

Stat 427/527: Advanced Data Analysis 1

Chapter 3: Two-Samples Inferences

Fall, 2018



Topics

Suppose you have collected data on one variable from two (independent) samples and you are interested in “comparing” the samples. What tools are good to use?

- ▶ Graphical displays that meaningfully compare independent populations.
- ▶ Assess the assumptions of the two-sample t-test.
- ▶ Decide whether the means between two populations are different.
- ▶ Recommend action based on analysis.

Example: Head Breadths

Want to compare the maximum head breadths (in millimeters) of modern day Englishmen with that of their ancient countrymen (Celts).

- ▶ The Celts were a vigorous race of people who once populated parts of England
—It is not entirely clear whether they simply died out or merged with other people who were the ancestors of those who live in England today.
- ▶ A goal of this study might be to shed some light on possible genetic links between the Celts and modern day Englishmen.
- ▶ The study is based on the maximum head breadths made on 16 unearthed Celtic skulls and on 18 of skulls of modern-day Englishmen.

Data

```
#### Example: Head Breadths
# unstacked data as two vectors
english<-c(141,148,132,138,154,142,150,146,155,158,150,140,
           147,148,144,150,149,145)
celts<-c(133, 138, 130, 138, 134, 127, 128, 138, 136,
         131,126, 120, 124, 132, 132, 125)
```

- ▶ Stack data, where one column contains both samples, with a second column of labels or **subscripts** to distinguish the samples.
- ▶ Interested in comparing the centers of the distributions, the spreads within each distribution, the distributional shapes, etc.

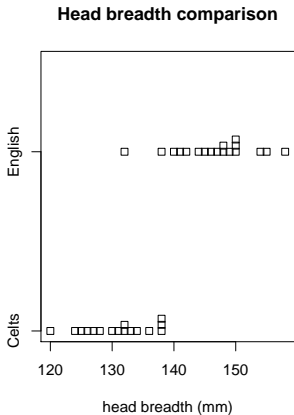
```
# stacked data as a vector of values and a vector of labels
HeadBreadth <- c(english, celts)
Group <- c(rep("English", length(english)), rep("Celts", length(celts)))
hb <- data.frame(HeadBreadth, Group)
hb
```

```
##      HeadBreadth  Group
## 1           141 English
## 2           148 English
## 3           132 English
## 4           138 English
## 5           154 English
## 6           142 English
## 7           150 English
## 8           146 English
## 9           155 English
## 10          158 English
## 11          150 English
## 12          140 English
## 13          147 English
## 14          148 English
## 15          144 English
```

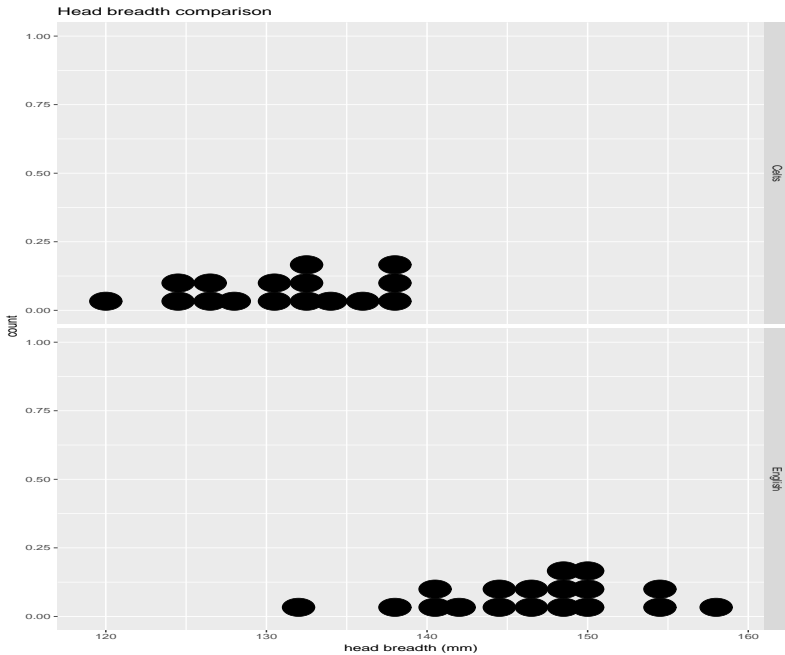
## 16	150	English
## 17	149	English
## 18	145	English
## 19	133	Celts
## 20	138	Celts
## 21	130	Celts
## 22	138	Celts
## 23	134	Celts
## 24	127	Celts
## 25	128	Celts
## 26	138	Celts
## 27	136	Celts
## 28	131	Celts
## 29	126	Celts
## 30	120	Celts
## 31	124	Celts
## 32	132	Celts
## 33	132	Celts
## 34	125	Celts

