Methods of Empirical Finance

Seminar (UE)

Christoph Huber
University of Innsbruck

Master in Banking and Finance
Winter term 2019/20 (this version: 2019-11-26)

(Non)Parametric Statistical Tests

Methods of Empirical Finance

(Non)Parametric Statistical Tests

Parametric vs. Non-parametric tests

Parametric:

- based on *parametric* families of *probability distributions*
- rely on assumptions that the data are drawn from a given prob. distribution (e.g. the underlying population is normally distributed)

Non-parametric:

- *distribution-free* test statistics
- do *not* rely on assumptions about the data's prob. distribution

(Non)Parametric Statistical Tests

Parametric vs. Non-parametric tests

When should I use a parametric test?

When should I use a *non*-parametric test?

- **scale/level of measurement** ?

  non-metric (nominal, ordinal) vs. metric (interval, ratio) data

- **the distribution of the data** ?

  normally distributed vs. non-normally distributed

(Non)Parametric Statistical Tests

Parametric vs. Non-parametric tests

We mostly use ...

... **parametric** test statistics if ...

- the level of measurement is *metric* (interval or ratio),
- the population is *normally distributed*,
- the sample size is *large*

... **non-parametric** test statistics if ...

- the level of measurement is *non-metric* (nominal or ordinal),
- the population is *non-normally distributed*,
- the sample size is *small*

(Non)Parametric Statistical Tests

As the appropriate test statistic depends on the data's level of measure, its (assumed) distribution, as well as its sample size, …

… it makes sense to always get an overview about the data first,

… and — if you want to use parametric test statistics, in particular —, to test for normality of your data
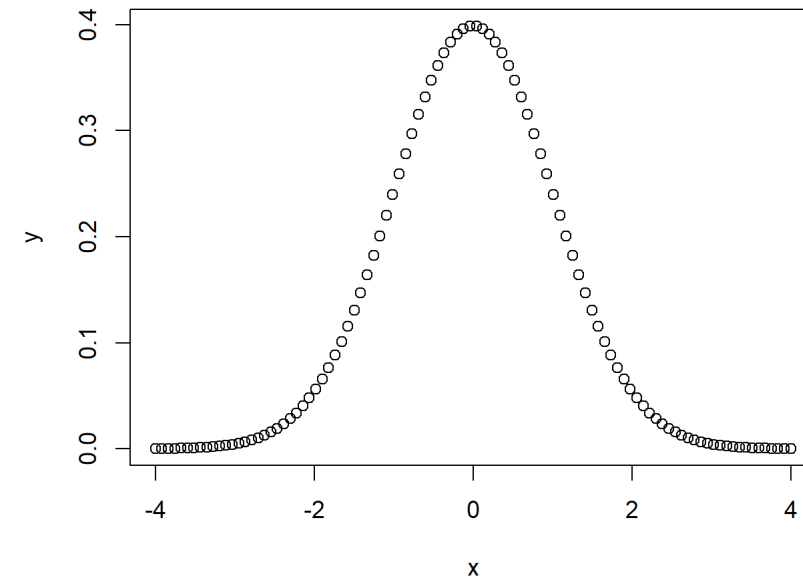
Testing for Normality

