C | A | U
Christian-Albrechts-Universität zu Kiel

09_parametrische_tests

F-Test, t-Test und Anova

Wiederholung: Grundgesamtheit und Stichprobe

Eigenschaften der Grundgesamtheit: Parameter

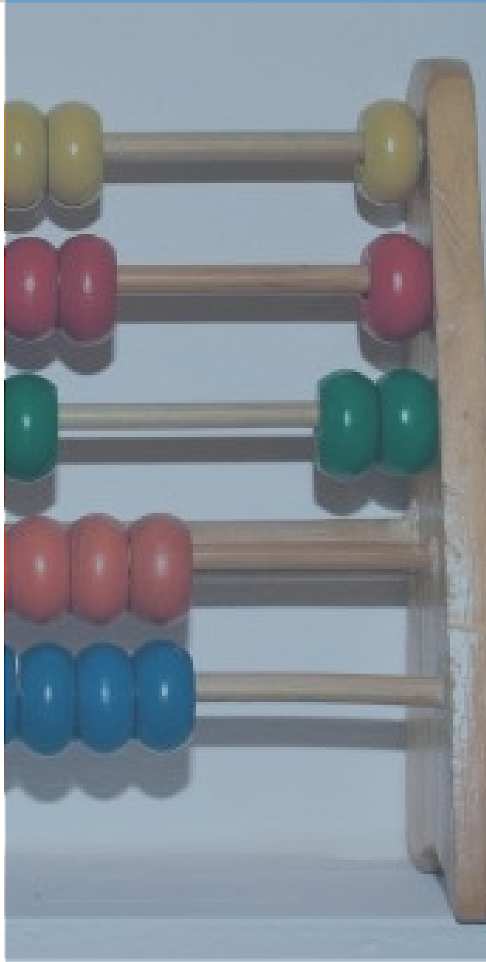
Parameter gibt es immer, sie sind feste Werte, aber sie sind unbekannt, meist auch nicht überprüfbar

Bsp:

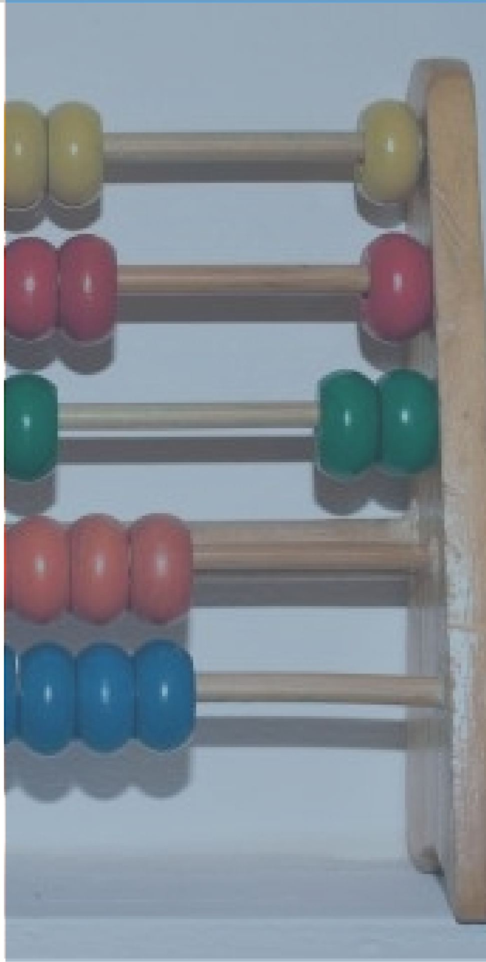
$$\mu : \textit{Arithm. Mittel der Grundgesamtheit}$$
$$\bar{x} : \textit{Arithm. Mittel der Stichprobe}$$

$$\sigma : \textit{Standardabweichung der Grundgesamtheit}$$
$$s : \textit{Standardabweichung der Stichprobe}$$

In statistischen Tests können immer nur die Eigenschaften der Stichprobe geprüft werden. Daher hängt die Qualität der Aussage immer von der Wahl der Stichprobe ab (Repräsentativität)



Parametrische Tests



parametrisch vs. nicht-parametrisch/parameterfrei:

parametrisch: Werte müssen bestimmter Verteilung folgen (z. B. Normalverteilung); Grundannahmen zur Verteilung sind notwendig

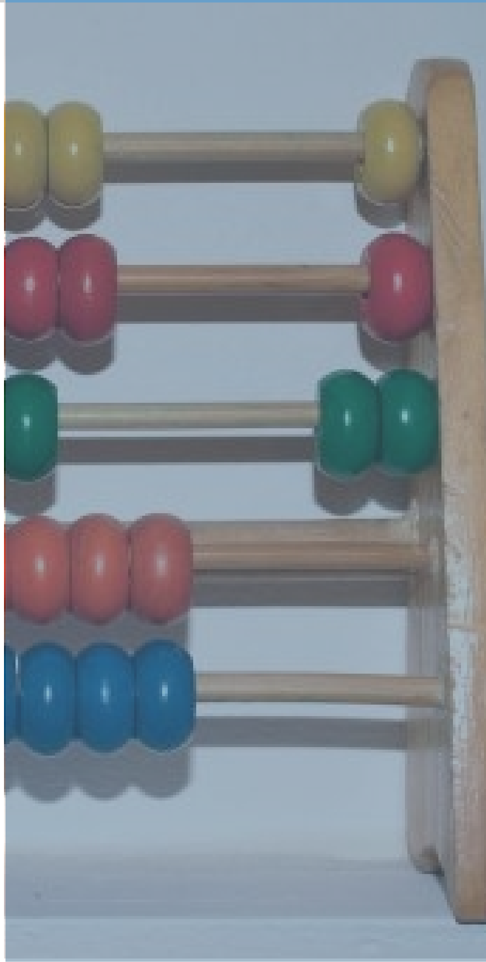
nicht-parametrisch: Annahmen zur Werteverteilung entfallen; keine Grundannahmen notwendig

Parametrische Tests, Vorteile - Nachteile:

Vorteil: Haben meist eine größere Power (Teststärke).

Nachteil: Sind nicht anwendbar, wenn keine Aussage über die Verteilung möglich ist oder die Verteilung nicht den für parametrische Tests gegebenen Anforderungen entspricht.
Es können meist keine relativ kleine Stichproben getestet werden.

Mögliche Voraussetzungen für parametrische Tests



Bestimmte Verteilung
Die Daten müssen bestimmten Verteilungen entsprechen, also aus Phänomenen bestimmter Art stammen.

Beispiel: t-Test: Normalverteilung

Bestimmte Ähnlichkeiten
Die Daten müssen einander in bestimmten Parametern gleichen.

Beispiel: Anova: Varianzhomogenität

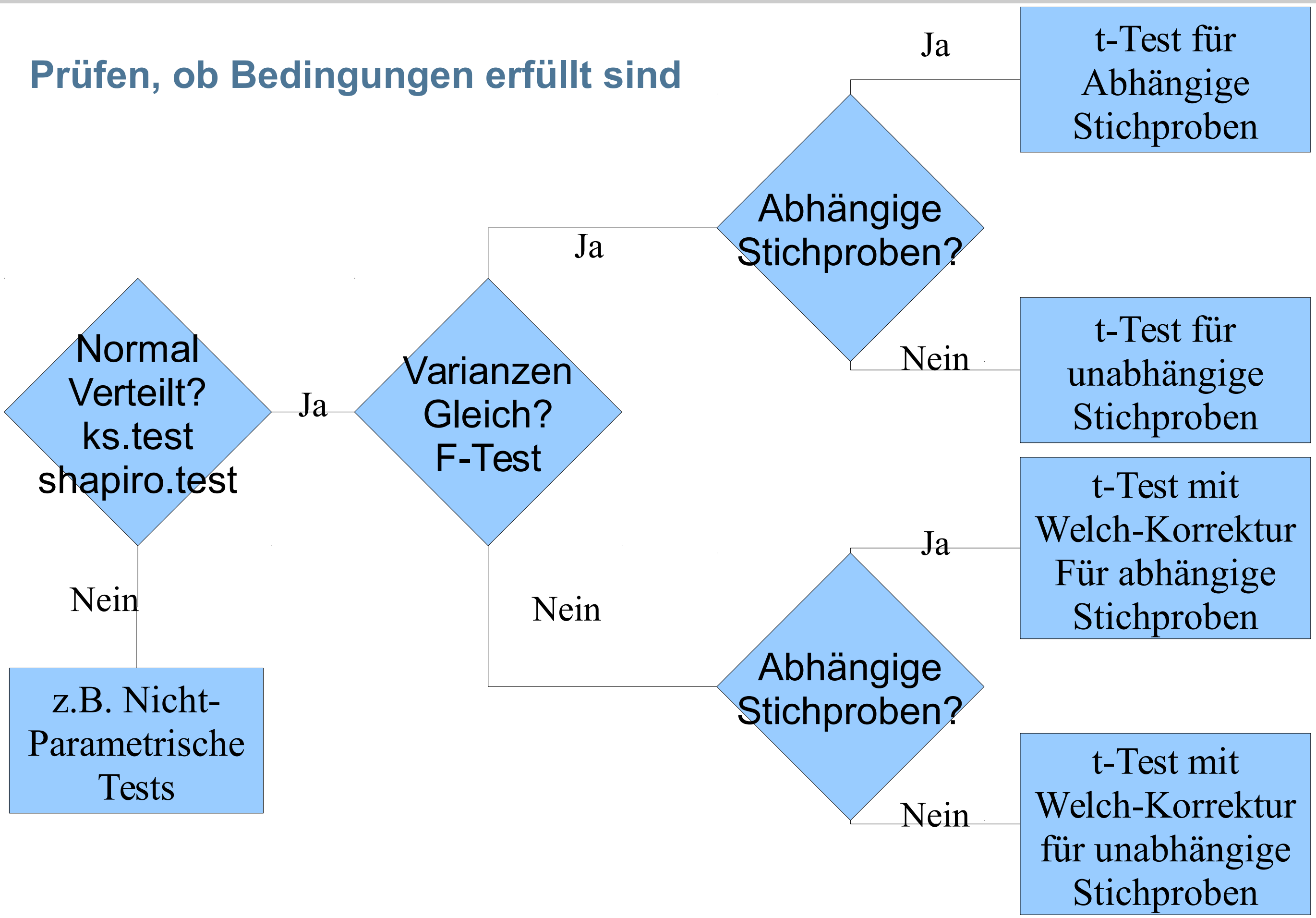
Bestimmte Skalierungen
Fast immer muss mind. 1 Variable intervallskaliert oder höher sein.