Plotting head breadth data:

```
#### Plotting head breadth data
# stripchart (dotplot) using R base graphics
stripchart(HeadBreadth~Group, method="stack", data=hb,
  main="Head breadth comparison", xlab="head breadth (mm)")
```



```
# stripchart (dotplot) using ggplot
library(ggplot2)
p <- ggplot(hb, aes(x = HeadBreadth))
p <- p + geom_dotplot(binwidth = 2)
p <- p + facet_grid(Group ~ .)
# rows are Group categories
p <- p + labs(title = "Head breadth comparison")
  + xlab("head breadth (mm)")
print(p)
```

Boxplots for comparison are most helpful when plotted in the same axes.

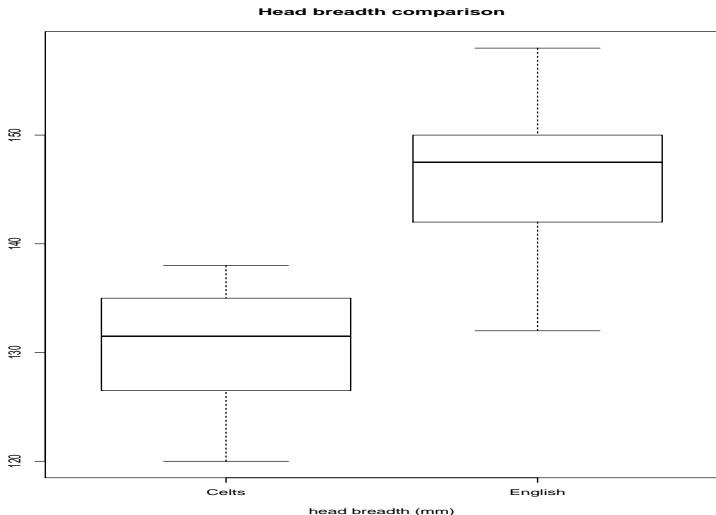


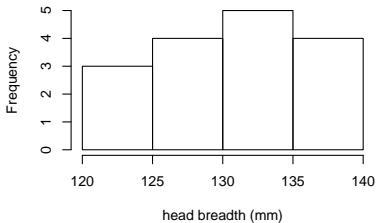
Figure 2: Boxplot of HeadBreadth Data

Histograms are hard to compare. But you can take a look of the shape of data.

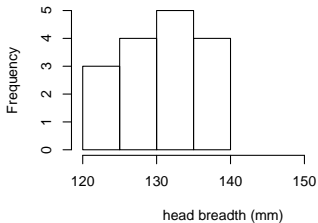
```
# histogram using R base graphics
par(mfcol=c(2,2))
hist(hb$HeadBreadth[(hb$Group == "Celts")],
     main = "Head breadth, Celts", xlab = "head breadth (mm)")
hist(hb$HeadBreadth[(hb$Group == "English")],
     main = "Head breadth, English", xlab = "head breadth (mm)")

# common x-axis limits based on the range of the entire data set
hist(hb$HeadBreadth[(hb$Group == "Celts")], xlim = range(hb$HeadBreadth))
hist(hb$HeadBreadth[(hb$Group == "English")], xlim = range(hb$HeadBreadth))
main = "Head breadth, English", xlab = "head breadth (mm)"
```

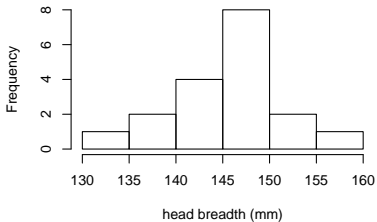
Head breadth, Celts



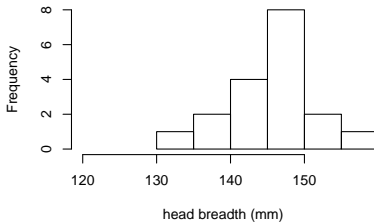
Head breadth, Celts



Head breadth, English



Head breadth, English



Summary Statistics

```
#### summaries by group  
# summary for separate vectors  
summary(english)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      132.0   142.5   147.5   146.5   150.0   158.0
```

```
summary(celts)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      120.0   126.8   131.5   130.8   134.5   138.0
```

```
# comparing spreads, an assumption of equal variances seems reasonable  
sd(english)  
  
## [1] 6.382421  
  
sd(celts)  
  
## [1] 5.434458  
  
IQR(english)  
  
## [1] 7.5  
  
IQR(celts)  
  
## [1] 7.75
```

```
# numerical summary of each column in data.frame hb by Group
by(hb, Group, summary)
```

```
## Group: Celts
##   HeadBreadth      Group
##   Min.   :120.0    Celts   :16
##   1st Qu.:126.8    English: 0
##   Median :131.5
##   Mean   :130.8
##   3rd Qu.:134.5
##   Max.   :138.0
## -----
## Group: English
##   HeadBreadth      Group
##   Min.   :132.0    Celts   : 0
##   1st Qu.:142.5    English:18
##   Median :147.5
##   Mean   :146.5
##   3rd Qu.:150.0
##   Max.   :158.0
```

