# I09 - Comparing means

STAT 5870 (Engineering)
Iowa State University

October 16, 2024

# 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 $\mu$:

- Frequentist: based on distribution of $\frac{\overline{Y}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
  - $p$-value for a hypothesis test, e.g. $H_0 : \mu = m$,
  - confidence interval for $\mu$,
- Bayesian: $\frac{\mu-\overline{y}}{s/\sqrt{n}} \sim t_{n-1}$
  - credible interval for $\mu$,
  - posterior model probability, e.g. $p(H_0|y)$, and
  - posterior probabilities, e.g. $P(\mu < m|y)$.

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

# Two means

Consider the model

$$Y_{g,i} \stackrel{ind}{\sim} N(\mu_g, \sigma^2)$$

for $g = 1, 2$ and $i = 1, \ldots, n_g$. and you are interested in the relationship between $\mu_1$ and $\mu_2$.

- Frequentist: based on distribution of

$$\frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2} \quad \text{where} \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

  - $p$-value for a hypothesis test, e.g. $H_0 : \mu_1 - \mu_2 = d$,
  - confidence interval for $\mu_1 - \mu_2$,

- Bayesian

$$\frac{\mu_1 - \mu_2 - (\overline{y}_1 + \overline{y}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

  - credible interval for $\mu_1 - \mu_2$,
  - probability statements, e.g. $P(\mu_1 - \mu_2 < d|y)$.

where $y = (y_{1,1}, \ldots, y_{1,n_1}, y_{2,1}, \ldots, y_{2,n_2})$.

# 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 observe the following summary statistics:

```
# 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
```

Let $Y_{g,i}$ be the sensitivity of the $i$th sensor in the $g$th group. Assume

$$Y_{g,i} \stackrel{ind}{\sim} N(\mu_g, \sigma^2).$$

# Frequentist analysis formulas

Consider the hypothesis $H_0 : \mu_1 = \mu_2$ or, equivalently, $H_0 : \mu_1 - \mu_2 = 0$. We calculate the $p$-value using

$$2P\left(T_{n_1+n_2-1} < -|t|\right) \quad \text{where} \quad t = \frac{\overline{y}_1 - \overline{y}_2 - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

We calculate a $100(1-a)\%$ confidence interval using

$$\overline{y}_1 - \overline{y}_2 \pm t_{n_1+n_2-2,1-a/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

# Frequentist analysis by hand

```
v    <- sm$n[1] + sm$n[2] - 2
diff <- sm$mean[1] - sm$mean[2]

# Calculate standard error
sp2  <- ( (sm$n[1]-1)*sm$sd[1]^2 + (sm$n[2]-1)*sm$sd[2]^2 ) / v # Pooled variance
sp   <- sqrt(sp2)
se   <- sp * sqrt(1/sm$n[1] + 1/sm$n[2])

# Two-sided p-value
2 * pt(-abs(diff / se), df = v)

[1] 0.01245222

# Equal-tail confidence interval
diff + c(-1,1) * qt(.975, df = v) * se

[1] -2.655043 -0.335885
```

# Bayesian analysis formulas

We calculate a $100(1-a)\%$ credible interval using

$$\overline{y}_1 - \overline{y}_2 \pm t_{n_1+n_2-2,1-a/2} s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}.$$

We calculate a posterior probability using

$$P(\mu_1-\mu_2<0|y) = P\left(T_{n_1+n_2-2} < \frac{0-(\overline{y}_1-\overline{y}_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right)$$

Thus, half the $p$-value corresponds to either $P(\mu_1-\mu_2<0|y)$ or $P(\mu_1-\mu_2>0|y)$.

# Analyses using t.test

```
(tt <- t.test(sensitivity ~ process, data = d2, var.equal = TRUE))


Two Sample t-test

data:  sensitivity by process
t = -2.5856, df = 54, p-value = 0.01245
alternative hypothesis: true difference in means between group P1 and group P2 is not equal to 0
95 percent confidence interval:
 -2.655043 -0.335885
sample estimates:
mean in group P1 mean in group P2
        7.743761         9.239224

# Since estimate of the difference is negative
# the following is P(mu_1 - mu_2 > 0)
tt$p.value / 2

[1] 0.006226109
```

# Unequal variances

Consider the model

$$Y_{g,i} \overset{ind}{\sim} N(\mu_g, {\sigma_g}^2)$$

for $g = 1, 2$ and $i = 1, \ldots, n_g$. and you are interested in the relationship between $\mu_1$ and $\mu_2$.
Frequentist:

$$\frac{\overline{Y}_1 - \overline{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \dot{\sim} t_v$$

using Satterthwaite approximation for the degrees of freedo $v$. But there is controversy in using this approximate distribution.

Bayesian:

$$\left.\frac{\mu_g - \overline{y}_g}{s_g/\sqrt{n_g}}\right| y \overset{ind}{\sim} t_{n_g - 1}$$

Simulate means separately and take the difference.