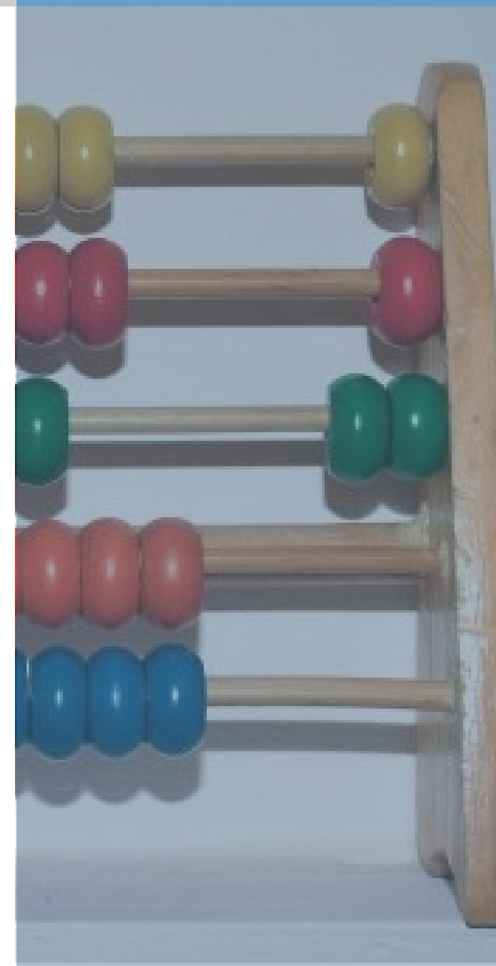
Beispiel F-Test: Test für Varianzen, daher mind. Intervalskaliert (abweichung vom arith. Mittel, dieses nur bei Intervall oder höher skalierten Variablen zu berechnen)

Christian-Albrechts-Universität zu Kiel

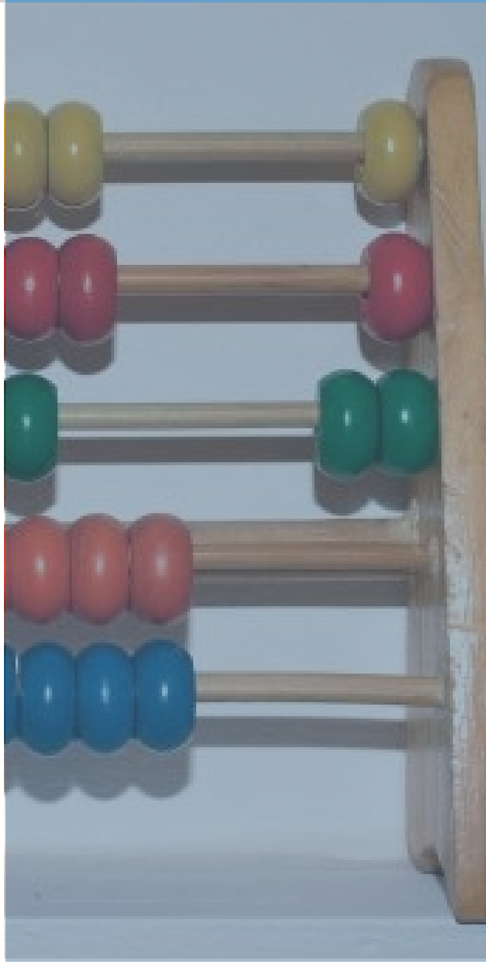
Möglicher Testbaum

Prüfen, ob Bedingungen erfüllt sind





Prüfen auf Normalverteilung ks.test



Der gute alte Kolmogorv-Smirnov

Möglich, wenn Parameter der Grundgesamtheit bekannt (μ, σ) sind oder approximiert werden können.

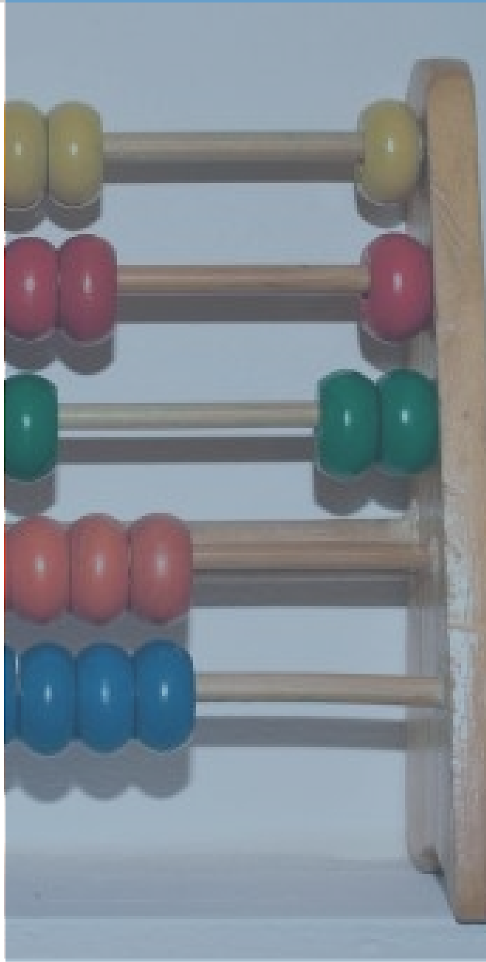
Bsp. Längen von Silex-Klingen (simuliert) normalverteilt?

```
klingenlaenge<-c(14.9, 24.0, 8.7, 29.3, 25.5, 23.9, 22.4, 12.7,
8.7, 25.1, 25.6, 14.7, 23.0, 23.2, 26.5, 11.1, 15.2, 20.6, 20.1,
25.1)

ks.test(klingenlaenge,"pnorm",mean(klingenlaenge),sd(klingenlaenge))

     One-sample Kolmogorov-Smirnov test

data:  klingenlaenge
D = 0.1966, p-value = 0.4221
alternative hypothesis: two-sided
```

Ergebnis nicht signifikant, Verteilung weicht nicht signifikant von Normalverteilung ab.

Aber: K-S-Test ist sehr konservativ, null-hypothese (gleichverteilung) wird nur abgelehnt, wenn sehr deutliche Unterschiede existieren.

Prüfen auf Normalverteilung shapiro.test



Der bessere Test auf Normalverteilung

Voraussetzungen: $X_1, \ldots, X_n$ ist unabhängige Stichprobe metrisch einer skalierter Variabe
H_0: die Grundgesamtheit ist normalverteilt
H_1: die Grundgesamtheit ist nicht normalverteilt

```
> shapiro.test(klingenlaenge)

    Shapiro-Wilk normality test

data:  klingenlaenge
W = 0.902, p-value = 0.04494
```

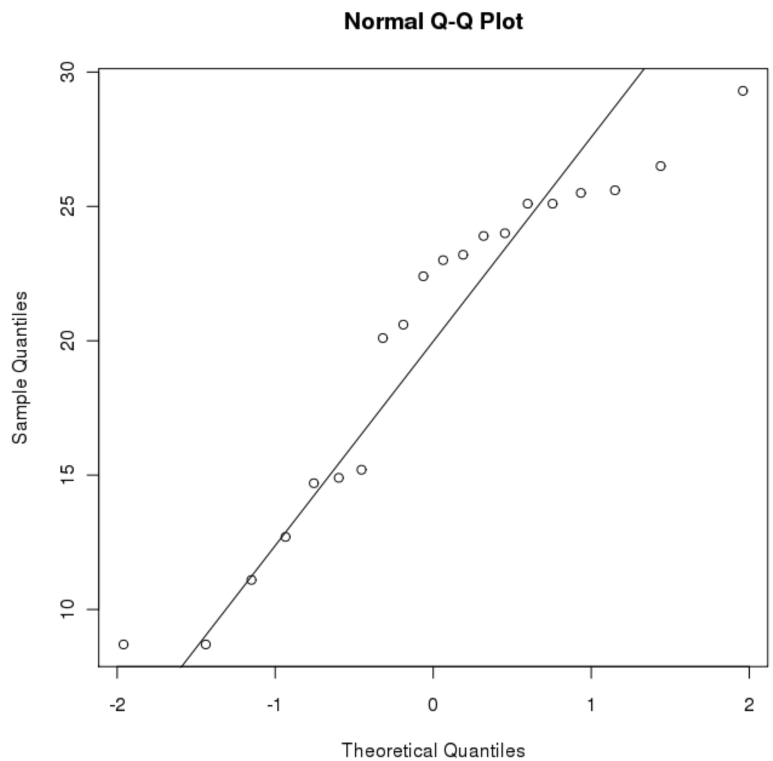
Aber: t-Test ist relativ robust, eine Nicht-Ablehnung durch den ks.test reicht eigentlich aus (bei größeren n-Werten).

