STAT 210
Applied Statistics and Data Analysis
Week 5 - Summary

Joaquin Ortega

King Abdullah University of Science and Technology

Reminders

The first exam will be on Saturday, October 22, 9 am - 12 noon.

The deadlline for the project email is September 30.

Video 16: Hypothesis Tests

A **hypothesis** is an assumption about the value or values of a parameter or parameters of the population.

Hypotheses are formulated in mutually exclusive pairs. If possible, hypotheses are chosen so that they are exhaustive (i.e., their union covers all possible outcomes of an experiment). This forces us to make a choice.

If θ is the parameter in question and Θ is the parameter space, then the **null hypothesis** defines the region $\{\theta \in \Theta_0\}$ while the **alternative hypothesis** describes $\{\theta \in \Theta_1\}$.

These regions are mutually exclusive: $\Theta_0 \cap \Theta_1 = \emptyset$

When a hypothesis completely specifies the distribution of the population, we say that the hypothesis is **simple**.

Any hypothesis that is not simple is **composite**.

We will only consider simple null hypotheses, while the alternatives will usually be composite.

Null hypothesis	Alternative hypothesis	Type of alternative
$H_0 : \theta = \theta_0$	$H_1 : \theta < \theta_0$	lower one-sided
	$H_1 : \theta > \theta_0$	upper one-sided
	$H_1 : \theta \neq \theta_0$	two-sided

In the Neyman-Pearson approach, one must choose one of the two alternatives using information from a sample, from which a test statistic is computed.

The sample space is split into two regions,

- $\mathcal{R}$, known as the **rejection** region, and
- $\mathcal{R}^c$, known as the **acceptance** region.

The value of θ that separates these two regions is known as the **critical** value.

The test statistic is computed from the sample, and if it falls in the acceptance region, the null hypothesis is accepted; otherwise, the alternative hypothesis is accepted.

R. Fisher proposed a different approach.

In this approach, one determines how much evidence there is in the sample against the null hypothesis.

The null hypothesis is not accepted, it is just 'not rejected'.

The test will determine if the sample collected can be due to chance alone under the null hypothesis. If this is not likely, the researcher has evidence to reject the null hypothesis.

A test with these characteristics is called a **significance test**.

Level and Power of a Test

Types of error

Our decision is always subject to error. There are two types of error:

Type I: Reject H_0 when it is true

Type II: Fail to reject H_0 when it is false.

State of Nature	H_0 not rejected	H_0 rejected
H_0 is true	No error	**Type I error**
H_0 is false	**Type II error**	No error

Types of error

Each error has a probability associated with it.

We denote by α the probability of a Type I error and by β the probability of a Type II error.

α is known as the **level** or **significance level** of the test.

For a fixed sample size, α and β go in opposite ways: the smaller α, the larger β, so we cannot reduce both at the same time unless we change the sample size.

The p-value for a hypothesis test is defined as the probability of observing a difference (or value) as extreme or more extreme than the observed difference (or value) under the assumption that the null hypothesis is true.

The smaller the p-value, the stronger the evidence against the null hypothesis.

Calculation of p-values for continuous distributions.

	p-value
$H_1 : \theta < \theta_0$	$P(t \leq t_{obs} \mid H_0)$
$H_1 : \theta > \theta_0$	$P(t \geq t_{obs} \mid H_0)$
$H_1 : \theta \neq \theta_0$	$2 \min\{P(t \leq t_{obs} \mid H_0), P(t \geq t_{obs} \mid H_0)\}$

Consider the test

$$H_0 : \theta = \theta_0 \qquad vs \qquad H_1 : \theta = \theta_1$$

where both hypotheses are simple. The **power** of this test is the probability of rejecting H_0 when H_1 is true.

$$P(\text{Reject } H_0 | \theta = \theta_1) = 1 - P(\text{Accept } H_0 | \theta = \theta_1) = 1 - \beta(\theta_1)$$

The power reflects the capacity of the test to detect the alternative hypothesis when it is true.

Observe that we need to know the sampling distribution for the test statistic **under the alternative hypothesis**.

If instead of a simple alternative we have a composite hypothesis $H_1 : \theta \in \Theta_1$, we consider the power as a function of θ for values $\theta \in \Theta_1$.

Shiny App

Shiny app for hypothesis tests:

https://andrewlau.shinyapps.io/Power/

Example (from Ugarte et al.)

