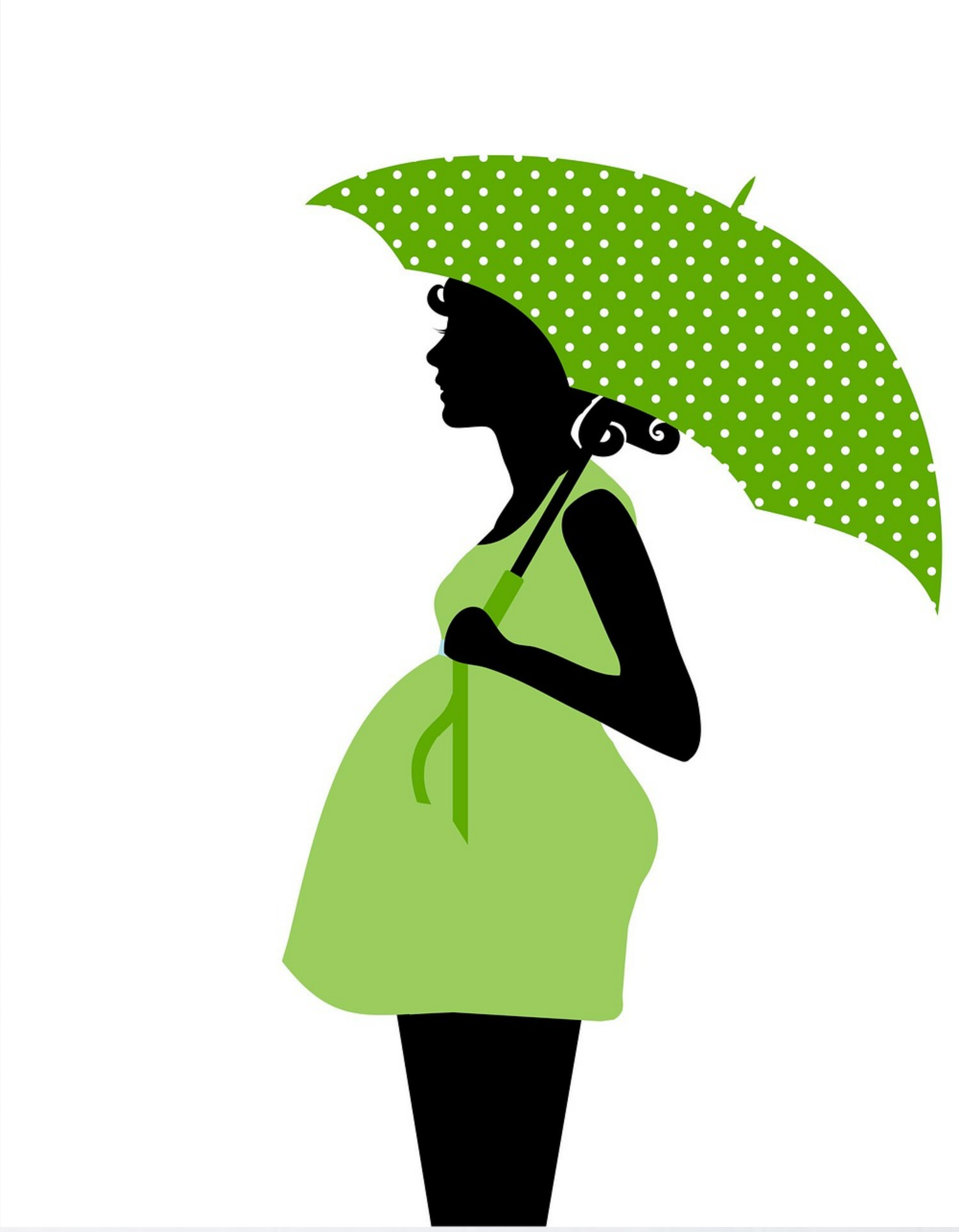


https://pixabay.com/en/woman-pregnancy-female-lady-163617/

weight gain during pregnancy and mother's age

2004 survey of births on North Carolina

gained	weight gain during pregnancy in pounds
mature	categorical variable with 'younger mom' and 'older mom'