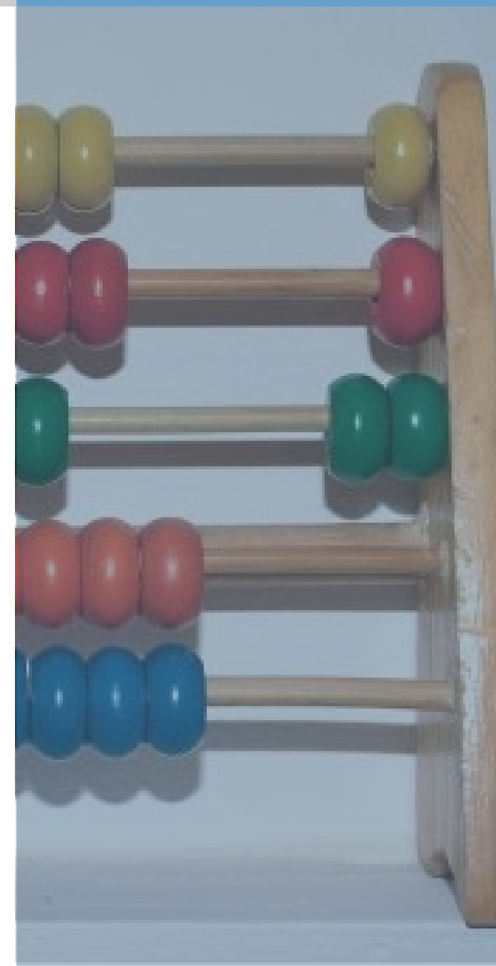
Prüfen auf Normalverteilung QQ-Plot

Graphische Kontrolle auf Normalverteilung

Abtrag der Quantile der Verteilung gegen die Quantile einer Normalverteilung

```
> qqnorm(klingenlaenge)
> qqline(klingenlaenge)
```



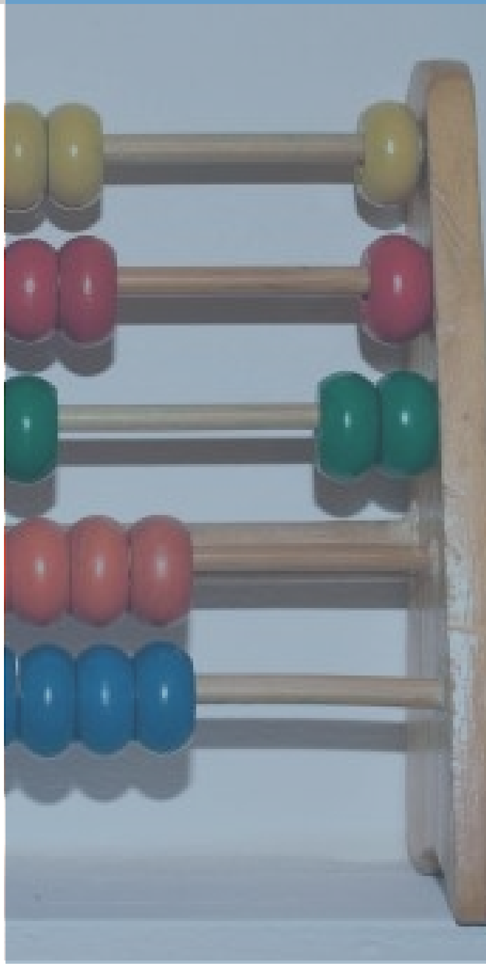


Aufgabe Prüfen auf Normalverteilung

Henkellängen von Amphoren der Dresseltypen 10 (Ihm 1978)

Gegeben sind die Henkellängen verschiedener Amphoren. Stellen Sie mit allen Ihnen zur Verfügung stehenden Mitteln fest, ob diese normalverteilt sind.

File: henkel_amphoren.csv



Lösung Prüfen auf Normalverteilung

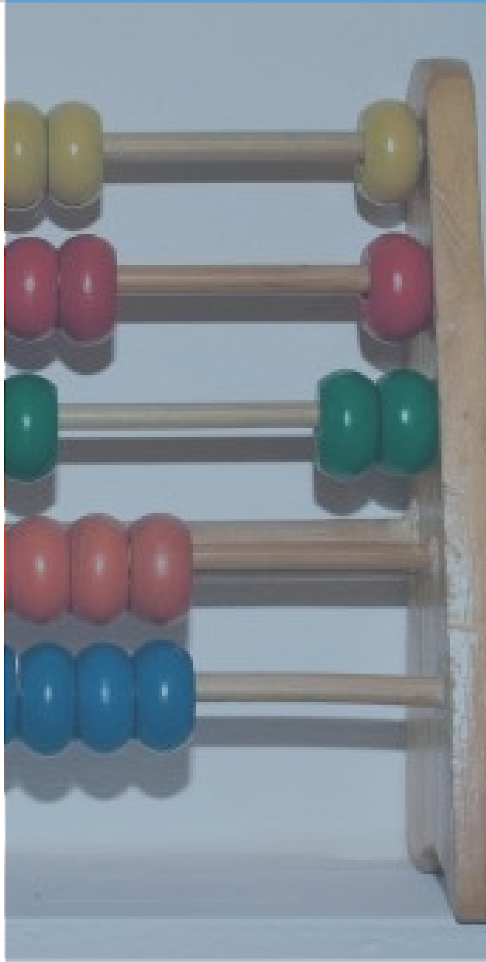
Henkellängen von Amphoren der Dresseltypen 10 (Ihm 1978)

Gegeben sind die Henkellängen verschiedener Amphoren. Stellen Sie mit allen Ihnen zur Verfügung stehenden Mitteln fest, ob diese normalverteilt sind.

File: henkel_amphoren.csv

```
> henkel<-read.csv2("henkel_amphoren.csv",row.names=1)
> shapiro.test(henkel$henkellaenge)
…
W = 0.8849, p-value = 0.02174
>
ks.test(henkel$henkellaenge,"pnorm",mean(henkel$henkellaenge),sd(hen
kel$henkellaenge))
…
D = 0.1542, p-value = 0.7286
…
> qqnorm(henkel$henkellaenge)
> qqline(henkel$henkellaenge)
```

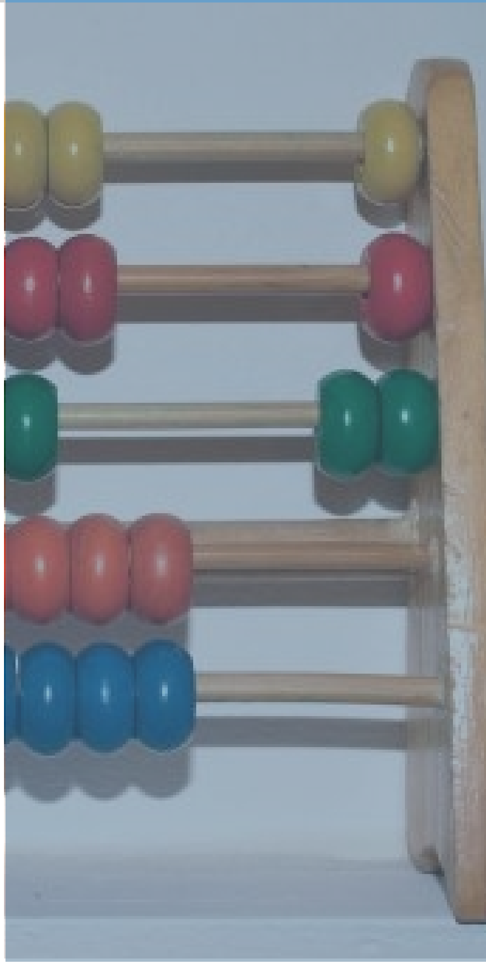
Nach ks.test normal, nach shapiro nicht. QQ zeigt zwei Ausreißer an den Enden.