Graphical observations:

- ▶ The dotplots, boxplots, and histograms indicate that the English and Celt samples are slightly skewed to the left.
- ▶ There are no outliers in either sample.
- ▶ It is not unreasonable to operationally assume that the population frequency curves (i.e., the histograms for the populations from which the samples were selected) for the English and Celtic head breadths are normal.
- ▶ The sample means and medians are close to each other in each sample, which is not surprising given the near symmetry and the lack of outliers.
- ▶ The data suggest that the typical modern English head breadth is greater than that for Celts. The data sets have relatively small variation/spreads, as measured by either the standard deviation or the IQR.

Paired Versus Independent Samples

Suppose you have two populations of interest, say populations 1 and 2

- ▶ Interested in comparing their (unknown) population means, μ_1 and μ_2 .
- ▶ Inferences on the unknown population means are based on samples from each population. In practice, most problems fall into one of two categories.

Independent samples where the sample taken from population 1 has no effect on which observations are selected from population 2, and vice versa.

Paired or dependent samples where experimental units are paired based on factors related or unrelated to the variable measured. Note that with paired data, the sample sizes are equal to the number of pairs.

Examples:

- ▶ The English and Celt head breadth samples are **independent**.
- ▶ Suppose you are interested in whether the CaCO_3 (calcium carbonate) level in the Atrisco well field, which is the water source for Albuquerque, is changing over time.
 - the CaCO_3 level was recorded at each of 15 wells at two time points.
 - These data are **paired**. The two samples are the observations at Times 1 and 2.
- ▶ To compare state incomes, a random sample of New Mexico households was selected, and an independent sample of Arizona households was obtained. It is reasonable to assume **independent** samples.

- ▶ Suppose you are interested in whether the husband or wife is typically the heavier smoker among couples where both adults smoke.
 - Data are collected on households.
 - Measure the average number of cigarettes smoked by each husband and wife within the sample of households.
 - These data are **paired**, i.e., you have selected husband wife pairs as the basis for the samples. It is reasonable to believe that the responses within a pair are related, or correlated.

CI of Two Independent Samples (equal variance)

Assume that the populations

- ▶ have normal frequency curves
- ▶ with **equal** population standard deviations, i.e., $\sigma_1 = \sigma_2$.
- ▶ Let (n_1, \bar{y}_1, s_1) and (n_2, \bar{y}_2, s_2) be the sample sizes, means and standard deviations from the two samples.

The standard CI for $\mu_1 - \mu_2$ is given by

$$\text{CI} = (\bar{y}_1 - \bar{y}_2) \pm t_{\text{crit}} SE_{\bar{Y}_1 - \bar{Y}_2}$$

where

$$SE_{\bar{Y}_1 - \bar{Y}_2} = s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

the **pooled variance estimator**,

$$s_{\text{pooled}}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

and the critical value t_{crit} for CI and tests is obtained in usual way from a t -table with $df = n_1 + n_2 - 2$.

$$s_{\text{pooled}}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

—The pooled estimator of variance is our best estimate of the common population variance.

— It is a weighted average of the two sample variances, with more weight given to the larger sample.

—If $n_1 = n_2$ then s_{pooled}^2 is the average of s_1^2 and s_2^2 .

Test of Two Independent Samples (equal variance)

$H_0 : \mu_1 - \mu_2 = 0$ ($\mu_1 = \mu_2$) against $H_A : \mu_1 - \mu_2 \neq 0$ ($\mu_1 \neq \mu_2$)

$$t_s = \frac{\bar{y}_1 - \bar{y}_2}{SE_{\bar{Y}_1 - \bar{Y}_2}}$$

where

$$SE_{\bar{Y}_1 - \bar{Y}_2} = s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

$$s_{\text{pooled}}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

and The critical value t_{crit} is obtained in usual way from a t -table with $df = n_1 + n_2 - 2$. Reject H_0 if $|t_s| > t_{\text{crit}}$.