Testing for Normality

Some useful functions for distributions in *R*

For normal distributions:

```
dnorm(x=0, mean=0, sd=1)     # prob. density function (pd
## [1] 0.3989423
```

```
x <- seq(-4, 4, length=100)
y <- dnorm(x)
plot(x, y)
```
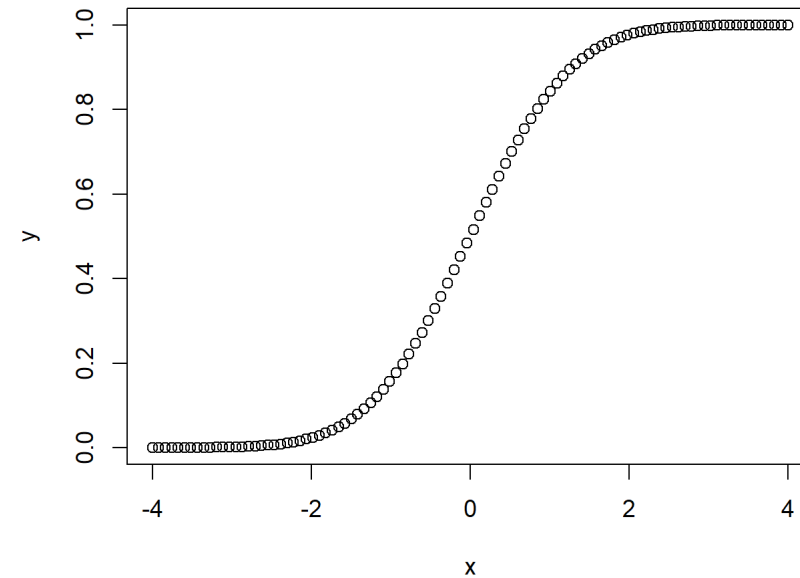
Testing for Normality

Some useful functions for distributions in *R*

```r
pnorm(74, mean=70, sd=2)     # cumulative density functio
```

```
## [1] 0.9772499
```

```r
y <- pnorm(x)
plot(x, y)
```

Testing for Normality

Some useful functions for distributions in *R*

```r
# e.g. find the Z-score of the 99th quantile of the standard normal distribution
qnorm(.99, mean=0, sd=1)    # (inverse cumulative density function (cdf))
```

```
## [1] 2.326348
```

```r
rnorm(5, mean = 10, sd = 2)    # generates normally distributed random variables
```

```
## [1] 12.441864 11.688925  7.233207 10.074942 11.343875
```

See here, for example.

Similarly, these functions are available for most other popular probability distributions.