Thea Dorn
Die neue F-Klasse
Wie die Zukunft von Frauen gemacht wird
SERIE PIPER

F-Test

F-Test [1]



Test auf Varianzgleichheit zweier Stichproben

Voraussetzungen: zwei unabhängig normalverteilte Stichproben
H_0: Beide Stichproben haben die gleiche Varianz (Streuung).
H_1: Eine Stichprobe hat eine größere Varianz (Streuung).

Idee: Wenn Varianzen gleich sind, müßte ihr Quotient 1 betragen.

$$s_1^2 = s_2^2 ; dann \frac{s_1^2}{s_2^2} = 1$$

Der Quotient wird mit nach Anzahl von Freiheitsgraden ($df_1 = n_1 - 1$, $df_2 = n_2 - 1$) und gewünschtem Signifikanzlevel mit einer Tabelle F-verteilter Werte verglichen (z.B. Shennan). Wenn der errechnete Quotient > Grenzwert, dann muß die H_0 verworfen werden, sonst ist sie beizubehalten.

Signifikant: ungleiche Varianzen
Nicht signifikant: es kann von gleichen Varianzen ausgegangen werden

F-Test [2]

Beispiel: Klingenlänge zweier Fundstellen, Berechnung von Hand

$$n\ Fundstelle\ 1 = 20\ ;\ \bar{x}\ Fundstelle\ 1 = 20.015$$

$$Varianz\ Fundstelle\ 1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{20}(x_i - 20.015)^2}{20-1} = 40.20871$$

$$n\ Fundstelle\ 2 = 25\ ;\ \bar{x}\ Fundstelle\ 2 = 20.492$$

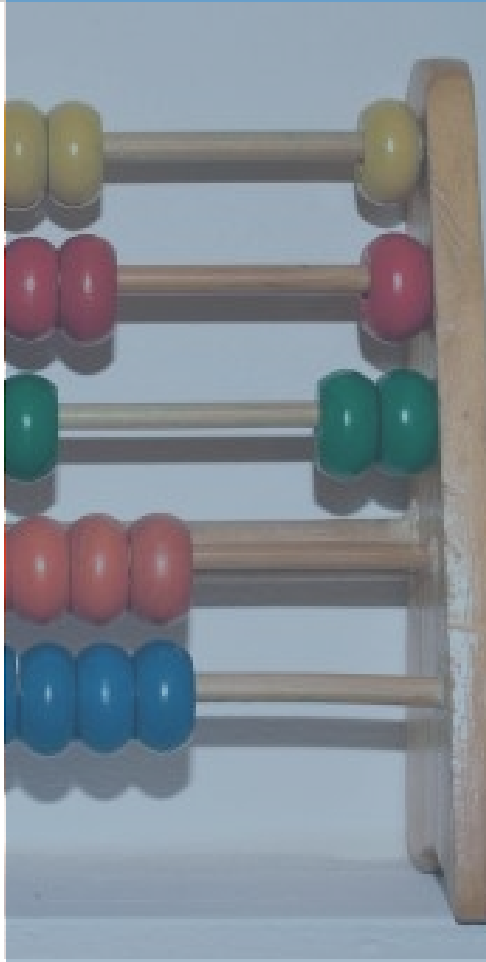
$$Varianz\ Fundstelle\ 1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{20}(x_i - 20{,}492)^2}{20-1} = 33.0641$$

$$F = \frac{s_1^2}{s_2^2} = \frac{40.20871}{33.0641} = 1{,}216083607$$

$$df_1 = 20 - 1 = 19,\ df_2 = 25 - 1 = 24\ ;\ Sign.level = 0.05$$

$$Grenzwert\ bei\ df_1 = 19,\ df_2 = 24,\ \alpha = 0.05 : 2.114$$

$$1{,}216083607 < 2.114,\ nicht\ signifikant\ ,\ die\ Varianzen\ unterscheiden\ sich\ nicht\ signifikant\ voneinander$$

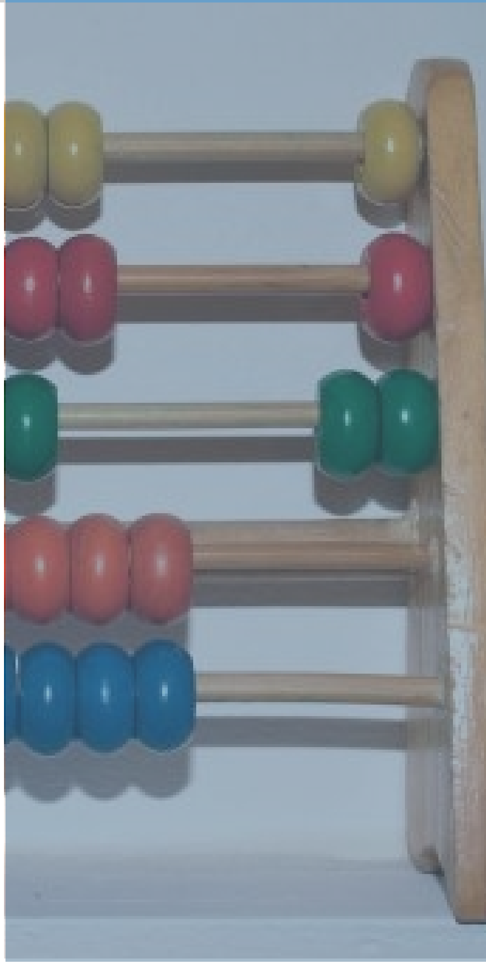


F-Test [3]

F-Test in R

```
> klingenlaenge <- read.csv2("klingenlaengen.csv",row.names=1)
> var.test(laenge~fundort,data=klingenlaenge)

     F test to compare two variances

data:  laenge by fundort
F = 1.2161, num df = 19, denom df = 24, p-value = 0.643
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5185518 2.9822271
sample estimates:
ratio of variances
          1.216084
```

Ergebnis nicht signifikant, Varianzen unterscheiden sich nicht signifikant voneinander

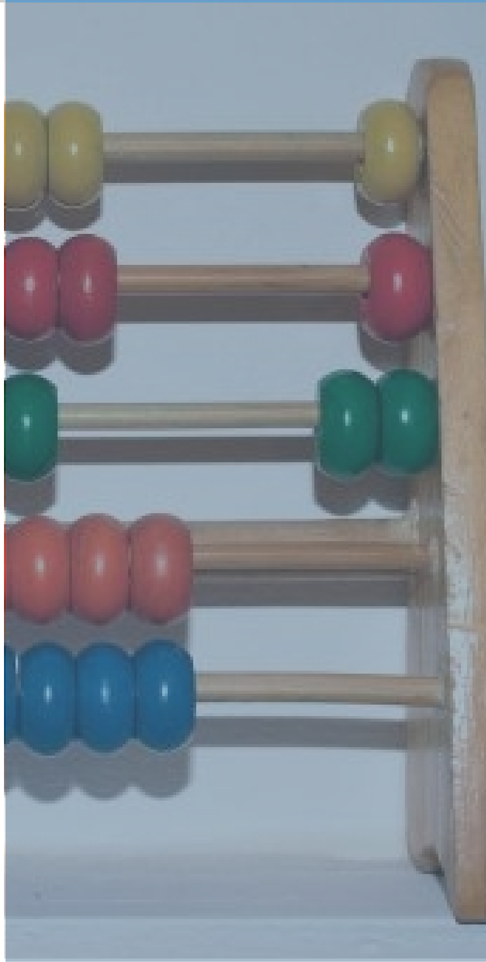


Aufgabe F-Test

(logarithmierte) Größe von zeremoniellen Erdwerken auf den Gesellschaftsinseln (Beispiel nach Shennan)

Gegeben sind die (logarithmierten) Größen von zeremoniellen Erdwerken in zwei verschiedenen Tälern auf den Gesellschaftsinseln.
Stellen Sie fest, ob die Varianzen in beiden Tälern als gleich anzusehen sind!

File: marae.csv



Aufgabe F-Test

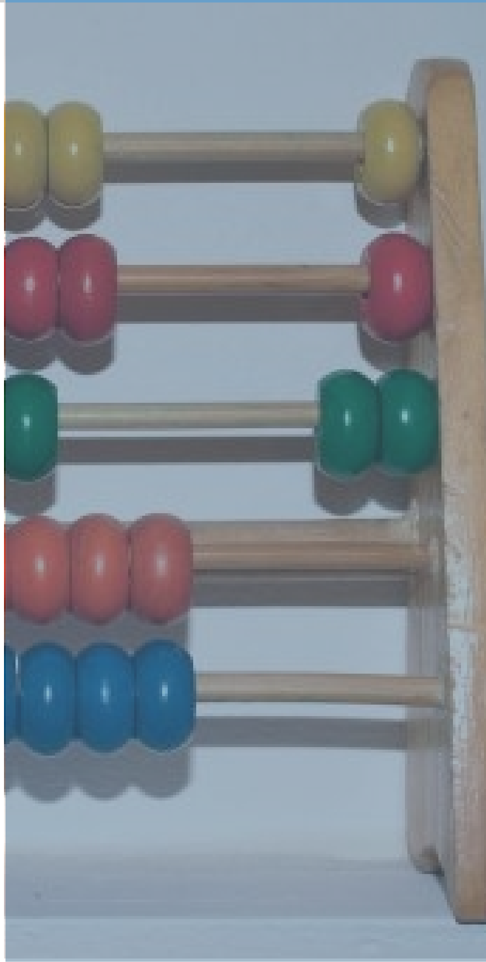
(logarithmierte) Größe von zeremoniellen Erdwerken auf den Gesellschaftsinseln (Beispiel nach Shennan)

Gegeben sind die (logarithmierten) Größen von zeremoniellen Erdwerken in zwei verschiedenen Tälern auf den Gesellschaftsinseln.
Stellen Sie fest, ob die Varianzen in beiden Tälern als gleich anzusehen sind!