Comments:

- ▶ The pooled CI and tests are sensitive to the normality and equal standard deviation assumptions.
- ▶ The observed data can be used to assess the reasonableness of these assumptions
 - look at boxplots and histograms to assess normality
 - check whether s_1 close to s_2 to have the first assess of the assumption $\sigma_1 = \sigma_2$.
 - Formal tests of these assumptions will be discussed later.

Satterthwaite's Method, unequal variances

- ▶ assumes normality
- ▶ does not require equal population standard deviations
- ▶ Satterthwaite's procedures are somewhat conservative. It adjusts the SE and df to account for unequal population variances.

$$CI = (\bar{y}_1 - \bar{y}_2) \pm t_{\text{crit}} SE_{\bar{Y}_1 - \bar{Y}_2}$$

$$H_0 : \mu_1 - \mu_2 = 0 \quad (\mu_1 = \mu_2) \text{ against } H_A : \mu_1 - \mu_2 \neq 0 \quad (\mu_1 \neq \mu_2)$$

$$t_s = \frac{\bar{y}_1 - \bar{y}_2}{SE_{\bar{Y}_1 - \bar{Y}_2}}$$

where

$$SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}},$$

and degrees of freedom:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}.$$

- ▶ Note that $df = n_1 + n_2 - 2$ when $n_1 = n_2$ and $s_1 = s_2$.
- ▶ The Satterthwaite and pooled variance procedures usually give similar results when $s_1 \doteq s_2$.
- ▶ The df formula for Satterthwaite's method is fairly complex, so when done by hand some use a conservative df formula that uses the minimum of $n_1 - 1$ and $n_2 - 1$ instead.

Head Breadths Example continued

- ▶ The English and Celts are independent samples.
- ▶ We looked at boxplots and histograms, which suggested that the normality assumption for the t -test is reasonable.
- ▶ The R output shows the English and Celt sample standard deviations and IQRs are fairly close, so the pooled and Satterthwaite results should be comparable.
- ▶ The pooled analysis is preferable here, but either is appropriate.

Note that R does the pooled and Satterthwaite (Welch) analyses. The output will contain a p -value for a two-sided test of equal population means and a CI for the difference in population means. If you include `var.equal = TRUE` you will get the pooled method, otherwise the output is for Satterthwaite's method.

We are interested in difference in mean head breadths between Celts and English.

► **1. Define the population parameters and hypotheses in words and notation**

Let μ_1 and μ_2 be the mean head breadth for the Celts and English, respectively.

In words: “The difference in population means between Celts and English is different from zero mm.”

In notation: $H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$.

Alternatively: $H_0 : \mu_1 - \mu_2 = 0$ versus $H_A : \mu_1 - \mu_2 \neq 0$.

► 2. Calculate summary statistics from sample

Mean, standard deviation, sample size:

```
#### Calculate summary statistics
m1 <- mean(celts)
s1 <- sd(celts)
n1 <- length(celts)
m2 <- mean(english)
s2 <- sd(english)
n2 <- length(english)
c(m1, s1, n1)
## [1] 130.750000 5.434458 16.000000
c(m2, s2, n2)
## [1] 146.500000 6.382421 18.000000
```

The pooled-standard deviation, standard error, and degrees-of-freedom are:

```
sdpool
<- sqrt(((n1 - 1) * s1^2 + (n2 - 1) * s2^2) /
  (n1 + n2 - 2))
sdpool
## [1] 5.956876
SEpool <- sdpool * sqrt(1 / n1 + 1 / n2)
SEpool
## [1] 2.046736
dfpool <- n1 + n2 - 2
dfpool
## [1] 32
t_pool <- (m1 - m2) / SEpool
t_pool
## [1] -7.69518
```

The Satterthwaite SE and degrees-of-freedom are:

```
SE_Sat <- sqrt(s1^2 / n1 + s2^2 / n2)
```

```
SE_Sat
```

```
## [1] 2.027043
```

```
df_Sat <- (SE_Sat^2)^2 / (s1^4 / (n1^2 * (n1 - 1)) +  
  s2^4 / (n2^2 * (n2 - 1)))
```

```
df_Sat
```

```
## [1] 31.9511
```

```
t_Sat <- (m1 - m2) / SE_Sat
```

```
t_Sat
```

```
## [1] -7.769937
```

► **3. Specify confidence level, calculate t-stat, CI limits, p-value**

Let us calculate a 95% CI for $\mu_1 - \mu_2$. Assuming equal variances, using pooled-variance procedure:

```
# var.equal = FALSE is the default
# two-sample t-test specifying two separate vectors
t.summary.eqvar <- t.test(celts, english,
var.equal = TRUE)
t.summary.eqvar
## data: celts and english
## t = -7.6952, df = 32, p-value = 9.003e-09
## alternative hypothesis: true difference in means
  is not equal to 0
## 95 percent confidence interval:
## -19.91906 -11.58094
## sample estimates:
## mean of x mean of y
## 130.75 146.50
```

