

I09 - Comparing means

STAT 401 (Engineering) - Iowa State University

February 26, 2018

One mean

Consider the model $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$. We have discussed a number of statistical procedures to draw inferences about μ :

- pvalue for a hypothesis test, e.g. $H_0 : \mu = \mu_0$,
- confidence interval for μ ,
- credible interval for μ ,
- posterior model probability, e.g. $p(H_0|y)$, and
- posterior probabilities, e.g. $P(\mu < \mu_0|y)$.

Now, we will consider what happens when you have multiple μ s.

Two means

Consider the model

$$Y_{i,j} \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma_i^2)$$

for $i = 1, 2$ and $j = 1, \dots, n_i$. and you are interested in the relationship between μ_1 and μ_2 . We can perform the following statistical procedures:

- pvalue for a hypothesis test, e.g. $H_0 : \mu_1 = \mu_2$,
- confidence interval for $\mu_1 - \mu_2$,
- credible interval for $\mu_1 - \mu_2$,
- posterior model probability, e.g. $p(H_0|y)$, and
- probability statements, e.g. $P(\mu_1 < \mu_2|y)$.

where $y = (y_1, y_2)$.

Simulating data

We will simulate the following data

$$Y_{i,j} \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma_i^2)$$

for $i = 1, 2, 3$ and $j = 1, \dots, n_i$. For the moment we will only use the first two groups, but later we will use all 3 groups.

```
set.seed(20170301)
d <- bind_rows(
  data_frame(process = "P1", sensitivity = rnorm(22, 7.8, 2.3)),
  data_frame(process = "P2", sensitivity = rnorm(34, 9.3, 2.3)),
  data_frame(process = "P3", sensitivity = rnorm( 7, 10.0, 2.3))

# readr::write_csv(d, path="sensitivity.csv")
```

Data example

Suppose you have two manufacturing processes to produce sensors and you are interested in the average sensitivity of the sensors.

So you run the two processes and record the sensitivity of each sensor in units of mV/V/mm Hg

(<http://www.ni.com/white-paper/14860/en/>). And you have the following summary statistics:

```
# d <- readr::read_csv("sensitivity.csv")
d2 <- d %>% filter(process %in% c("P1", "P2"))
```

```
sm <- d2 %>%
  group_by(process) %>%
  summarize(
    n    = n(),
    mean = mean(sensitivity),
    sd   = sd(sensitivity)
  )
sm
```

```
# A tibble: 2 x 4
  process     n mean   sd
  <chr>   <int> <dbl> <dbl>
1 P1       22  7.74  1.87
2 P2       34  9.24  2.26
```

Pvalues and confidence intervals

Because there is no indication that you have any expectation regarding the sensitivities of process 1 compared to process 2, we will conduct a two-sided **two-sample t-test** assuming the variances are not equal, i.e.

$$Y_{i,j} \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma_i^2)$$

and

$$H_0 : \mu_1 = \mu_2 \quad \text{and} \quad H_A : \mu_1 \neq \mu_2$$

```
t.test(sensitivity ~ process, data = d2)
```

Welch Two Sample t-test

data: sensitivity by process

t = -2.6932, df = 50.649, p-value = 0.009571

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2.610398 -0.380530

sample estimates:

mean in group P1 mean in group P2

7.743761

9.239224

Posteriors for μ

When

$$Y_{i,j} \stackrel{ind}{\sim} N(\mu_i, \sigma_i^2)$$

Recall that

$$\mu_i | y_i \stackrel{ind}{\sim} t_{n_i-1}(\bar{y}_i, s_i^2/n_i)$$

and that a draw for μ_i can be obtained by taking

$$\bar{y}_i + T_{n_i-1} s_i / \sqrt{n_i}, \quad T_{n_i-1} \stackrel{ind}{\sim} t_{n_i-1}(0, 1).$$