File: marae.csv

```
> marae<-read.csv2("marae.csv")
> var.test(groesse~tal,data=marae)

     F test to compare two variances

data:  groesse by tal
F = 1.4132, num df = 13, denom df = 9, p-value = 0.6124
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3689233 4.6805505
sample estimates:
ratio of variances
          1.413196
```

Die Varianzen sind als gleich anzusehen!



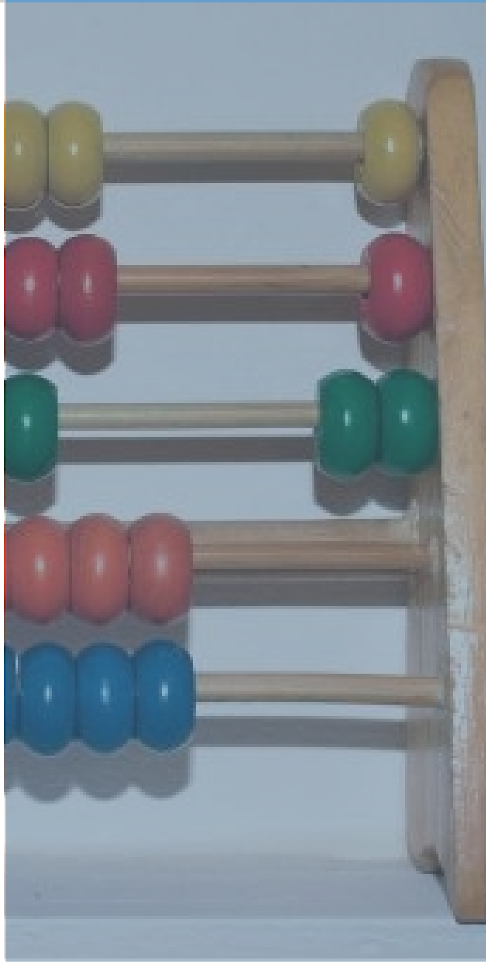
T-Test

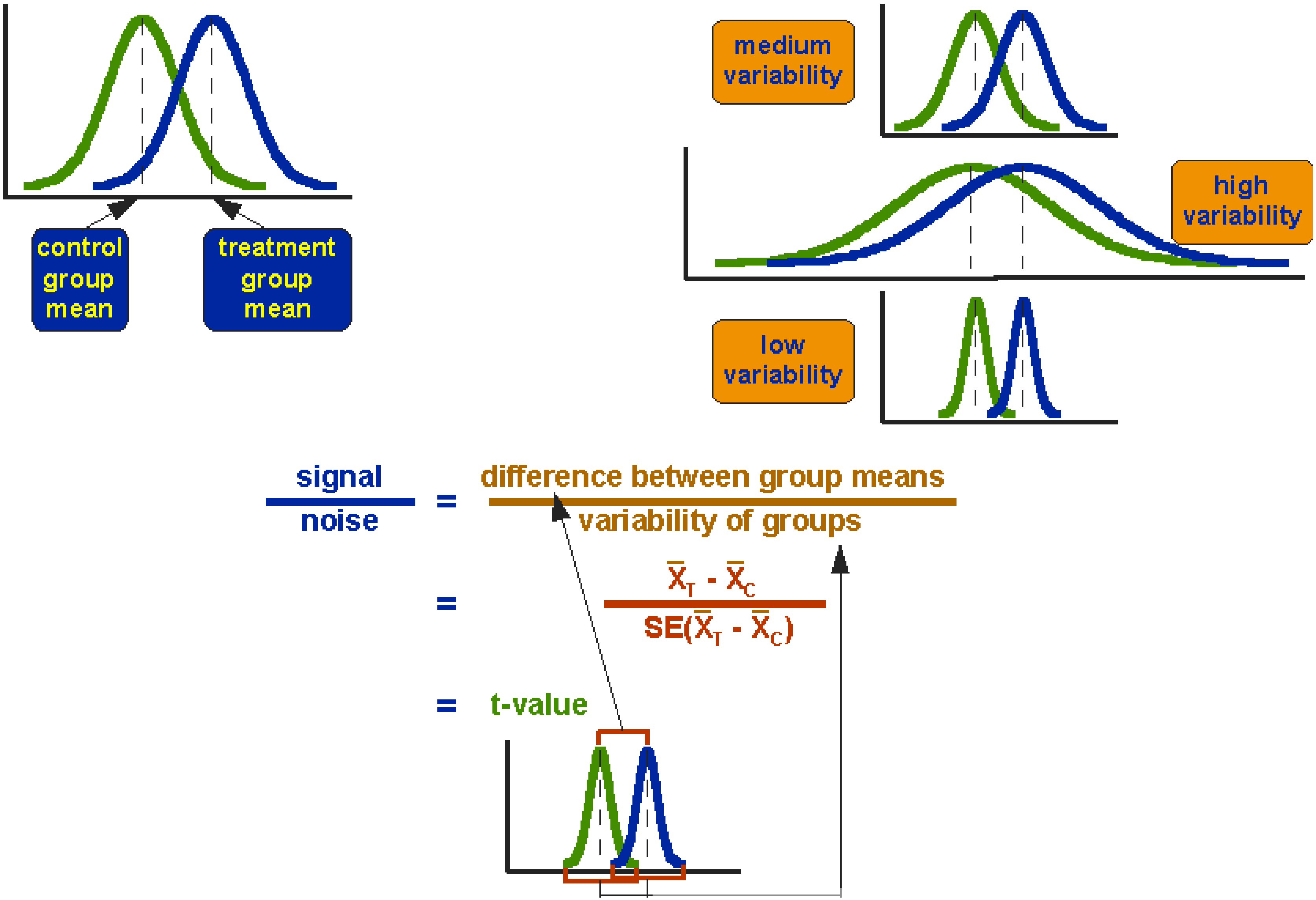
t-Test, gleiche Varianzen [1]

Test für den Vergleich der Mittelwerte zweier Stichproben
Wenn die Mittelwerte sich signifikant unterscheiden, stammen die Stichproben von unterschiedlichen Populationen

Voraussetzungen: zwei unabhängig normalverteilte Stichproben mit gleicher Varianz
H_0: Die Populationen beider Stichproben haben gleichen Mittelwert
H_1: Die Populationen beider Stichproben haben verschiedenen Mittelwert

Idee: Wenn die Mittelwerte zweier Stichproben innerhalb des Standard-Fehlers der Schätzungen der Differenz der Mittelwerte der zugehörigen Populationen liegen, dann können beide Populationen gleich sein, sonst nicht.





Quelle: http://www.socialresearchmethods.net/kb/stat_t.php

t-Test, gleiche Varianzen [2]

Beispiel: Klingenlänge zweier Fundstellen, Berechnung von Hand

$$n\,Fundstelle\,1=20\,;\,\bar{x}\,Fundstelle\,1=20.015\,;\,s_1^2=40.20871$$

$$n\,Fundstelle\,2=25\,;\,\bar{x}\,Fundstelle\,2=20.492\,s_2^2 33.0641$$

$$t=\frac{\bar{x}_1-\bar{x}_2}{s_{\bar{x}_1-\bar{x}_2}}$$

$$s_{\bar{x}_1-\bar{x}_2}=\sqrt{\frac{(n_1-1)*s_1^1+(n_2-1)*s_2^2}{n_1+n_2-2}}*\sqrt{\frac{n_1+n_2}{n_1*n_2}}=\sqrt{\frac{(20-1)*40.20871+(25-1)*33.0641}{20+25-2}}*\sqrt{\frac{20+25}{20*25}}$$

$$s_{\bar{x}_1-\bar{x}_2}=\sqrt{\frac{1557{,}50389}{43}}*\sqrt{\frac{45}{500}}=\sqrt{36{,}221020698}*\sqrt{0{,}09}=1{,}805517062$$