Not assuming equal variances, Satterthwaite (Welch):

```
# two-sample t-test with unequal variances
(Welch = Satterthwaite)
# specified using data.frame and a formula,
HeadBreadth by Group
t.summary.uneqvar <- t.test(HeadBreadth ~ Group,
data = hb, var.equal = FALSE)
t.summary.uneqvar
## Welch Two Sample t-test
## data: HeadBreadth by Group
## t = -7.7699, df = 31.951, p-value = 7.414e-09
## alternative hypothesis: true difference in means
  is not equal to 0
## 95 percent confidence interval:
## -19.8792 -11.6208
## sample estimates:
## mean in group Celts mean in group English
## 130.75 146.50
```


Both will give the same results

```
t.summary.uneqvar <- t.test(HeadBreadth ~ Group,  
  data = hb, var.equal = FALSE)  
t.summary.uneqvar2 <- t.test(celts, english,  
  var.equal = FALSE)  
> t.summary.uneqvar2
```

- ▶ **4. Summarize in words** (Using the pooled-variance results.)
1. The pooled analysis strongly reject $H_0 : \mu_1 - \mu_2 = 0$
 - the t -statistic is -7.7
 - two-sided p -value is 9×10^{-9}
 - Because the p -value < 0.05 , consider other assumptions are all met, we reject the Null hypothesis in favor of the Alternative hypothesis, concluding that the difference in population mean head breadths between the Celts and English are different.
 2. We are 95% confident that the difference in population means, $\mu_1 - \mu_2$, is between -19.9 and -11.6 mm. That is, we are 95% confident that the population mean head breadth for Englishmen (μ_2) exceeds the population mean head breadth for Celts (μ_1) by between 11.6 and 19.9 mm.

3. The CI interpretation is made easier by recognizing that we concluded the population means are different, so the direction of difference must be consistent with that seen in the observed data, where the sample mean head breadth for Englishmen exceeds that for the Celts. Thus, the limits on the CI for $\mu_1 - \mu_2$ tells us how much smaller the mean is for the Celts (that is, between 11.6 and 19.9 mm).

► **5. Check assumptions**

We can test the assumption that the distribution of $\bar{Y}_1 - \bar{Y}_2$ is normal using the bootstrap in the following function.

The assumption of equal population variances will be left to a later chapter.

Run the code in the class to make sure you can do it.

```
#### Visual comparison of whether sampling distribution
is close to Normal via Bootstrap
# a function to compare the bootstrap
sampling distribution
# of the difference of means from two samples with
# a normal distribution with mean and
SEM estimated from the data
bs.two.samp.diff.dist <- function(dat1, dat2, N = 1e4) {
  n1 <- length(dat1);
  n2 <- length(dat2);
  # resample from data
  sam1 <- matrix(sample(dat1, size = N * n1,
    replace = TRUE), ncol=N);
  sam2 <- matrix(sample(dat2, size = N * n2,
    replace = TRUE), ncol=N);
  # calculate the means and take difference between
  populations
  sam1.mean <- colMeans(sam1);
```

```
sam2.mean <- colMeans(sam2);
diff.mean <- sam1.mean - sam2.mean;
# save par() settings
old.par <- par(no.readonly = TRUE)
# make smaller margins
par(mfrow=c(3,1), mar=c(3,2,2,1), oma=c(1,1,1,1))
# Histogram overlaid with kernel density curve
hist(dat1, freq = FALSE, breaks = 6
      , main = paste("Sample 1", "\n"
                    , "n =", n1
                    , ", mean =", signif(mean(dat1), digits = 5)
                    , ", sd =", signif(sd(dat1), digits = 5))
      , xlim = range(c(dat1, dat2)))
points(density(dat1), type = "l")
rug(dat1)

hist(dat2, freq = FALSE, breaks = 6
      , main = paste("Sample 2", "\n"
```

```

      , "n =", n2
, ", mean =", signif(mean(dat2), digits = 5)
, ", sd =", signif(sd(dat2), digits = 5))
  , xlim = range(c(dat1, dat2)))
points(density(dat2), type = "l")
rug(dat2)

hist(diff.mean, freq = FALSE, breaks = 25
      , main = paste("Bootstrap sampling distribution
of the difference in means", "\n"
, "mean =", signif(mean(diff.mean), digits = 5)
, ", se =", signif(sd(diff.mean), digits = 5)))
# overlay a density curve for the sample means
points(density(diff.mean), type = "l")
# overlay a normal distribution, bold and red
x <- seq(min(diff.mean), max(diff.mean), length = 1000)
points(x, dnorm(x, mean = mean(diff.mean),
sd = sd(diff.mean))

