431 Class 16

Thomas E. Love

2018-10-25

Today's Agenda

- Lots of Discussion
- Comparing Population Means with Independent Samples

Today's R Setup

```
library(boot); library(Hmisc); library(broom)
library(tidyverse) # always load tidyverse last

source("Love-boost.R") # script from our Data page
```

```
dm192 <- read.csv("data/dm192.csv") %>% tbl_df
```

Independent Samples Study Designs

- Independent samples designs do not impose a matching, but instead sample two unrelated sets of subjects, where each group receives one of the two exposures.
- The two groups of subjects are drawn independently from their separate populations of interest.
- One obvious way to tell if we have an independent samples design is that this design does not require the sizes of the two exposure groups to be equal.

The best way to establish whether a study uses paired or independent samples is to look for the **link** between the two measurements that creates paired differences.

- Deciding whether or not the samples are paired (matched) is something we do before we analyze the data.

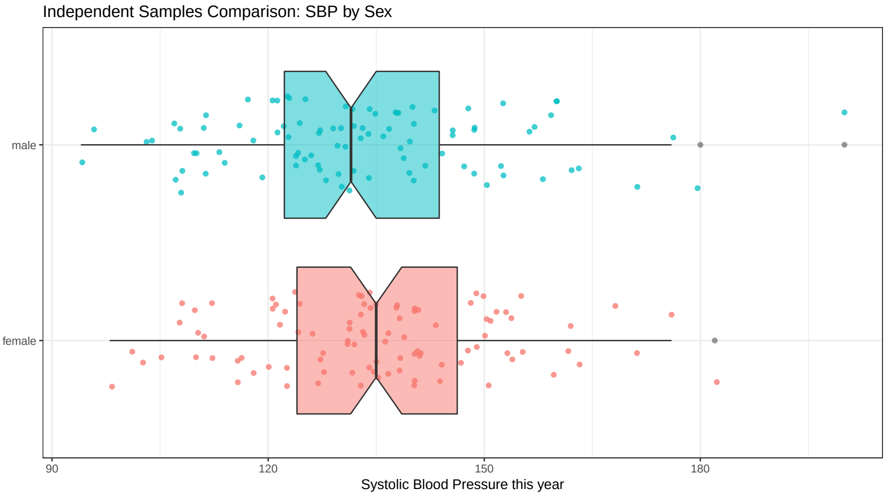
What if the Samples Aren't Paired?

In the dm192 frame, we might also consider looking at a different kind of comparison, perhaps whether the average systolic blood pressure is larger in male or in female adults in NE Ohio living with diabetes.

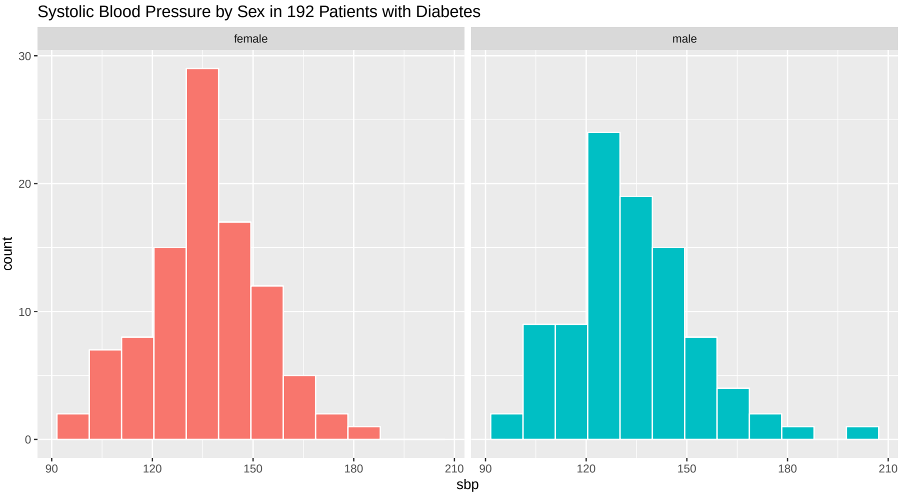
```
dm_second <- select(dm192, pt.id, sex, sbp)
summary(dm_second)
```

```
     pt.id            sex          sbp       
 Min.   :  1.00   female:98   Min.   : 94.0  
 1st Qu.: 48.75   male  :94   1st Qu.:123.0  
 Median : 96.50               Median :133.0  
 Mean   : 96.50               Mean   :134.2  
 3rd Qu.:144.25               3rd Qu.:144.5  
 Max.   :192.00               Max.   :200.0  
```