# Analyses using t.test

```
t.test(sensitivity ~ process,
       data = d2,
       var.equal = FALSE)      # this was the default

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 between group P1 and group P2 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
```

# Posterior for $\mu_1, \mu_2$

Assume

$$Y_{g,i} \stackrel{ind}{\sim} N(\mu_g, \sigma_g^2) \quad \text{and} \quad p(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \propto \frac{1}{\sigma_1^2}\frac{1}{\sigma_2^2}.$$

Then

$$\mu_g|y \stackrel{ind}{\sim} t_{n_g-1}(\overline{y}_g, s_g^2/n_g)$$

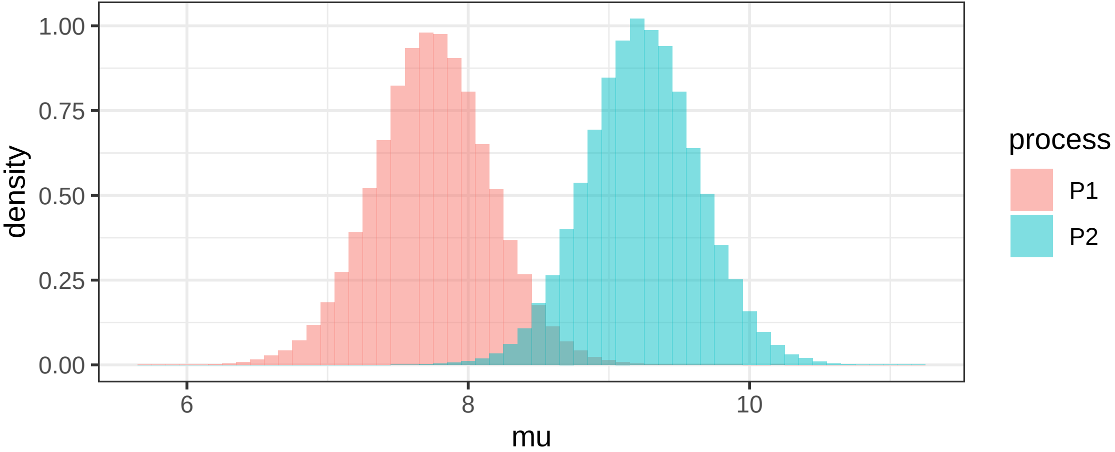
and a draw for $\mu_g$ can be obtained by taking

$$\overline{y}_g + T_{n_g-1} s_g/\sqrt{n_g}, \quad T_{n_g-1} \stackrel{ind}{\sim} t_{n_g-1}(0,1).$$

# Bayesian analysis in R

```
nr = 1e5
sims <- bind_rows(
  tibble(          # tibble is just a special 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])),
  tibble(
    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



# 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 smaller than mu2
mean(d3$diff < 0)

[1] 0.99409
```

# Three or more means

Now, let's consider the more general problem of

$$Y_{g,i} \stackrel{ind}{\sim} N(\mu_g, \sigma_g^2)$$

for $g = 1, 2, \ldots, G$ and $i = 1, \ldots, n_g$ and you are interested in the relationship amongst the $\mu_g$.

We can perform the following statistical procedures:

- Frequentist:
  - $p$-value for test of $H_0 : \mu_g = \mu$ for all $g$,
  - confidence interval for $\mu_g - \mu_{g'}$,
- Bayesian: based on posterior for $\mu_1, \ldots, \mu_G$
  - credible interval for $\mu_g - \mu_{g'}$ ,
  - probability statements, e.g. $P(\mu_g < \mu_{g'}|y)$

  where $g$ and $g'$ are two different groups.

# 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:

```
# 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
```

# $p$-values

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

$$H_0 : \mu_g = \mu \,\forall\, g$$

```
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
```