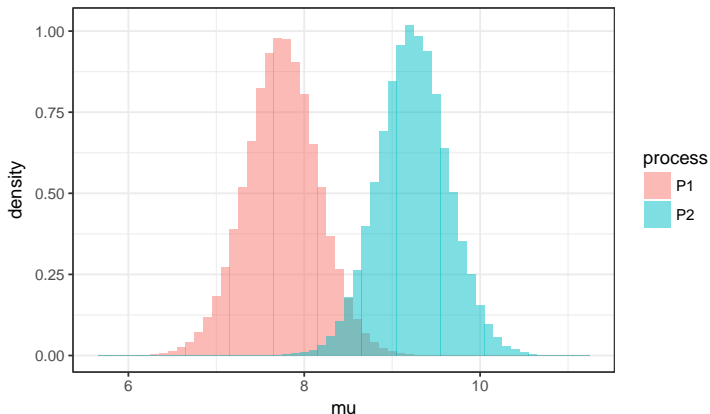
Simulations:

```
nr = 1e5
sims <- bind_rows(
  data_frame(
    rep = 1:nr,
    process = "P1",
    mu = sm$mean[1] + rt(nr, df = sm$n[1]-1) * sm$sd[1] / sqrt(sm$n[1])),
  data_frame(
    rep = 1:nr,
    process = "P2",
    mu = sm$mean[2] + rt(nr, df = sm$n[2]-1) * sm$sd[2] / sqrt(sm$n[2]))
)
```

We can use these draws to compare the posteriors

Compare posterior histograms for μ_1 and μ_2 .

```
ggplot(sims, aes(x=mu, y=..density.., fill=process, group=process)) +  
  geom_histogram(position = "identity", alpha=0.5, binwidth=0.1) +  
  theme_bw()
```



Credible interval for the difference

To obtain statistical inference on the difference, we use the samples and take the difference

```
d3 <- sims %>%
  spread(process, mu) %>%
  mutate(diff = P1-P2)

# Bayes estimate for the difference
mean(d3$diff)

[1] -1.493267

# Estimated 95% equal-tail credible interval
quantile(d3$diff, c(.025,.975))

      2.5%      97.5%
-2.6339752 -0.3483025

# Estimate of the probability that mu1 is larger than mu2
mean(d3$diff > 0)

[1] 0.00591
```

Three or more means

Now, let's consider the more general problem of

$$Y_{i,j} \stackrel{ind}{\sim} N(\mu_i, \sigma_i^2)$$

for $i = 1, 2, \dots, I$ and $j = 1, \dots, n_i$ and you are interested in the relationship amongst the μ_i .

We can perform the following statistical procedures:

- pvalue for a hypothesis test, e.g. $H_0 : \mu_i = \mu$ for all i ,
- confidence interval for $\mu_i - \mu_j$ for a specified i and j ,
- credible interval for $\mu_i - \mu_j$ for a specified i and j ,
- posterior model probability, e.g. $p(H_0|y)$, and
- probability statements, e.g. $P(\mu_i < \mu_j|y)$ for a specified i and j .

Data example

Suppose you have three manufacturing processes to produce sensors and you are interested in the average sensitivity of the sensors.

So you run the three processes and record the sensitivity of each sensor in units of mV/V/mm Hg (<http://www.ni.com/white-paper/14860/en/>). And you have the following summary statistics:

```
sm <- d %>%
  group_by(process) %>%
  summarize(
    n      = n(),
    mean   = mean(sensitivity),
    sd     = sd(sensitivity)
  )
sm
```

```
# A tibble: 3 x 4
  process      n mean   sd
<chr>    <int> <dbl> <dbl>
1 P1         22  7.74  1.87
2 P2         34  9.24  2.26
3 P3          7 10.8   1.96
```

Pvalues

When there are lots of means, the first null hypothesis is typically

$$H_0 : \mu_i = \mu \forall i$$

```
oneway.test(sensitivity ~ process, data = d)
```

One-way analysis of means (not assuming equal variances)

```
data:  sensitivity and process
F = 7.6287, num df = 2.000, denom df = 17.418, p-value = 0.004174
```

Then we typically look at pairwise differences:

```
pairwise.t.test(d$sensitivity,
               d$process,
               pool.sd = FALSE,
               p.adjust.method = "none")
```

Pairwise comparisons using t tests with non-pooled SD

```
data:  d$sensitivity and d$process
```