Our comparison now is between females and males



Another Way to Picture Two Independent Samples

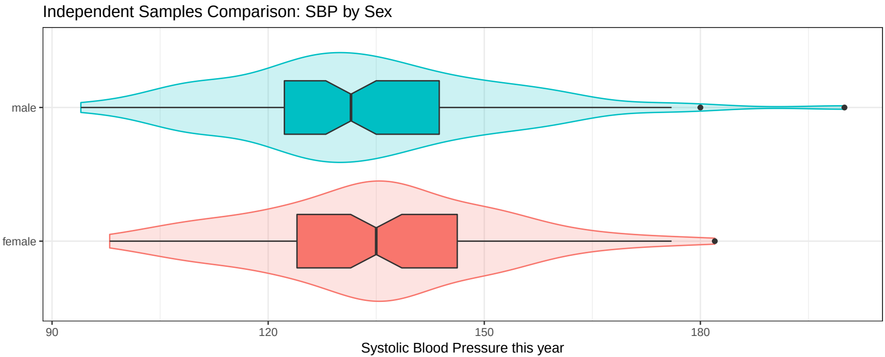


Numerical Summary for Two Independent Samples

```
mosaic::favstats(sbp ~ sex, data = dm_second)
```

```
     sex min     Q1 median     Q3 max     mean
1 female  98 124.00  135.0 146.25 182 135.1327
2   male  94 122.25  131.5 143.75 200 133.2447
        sd  n missing
1 16.75637 98       0
2 18.82785 94       0
```

Systolic BP, within groups defined by sex



Group	n	mean	median	sd
Males	94	133.2	131.5	18.8
Females	98	135.1	135.0	16.8

Hypotheses Under Consideration

The hypotheses we are testing are:

- H_0: mean in population 1 = mean in population 2 + hypothesized difference Δ_0 vs.
- H_A: mean in population 1 $\neq$ mean in population 2 + hypothesized difference Δ_0,

where Δ_0 is almost always zero. An equivalent way to write this is:

- $H_0 : \mu_1 = \mu_2 + \Delta_0$ vs.
- $H_A : \mu_1 \neq \mu_2 + \Delta_0$

Yet another equally valid way to write this is:

- $H_0 : \mu_1 - \mu_2 = \Delta_0$ vs.
- $H_A : \mu_1 - \mu_2 \neq \Delta_0$,

where, again Δ_0 is almost always zero.

Testing Options for Independent Samples

1. Pooled t test or Indicator Variable Regression Model (t test assuming equal population variances)
2. Welch t test (t test without assuming equal population variances)
3. Wilcoxon-Mann-Whitney Rank Sum Test (non-parametric test not assuming populations are Normal)
4. Bootstrap confidence interval for the difference in population means

Assumptions of the Pooled T test

The standard method for comparing population means based on two independent samples is based on the t distribution, and requires the following assumptions:

1. [Independence] The samples for the two groups are drawn independently.
2. [Random Samples] The samples for each of the groups are drawn at random from the populations of interest.
3. [Normal Population] The two populations are each Normally distributed
4. [Equal Variances] The population variances in the two groups being compared are the same, so we can obtain a pooled estimate of their joint variance.

The Pooled Variances t test in R

Also referred to as the t test assuming equal population variances:

```
t.test(sbp ~ sex, data = dm_second, var.equal = TRUE)
```

```
Two Sample t-test


data:  sbp by sex
t = 0.73467, df = 190, p-value = 0.4634
alternative hypothesis:
    true difference in means is not equal to 0
95 percent confidence interval: -3.181093  6.957037
sample estimates:
mean in group female   mean in group male
            135.1327             133.2447
```

Note: CI shows estimate for $\mu_{female} - \mu_{male}$ here.

Regressing sbp on sex yields pooled t test!

```r
m1 <- lm(sbp ~ sex, data = dm_second)
summary(m1)
```

```
Call: lm(formula = sbp ~ sex, data = dm_second)

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept)    135.133      1.798  75.152   <2e-16 ***
sexmale         -1.888      2.570  -0.735    0.463
---
Sig. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.8 on 190 degrees of freedom
Multiple R-squared:  0.0028,    Adjusted R-squared:  -0.0024
F-statistic: 0.5397 on 1 and 190 DF,  p-value: 0.4634
```

`broom::tidy` to summarize the regression, plus CI for difference in population means (edited output)

```r
m1 <- lm(sbp ~ sex, data = dm_second)
broom::tidy(m1, conf.int = TRUE, conf.level = 0.95)
```

term	estimate	std.error	statistic	p.value
(Intercept)	135.13	1.80	75.15	2.86e-143
sexmale	-1.89	2.57	-0.73	4.63e-01

term	conf.low	conf.high
(Intercept)	131.6	138.7
sexmale	-6.96	3.18

This indicator shows the effect of being Male, so the displayed CI estimates $\mu_{male} - \mu_{female}$. Invert the signs to get the $\mu_{female} - \mu_{male}$ estimate.

