WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

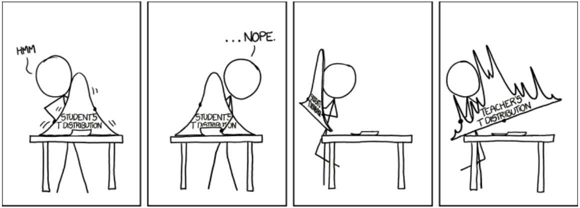
Univariate Tests

Class 3: Marketing Research

Service and Digital Marketing
March 20, 2017

Chapter objectives

- Be able to identify situations where it is useful to test hypothesis related to differences as well as to understand the differences in use of parametric and nonparametric tests
- Be able to give a short and accurate numerical and graphical summary of the data used for statistical analysis
- Be able to conduct analysis of variance and know how to interpret the results
- R: working with data frames, advanced graph annotations, add-on packages



Overview of univariate tests

Measurement level	One sample	Two samples: Independent	Two samples: Dependent
Nominal	binom. test, $\boxed{\chi^2}$, z-test	$\boxed{\chi^2}$	McNemar
Ordinal	KS-test	**U-test**	Wilcoxon
Interval (or Ratio)	T-test, Z-test	**T-test**, Z-test	T-test

Measurement level	K-samples: Independent	K-samples: Dependent
Nominal	χ^2	Cochran
Ordinal	**Kruskal**	Friedman
Interval (or Ratio)	**ANOVA**	repeated ANOVA

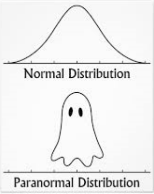
Student t-test for independent samples (1)

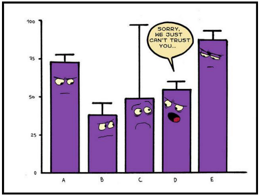
- General: to examine if there is a **difference in means** between **two independent groups** (= observations are drawn from different populations, e.g. males and females, buyers and non-buyers, ...)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \quad with\ df = n_1 + n_2 - 2$$

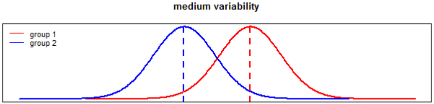
Assumptions

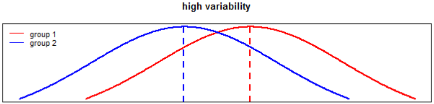
1. Interval scaled variables & two independent samples
2. Normal distributed data in each group (two sample means)
3. Equal group variances (one pooled variance)

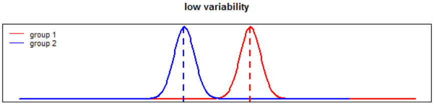




Student t-test for independent samples (2)







- t-test judges the difference between two group means relative to their variability

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{effect\ size}{noise}$$

- small differences + loads of variability $\rightarrow$ hard to detect
- large differences + low variability $\rightarrow$ easy to detect

Student t-test for independent samples (3)

Example: human weight data

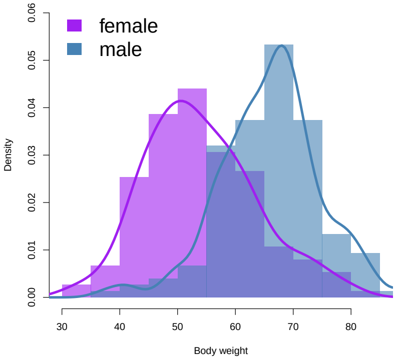
Are males and females differing significantly within their body weight?

```
# We simulate 150 random observations for two samples from
# two different normal distributions
> set.seed(1234)
> female <- rnorm(150, 55, 10)   # mean=55, sd=10
> male <- rnorm(150, 65, 8.5)    # mean=65, sd=8.5
> sex <- rep(c("female", "male"), each=150)
> wdata <- data.frame("group"=sex, "weight"=c(female,male))
> head(wdata)

# calculate mean and standard deviation for the two samples
> mean. <- tapply(wdata$weight, wdata$group, mean)
> sd. <-  tapply(wdata$weight, wdata$group, sd)
> rbind(mean., sd.)

         female      male
mean. 54.021515 66.053194
sd.    9.599219  8.824678
```