# Pairwise differences

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 $\mu$

When

$$Y_{g,i} \stackrel{ind}{\sim} N(\mu_g, \sigma_g^2),$$

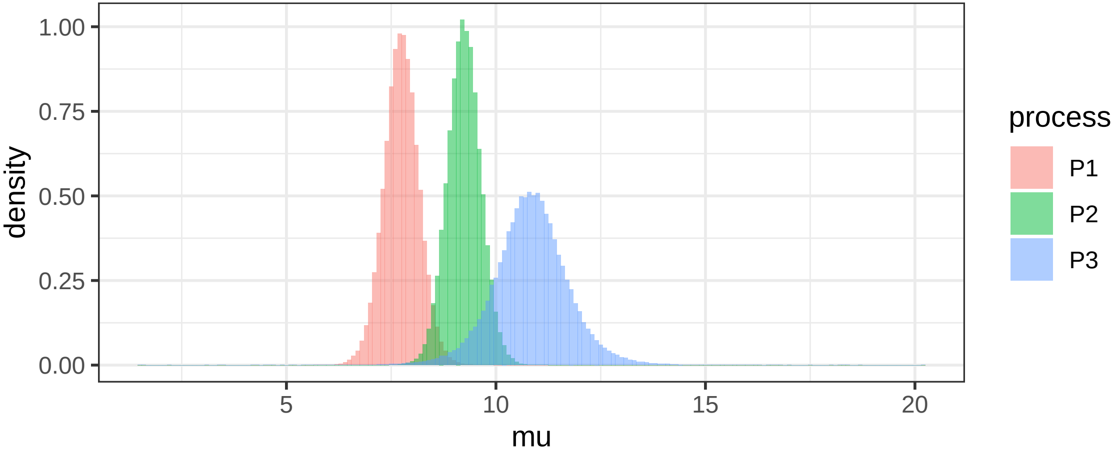
we have

$$\mu_g | y \stackrel{ind}{\sim} t_{n_g-1}(\overline{y}_g, s_g^2/n_g)$$

and that a draw for $\mu_g$ can be obtained by taking

$$\overline{y}_g + T_{n_g-1} s_g/\sqrt{n_g}, \quad T_{n_g-1} \stackrel{ind}{\sim} t_{n_g-1}(0,1).$$

# Compare posteriors



# 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         0.0059 (-2.63,-0.35)
2 mu1-mu3         0.0037 (-5.06,-1.11)
3 mu2-mu3         0.0493 (-3.56,0.37)
```

# Common variance model

In the model

$$Y_{g,i} \stackrel{ind}{\sim} N(\mu_g, \sigma_g^2)$$

we can calculate a $p$-value for the following null hypothesis:

$$H_0 : \sigma_g = \sigma \quad \text{for all} \quad g$$

```
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
```

This may give us reason to proceed as if the variances is the same in all groups, i.e.

$$Y_{g,i} \stackrel{ind}{\sim} N(\mu_g, \sigma^2).$$

This assumption is common when the number of observations in the groups is small.

# Comparing means when the variances are equal

Assuming $Y_{g,i} \stackrel{ind}{\sim} N(\mu_g, \sigma^2)$, we can test

$$H_0 : \mu_g = \mu \,\forall\, g$$

```
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, i.e. $H_0 : \mu_g = \mu_{g'}$.

```
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
```

# Posteriors for $\mu$

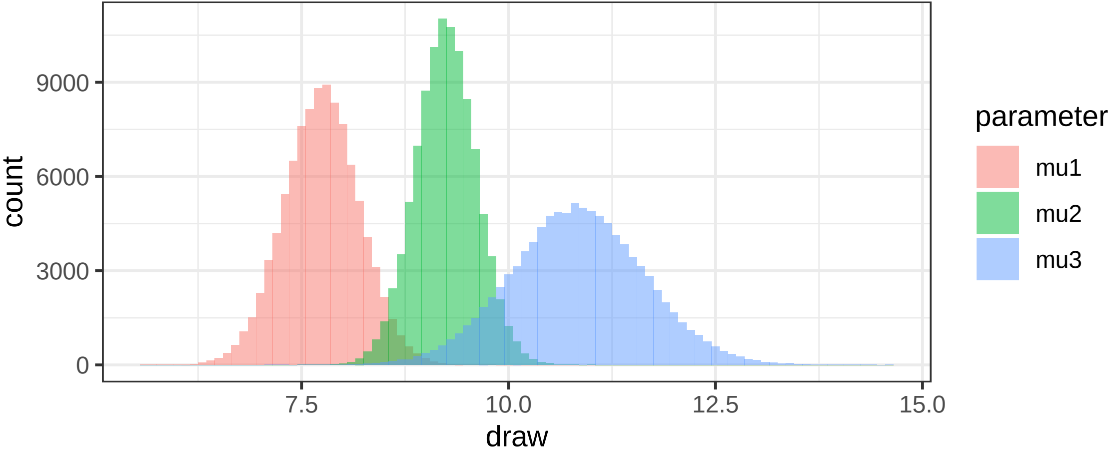
If $Y_{g,i} \stackrel{ind}{\sim} N(\mu_g, \sigma^2)$ and we use the prior $p(\mu_1, \ldots, \mu_G, \sigma^2) \propto 1/\sigma^2$, then

$$\mu_g|y, \sigma^2 \stackrel{ind}{\sim} N(\overline{y}_g, \sigma^2/n_g) \quad \sigma^2|y \sim IG\left(\frac{n-G}{2}, \frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{n_g}(y_{g,i} - \overline{y}_g)^2\right)$$

where $n = \sum_{g=1}^{G} n_g$. and thus, we obtain joint samples for $\mu$ by performing the following

1. $\sigma^{2(m)} \sim p(\sigma^2|y)$
2. For $g = 1, \ldots, G$, $\mu_g \sim p(\mu_g|y, \sigma^{2(m)})$.

# Compare 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         0.0059 (-2.65,-0.35)
2 mu1-mu3         0.0007 (-4.92,-1.26)
3 mu2-mu3         0.036  (-3.34,0.15)
```

# Summary

Multiple (independent) normal means

- $p$-values
- confidence intervals
- posterior densities
- credible intervals
- posterior probabilities