```

```
, type = "l", lwd = 2, col = "red")  
# place a rug of points under the plot  
rug(diff.mean)  
# restore par() settings  
par(old.par)  
}
```

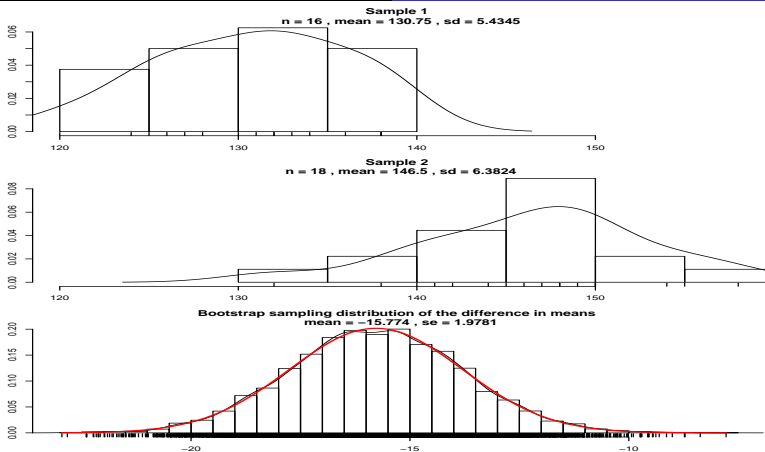



Figure 3: Bootstrap sampling distribution of the differences in means

The distribution of difference in means in the third plot looks very close to normal.

One-Sided Tests

$$H_0 : \mu_1 - \mu_2 = 0 \text{ v.s. } H_A : \mu_1 - \mu_2 < 0$$

or $H_A : \mu_1 - \mu_2 > 0$ (i.e., $\mu_1 > \mu_2$).

- ▶ Use the t -statistic, with the appropriate tail of the t -distribution to define critical values and p-values.
- ▶ One-sided two-sample tests are directly implemented in R, by specifying the type of test with *alternative* = "less" or *alternative* = "greater". One-sided confidence bounds are given with the one-sided tests.

Paired Analysis

Paired data

Table 1: Example of paired data

Unit	Time 1	Time 2
1	Y_{11}	Y_{12}
2	Y_{21}	Y_{22}
	...	
n	Y_{n1}	Y_{n2}

Within each pair, compute the difference $d = Y_1 - Y_2$:

$$d_1 = Y_{11} - Y_{12}$$

$$d_2 = Y_{21} - Y_{22}$$

$$\vdots$$

$$d_n = Y_{n1} - Y_{n2}$$

- ▶ If the Y_1 data are from a population with mean μ_1 and the Y_2 data are from a population with mean μ_2 , then the d s are a sample from a population with mean $\mu_d = \mu_1 - \mu_2$.
- ▶ If the sample of differences comes from a normal population, then we can use standard one-sample techniques on d_1, \dots, d_n to test $\mu_d = 0$ (that is, $\mu_1 = \mu_2$), and to get a CI for $\mu_d = \mu_1 - \mu_2$.

- ▶ Let $\bar{d} = n^{-1} \sum_i d_i = \bar{Y}_1 - \bar{Y}_2$ be the sample mean of the differences (which is also the mean difference)
- ▶ Let s_d be the sample standard deviation of the differences.
- ▶ The standard error of \bar{d} is $SE_{\bar{d}} = s_d/\sqrt{n}$, where n is the number of pairs. $df = n - 1$.

The paired t -test (two-sided) CI for μ_d is given by $\bar{d} \pm t_{\text{crit}} SE_{\bar{d}}$.

Test $H_0 : \mu_d = 0$ ($\mu_1 = \mu_2$) **against** $H_A : \mu_d \neq 0$ ($\mu_1 \neq \mu_2$)



$$t_s = \frac{\bar{d} - 0}{SE_{\bar{d}}}$$

- ▶ compute a p-value as in a two-sided one-sample test. One-sided tests are evaluated in the usual way for one-sample tests on means.
- ▶ A graphical analysis of paired data focuses on the **sample of differences**, and not on the original samples. In particular, the normality assumption is assessed on the sample of differences.

Example: Paired Analysis of Data on Twins

Burt (1966) presented data on IQ scores for identical twins that were raised apart, one by foster parents and one by the genetic parents.