Student t-test for independent samples (4)

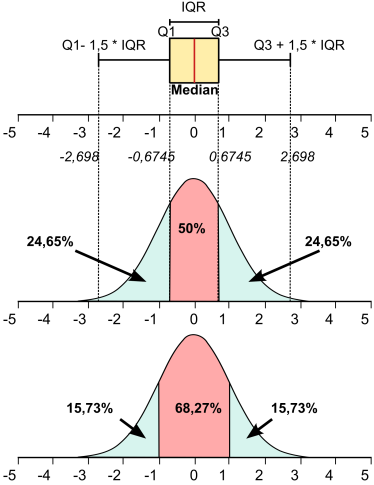
Assumption 2: Normal distribution within both groups. (✓)

```
# a more advanced histogramm:

> hist(female, col=adjustcolor("purple", 0.6),
+ main="", breaks=15, border=NA, prob=TRUE,
+ xlab="Body weight", ylim=c(0,0.06))

> hist(male, col=adjustcolor("steelblue", 0.6),
+ add=T, breaks=15, border=NA, prob=TRUE)

lines(density(female, n=150, cut=5)$x,
+ density(female, n=150, cut=5)$y, lwd=4,
+ col="purple")
lines(density(male, n=150, cut=5)$x,
+ density(male, n=150, cut=5)$y, lwd=4,
+ col="steelblue")

# add a legend to the plot
legend("topleft", legend=c("female","male"),
+ fill=c("purple","steelblue"), border=NA,
+ bty="n", cex=2)
```



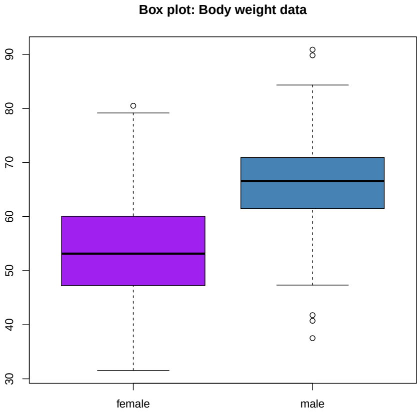
Excursus: Boxplot

- Graphical method for depicting groups of numerical data through their quartiles.
- The spacings between the different parts of the box indicate the degree of dispersion (spread) and skewness in the data, and shows outliers.
- The bottom and top of the box are always the first (25th) and third (75th) percentile, the band inside the box is always the second (50th) percentile (the median)
- The whiskers are represented as the extend of 1.5 times the difference between the 25th and the 75th percentile



Student t-test for independent samples (5)

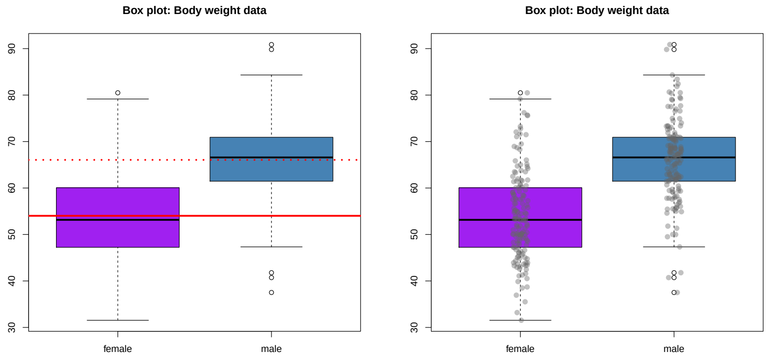
```
> boxplot(weight ~ sex, main="Boxplot: Body weight data",
+ col=c("purple","steelblue"), data=wdata)
```



Student t-test for independent samples (6)

```
# adding information 1: boxplot with data mean
> boxplot(weight ~ sex, main="Boxplot: Body weight data",
+ col=c("purple","steelblue"), data=wdata)
# mean of female group
> abline(h=tapply(wdata$weight, wdata$group, mean)[1], cex=3,
+ lwd=3, col="red")
# mean of male group
> abline(h=tapply(wdata$weight, wdata$group, mean)[2], cex=3,
+ lwd=3, col="red", lty="dotted")

# adding information 2: boxplot with data points
> boxplot(weight ~ sex, main="Boxplot: Body weight data",
+ col=c("purple","steelblue"), data=wdata)
> stripchart(weight ~ sex, vertical = TRUE, data = wdata,
+ method = "jitter", jitter=0.05, add = TRUE, pch = 16,
+ col = adjustcolor("grey40",0.4), cex=1.25)
```

Student t-test for independent samples (7)



Student t-test for independent samples (8)

Assumption 3: Homogeneity of variances.