Testing for Normality

'Descriptive': Histograms and normal density

```
hist(data$PCT_WOMEN_ON_BOARD, prob = TRUE, breaks = 20)
curve(dnorm(x,
            mean = mean(data$PCT_WOMEN_ON_BOARD, na.rm = TRUE),
            sd = sd(data$PCT_WOMEN_ON_BOARD, na.rm = TRUE)),
      col = ubluelight, lwd = 2,
      add = TRUE, yaxt="n")
```
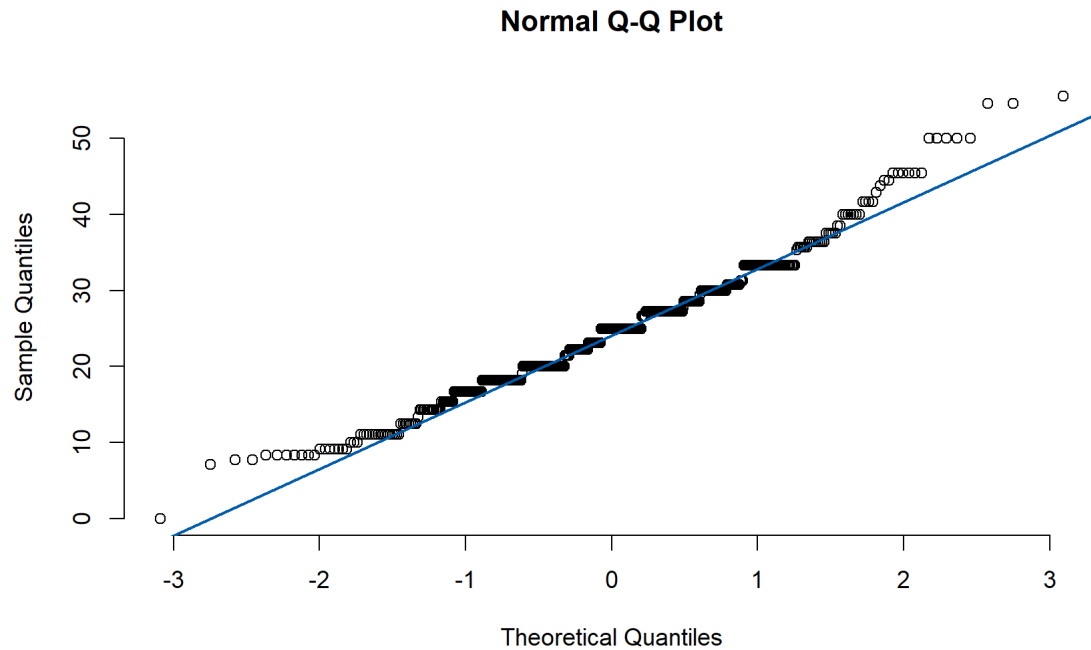
Histogram of data$PCT_WOMEN_ON_BOARD
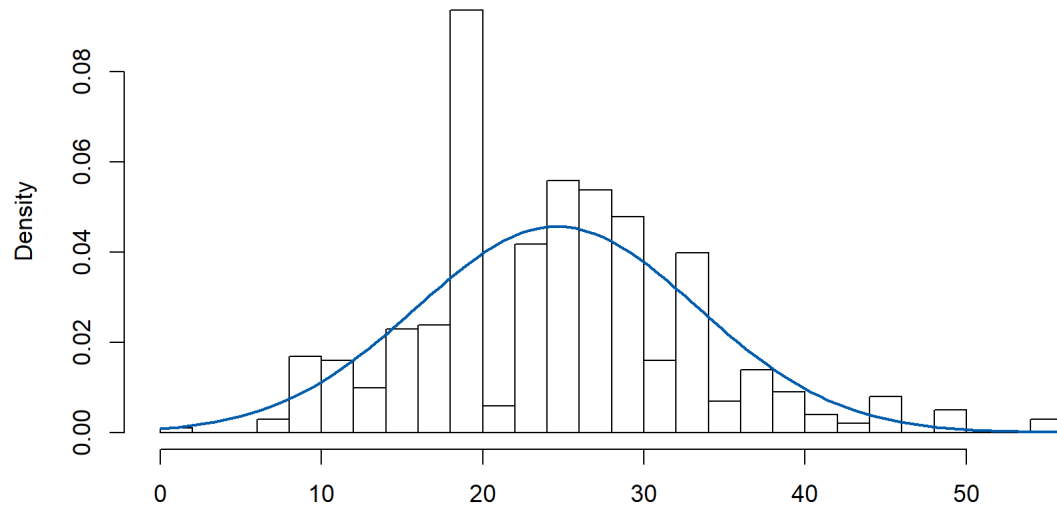
Testing for Normality

'Descriptive': QQ plots

```
qqnorm(data$PCT_WOMEN_ON_BOARD, frame = FALSE)
qqline(data$PCT_WOMEN_ON_BOARD, col = ubluelight, lwd = 2)
```


Normal Q-Q Plot

Testing for Normality

Kolmogorov-Smirnov Goodness-of-Fit Test

> tests whether the empirical distribution function of the sample and the cumulative distribution function of some reference distribution are equal in location and shape

- **nature of independent variable(s):** no independent variables, 1 population
- **scale of dependent variable:** ordinal or interval

- **hypothesis:**
  - H_0: $F(x) = F_0(x)$ for all values of x
  - H_1: $F(x) \neq F_0(x)$ or $F(x) > F_0(x)$ or $F(x) < F_0(x)$

universität
innsbruck

Testing for Normality

Kolmogorov-Smirnov Goodness-of-Fit Test — Example

- Test *whether the variable* `CHG_PCT_YTD` *is normally distributed*

```
ks.test(data$PCT_WOMEN_ON_BOARD,
        "pnorm",
        mean = mean(data$PCT_WOMEN_ON_BOARD, na.rm=TRUE),
        sd = sd(data$PCT_WOMEN_ON_BOARD, na.rm=TRUE))
```

```
##
##      One-sample Kolmogorov-Smirnov test
##
## data:  data$PCT_WOMEN_ON_BOARD
## D = 0.075887, p-value = 0.006166
## alternative hypothesis: two-sided
```