$$t=\frac{20.015-20.492}{1{,}805517062}=-0{,}264190248$$

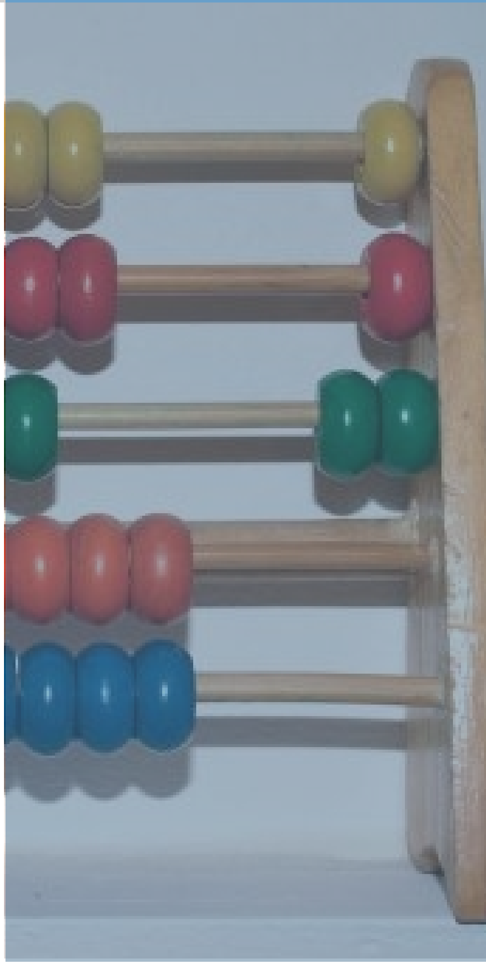
$$df=(n_1-1)+(n_2-1)=n_1+n_2-2=20+25-2=43$$

Nachschlagen in Tabelle (z.B. Shennan): df = 43; Sign.level = 0.05;
Mögliche Unterschiede: größer-kleiner, daher 2-seitige Fragestellung
Grenzwert = 2.021

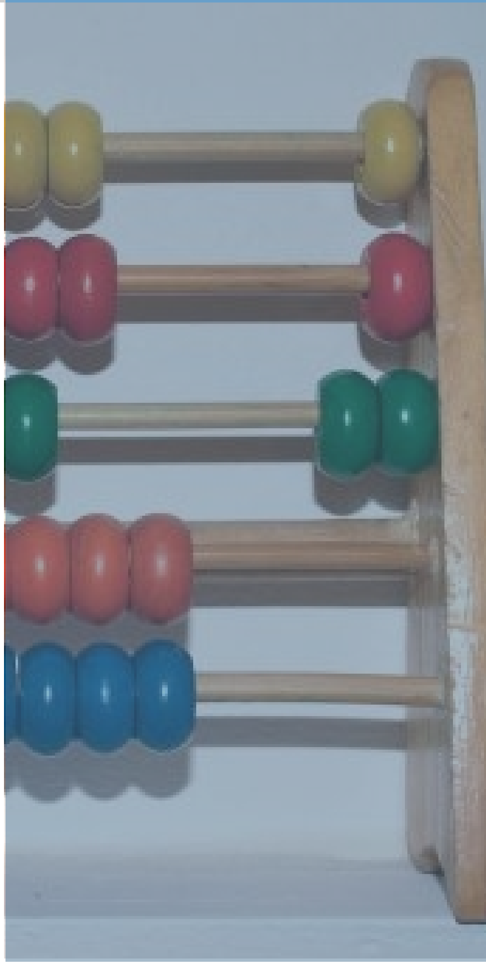
$$-0{,}264190248<2.021$$

$$nicht\ signifikant\,,\ Null-Hypothese\ kann\ nicht\ abgelehnt\ werden$$

Es besteht kein signifikanter Unterschied zwischen den beiden Stichproben bezüglich der Mittelwerte, sie könnten aus der gleichen Population stammen!



t-Test, gleiche Varianzen [3]



Test für den Vergleich der Mittelwerte zweier Stichproben
Wenn die Mittelwerte sich signifikant unterscheiden, stammen die Stichproben von unterschiedlichen Populationen

Voraussetzungen: zwei unabhängig normalverteilte Stichproben mit gleicher Varianz
H_0: Die Populationen beider Stichproben haben gleichen Mittelwert
H_1: Die Populationen beider Stichproben haben verschiedenen Mittelwert

```
> t.test(laenge~fundort,data=klingenlaenge,var.equal=T)

     Two Sample t-test

data:  laenge by fundort
t = -0.2642, df = 43, p-value = 0.7929
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.118172  3.164172
sample estimates:
mean in group site1 mean in group site2
             20.015              20.492
```

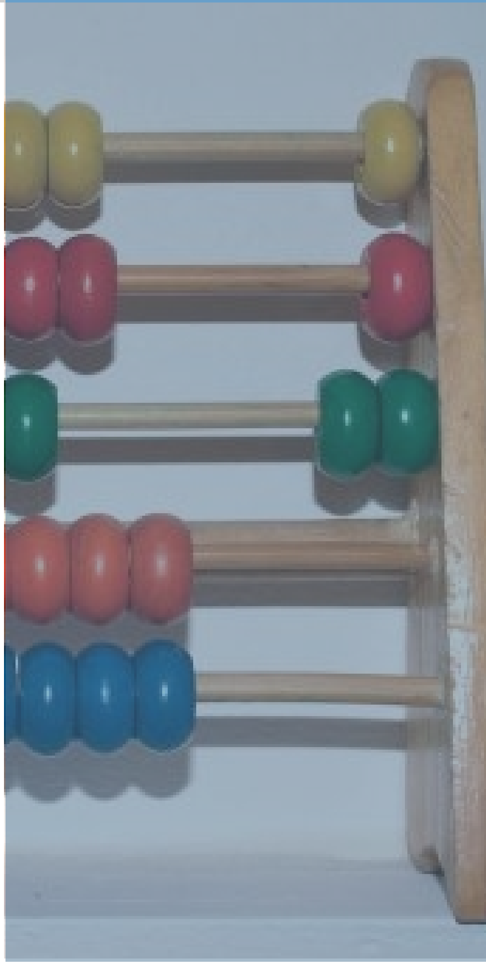
Kein Signifikantes Ergebnis, die beiden Stichproben unterscheiden sich nicht signifikant hinsichtlich ihrer Mittelwerte.

Aufgabe t-Test

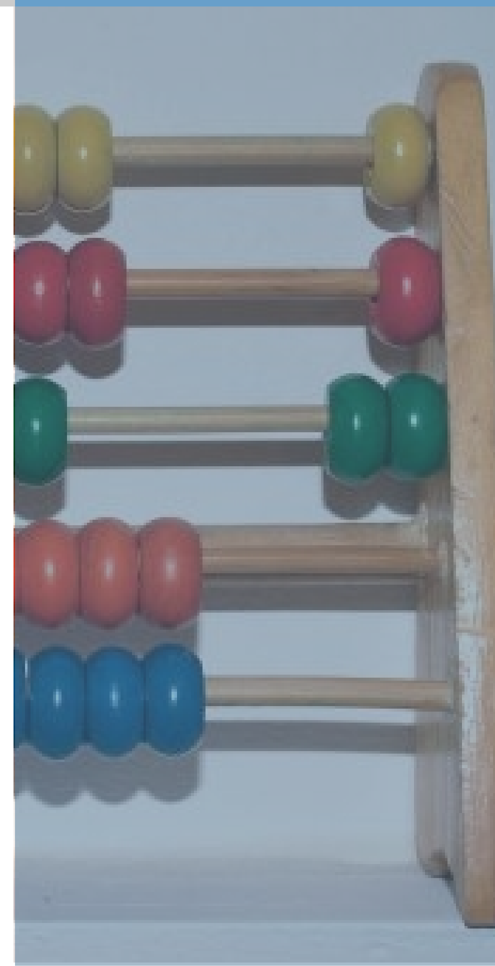
(logarithmierte) Größe von zeremoniellen Erdwerken auf den Gesellschaftsinseln (Beispiel nach Shennan)

Gegeben sind wiederum die (logarithmierten) Größen von zeremoniellen Erdwerken in zwei verschiedenen Tälern auf den Gesellschaftsinseln. Stellen Sie fest, ob sich die Mittelwerte signifikant unterscheiden!

File: marae.csv



Aufgabe t-Test



(logarithmierte) Größe von zeremoniellen Erdwerken auf den Gesellschaftsinseln (Beispiel nach Shennan)

Gegeben sind wiederum die (logarithmierten) Größen von zeremoniellen Erdwerken in zwei verschiedenen Tälern auf den Gesellschaftsinseln. Stellen Sie fest, ob sich die Mittelwerte signifikant unterscheiden!