1. H_0: **both samples have equal variances**. (✓)
2. H_1 : the variances differ.

```
> mean. <- tapply(wdata$weight, wdata$group, mean)
> sd. <-  tapply(wdata$weight, wdata$group, sd)
> rbind(mean., sd.)

         female      male
mean. 54.021515 66.053194
sd.    9.599219  8.824678

# variance homogeneity can be assessed using the "Bartlett-test"
# (is acting like a "model check", so we want to stay with H0)
> bartlett.test(wdata$weight ~ wdata$group)

        Bartlett test of homogeneity of variances

data:  wdata$weight by wdata$group
Bartlett's K-squared = 1.0498, df = 1, p-value = 0.3055
```

Student t-test for independent samples (9)

Independent samples Student t-test:

1. H_0: the means of both samples are equal
   (there is no difference between the means of the two samples)
2. H_1: the means of both samples differ

```
> diff. <- mean.[1] - mean.[2]
# female sample weight is lower than the male sample weight (negative sign)
> n1 <- 150
> n2 <- 150
# we use the pooled "sample" variance, i.e., 1/(n-1) instead of 1/n
> var. <- sqrt((1/n1 + 1/n2) * ((n1-1)*sd.[1]^2 + (n2-1)*sd.[2]^2)/(n1+n2-2))
# calculate the ratio of difference to the variance = test statistic
> t. <- diff./var.
# calculate the degrees of freedom
> df. <- 150 + 150 - 2
# compare the test statistic to the Student-t distribution with df=298
> pt(t., df.)
0

# summarize the results
> cbind(diff., var., t., df.)
     diff.      var.         t. df.
 -12.03168 1.064644 -11.30113 298
```

Student t-test for independent samples (10)

```
# compare your result to the built-in function t.test
# use the argument "var.equal=TRUE" as the Bartlett-test states:
# H0: the group variances are equal

> t.test(wdata$weight ~ wdata$group, var.equal = TRUE)

        Two Sample t-test

data:  wdata$weight by wdata$group
t = -11.301, df = 298, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -14.126852  -9.936506
sample estimates:
mean in group female   mean in group male
            54.02152             66.05319

# H1: the weights of female and male persons differ significantly
# (female weight less than men)
```

Mann-Whitney U (Wilcoxon rank-sum) test (1)

Alternative (**non parametrical**) test, if assumption 2 (normal distribution within groups) and/or assumption 3 (equal variances) do **not** hold.

- compares the difference in the location (**mean ranks**) of **two independent samples**
- has greater efficiency than the t-test on non-normal distributions
- particularly suitable for small sample sizes ($n_1,\ n_2 < 10$)

Example: Rankings of 12 body weight observations

weight	47	48	45	60	45	55	67	61	55	68	60	63
group	F	F	F	F	F	F	M	M	M	M	M	M
Rank	3	4	1.5	7.5	1.5	5.5	11	9	5.5	12	7.5	10

$$Mann - Whitney\ U = n_1 n_2 + \frac{n_2(n_2+1)}{2} + \sum_{i=n_2+1}^{n_2} - R_i$$

with n_1 the size of the group with the smaller rank-sum, n_2 the size of the group with the higher rank-sum, and R_i the higher rank-sum of the two groups.