Testing for Normality

Shapiro-Wilk Test

> tests whether a sample $x_1, \ldots, x_n$ is drawn from a normally distributed population

- **nature of independent variable(s):** no independent variables, 1 population
- **scale of dependent variable:** ordinal or interval

- **hypothesis:**
    - H_0: $F(x) = F_0(x)$ for all values of x
    - H_1: $F(x) \neq F_0(x)$ or $F(x) > F_0(x)$ or $F(x) < F_0(x)$

universität
innsbruck

Testing for Normality

Shapiro-Wilk Test — Example

- Test *whether the variable* `CHG_PCT_YTD` *is normally distributed*

```
shapiro.test(data$CHG_NET_YTD)
```

```
##
##      Shapiro-Wilk normality test
##
## data:  data$CHG_NET_YTD
## W = 0.37452, p-value < 2.2e-16
```

universität
innsbruck

One-Sample Test Statistics

One-Sample Test Statistics

χ^2 Goodness of Fit Test

> tests whether the observed proportions for a categorical variable of a single population differ significantly from hypothesized (expected) proportions

- **nature of independent variable(s):** no independent variables, 1 population
- **scale of dependent variable:** categorical/nominal

- **hypothesis:**
  - H_0: $p_i = p_{i,0} \forall i \in [1, \ldots, k]$
  - H_1: $p_i \neq p_{i,0} \forall i \in [1, \ldots, k]$ ('The observed proportions ($p_{i,0}$) differ from hypothesized (expected) proportions (p_i).')

One-Sample Test Statistics

χ^2 Goodness of Fit Test — Example

- We use the data from `sp500_data.csv`.
- Test *whether the proportion of companies in the S&P 500 with a female CEO is different from 20%.*

```r
females <- sum(data$FEMALE_CEO_OR_EQUIVALENT, na.rm=TRUE)
total <- sum(!is.na(data$FEMALE_CEO_OR_EQUIVALENT))

x <- c(total - females, females)
x
```

```
## [1] 479  23
```

```r
chisq.test(x, p = c(0.8, 0.2))
```

```
##
##      Chi-squared test for given probabilities
##
## data:  x
## X-squared = 74.586, df = 1, p-value < 2.2e-16
```

One-Sample Test Statistics

One-Sample Wilcoxon Signed Rank Test

> tests whether the population from which the data were sampled is symmetric or not about the default value

- **nature of independent variable(s):** no independent variables, 1 population
- **scale of dependent variable:** ordinal or interval

One-Sample Test Statistics

One-Sample Wilcoxon Signed Rank Test — Example

- Test *whether the median yearly net change in prices is different from 0.* I.e., whether the population from which the data were sampled is centered around 0.

```r
wilcox.test(data$CHG_NET_YTD, mu = 0)
```

```
## 
## 	Wilcoxon signed rank test with continuity correction
## 
## data:  data$CHG_NET_YTD
## V = 115647, p-value < 2.2e-16
## alternative hypothesis: true location is not equal to 0
```

universität
innsbruck

One-Sample Test Statistics

One-Sample t-Test

> evaluates whether the sample mean (μ) of a test variable significantly differs from a constant test value μ_0

- **nature of independent variable(s):** no independent variables, 1 population
- **scale of dependent variable:** interval or ratio

- **hypothesis:**
  - H_0: $\mu = \mu_0$
  - H_1: $\mu \neq \mu_0$ or $\mu > \mu_0$ or $\mu < \mu_0$ ('The sample mean μ significantly differs from the constant test value μ_0.')