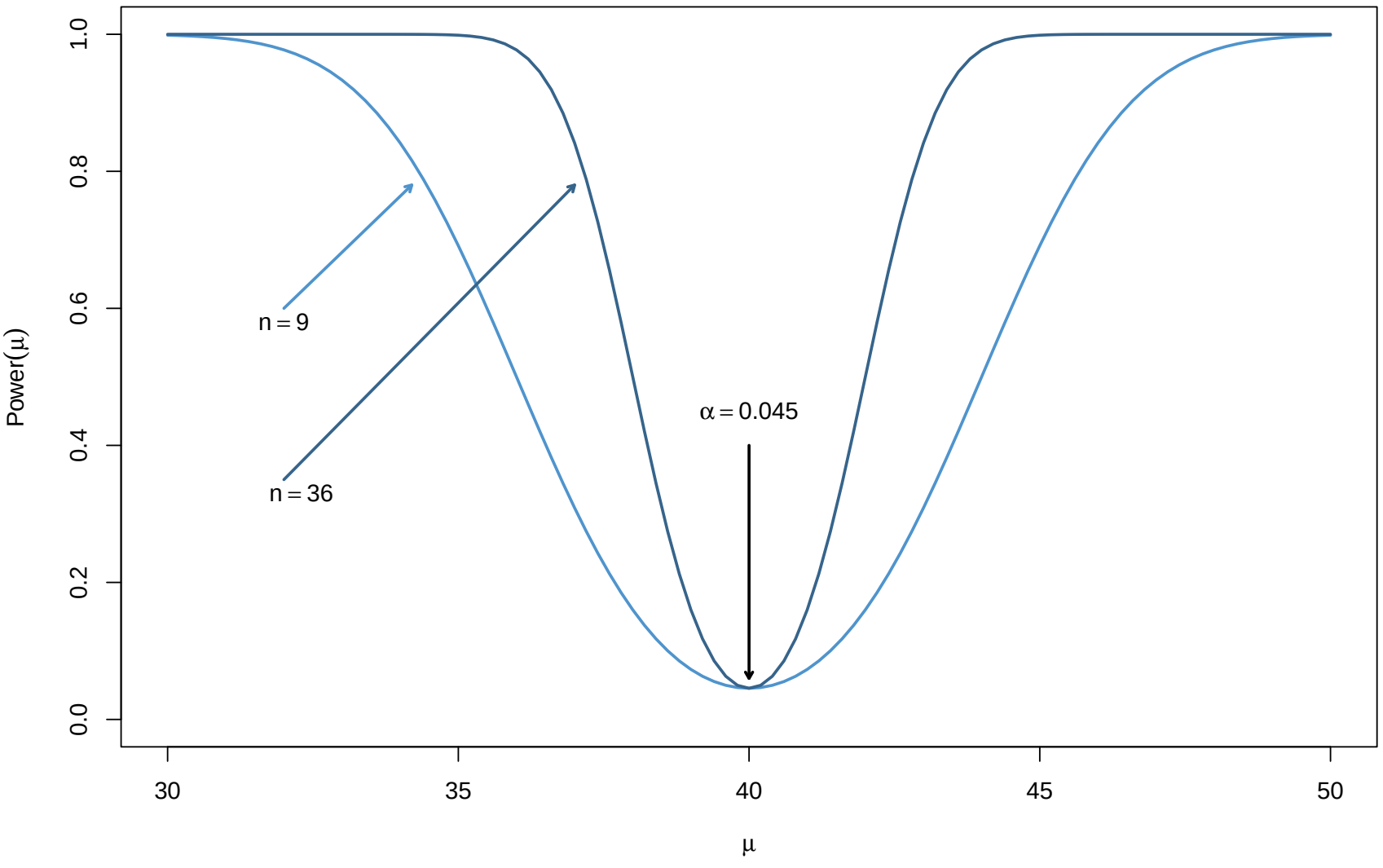
Test the null hypothesis that for a certain age group, the mean score on an achievement test is equal to 40 against the alternative that it is not equal to 40. Scores follow a normal distribution with $\sigma = 6$.

a. Find the probability of type I error for $n = 9$ if the null hypothesis is rejected when the sample mean is less than 36 or greater than 44.

b. Find the probability of type I error for $n = 36$ if the null hypothesis is rejected when the sample mean is less than 38 or greater than 42.

c. Plot the power functions for $n = 9$ and $n = 36$ for values of μ between 30 and 50.