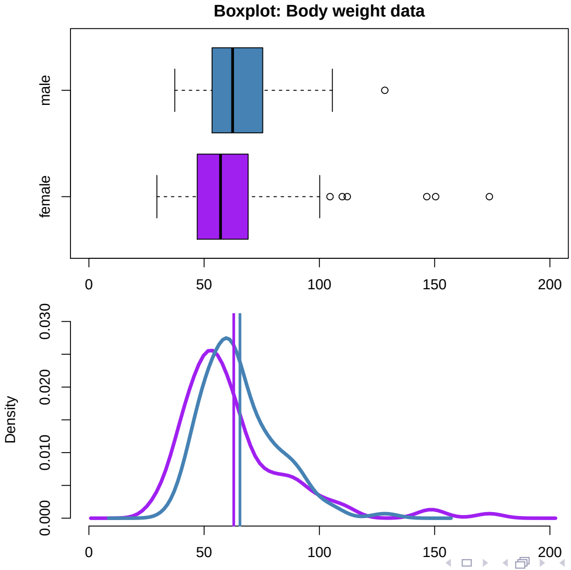
Mann-Whitney U (Wilcoxon rank-sum) test (2)

Example: skewed data, unequal variances

Assumption 2 (normal distribution within both groups) does not hold.

```
# simulate skewed data (non-central Student-t distribution)
> set.seed(1234)
> female <- rt(100, 5, ncp=55)
> male <- rt(100, 10, ncp=60)
> sex <- rep(c("female", "male"), each=100)
> wdata2 <- data.frame("group"=sex, "weight"=c(female,male))
> mean. <- tapply(wdata2$weight, wdata2$group, mean)
> median. <- tapply(wdata2$weight, wdata2$group, median)
> sd. <-  tapply(wdata2$weight, wdata2$group, sd)
> IQR. <-  tapply(wdata2$weight, wdata2$group, IQR)
> rbind(mean., median., sd., IQR.)

          female     male              female     male
mean.   62.84184 65.53958     sd.    24.56409 16.27775
median. 57.11752 62.35771     IQR.   21.46278 21.49835
```

Mann-Whitney U (Wilcoxon rank-sum) test (3)



Mann-Whitney U (Wilcoxon rank-sum) test (4)

```
> wilcox.test(wdata2$weight ~ wdata2$group) # wilcoxon rank-sum test

        Wilcoxon rank sum test with continuity correction

data:  wdata2$weight by wdata2$group
W = 4042, p-value = 0.01931
alternative hypothesis: true location shift is not equal to 0

> t.test(wdata2$weight ~ wdata2$group, var.equal=TRUE)  # Student t-test
> t.test(wdata2$weight ~ wdata2$group, var.equal=FALSE) # Welch t-test

        Welch Two Sample t-test

data:  wdata2$weight by wdata2$group
t = -0.91548, df = 171.89, p-value = 0.3612
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.514307  3.118820
sample estimates:
mean in group female   mean in group male
            62.84184             65.53958
```

Analysis of Variance (1)

- when comparing more than two groups, Student t-test becomes inefficient as one has to conduct tests on all possible group comparisons, i.e., $\binom{K}{2}$, with K the number of groups.
- **F-test** in Analysis of Variance (ANOVA) simultaneously compares all group-means (**omnibus-test**).
- The F-value is the proportion of between groups and within groups variance,

$$SS_{total} = SS_{between} + SS_{within}$$

  and shows the effect of the factor (c.f. appendix for the variance decomposition procedure)

$$F = \frac{SS_{Between}/(K-1)}{SS_{Within}/(N-K)} = \frac{MS_{between}}{MS_{within}}$$

  with $df_1 = K - 1$ and $df_2 = N - K$
- If the test is significant, we can conclude that **at least one** of the group-means differs from the other group-means.

Variance decomposition (c.f. Appendix)

	df	SS	MS	F
between	$K-1$	$\sum_{j=1}^{J}\sum_{i=1}^{K}(\bar{Y}_{i.}-\bar{Y}_{..})^2$	$MS_{betw} = \frac{SS_{betw}}{df_{betw}}$	$\frac{MS_{between}}{MS_{within}}$
within	$N-K$	$\sum_{i=1}^{K}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{i.})^2$	$MS_{within} = \frac{SS_{within}}{df_{within}}$	
total	$N-1$	$\sum_{i=1}^{K}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{..})^2$	$MS_{error} = \frac{SS_{total}}{df_{total}}$	

Analysis of Variance (2)

Overview of ANOVA methods:

- **ANOVA as test of difference** (in group means): **one-way ANOVA**
- **ANOVA for complete factorial designs**: analysis of variance with more than one independent nominal scaled variable (factor): **n-way ANOVA**
- **Analysis of covariance**: analysis of variance with an additional continuous independent variable (covariate): **ANCOVA**
- **ANOVA using repeated measures**: analysis of variance for several measurement points per observation (related samples)
- **Multivariate ANOVA**: analysis of variance for several dependent variables: **MANOVA**