- **assumptions:**
  - the observations in the sample are independent of each other
  - the observations follow a normal distribution with mean μ and variance σ^2

One-Sample Test Statistics

One-Sample t-Test — Example

- Test *whether the mean yearly net change in prices is different from 0.*

```
summary(data$CHG_NET_YTD)
```

```
##     Min.  1st Qu.  Median    Mean  3rd Qu.    Max.     NA's
## -146.950    5.395  15.170  28.076   33.490 1215.650        6
```

```
t.test(data$CHG_NET_YTD, mu = 0)
```

```
##
##     One Sample t-test
##
## data:  data$CHG_NET_YTD
## t = 9.1499, df = 498, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   22.04771 34.10538
## sample estimates:
## mean of x
##   28.07655
```

But: Do our assumptions hold? Is the data *normally distributed*?

Unpaired Two-Sample Test Statistics

Unpaired Two-Sample Test Statistics

Pearson's χ^2 Test for $r \times$ c tables

> tests whether there is a relationship between two categorical variables (or tests whether or not the r samples are homogeneous regarding the proportion of observations in each of the c categories; χ^2 test for homogeneity)

- **nature of independent variable(s):** 1 independent variables with 2 levels (independent groups)
- **scale of dependent variable:** categorical/nominal

- **hypothesis:**
  - H_0: $O_{ij} = E_{ij}$ for all cells
  - H_1: $O_{ij} \neq E_{ij}$ for at least one cell

Unpaired Two-Sample Test Statistics

Pearson's χ^2 Test for $r \times c$ tables — Example

- Test whether there is a relationship between the company's industry group and the percentage of women on its board.

```
chisq.test(data$INDUSTRY_GROUP, data$PCT_WOMEN_ON_BOARD)
```

```
##
##      Pearson's Chi-squared test
##
## data:  data$INDUSTRY_GROUP and data$PCT_WOMEN_ON_BOARD
## X-squared = 2532.7, df = 2419, p-value = 0.05262
```

Unpaired Two-Sample Test Statistics

Wilcoxon Rank Sum Test (Mann-Whitney U-test)

> tests whether two independent samples are drawn from the same population

- **nature of independent variable(s):** 1 independent variables with 2 levels (independent groups)
- **scale of dependent variable:** ordinal or interval

Unpaired Two-Sample Test Statistics

Wilcoxon Rank Sum Test — Example

- Test whether the percentage of women on the board of financial companies are drawn from the same distribution as the pct. of women on board of industrial companies.

```
wilcox.test(PCT_WOMEN_ON_BOARD ~ INDUSTRY_SECTOR,
            data = data %>% filter(INDUSTRY_SECTOR == "Financial" | INDUSTRY_SECTOR == "Industrial"))
```

```
##
##      Wilcoxon rank sum test with continuity correction
##
## data:  PCT_WOMEN_ON_BOARD by INDUSTRY_SECTOR
## W = 3877, p-value = 0.04192
## alternative hypothesis: true location shift is not equal to 0
```

Unpaired Two-Sample Test Statistics

Kolmogorov-Smirnov Test

> tests whether two samples are drawn from the same distribution, or, more precisely, whether two independent empirical distributions are equal in location and shape

- **nature of independent variable(s):** 1 independent variables with 2 levels (independent groups)
- **scale of dependent variable:** ordinal or interval

Unpaired Two-Sample Test Statistics

Kolmogorov-Smirnov Test

- Test whether the percentage of women on the board of financial companies are drawn from the same distribution as the pct. of women on board of industrial companies or — more precisely — whether the the empirical distributions are equal in location and shape.

```r
financial <- data %>% filter(INDUSTRY_SECTOR == "Financial") %>% .$PCT_WOMEN_ON_BOARD
industrial <- data %>% filter(INDUSTRY_SECTOR == "Industrial") %>% .$PCT_WOMEN_ON_BOARD

ks.test(financial, industrial)
```

```
##
##      Two-sample Kolmogorov-Smirnov test
##
## data:  financial and industrial
## D = 0.23232, p-value = 0.02783
## alternative hypothesis: two-sided
```

Unpaired Two-Sample Test Statistics

Independent Sample t-Test

> evaluates whether the sample means (μ_1 and μ_2) of two independent variables significantly differ