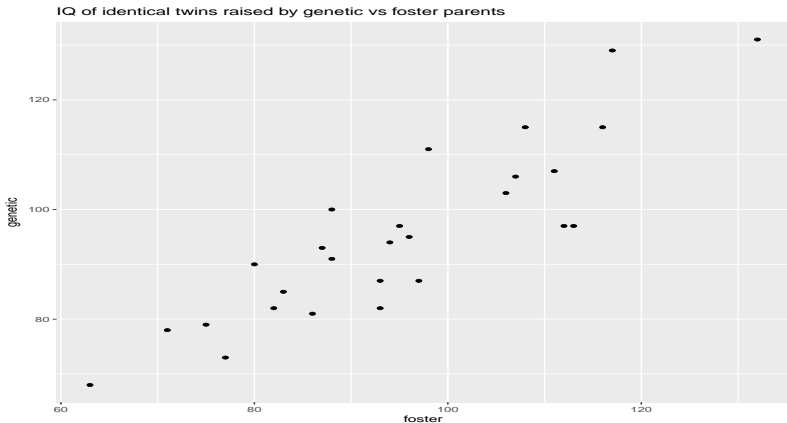
- ▶ Assuming the data are a random sample of twin pairs, consider comparing the population mean IQs for twins raised at home to those raised by foster parents.
- ▶ Let μ_f = population mean IQ for twin raised by foster parents
- ▶ μ_g = population mean IQ for twin raised by genetic parents.

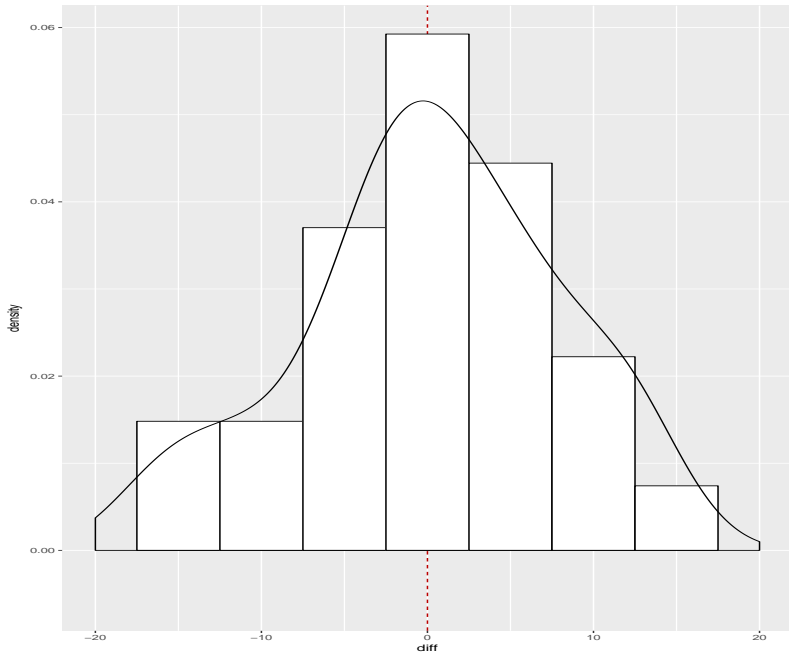
```
# Data and numerical summaries
foster <- c(82, 80, 88, 108, 116, 117, 132, 71, 75,
           93, 95, 88, 111, 63, 77, 86, 83, 93,
           97, 87, 94, 96, 112, 113, 106, 107, 98)
genetic <- c(82, 90, 91, 115, 115, 129, 131, 78, 79,
             82, 97, 100, 107, 68, 73, 81, 85, 87, 87,
             93, 94, 95, 97, 97, 103, 106, 111)
diff <- genetic - foster
iq <- data.frame(foster, genetic, diff)
> iq
```