One-way ANOVA (1)

- General: to find out what is the effect of one ore more discrete (i.e., grouping) variables on an (at least) interval-scaled variable.
- when the number of independent samples > 2
- the grouping variable is now called **factor** and the different values of the factor are referred to as **levels** (i.e., groups = factor levels)

Assumptions

1. One interval scaled variable & more than two independent samples (i.e. factor levels of the independent variable)
2. Normal distribution within each group (= normally distributed errors)
3. Equal error variances across all factor levels

Note: If there are only two groups in one-way ANOVA, then $F = t^2$ of the t-test.

One-way ANOVA (2)

Example: human weight data (cont.)

H_0: all category means are equal.
H_1: the category mean differs.

```
> summary(aov(wdata$weight ~ wdata$group))

Analysis of Variance Table
Response: wdata$weight
              Df Sum Sq Mean Sq F value    Pr(>F)
wdata$group    1  10857   10857  127.72 < 2.2e-16 ***
Residuals    298  25333      85
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- the significant F-test result rejects the **global** H_0, that the means are the same across the groups being compared (Note: t was -11.30113, F is 127.72)

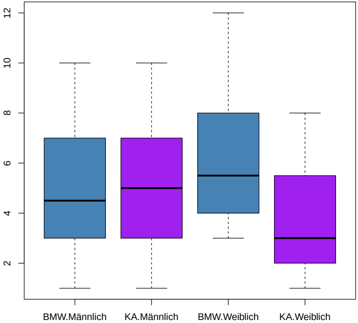
```
# check
> t.^2
127.7156
```

One-way ANOVA (3)

Example: road traffic aggression (`agg.Rdata`)

At the intersection of two roads a car is blocking the free side lane while the traffic light is green. Researchers recorded the honking behavior (frequency, duration till the first honk, overall duration of honking) for several days using two different "blocking" car types (Ford KA, BWM X5). In addition the gender of the person that has been blocked was recorded.

```
> load("agg.RData")
# factor levels
# (= the values of the nominal variable are labeled)
> head(agg$Auto)
BMW BMW BMW BMW BMW BMW
Levels: BMW KA
> head(agg$Geschlecht)
Männlich Männlich Männlich Männlich Männlich Männlich
Levels: Männlich Weiblich

> agg$g.new <- paste(agg$Auto, agg$Geschlecht, sep=".")
> tapply(agg$dauer, agg$g.new,
+ mean, na.rm=T)
BMW.Männlich BMW.Weiblich  KA.Männlich  KA.Weiblich
    5.131579     6.214286     5.297297     3.800000

> boxplot(agg$dauer ~ agg$g.new, col=c("steelblue","purple"))
```



One-way ANOVA (4)

Assumption 3: Equal error variances across factor levels.

1. H_0**: error variances are equal**. (✓)
2. H_1 : error variances differ.

```
> bartlett.test(agg$dauer ~ agg$g.new)

        Bartlett test of homogeneity of variances

data:  agg$dauer by agg$g.new
Bartlett's K-squared = 3.2684, df = 3, p-value = 0.3521
```

One-way ANOVA (5)

one-way ANOVA

H_0: there are no differences in the honking duration across groups.
H_1: there are differences in the honking duration across groups.

```
> summary(aov(agg$dauer ~ agg$g.new))
Analysis of Variance Table

Response: agg$dauer
           Df Sum Sq Mean Sq F value  Pr(>F)
agg$g.new   3  52.63 17.5426  2.6632 0.05178 .
Residuals 105 691.63  6.5869
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- the F-test result is not significant!
  (There are no differences in the honking duration across groups).

n-way ANOVA (1)

Full factorial design

When the experiment includes observations at all combinations of levels of each factor, it is termed **full factorial**. Factorial experiments are more efficient than a series of single factor experiments.

Fractional factorial design

Full factorial experiments can require many runs. The solution to this problem is to use only a **fraction** of the runs specified by the full factorial design.