- **nature of independent variable(s):** 1 independent variables with 2 levels (independent groups)
- **scale of dependent variable:** interval or ratio

- **hypothesis:**
  - H_0: $\mu_1 = \mu_2$
  - H_1: $\mu_1 \neq \mu_2$ or $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$

- **assumptions:**
  - all observations from both groups are independent of each other
  - the observations follow a normal distribution with mean μ and variance σ^2
  - the two independent samples have the same standard deviation, otherwise Welch's approximation should be applied

Unpaired Two-Sample Test Statistics

Independent Sample t-Test — Example

- Test whether the mean percentage of woman is different in financial companies compared to industrial companies.

```
financial <- data %>% filter(INDUSTRY_SECTOR == "Financial") %>% .$PCT_WOMEN_ON_BOARD
industrial <- data %>% filter(INDUSTRY_SECTOR == "Industrial") %>% .$PCT_WOMEN_ON_BOARD

t.test(financial, industrial, var.equal=TRUE)
```

```
##
##      Two Sample t-test
##
## data:  financial and industrial
## t = 2.0421, df = 163, p-value = 0.04275
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.07825703 4.65716398
## sample estimates:
## mean of x mean of y
##   25.40932  23.04161
```

Unpaired Two-Sample Test Statistics

Independent Sample t-Test — Example

- What if the two samples have different σs?

```
t.test(financial, industrial, var.equal=FALSE)
```

```
##
##      Welch Two Sample t-test
##
## data:  financial and industrial
## t = 2.0669, df = 145.08, p-value = 0.04052
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1036396 4.6317814
## sample estimates:
## mean of x mean of y
##  25.40932  23.04161
```

universität
innsbruck

Paired Two-Sample Test Statistics

Paired Two-Sample Test Statistics

Wilcoxon Signed Rank Test

> tests whether two *dependent* samples represent two different populations or whether the two *dependent* samples are drawn from the same population

- **nature of independent variable(s):** 1 independent variables with 2 levels (*dependent* groups)
- **scale of dependent variable:** ordinal or interval

- **assumptions:**
  - the data are paired and come from the same population
  - each pair is chosen randomly and independently

Paired Two-Sample Test Statistics

Wilcoxon Signed Rank Test — Example

- Test whether the percentage of women on the board of financial companies in 2019 are drawn from the same distribution as the pct. of women on board of financial companies in 2014.

```
wilcox.test()
```

Paired Two-Sample Test Statistics

Dependent t-Test for Paired Samples

> evaluates whether the sample means (μ_1 and μ_2) of two *independent* samples differ significantly

- **nature of independent variable(s):** 1 independent variables with 2 levels (*dependent* groups)
- **scale of dependent variable:** interval or ratio

- **hypothesis:**
    - H_0: $\mu = \mu_0$
    - H_1: $\mu \neq \mu_0$ or $\mu > \mu_0$ or $\mu < \mu_0$ ('...')

- **assumptions:**
    - the observations follow a normal distribution with mean μ and variance σ^2
    - the two samples have the same standard deviation (homogeneity of variance)

Paired Two-Sample Test Statistics

Dependent t-Test for Paired Samples

- Test whether the mean percentage of women on the board of financial companies in 2019 is different to the mean pct. of women on board of financial companies in 2014.

```
t.test()
```

n-Sample Generalizations: more than 2 groups

n-Sample Generalizations

Kruskal-Wallis Test

> tests whether two or more independent samples originate from the same distribution, i.e., extends the Mann-Whitney U-test when there are *more than two groups* (for two groups, both statistics are identical)