Example

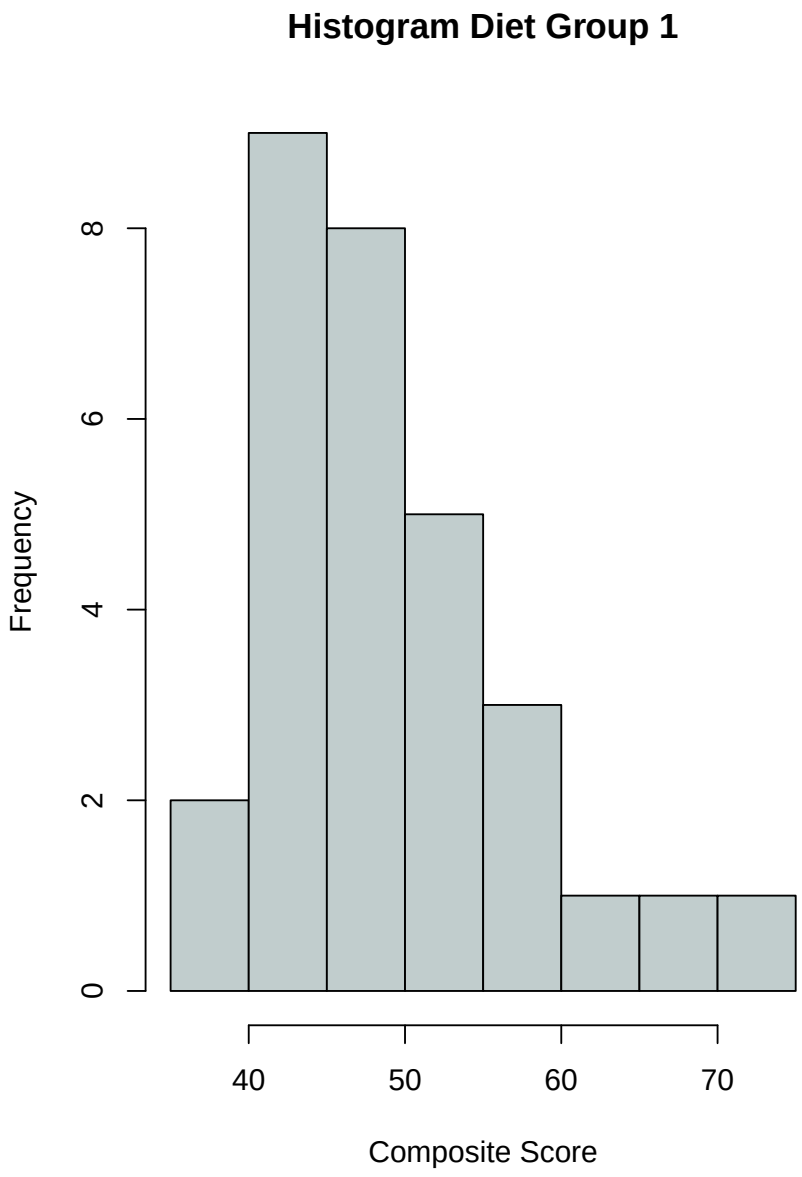


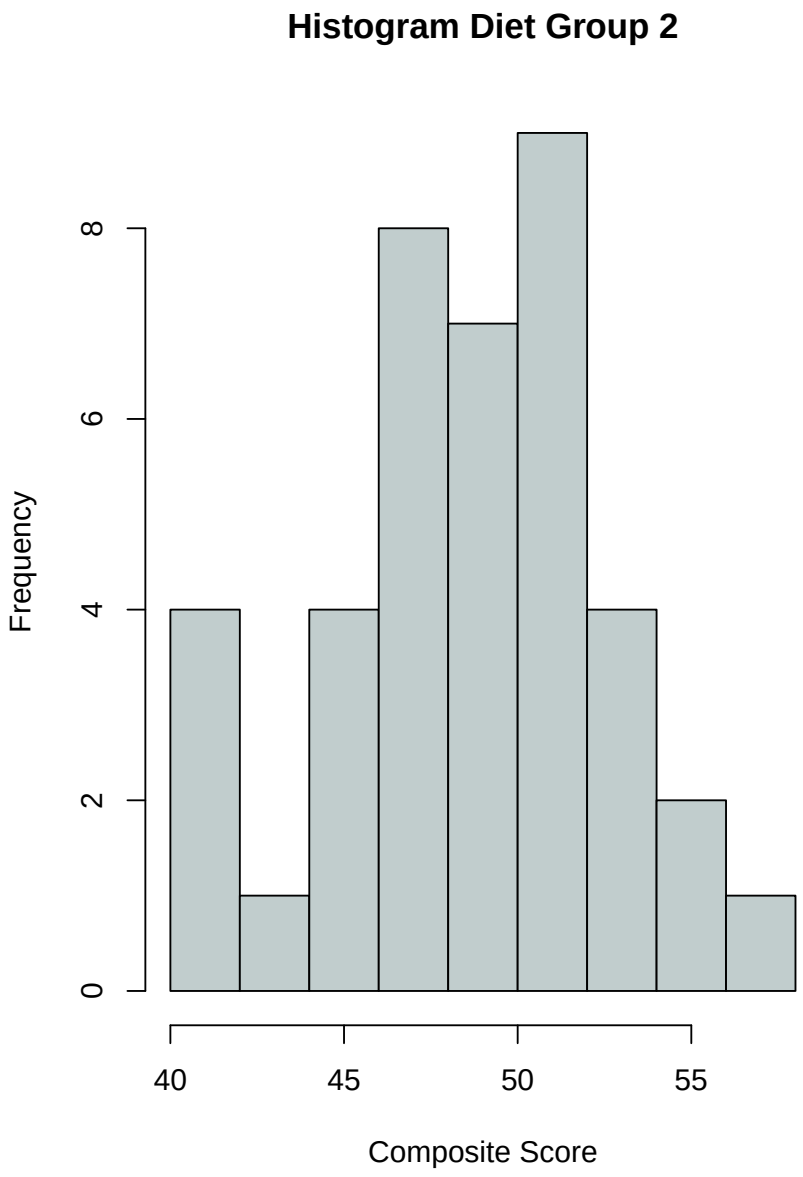
Video 17: Comparing Two Populations

Graphical tools

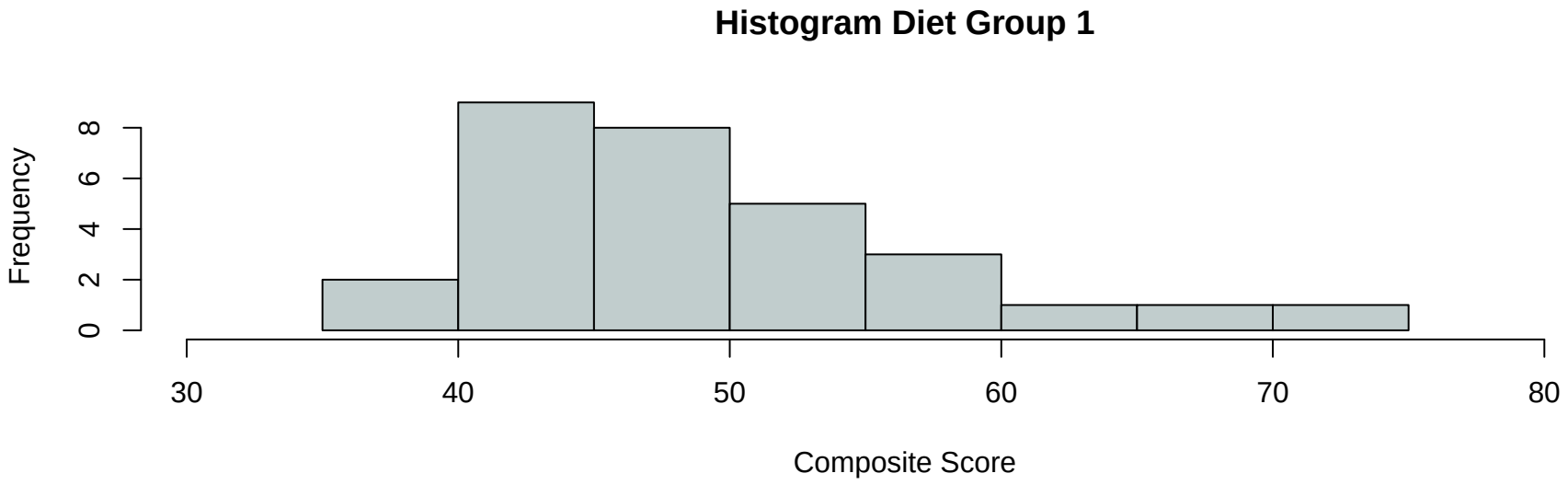
- Use of histograms and boxplots to compare samples from different populations.
- Uniform scales are necessary to make accurate comparisons.
- Outliers can have an important effect on the charts.