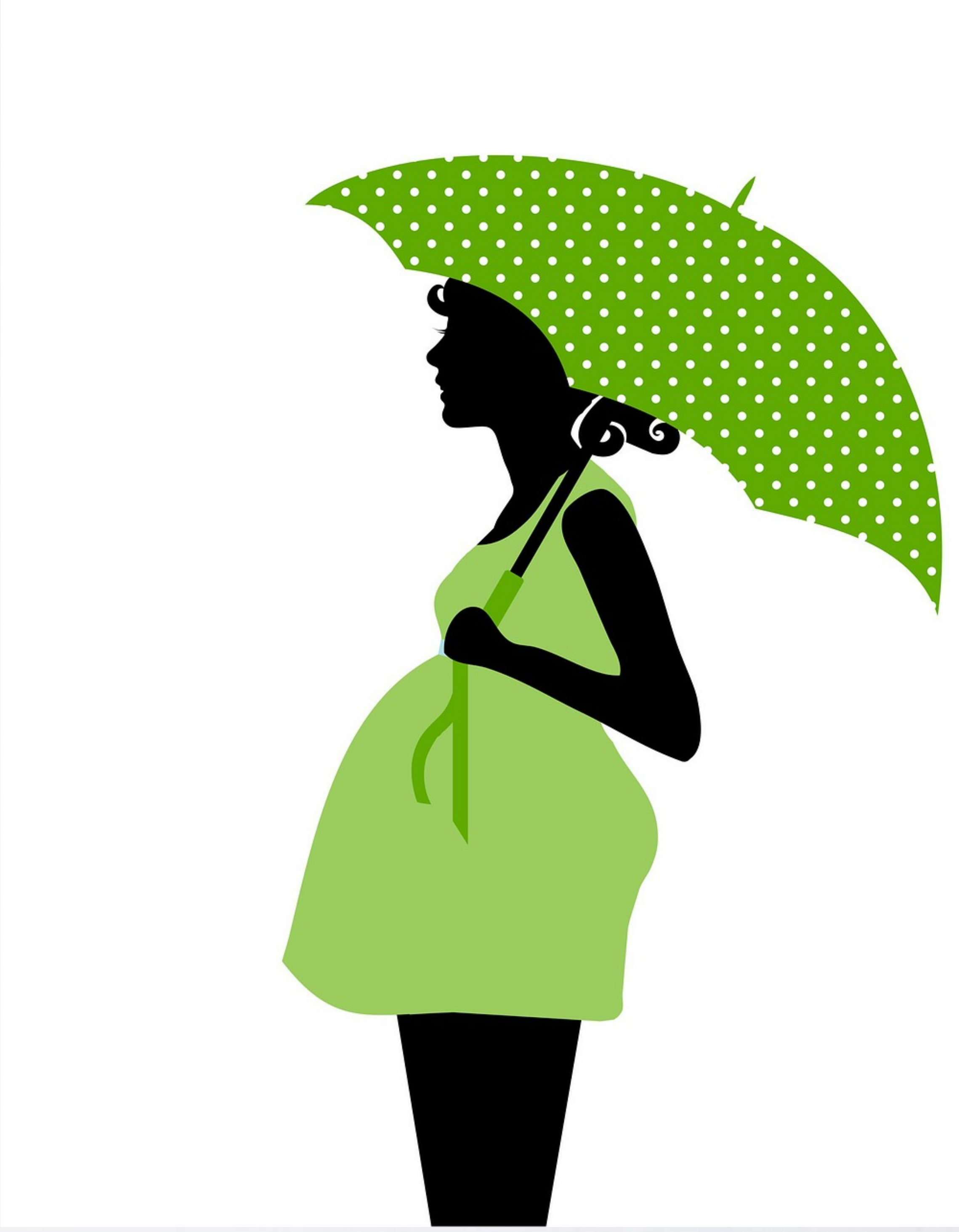
https://pixabay.com/en/woman-pregnancy-female-lady-163617/

weight gain during pregnancy and mother's age

2004 survey of births on North Carolina

gained	weight gain during pregnancy in pounds
mature	categorical variable with 'younger mom' and 'older mom'

is the average weight gain of older mothers different from the average weight gain of younger mothers?