File: marae.csv

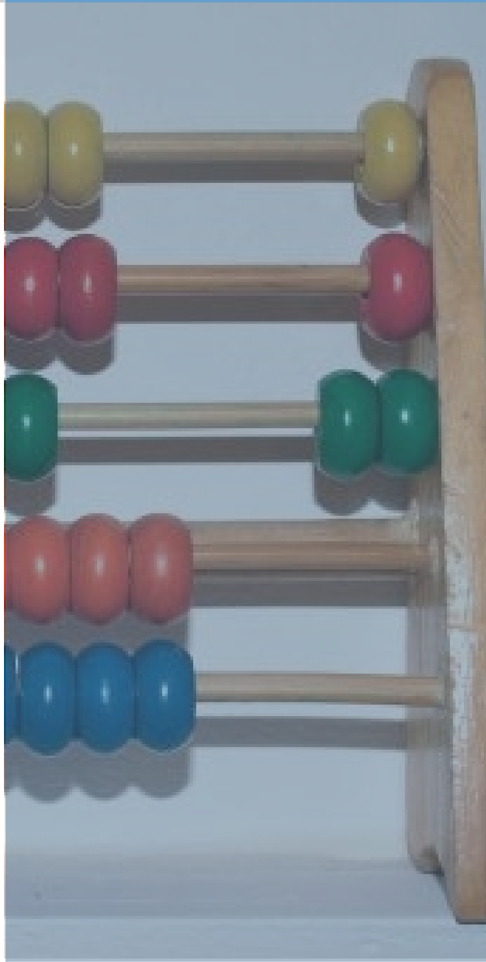
```
> marae<-read.csv2("marae.csv")
> t.test(groesse~tal,data=marae)

     Welch Two Sample t-test

data:  groesse by tal
t = -1.026, df = 21.309, p-value = 0.3164
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4193637  0.1421065
sample estimates:
mean in group tal1 mean in group tal2
          2.134071           2.272700
```

Welch-Test (t-Test für ungleiche Varianzen mit Welch-Korrektur)

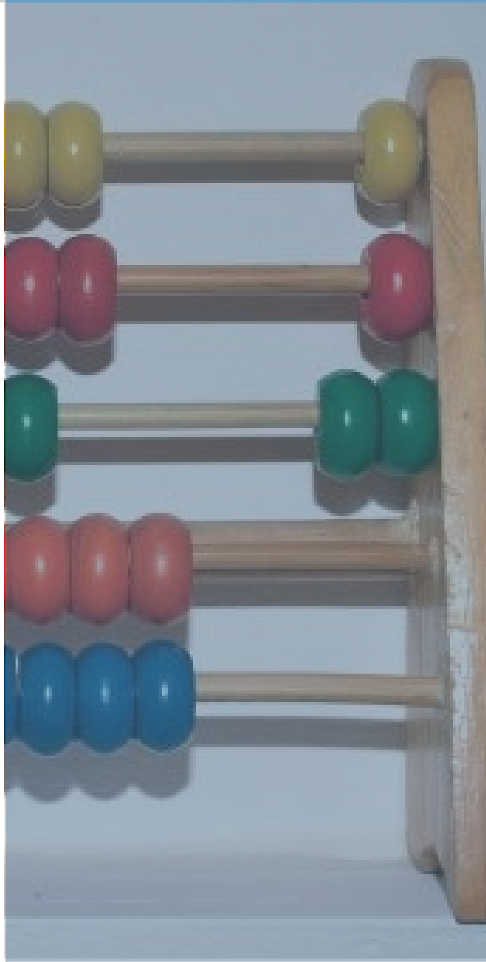
Wenn die Gleichheitsvoraussetzung nicht erfüllt ist

Annährung der Ergebnisse durch Korrektur der Freiheitsgrade

```
> klingenlaenge3<-read.csv2("klingenlaenge3.csv",row.names=1)
>
var.test(klingenlaenge3$x,klingenlaenge$laenge[klingenlaenge$fundort
=="site1"])
...
F = 0.3535, num df = 19, denom df = 19, p-value = 0.02854
>
t.test(klingenlaenge3,klingenlaenge$laenge[klingenlaenge$fundort=="s
ite1"],var.equal=F)

     Welch Two Sample t-test

data:  klingenlaenge3 and klingenlaenge$laenge[klingenlaenge$fundort
== "site1"]
t = 1.8368, df = 30.941, p-value = 0.07586
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3346190  6.3946190
sample estimates:
mean of x mean of y
   23.045    20.015
```



Wiederholung: Unabhängige – Abhängige Stichprobe



Abhängige Stichproben:
Das Ergebnis der einen Stichprobe ist teilweise von der anderen abhängig (Untersuchung von Patienten vor/nach Einnahme eines Medikamentes)

```
t.test(variable1,variable2,var.equal=F,paired=T)
```
oder
```
t.test(variable1,variable2,var.equal=T,paired=T)
```

Unabhängige Stichproben:
Das Ergebnis der einen Stichprobe ist nicht von der anderen abhängig (Untersuchung von zwei Gräberfeldern)

```
t.test(variable1,variable2,var.equal=F)
```
oder
```
t.test(variable1,variable2,var.equal=T)
```

Multiples Testen [1]

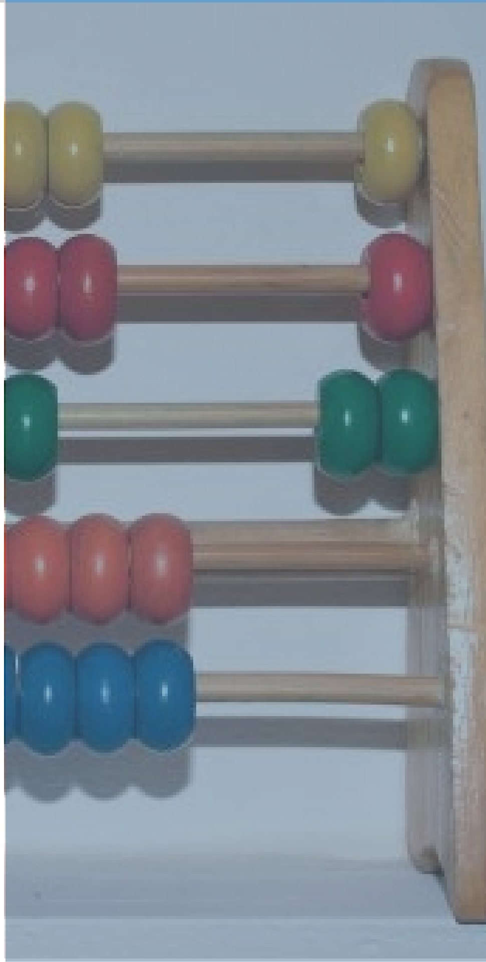
Das folgende trifft auf alles statistischen Tests zu! Was, wenn man mehr als zwei Gruppen vergleichen will?

Gegeben sind hypothetische Größen von TRB-Siedlungen in Schleswig-Holstein nach Feuchtigkeit des Gebietes und Region.

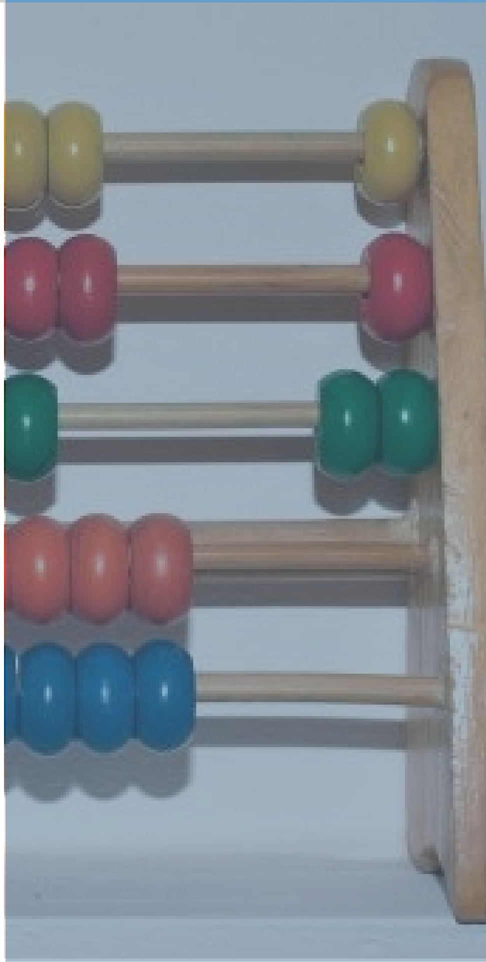
```
> settlements<-read.csv2("settlements.csv",row.names=1)
> settlements
        size                regions wetness
1  20.1956525              Stormarn   humid
2  37.5501213            Neumünster   humid
3   0.5980919            Neumünster    arid
…
```

Frage: Unterscheiden sich die Siedlungsgrößen signifikant je nachdem, wie feucht der Untergrund ist?

Wie ist vorzugehen?



Multiples Testen [2]



Wohl intuitive, aber falsche Lösung!

Man testet die verschiedenen Gruppen jeweils miteinander, ob es einen statistisch signifikanten Unterschied gibt.