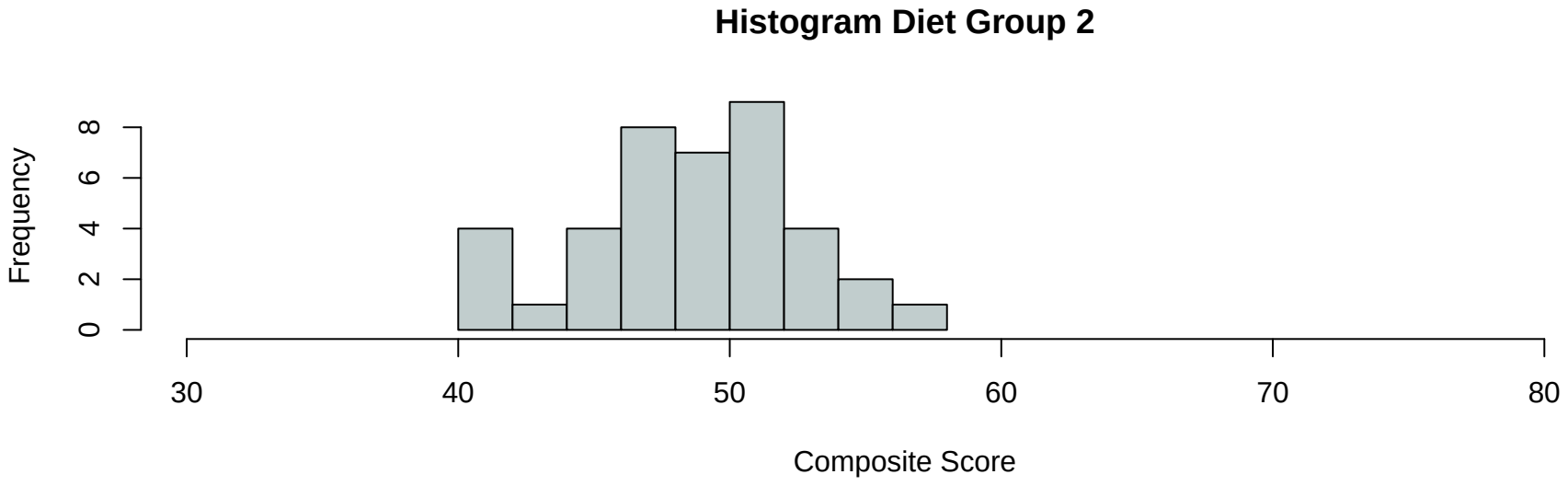
Graphical tools





Graphical tools

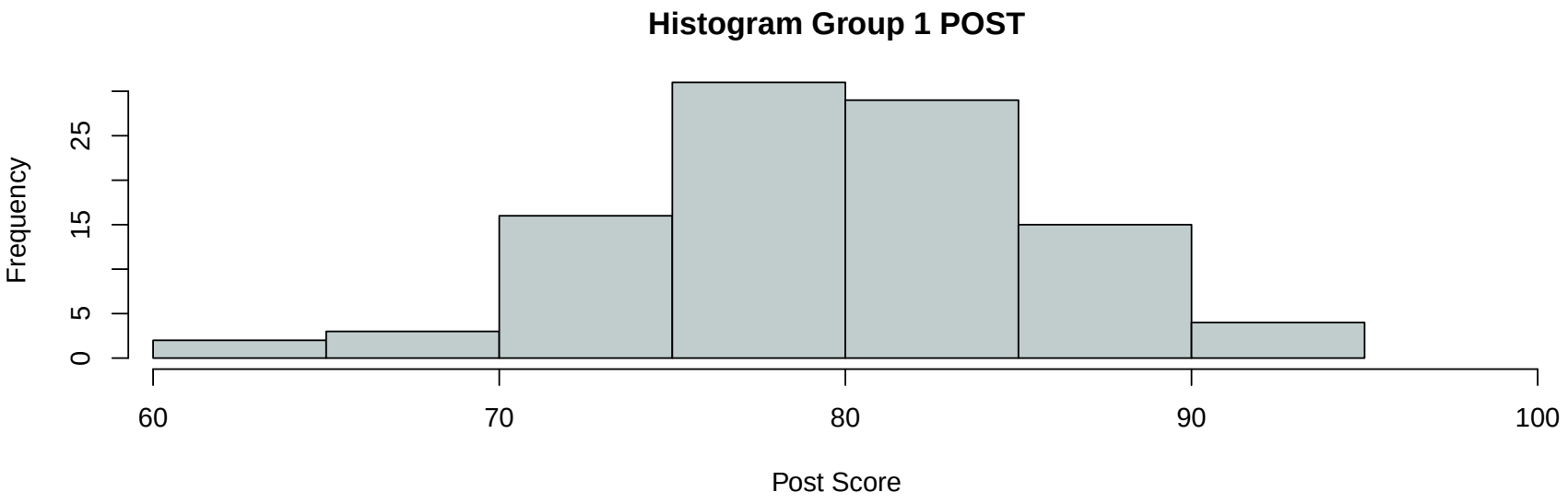


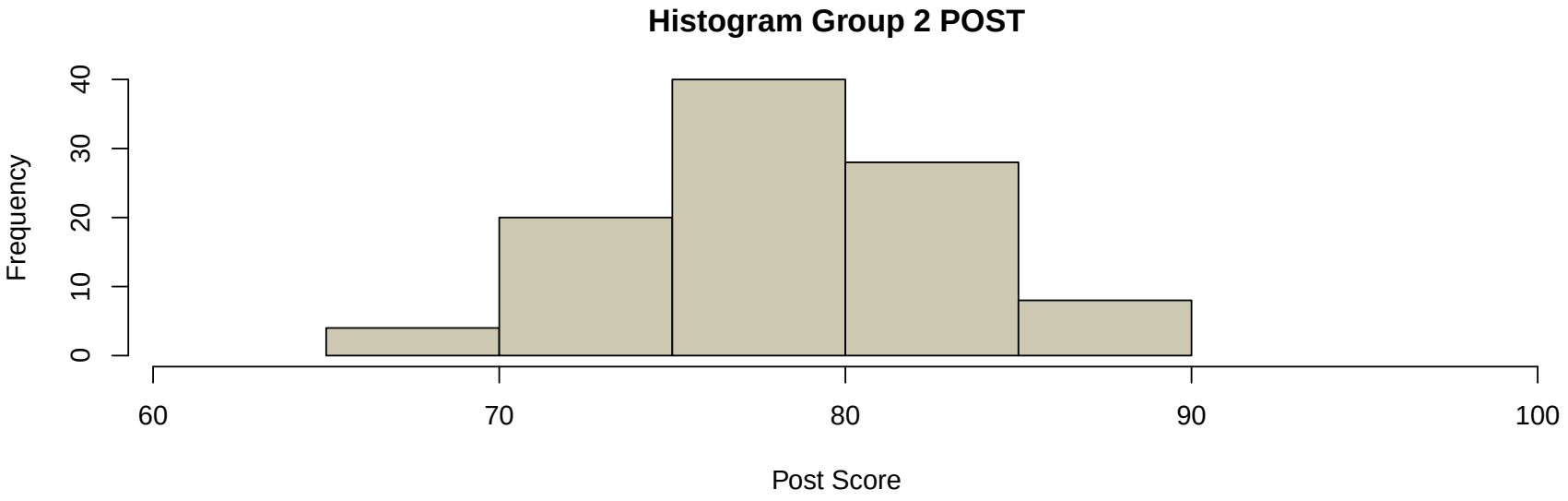


Comparing before and after measurements

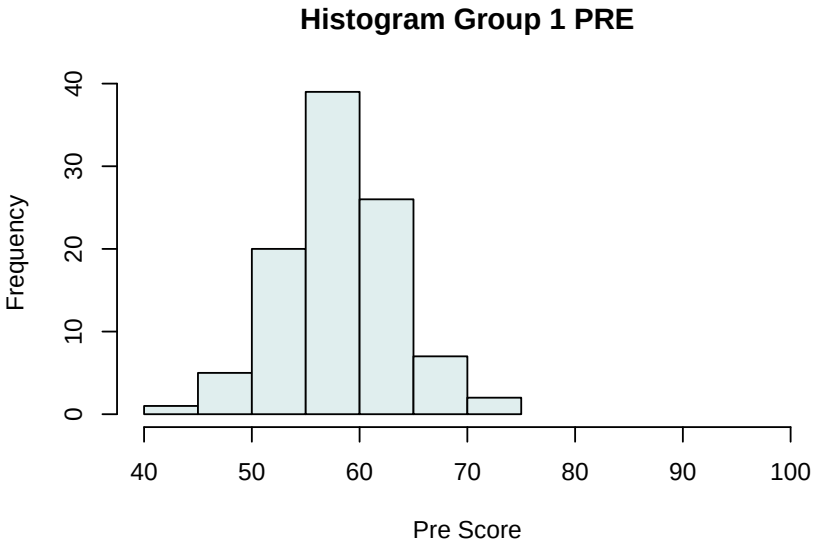
- When data measure the effect of a treatment, i.e., we have data before and after treatment for the same subject, looking at the final scores is not enough: we need to look at the baseline.
- In this case, improvement may be a better choice.