	P1	P2
P2	0.0096	-
P3	0.0045	0.0870

P value adjustment method: none

Posteriors for μ

Recall that

$$\mu_i | y_i \stackrel{\text{ind}}{\sim} t_{n_i-1}(\bar{y}_i, s_i^2/n_i)$$

and that a draw for μ_i can be obtained by taking

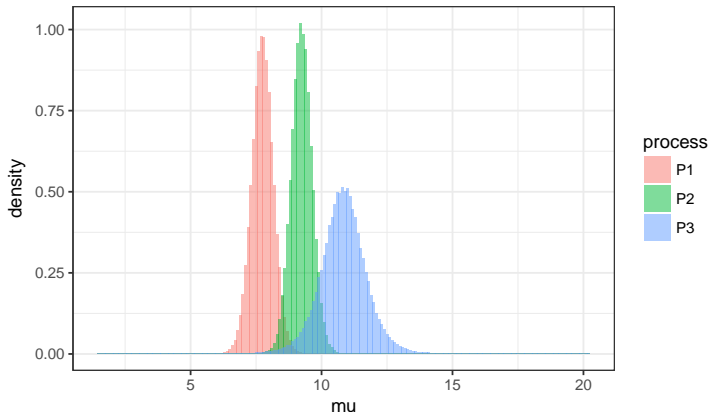
$$\bar{y}_i + T_{n_i-1} s_i / \sqrt{n_i}, \quad T_{n_i-1} \stackrel{\text{ind}}{\sim} t_{n_i-1}(0, 1).$$

```
sims <- bind_rows(
  sims, # groups 1 and 2
  data_frame(
    rep = 1:nr,
    process = "P3",
    mu = sm$mean[3] + rt(nr, df = sm$n[3]-1) * sm$sd[3] / sqrt(sm$n[3]))
)
```

We can use these draws to compare the posteriors

We can obtain posteriors for μ and plot histograms (or smoothed histograms) to compare the posteriors.

```
ggplot(sims, aes(x=mu, y=..density.., fill=process, group=process)) +  
  geom_histogram(position = "identity", alpha=0.5, binwidth=0.1) +  
  theme_bw()
```



Credible intervals for differences

Use the simulations to calculate posterior probabilities and credible intervals for differences.

```
# Estimate of the probability that one mean is larger than another
sims %>%
  spread(process, mu) %>%
  mutate('mu1-mu2' = P1-P2,
         'mu1-mu3' = P1-P3,
         'mu2-mu3' = P2-P3) %>%
  select('mu1-mu2', 'mu1-mu3', 'mu2-mu3') %>%
  gather(comparison, diff) %>%
  group_by(comparison) %>%
  summarize(probability = mean(diff>0) %>% round(4),
           lower = quantile(diff, .025) %>% round(2),
           upper = quantile(diff, .975) %>% round(2)) %>%
  mutate(credible_interval = paste("(", lower, ", ", upper, ")", sep="")) %>%
  select(comparison, probability, credible_interval)
```

```
# A tibble: 3 x 3
  comparison probability credible_interval
  <chr>         <dbl> <chr>
1 mu1-mu2      0.00590 (-2.63,-0.35)
2 mu1-mu3      0.00400 (-5.07,-1.11)
3 mu2-mu3      0.0492  (-3.55,0.36)
```

Common variance model

Especially when there is a small amount of data, it is common to assume that each group has the same variability. So we typically assume

$$Y_{ij} \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma^2)$$

We can calculate a pvalue for the following null hypothesis:

$$H_0 : \sigma_i = \sigma \forall i$$

```
bartlett.test(sensitivity ~ process, data = d)
```

```
Bartlett test of homogeneity of variances
```

```
data: sensitivity by process
```

```
Bartlett's K-squared = 0.90949, df = 2, p-value = 0.6346
```


Comparing means when the variances are equal

Now we can test

$$H_0 : \mu_i = \mu \forall i$$

within the model $Y_{ij} \overset{ind}{\sim} N(\mu_i, \sigma^2)$.

```
oneway.test(sensitivity ~ process, data = d, var.equal = TRUE)
```