- **nature of independent variable(s):** 1 independent variables with two or more levels (independent groups)
- **scale of dependent variable:** ordinal or interval

n-Sample Generalizations

Kruskal-Wallis Test — Example

- Test whether the percentage of women on the board of companies from different industry sectors are drawn from the same distribution.

```
kruskal.test(PCT_WOMEN_ON_BOARD ~ INDUSTRY_SECTOR,
             data = data)
```

```
##
##      Kruskal-Wallis rank sum test
##
## data:  PCT_WOMEN_ON_BOARD by INDUSTRY_SECTOR
## Kruskal-Wallis chi-squared = 9.3157, df = 8, p-value = 0.3164
```

n-Sample Generalizations

Jonkheere-Terpstra Test (Jonkheere's trend test)

> tests whether three or more independent samples originate from the same distribution, (just as the Kruskal-Wallis test) but with an *ordered* alternative hypothesis

- **nature of independent variable(s):** 1 independent variables with three or more levels (independent groups)
- **scale of dependent variable:** ordinal or interval

- **hypothesis:**
  - H_0: $\theta_1 = \theta_2 = \ldots = \theta_k$
  - H_1: $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$

- **assumptions and limitations:**
  - all observations from all k groups are independent of each other
  - hypothesized ordering of groups' rank-sums required

universität
innsbruck

n-Sample Generalizations

Jonkheere-Terpstra Test — Example

- Test whether the percentage of women on the board of companies from financial companies is larger than the percentage in industrial companies, which is then larger than the respective percentage in basic materials companies.

```r
library(clinfun)
groups <- data %>% filter(INDUSTRY_SECTOR == "Financial" |
                              INDUSTRY_SECTOR == "Industrial" |
                              INDUSTRY_SECTOR == "Basic Materials")
groups$INDUSTRY_SECTOR <- factor(groups$INDUSTRY_SECTOR, levels = c("Financial", "Industrial", "Basic Materials"), ordere
jonckheere.test(groups$PCT_WOMEN_ON_BOARD,
                groups$INDUSTRY_SECTOR)
```

```
##
##      Jonckheere-Terpstra test
##
## data:
## JT = 4237, p-value = 0.1625
## alternative hypothesis: two.sided
```

Permutation Tests

Permutation Tests

Consider the following hypothetical problem:

- Ten subjects have been randomly assigned to one of two treatment conditions (A or B) and an outcome variable (score) has been recorded.

Treatment A	Treatment B
40	57
57	64
45	55
55	62
58	65

- First, let's use a parametric approach to compare Treatment A and Treatment B. $\rightarrow$ t-test:
  - assume normal distributions, independent groups, and equal variances

```r
a <- c(40, 57, 45, 55, 58)
b <- c(57, 64, 55, 62, 65)
t.test(a, b, var.equal = TRUE)
```

```
##
##      Two Sample t-test
##
## data:  a and b
## t = -2.345, df = 8, p-value = 0.04705
## alternative hypothesis: true difference in means is not equal t
## 95 percent confidence interval:
##  -19.0405455  -0.1594545
## sample estimates:
## mean of x mean of y
##      51.0      60.6
```

universität
innsbruck

Permutation Tests

A permutation test takes a different approach: If the two treatments are truly equivalent, the label (Treatment A or Treatment B) assigned to an observed score is arbitrary.

1. Calculate the observed t-statistic, as in the parametric approach; call this t0.

2. Place all 10 scores in a single group.

3. Randomly assign five scores to Treatment A and five scores to Treatment B.

4. Calculate and record the new observed t-statistic.

5. Repeat steps 3–4 for every possible way of assigning five scores to Treatment A and five scores to Treatment B. There are 252 such possible arrangements.

6. Arrange the 252 t-statistics in ascending order. This is the empirical distribution, based on (or conditioned on) the sample data.

7. If t0 falls outside the middle 95 percent of the empirical distribution, reject the null hypothesis that the population means for the two treatment groups are equal at the 0.05 level of significance.

Permutation Tests

- In both the permutation and the parametric approach, the exact same t-statistic is calculated.

- In the permutation approach, however, the statistic is compared not to a theoretical distribution (e.g. an assumed normal distribution) but to an *empirical distribution* created from *permutations of the observed data*.