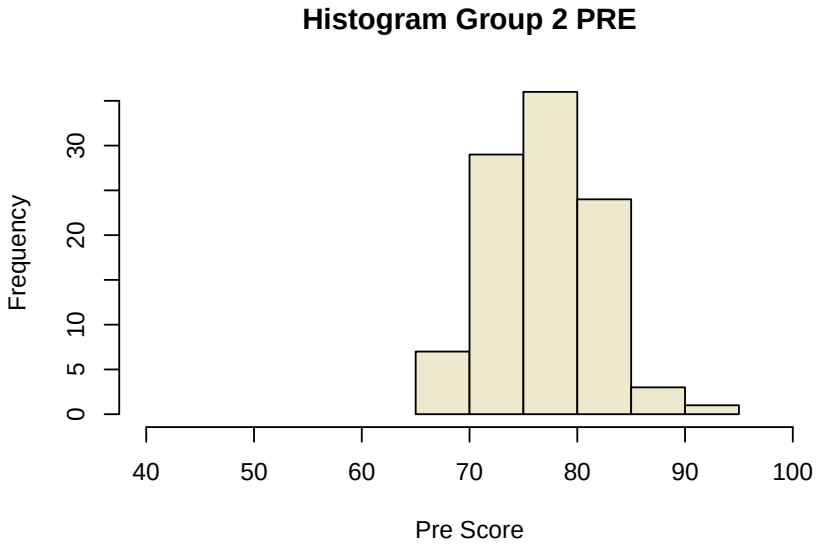
Graphical tools

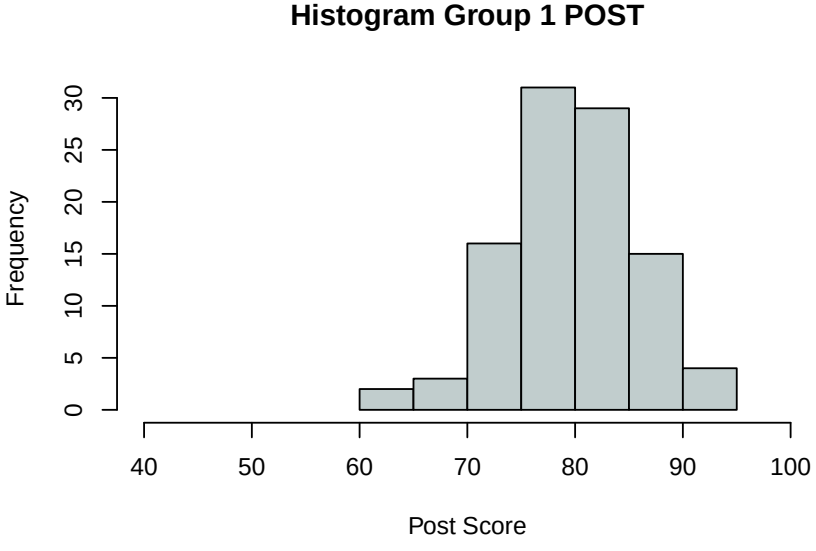


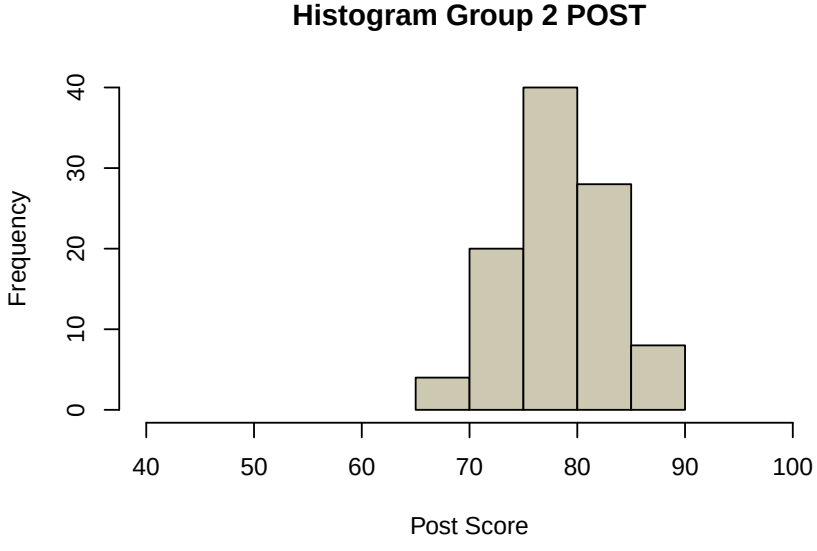


Comparing before and after measurements

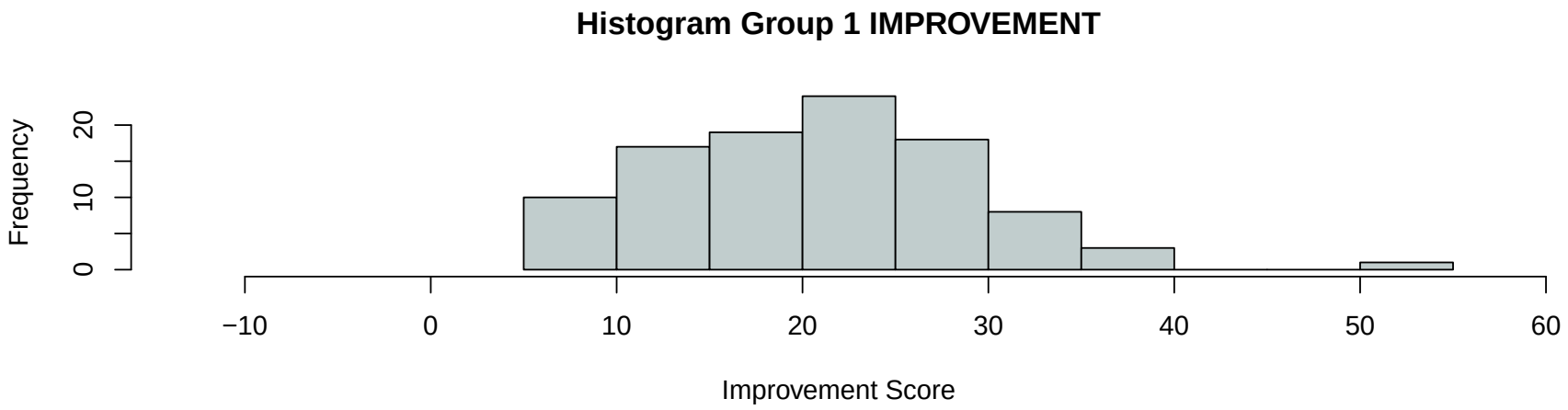


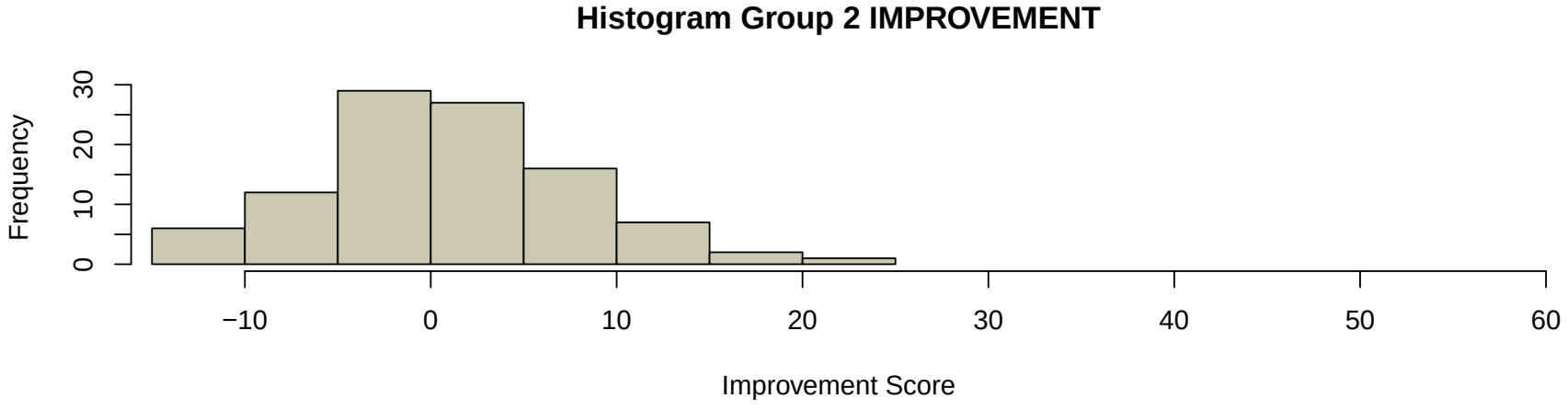






Comparing before and after measurements

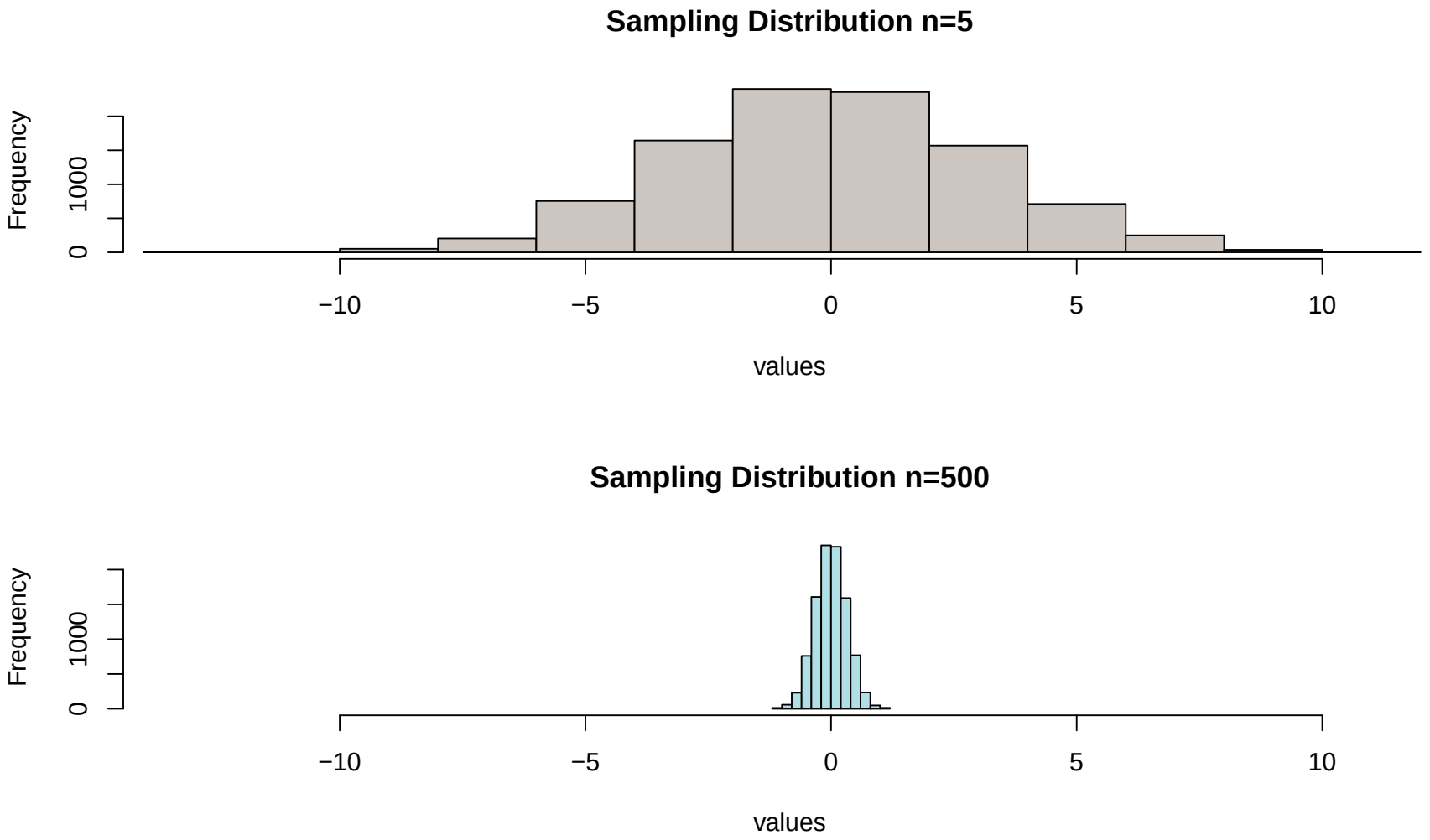




Uncertainty Arising from Random Sampling

- We looked at the difference between sample means through a simulation study.
- Increasing sample size decreases variability in the differences of the means

Uncertainty Arising from Random Sampling



Video 18: Two-Sample Problems

Comparing two population means

The usual test for comparing the means from two populations is the two-sample t test.