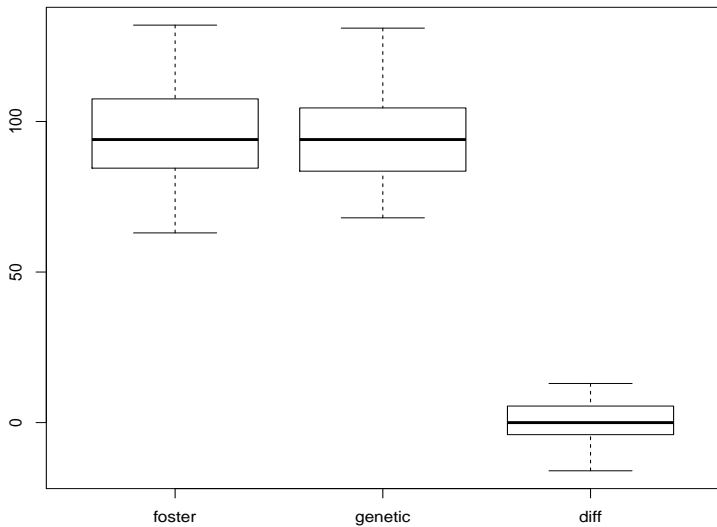
	foster	genetic	diff
1	82	82	0
2	80	90	10
3	88	91	3
4	108	115	7
5	116	115	-1
6	117	129	12
7	132	131	-1

```
# scatterplot of foster and genetic IQs, with 1:1 line
p <- ggplot(iq, aes(x = foster, y = genetic))
# draw a 1:1 line, dots above line
indicate "genetic > foster"
p <- p + geom_abline(intercept=0, slope=1, alpha=0.2)
p <- p + geom_point()
p <- p + geom_rug()
# make the axes square so it's a fair visual comparison
p <- p + coord_equal()
p <- p + scale_x_continuous(limits=axis.lim)
p <- p + scale_y_continuous(limits=axis.lim)
p <- p + labs(title = "IQ of identical twins raised
by genetic vs foster parents")
print(p)
```


This plot of IQ scores shows that scores are related within pairs of twins.







- ▶ Given the sample of differences, neither boxplot nor histogram showed marked deviation from normality.
- ▶ The boxplot is centered at zero, so one would not be too surprised if the test result is insignificant.

One-sample CI and test

```
# one-sample t-test of differences (paired t-test)
t.summary <- t.test(iq$diff)
t.summary
## One Sample t-test
## data: iq$diff
## t = 0.12438, df = 26, p-value = 0.902
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -2.875159 3.245529
## sample estimates:
## mean of x
## 0.1851852
```

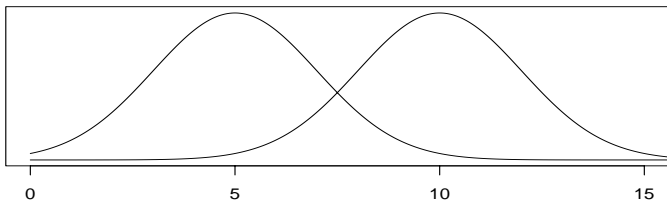
The hypothesis is $\mu_d = \mu_g - \mu_f = 0$. The p-value for this test is p-value = 0.902. We do not have sufficient evidence to claim that the population mean IQs for twins raised apart are different. This is consistent with the CI for μ_d given by $[-2.875159, 3.245529]$, which covers zero.

Alternatively, we can generate the test and CI directly from the raw data in two columns, specifying *paired = TRUE*. This gives the following output, which leads to identical conclusions to the earlier analysis.

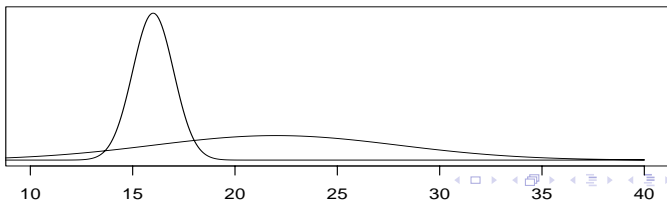
```
# two-sample paired t-test
t.summary <- t.test(iq$genetic, iq$foster, paired=TRUE)
t.summary
# two-sample paired t-test
t.summary <- t.test(iq$genetic, iq$foster, paired=TRUE)
t.summary
## Paired t-test
##
## data: iq$genetic and iq$foster
## t = 0.12438, df = 26, p-value = 0.902
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
## -2.875159 3.245529
## sample estimates:
## mean of the differences
## 0.1851852
```


Shall we compare means?

Normal Distributions with Identical Variances



Normal Distributions with Different Variances



When the two (normal) populations have equal/similar spreads/variance

- ▶ In such cases the difference between the two population means is equal to the difference between any fixed percentile for the two distributions, so the mean difference is a natural measure of difference.
- ▶ The mean is the most common feature on which two distributions are compared.

When the two (normal) populations have unequal spreads/variance

- ▶ Consider the hypothetical scenario depicted in the bottom panel above, where the population mean lifetimes using two distinct drugs for a fatal disease are $\mu_1 = 16$ months from time of diagnosis and $\mu_2 = 22$ months from time of diagnosis, respectively.
- ▶ The standard deviations under the two drugs are $\sigma_1 = 1$ and $\sigma_2 = 6$, respectively.
- ▶ The second drug has the higher mean lifetime, but at the expense of greater risk.
—For example, the first drug gives you a 97.7% chance of living at least 14 months, whereas the second drug only gives you a 90.8% chance of living at least 14 months. Which drug is best?
- ▶ It depends on what is important to you, a higher expected lifetime or a lower risk of dying early.