Results for the SBP and Sex Study

Procedure	p for $H_0 : \mu_F = \mu_M$	95% CI for $\mu_F - \mu_M$
Pooled t test	0.463	(-3.2, 7.0)

What conclusions should we draw, at $\alpha = 0.05$?

Assumptions of the Welch t test

The Welch test still requires:

1. [Independence] The samples for the two groups are drawn independently.
2. [Random Samples] The samples for each of the groups are drawn at random from the populations of interest.
3. [Normal Population] The two populations are each Normally distributed

But it doesn't require:

4. [Equal Variances] The population variances in the two groups being compared are the same.

Welch's t test is the default choice in R.

Welch t test not assuming equal population variances

```
t.test(sbp ~ sex, data = dm_second)
```

```

    Welch Two Sample t-test

data:  sbp by sex
t = 0.73288, df = 185.39, p-value = 0.4646
alternative hypothesis: true difference in means is not equal
95 percent confidence interval:
 -3.194236  6.970180
sample estimates:
mean in group female   mean in group male
            135.1327             133.2447
```

Results for the SBP and Sex Study

Procedure	p for $H_0 : \mu_F = \mu_M$	95% CI for $\mu_F - \mu_M$
Pooled t test	0.463	(-3.2, 7.0)
Welch t test	0.465	(-3.2, 7.0)

What conclusions should we draw, at $\alpha = 0.05$?

Assumptions of the Wilcoxon-Mann-Whitney Rank Sum Test

The Wilcoxon-Mann-Whitney Rank Sum test still requires:

1. [Independence] The samples for the two groups are drawn independently.
2. [Random Samples] The samples for each of the groups are drawn at random from the populations of interest.

But it doesn't require:

3. [Normal Population] The two populations are each Normally distributed
4. [Equal Variances] The population variances in the two groups being compared are the same.

It also doesn't really compare population means.

Wilcoxon-Mann-Whitney Rank Sum Test

```r
wilcox.test(sbp ~ sex, data = dm_second, conf.int = TRUE)
```

```
    Wilcoxon rank sum test with continuity
    correction

data:  sbp by sex
W = 5035.5, p-value = 0.2649
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 -2.000061  7.999993
sample estimates:
difference in location
              2.999918
```

Results for the SBP and Sex Study

Procedure	p for $H_0 : \mu_F = \mu_M$	95% CI for $\mu_F - \mu_M$
Pooled t test	0.463	(-3.2, 7.0)
Welch t test	0.465	(-3.2, 7.0)
Rank Sum test	0.265	(-2.0, 8.0)

What conclusions should we draw, at $\alpha = 0.05$?

The Bootstrap

This bootstrap approach to comparing population means using two independent samples still requires:

1. [Independence] The samples for the two groups are drawn independently.
2. [Random Samples] The samples for each of the groups are drawn at random from the populations of interest.

but does not require either of the other two assumptions:

3. [Normal Population] The two populations are each Normally distributed
4. [Equal Variances] The population variances in the two groups being compared are the same.

The bootstrap procedure I use in R was adapted from Frank Harrell and colleagues. http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/BootstrapMeansSoftware

The `bootdif` function

The procedure requires the definition of a function, which I have adapted a bit, called `bootdif`, which is part of the `Love-boost.R` script we loaded earlier.

As in our previous bootstrap procedures, we are sampling (with replacement) a series of many data sets (default: 2000).

- Here, we are building bootstrap samples based on the SBP levels in the two independent samples (M vs. F).
- For each bootstrap sample, we are calculating a mean difference between the two groups (M vs. F).
- We then determine the 2.5th and 97.5th percentile of the resulting distribution of mean differences (for a 95% confidence interval).

Using the `bootdif` function to compare means based on independent samples

So, to compare systolic BP (our outcome) across the two levels of sex (our grouping factor) for the adult patients with diabetes in NE Ohio, run…

```r
set.seed(4314); bootdif(dm_second$sbp, dm_second$sex)
```

```
Mean Difference           0.025           0.975
      -1.887972       -6.977860        2.917249
```

- The two columns must be separated here with a comma rather than a tilde (~), and are specified using `data$variable` notation.
- This CI estimates $\mu_{male} - \mu_{female}$: observe the listed sample mean difference for the necessary context. Invert the signs, as before, to estimate $\mu_{female} - \mu_{male}$.