Suppose we have samples

$$x_1, x_2, \ldots, x_{n_1} \quad \text{and} \quad y_1, y_2, \ldots, y_{n_2}$$

and assume they come from the normal distributions

$$N(\mu_1, \sigma_1^2) \quad \text{and} \quad N(\mu_2, \sigma_2^2)$$

We want to test the null hypothesis

$$H_0 : \mu_1 = \mu_2.$$

Let $\bar{x}$ and $\bar{y}$ be the sample means for each sample, and s_x^2 and s_y^2 be the corresponding variances.

The test statistic, in this case, is the difference between the sample means divided by the standard error of the difference of the means:

$$\frac{\bar{x} - \bar{y}}{SEDM}$$

The standard error of the difference of the means is calculated by using the fact that for two independent random variables, the variance of their difference is the sum of the variances.

Comparing two population means

Two cases:

1.- Equal variances

```
t.test(sample1, sample2, var.equal = TRUE)
with(example.df,t.test(data ~ group, var.equal = TRUE))
```

2.- Unequal variances (default)

```
t.test(sample1, sample2, var.equal = FALSE)
with(example.df,t.test(data ~ group))
```

```
with(fish.df,t.test(fish ~ group, var.equal = TRUE))
```

```
##
##  Two Sample t-test
##
## data:  fish by group
## t = 0.37877, df = 68, p-value = 0.706
## alternative hypothesis: true difference in means betwee
## 95 percent confidence interval:
##  -2.489833  3.656499
## sample estimates:
## mean in group 1 mean in group 2
##        49.63333        49.05000
```

```
with(fish.df,t.test(fish ~ group))
```

```
##
##  Welch Two Sample t-test
##
## data:  fish by group
## t = 0.34458, df = 38.246, p-value = 0.7323
## alternative hypothesis: true difference in means betwee
## 95 percent confidence interval:
##  -2.843034  4.009701
## sample estimates:
## mean in group 1 mean in group 2
##        49.63333        49.05000
```

Comparing Variances

The usual test for comparing variances is known as Fisher's F test.

The test is based on the quotient between the two variances, with the condition that the bigger variance must be on top.

The distribution for this quotient is known as Fisher's F distribution. It has two parameters that correspond to the degrees of freedom of the numerator and the denominator.

```
var.test(sample1,sample2)
with(example.df,var.test(data ~ group))
```

```
options(width=50)
var.test(fish1,fish2)
```

```
##
##  F test to compare two variances
##
## data:  fish1 and fish2
## F = 4.712, num df = 29, denom df = 39,
## p-value = 1.034e-05
## alternative hypothesis: true ratio of variances is not e
## 95 percent confidence interval:
##  2.401777 9.577864
## sample estimates:
## ratio of variances
##           4.711972
```

Nonparametric Test: Wilcoxon's test

The test statistic W is calculated as follows:

- Mix the two samples and order them.
- Calculate the ranks for the elements of each group and sum the ranks for one of the groups

The problem now is reduced to sampling n_1 values from the numbers 1 to $n_1 + n_2$.

```
wilcox.test(sample1,sample2)
with(example.df,wilcox.test(data ~ group))
```

Nonparametric Test: Wilcoxon's test

```
options(width=50)
wilcox.test(fish1,fish2)
```

```
## Warning in wilcox.test.default(fish1, fish2):
## cannot compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity
##  correction
##
## data:  fish1 and fish2
## W = 561.5, p-value = 0.6512
## alternative hypothesis: true location shift is not equal
```

We have measurements for the same subjects before and after 'treatment'.

These measurements are usually highly correlated.

Analyzing this data as if the samples were independent could lead to severe errors.

Remember that we showed that

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

If the covariance between X and Y is positive, by the equation above, the variance of the difference will be smaller than the sum of the individual variances.

This is an advantage because it makes easier it to detect significant differences between the means.

Tests for Paired Data

- t.test

```
t.test(sample.pre, sample.post, paired = TRUE)
with(example.df, t.test(data ~ group), paired = TRUE)
```

- Nonparametric test

```
wilcox.test(sample1,sample2, paired = TRUE)
with(example.df,wilcox.test(data ~ group,
                            paired = TRUE))
```

```
library(MASS); options(width = 50)
with(shoes,t.test(A, B))
```

```
##
##  Welch Two Sample t-test
##
## data:  A and B
## t = -0.36891, df = 17.987, p-value = 0.7165
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  -2.745046  1.925046
## sample estimates:
## mean of x mean of y
##     10.63     11.04
```

```
options(width = 50)
with(shoes,t.test(A, B, paired = TRUE))
```

```
##
##  Paired t-test
##
## data:  A and B
## t = -3.3489, df = 9, p-value = 0.008539
## alternative hypothesis: true difference in means is not
## 95 percent confidence interval:
##  -0.6869539 -0.1330461
## sample estimates:
## mean of the differences
##                   -0.41
```

```
with(shoes,wilcox.test(A, B))
```

```
## Warning in wilcox.test.default(A, B): cannot
## compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity
##  correction
##
## data:  A and B
## W = 42.5, p-value = 0.5966
## alternative hypothesis: true location shift is not equal
```

```
with(shoes,wilcox.test(A, B, paired = TRUE))
```

```
## Warning in wilcox.test.default(A, B, paired =
## TRUE): cannot compute exact p-value with ties

##
##  Wilcoxon signed rank test with continuity
##  correction
##
## data:  A and B
## V = 3, p-value = 0.01431
## alternative hypothesis: true location shift is not equal
```