- If this empirical distribution contains *all possible permutations* of the data, the permutation test is called an 'exact' test.

- If there are too many possible permutations to compute the test statistic in appropriate time, we ca use (Monte Carlo) *simulations*.

Permutation Tests

In *R*, we can use the `coin` package to run permutation tests as alternatives for a number of tests.

```r
library(coin)
```

```
## Loading required package: survival
```

E.g.

```r
df <- data.frame(value = c(a, b), treatment = c(rep("A", 5), rep("B", 5)))
oneway_test(value ~ treatment, data = df)
```

```
##
##     Asymptotic Two-Sample Fisher-Pitman Permutation Test
##
## data:  value by treatment (A, B)
## Z = -1.9147, p-value = 0.05553
## alternative hypothesis: true mu is not equal to 0
```

See here, for example.

(Non)parametric Correlation Measures

- Parametric: Pearson's ρ

- Non-parametric:

  - Spearman rank correlation ρ: i.e., the Pearson correlation coefficient between the ranked variables

  - Kendall's τ: similar to Spearman's ρ; measures *ordinal association* between two measured quantities

The argumentations for using non-parametric *or* parametric correlation measures are very similar to the ones for using non_parametric or parametric test statistics.

Things to consider are: level of measurement, assumptions about the distribution, assumptions about a linear or non-linear relationship, etc.
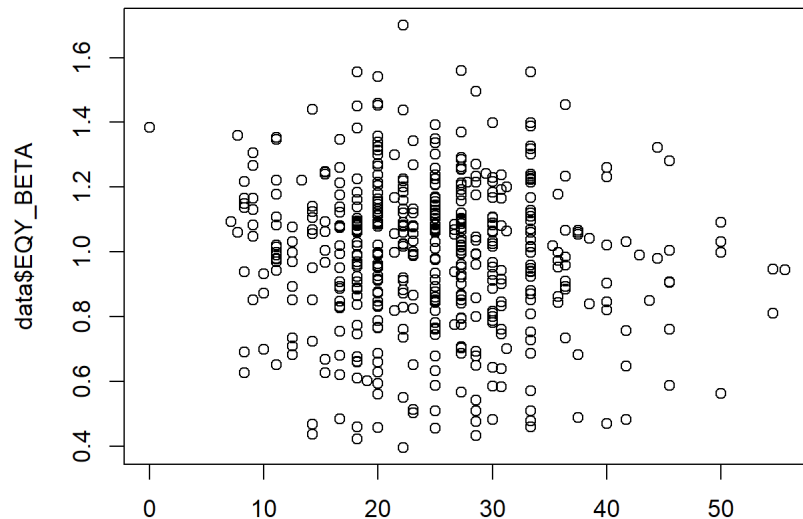
(Non)parametric Correlation Measures

Pearson's ρ

Linear relation between `PCT__WOMEN_ON_BOARD` and `EQY_BETA`?

```
cor(data$PCT_WOMEN_ON_BOARD, data$EQY_BETA, use="complete.obs")
```

```
## [1] -0.08745765
```

For confidence intervals and a significance test, use `cor.test()`.

(Non)parametric Correlation Measures

Spearman rank correlation

```
cor(data$PCT_WOMEN_ON_BOARD, data$EQY_BETA, use="complete.obs", method="spearman")
```

```
## [1] -0.08065364
```

Kendall's τ

```
cor(data$PCT_WOMEN_ON_BOARD, data$EQY_BETA, use="complete.obs", method="kendall")
```

```
## [1] -0.05475672
```

See here (Wikipedia) for comparisons between Spearman and Pearson correlations.

universität
innsbruck

Your turn

- Load the data in `sp500_data.csv` again

- Think about interesting questions and formulate hypotheses you could test with these data

- Test your hypothesis(-es) using appropriate tests

- Can you reject the H_0?

- How do you interpret your results?

- Comment on upcoming issues regarding the statistical power, multiple hyptheses, normality assumptions, etc.