https://pixabay.com/en/woman-pregnancy-female-lady-163617/

model and priors

- random sample from two populations

model and priors

- random sample from two populations

$$Y_{O,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu + \alpha/2, \sigma)$$

model and priors

- random sample from two populations

$$Y_{O,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu + \alpha/2, \sigma)$$

$$Y_{Y,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu - \alpha/2, \sigma)$$

model and priors

- random sample from two populations

$$Y_{O,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu + \alpha/2, \sigma)$$

$$Y_{Y,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu - \alpha/2, \sigma)$$

- difference in population means $\mu + \alpha/2 - (\mu + \alpha/2) = \alpha$

model and priors

- random sample from two populations

$$Y_{O,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu + \alpha/2, \sigma)$$

$$Y_{Y,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu - \alpha/2, \sigma)$$

- difference in population means $\mu + \alpha/2 - (\mu + \alpha/2) = \alpha$
- $H_1 : \alpha = 0$ versus $H_2 : \alpha \neq 0$

model and priors

- random sample from two populations

$$Y_{O,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu + \alpha/2, \sigma)$$

$$Y_{Y,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu - \alpha/2, \sigma)$$

- difference in population means $\mu + \alpha/2 - (\mu + \alpha/2) = \alpha$
- $H_1 : \alpha = 0$ versus $H_2 : \alpha \neq 0$
- μ overall average weight gain for all women

model and priors

- random sample from two populations

$$Y_{O,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu + \alpha/2, \sigma)$$

$$Y_{Y,i} \overset{\text{iid}}{\sim} \mathsf{N}(\mu - \alpha/2, \sigma)$$

- difference in population means $\mu + \alpha/2 - (\mu + \alpha/2) = \alpha$
- $H_1 : \alpha = 0$ versus $H_2 : \alpha \neq 0$
- μ overall average weight gain for all women
- need prior distributions for α, μ, σ^2 under both hypotheses

Bayes factor

- Bayes factor

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

Bayes factor

- Bayes factor

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

Bayes factor

- Bayes factor

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

- under H_2: $\delta = \alpha/\sigma^2 \sim \mathsf{C}(0, r^2)$

Bayes factor

- Bayes factor

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

- under H_2: $\delta = \alpha/\sigma^2 \sim \mathsf{C}(0, r^2)$
- under both H_1 and H_2 use reference prior $p(\mu, \sigma^2) \propto 1/\sigma^2$

Bayes factor

- Bayes factor

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

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

- under H_2: $\delta = \alpha/\sigma^2 \sim \mathsf{C}(0, r^2)$
- under both H_1 and H_2 use reference prior $p(\mu, \sigma^2) \propto 1/\sigma^2$
- Jeffreys-Zellner-Siow or 'JZS'

R code

```r
library(statsr)
data(nc)
bayes_inference(y=gained, x=mature, data=nc,type='ht',
                statistic='mean',  alternative='twosided', null=0,
                prior='JZS', r=1, method='theo', show_summ=FALSE)

## Hypotheses:
## H1: mu_mature mom  = mu_younger mom
## H2: mu_mature mom != mu_younger mom
##
## Priors: P(H1) = 0.5  P(H2) = 0.5
##
## Results:
## BF[H1:H2] = 5.7162
## P(H1|data) = 0.8511
## P(H2|data) = 0.1489
##
```

summary

- compared two independent means using Bayesian hypothesis testing

summary

- compared two independent means using Bayesian hypothesis testing
- assumption that variances are the same in the two groups

summary

- compared two independent means using Bayesian hypothesis testing
- assumption that variances are the same in the two groups
- Behren-Fisher problem

summary

- compared two independent means using Bayesian hypothesis testing
- assumption that variances are the same in the two groups
- Behren-Fisher problem

summary

- compared two independent means using Bayesian hypothesis testing
- assumption that variances are the same in the two groups
- Behren-Fisher problem

next: putting it all together and summarizing results