One-way analysis of means

```
data:  sensitivity and process
F = 6.7543, num df = 2, denom df = 60, p-value = 0.002261
```

Then we typically look at pairwise differences:

```
pairwise.t.test(d$sensitivity, d$process, p.adjust.method = "none")
```

Pairwise comparisons using t tests with pooled SD

```
data:  d$sensitivity and d$process
```

```
      P1      P2
P2 0.0116 -
P3 0.0012 0.0720
```

```
P value adjustment method: none
```

Posteriors for μ

If $Y_{ij} \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma^2)$ and we use the prior $p(\mu_1, \dots, \mu_I, \sigma^2) \propto 1/\sigma^2$, then

$$\mu_i | y_i, \sigma^2 \stackrel{\text{ind}}{\sim} N(\bar{y}_i, \sigma^2/n_i) \quad \sigma^2 | y \sim IG \left(\frac{n - I}{2}, \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \right)$$

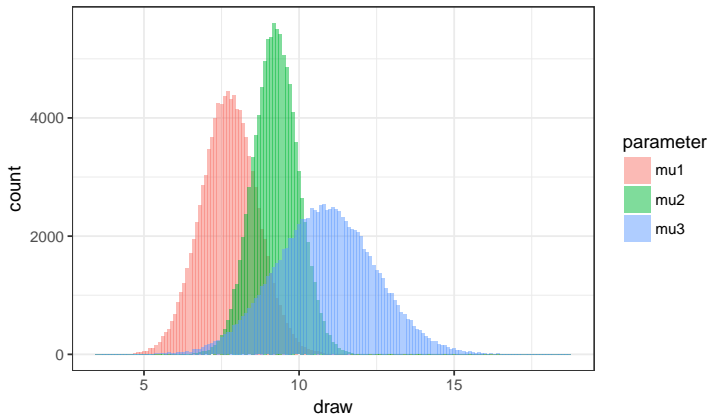
where $n = \sum_{i=1}^I n_i$. and thus, we obtain joint samples for μ by performing the following

1. $\sigma^{2(m)} \sim p(\sigma^2 | y)$
2. For $i = 1, \dots, I$, $\mu_i \sim p(\mu | y_i, \sigma^{2(m)})$.

```
nr = 1e5
sims <- data.frame(rep = 1:nr,
  sigma = 1/sqrt( rgamma(nr,
    shape = sum(sm$n-1)/2,
    rate = sum((sm$n-1)*sm$sd^2)/2 )) %>%
mutate(
  mu1 = rnorm(nr, mean = sm$mean[1], sd = sigma / sqrt(sm$n[1])),
  mu2 = rnorm(nr, mean = sm$mean[2], sd = sigma / sqrt(sm$n[2])),
  mu3 = rnorm(nr, mean = sm$mean[3], sd = sigma / sqrt(sm$n[3])))
```

We can use these draws to compare the posteriors

We can obtain posteriors for μ and plot histograms (or smoothed histograms) to compare the posteriors.



Credible interval for the differences

To compare the means, we compare the samples drawn from the posterior.

```
sims %>%
  mutate('mu1-mu2' = mu1-mu2,
         'mu1-mu3' = mu1-mu3,
         'mu2-mu3' = mu2-mu3) %>%
  select('mu1-mu2', 'mu1-mu3', 'mu2-mu3') %>%
  gather(comparison, diff) %>%
  group_by(comparison) %>%
  summarize(probability = mean(diff>0) %>% round(4),
           lower = quantile(diff, .025) %>% round(2),
           upper = quantile(diff, .975) %>% round(2)) %>%
  mutate(credible_interval = paste("(", lower, ", ", upper, ")", sep="")) %>%
  select(comparison, probability, credible_interval)
```

```
# A tibble: 3 x 3
  comparison probability credible_interval
  <chr>          <dbl> <chr>
1 mu1-mu2      0.0979 (-3.78,0.79)
2 mu1-mu3      0.0468 (-6.74,0.55)
3 mu2-mu3      0.181  (-5.08,1.85)
```