- to test the effect of different levels of several factors on one interval scaled variable
- to test if combinations of the independent variable may have a significant influence on the dependent variable → **interaction effects** are possible
  (= the effect of a certain variable will depend on the level of another variable)

n-way ANOVA (2)

Several hypothesis are tested simultaneously:

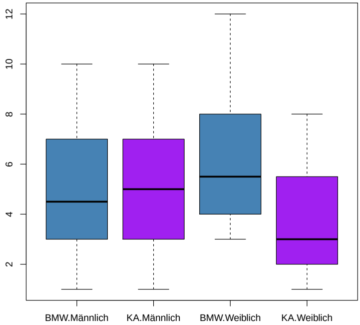
- H_{1_a} and H_{1_b} are for testing **main effects**
  (i.e., that population means of the outcome are different for at least one level of the factor ignoring the other factor)
- H_{1_c} is for testing the **interaction effect**
  (i.e., combinations of factor levels of the main effects do produce different population means)

Example: road traffic aggression (cont.) (`agg.Rdata`)

H_{1_a}: there are differences in honking duration between car types
H_{1_b}: there are differences in honking duration between the gender
H_{1_c}: there exists an interaction between car type and gender in honking duration

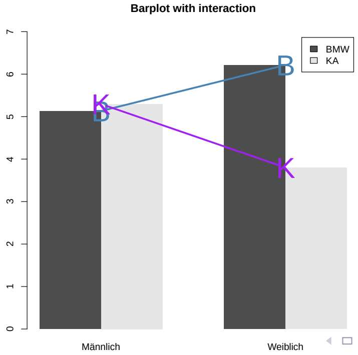
n-way ANOVA (3)

```
# we are now investigating column
# and row means
> tapply(agg$dauer, list(agg$Auto,
+ agg$Geschlecht), mean, na.rm=T)
    Männlich Weiblich
BMW 5.131579 6.214286
KA  5.297297 3.800000

> boxplot(agg$dauer ~
+ agg$Auto + agg$Geschlecht,
+ col=c("steelblue","purple"))
```

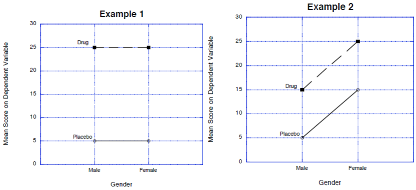


n-way ANOVA (4)

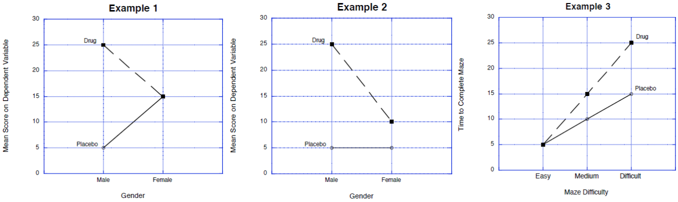
```
# interaction illustration: barplot using the group means
means <- tapply(agg$dauer, list(agg$Auto,agg$Geschlecht), mean, na.rm=T)
df.bar <- barplot(means, beside=T, border=NA, legend=T, ylim=c(0,7), main="")
# define coordinates for the lines to be added
df.bar[,1] <- 2; df.bar[,2] <- 5
points(df.bar[1,], means[1,], type="b", cex=3, lwd=3, col="steelblue", pch="B")
points(df.bar[2,], means[2,], type="b", cex=3, lwd=3, col="purple", pch="K")
```



- No interaction present (main effects only)



- Interaction present



n-way ANOVA (5)

Assumption 3: Equal error variances across factor levels.

1. H_0: **error variances are equal across all samples (= factor combinations).** (✓)
2. H_1: error variances differ.

```
# load the add-on package "car" first
library(car)

# if the package is not already installed
# install.packages("car")
# and load the add-on package afterwards

# the "Companion to Applied Regression" entails the Levene-test:
leveneTest(agg$dauer ~ agg$Auto * agg$Geschlecht)

Levene's Test for Homogeneity of Variance (center = median)
       Df F value Pr(>F)
group   3  1.5443 0.2074
      105
```

n-way ANOVA (6)

n-way Anova:

H_{1_a}: there are differences in honking duration between car types
H_{1_b}: there are differences in honking duration between the gender
H_{1_c}: there exists an interaction between car type and gender in honking duration

```
> summary(aov(agg$dauer ~ agg$Auto * agg$Geschlecht))

                         Df Sum Sq Mean Sq F value  Pr(>F)
agg$Auto                  1  11.53  11.529  1.7504 0.18871
agg$Geschlecht            1   3.01   3.012  0.4573 0.50039
agg$Auto:agg$Geschlecht   1  38.09  38.086  5.7821 0.01794 *
Residuals               105 691.63   6.587
```

Gender in combination with car type does have an influence on the honking duration (as seen in the graphs: women are honking significantly longer at the BMW)

Note: interaction effect **is significant**: the corresponding main effects are not allowed to be interpreted any longer (= some changes in both explanatory variables must have an effect on the outcome, regardless of the main effect p-values).

Summary

Parametric	Non-parametric alternative
one sample t-test	**K-S test**
interval (ordinal) scaled response normally distributed response	rank (ordinal) scaled response
independent (two sample) t-test	**Mann & Whitney U-test**
two independent samples interval (ordinal) scaled responses normally distributed response (each group) equal group variances	two independent samples rank (ordinal) scaled responses
one-way ANOVA	**Kruskal-Wallis test**
two or more independent samples interval (ordinal) scaled responses normally distributed response (each group) equal group variances	two or more independent samples rank (ordinal) scaled responses

See learning platform!

Submission deadline: 26. 03. at 23:00 pm

(via the learning platform `www.learn.wu.ac.at`)

Oral presentation of solutions on Monday!

(Recap: random selection of four students to present their solution).

Degrees of freedom

Recap: degrees of freedom

Describes the number of values in the final calculation of a statistic that are free to vary. In general each parameter being estimated costs a degree of freedom.

- **Independent t-Test**: Two group means are compared:
  df = N (total sample size) – K (number of group means estimated)
  $\rightarrow$ df = n1 + n2 - 2

- **Anaylsis of variance**:
  Error variance: K-group means are estimated:
  df = N (total sample size) – K (number of groups means estimated)

  Explained variance: estimation of variability across k group means:
  df = K (group means) – 1 (variance across group means)

Variance decomposition (1)

One-way analysis of variance

$$SS_{total} = SS_{between} + SS_{within}$$

Group 1	Group 2	Group 3
y_{11}	y_{21}	y_{31}
y_{12}	y_{22}	y_{32}
y_{13}	y_{23}	y_{33}
y_{14}	y_{24}	y_{34}
$y_{(i=1)5}$	$y_{(i=2)5}$	$y_{(i=3)5}$
$y_{(i=1)(j=6)}$	$y_{(i=2)(j=6)}$	$y_{(i=3)(j=6)}$

between variation = K

within variation = J

$$\sum_{i=1}^{K}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{..})^2 = \sum_{j=1}^{J}\sum_{i=1}^{K}(\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^{K}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i.})^2$$

Variance decomposition (2)

	Group 1	Group 2	Group 3
	643	469	484
	655	427	456
	702	525	402
$\bar{Y}_{i.} =$	666.67	473.67	447.33

$$\bar{Y}_{..} = \frac{643 + 655 + 702 + 469 + ... + 402}{9} = 529.22$$

$$\begin{aligned} SS_{total} &= \sum_{i=1}^{K} \sum_{j=1}^{J} (Y_{ij} - \bar{Y}_{..})^2 \\ &= (643 - 529.22)^2 + (655 - 529.22)^2 + \ldots + (402 - 529.22)^2 \\ &= 9603.55 \end{aligned}$$

Variance decomposition (3)

$$SS_{between} = \sum_{j=1}^{J}\sum_{i=1}^{K}(\bar{Y}_{i.} - \bar{Y}_{..})^2$$
$$= 3 \times ((666.67 - 529.22)^2 + (473.67 - 529.22)^2 + (447.33 - 529.22)^2)$$
$$= 86049.55$$

$$SS_{within} = \sum_{i=1}^{K}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i.})^2$$
$$= (643 - 666.67)^2 + (655 - 666.67)^2 + (702 - 666.67)^2 + (469 - 473)^2$$
$$+ (427 - 473.67)^2 + (535 - 473.67)^2 + (484 - 447.33)^2 \ldots$$
$$= 10254$$