Results for the SBP and Sex Study

Procedure	p for $H_0 : \mu_F = \mu_M$	95% CI for $\mu_F - \mu_M$
Pooled t test	0.463	(-3.2, 7.0)
Welch t test	0.465	(-3.2, 7.0)
Rank Sum test	0.265	(-2.0, 8.0)
Bootstrap CI	$p > 0.05$	(-2.9, 7.0)

What conclusions should we draw, at $\alpha = 0.05$?

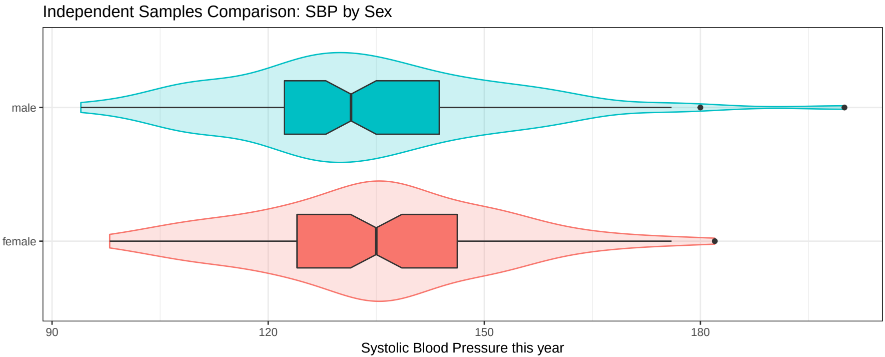
Which Method Should We Use?

1. Plot the distributions of the two independent samples.
2. Does it seem reasonable to assume that **each** distribution (here, both sbp in males and sbp in females) follows an approximately Normal distribution?

- If Yes, Normal models seem appropriate, then
  - use the pooled t test (or indicator variable regression) if the sample sizes are nearly the same, or if the sample variances are quite similar
  - use the Welch's t test, otherwise (this is the default R choice)
- If No, Normal models don't seem appropriate, then
  - compare means using the bootstrap via `bootdif`, or
  - compare pseudo-medians using the rank sum test

What did we see in our systolic BP data?

Systolic BP, within groups defined by sex



Group	n	mean	median	sd
Males	94	133.2	131.5	18.8
Females	98	135.1	135.0	16.8

Formatting the Data (Wide vs. Long)

Wide format (most appropriate for paired/matched samples)

subject	treatment1	treatment2
A	140	150
B	135	145
C	128	119

Long format (most appropriate for independent samples)

subject	sbp	group
A	140	treatment1
A	150	treatment2
B	135	treatment1
B	145	treatment2
C	128	treatment1
C	119	treatment2

Suppose you have a wide data set...

```
tempdat_wide <- data_frame(
  subject = c("A", "B", "C"),
  treatment_1 = c(140, 135, 128),
  treatment_2 = c(150, 145, 119)
)

tempdat_wide
```

```
# A tibble: 3 x 3
  subject treatment_1 treatment_2
  <chr>         <dbl>       <dbl>
1 A               140         150
2 B               135         145
3 C               128         119
```

Gather the Data from Wide to Long

```
tempdat_long <- tempdat_wide %>%
  gather(treatment_1, treatment_2,
         key = "group", value = "sbp")
tempdat_long
```

```
# A tibble: 6 x 3
  subject group         sbp
  <chr>   <chr>       <dbl>
1 A       treatment_1   140
2 B       treatment_1   135
3 C       treatment_1   128
4 A       treatment_2   150
5 B       treatment_2   145
6 C       treatment_2   119
```

Spread the Data from Long to Wide

```
tempdat_wide2 <- tempdat_long %>% spread(group, sbp)

tempdat_wide2
```

```
# A tibble: 3 x 3
  subject treatment_1 treatment_2
  <chr>         <dbl>       <dbl>
1 A               140         150
2 B               135         145
3 C               128         119
```

A Few Reminders About Significance

- **A significant effect is not necessarily the same thing as an interesting effect.** For example, results calculated from large samples are nearly always "significant" even when the effects are quite small in magnitude. Before doing a test, always ask if the effect is large enough to be of any practical interest. If not, why do the test?
- **A non-significant effect is not necessarily the same thing as no difference.** A large effect of real practical interest may still produce a non-significant result simply because the sample is too small.
- **There are assumptions behind all statistical inferences.** Checking assumptions is crucial to validating the inference made by any test or confidence interval.

Next Time

- Power and Sample Size Considerations