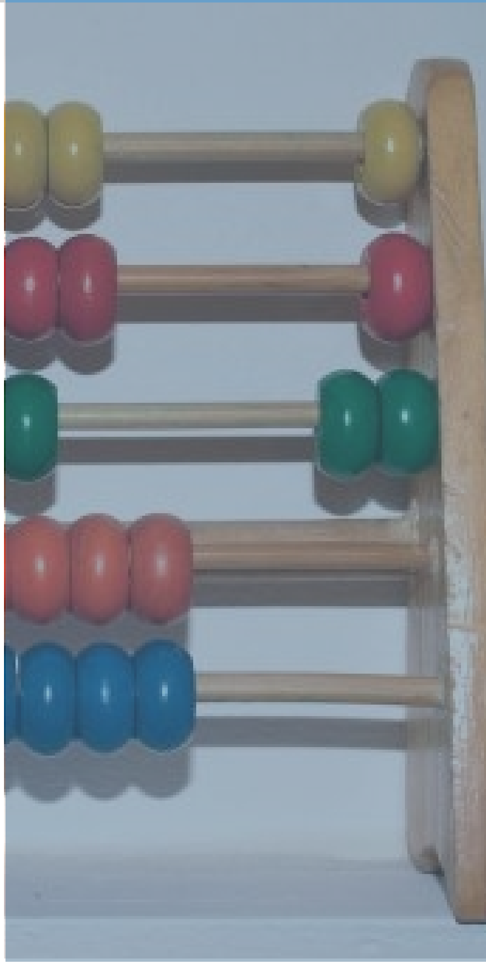
Problem: Je öfter man testet, um so höher wird die Wahrscheinlichkeit, das ein signifikantes Ergebnis "per Zufall" entsteht.

Beispiel: Wir testen 3 Gruppen gegeneinander, brauche dazu 3 Tests:
Rest ↔ Arid, Rest ↔ Medium, Rest ↔ Humid

Die Wahrscheinlichkeit, dass die Alternativhypothese trotz signifikantem Ergebnis falsch ist, liegt bei jedem Test bei 0.05 %. Das ergibt bei 3 Tests eine Fehlerwahrscheinlichkeit von 3*0,05=0,15%.

Bei 100 Tests wären 5 Tests wahrscheinlich falsch signifikant!

Multiples Testen [3]



Lösungsmöglichkeit 1 (kann auch für alle anderen Tests verwand werden): Korrektur der p-Werte, z.B. Bonferroni-Korrektur

Da die gesamte Prozedur als ein Test aufgefasst werden muss:
Teilen der Signifkanzschwelle für die Einzeltests durch Anzahl der Tests ergibt globales Signifikanzniveau für jeden Test.

Beispiel: Vergleichen wir 3 Gruppen miteinander, werden 3 Tests durchgeführt. Signifikant werden nur die p-Werte gewertet, die kleiner als 0.05/3=0,016666667 oder niedriger ausfallen.

t-Test
Medium ↔ Humid : p-Wert = 0.287
Humid ↔ Arid: p-Wert = 1.939e-07
Arid ↔ Medium: p-Wert = 3.436e-05

```
> p.adjust(c(0.287,1.939e-07,3.436e-05),method="bonferroni")
[1] 8.6100e-01 5.8170e-07 1.0308e-04
```

Das erste Ergebnis ist bei drei Tests nicht signifikant.

nach

ANOVA

ANOVA [1]

Lösungsmöglichkeit 2 für den t-Test
Vergleich von mehreren Gruppen miteinander

Abhängige Variable: die zu testende Größe (hier size)
Unabhängige Variable: die zu testende Gruppierungsvariable (hier wetness)

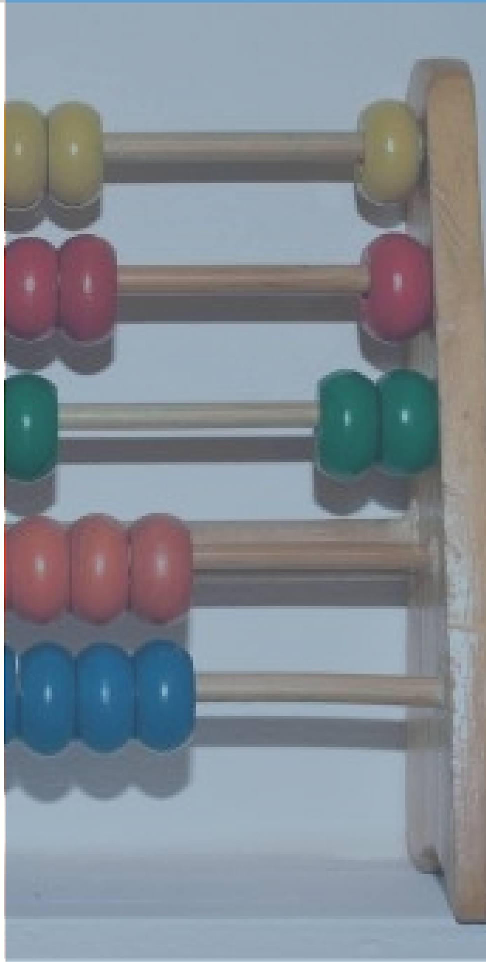
Graphische Darstellung:

```
> boxplot(size~wetness, data=settlements)
```

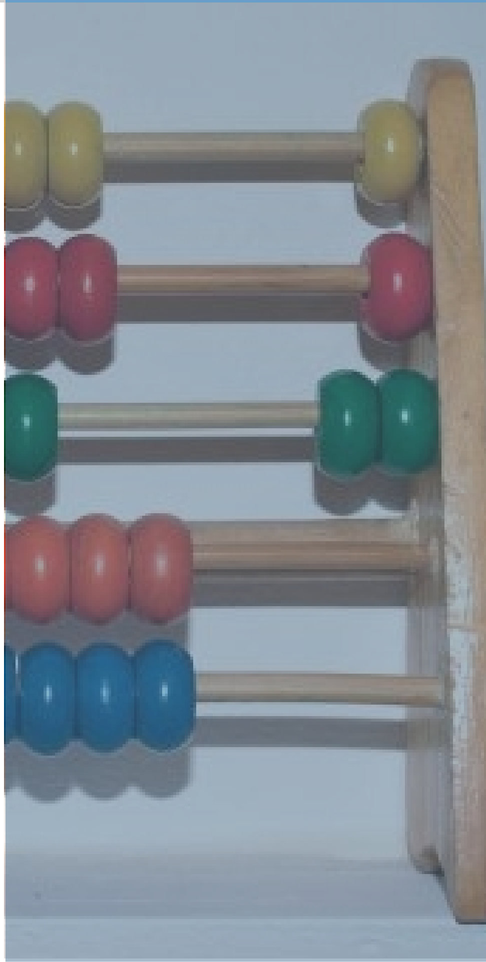
ANOVA

```
> summary(aov(size~wetness, data=settlements))
            Df Sum Sq Mean Sq F value  Pr(>F)
wetness      2   3564  1782.1   23.12 1.7e-07 ***
Residuals   42   3238    77.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Es gibt signifikante Unterschiede, nur wo?



ANOVA [2]



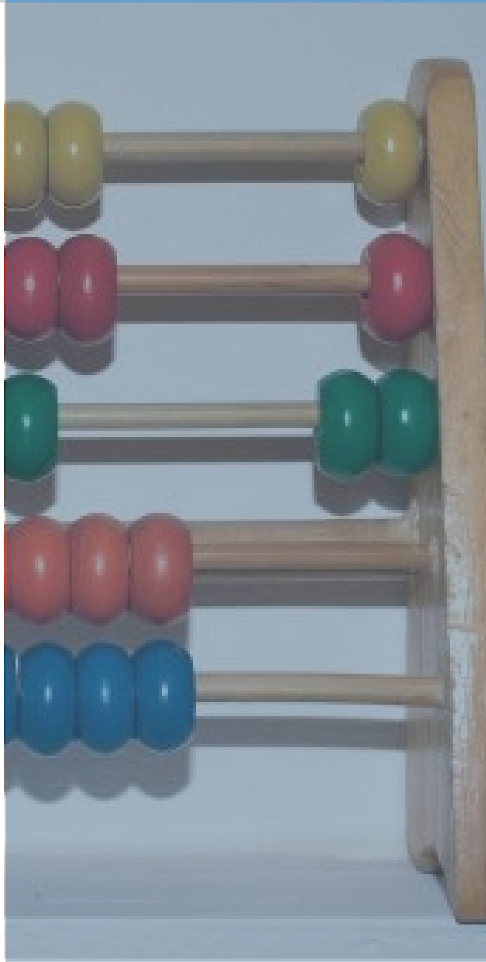
Detaillierte Darstellung der Unterschiede

Zur umfassenden Ausgabe wird der Datensatz als lineares Modell aufgefasst:

```
> summary.lm(aov(size~wetness, data=settlements))
…

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      9.388      2.267   4.141 0.000163 ***
wetnesshumid    20.387      3.206   6.359 1.21e-07 ***
wetnessmedium   16.880      3.206   5.265 4.48e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.78 on 42 degrees of freedom
Multiple R-squared:  0.524,  Adjusted R-squared:  0.5013
F-statistic: 23.12 on 2 and 42 DF,  p-value: 1.697e-07
```

Die erste Gruppe (arid) wird als Kontrollgruppe aufgefasst, die anderen Gruppen werden hiermit verglichen. Uns interessiert, wie feuchte und trockene Böden sich von mittelfeuchten Böden unterscheiden, daher muß die Kontrollgruppe auf medium umgestellt werden.

ANOVA [3]



Definieren der Kontrollgruppe

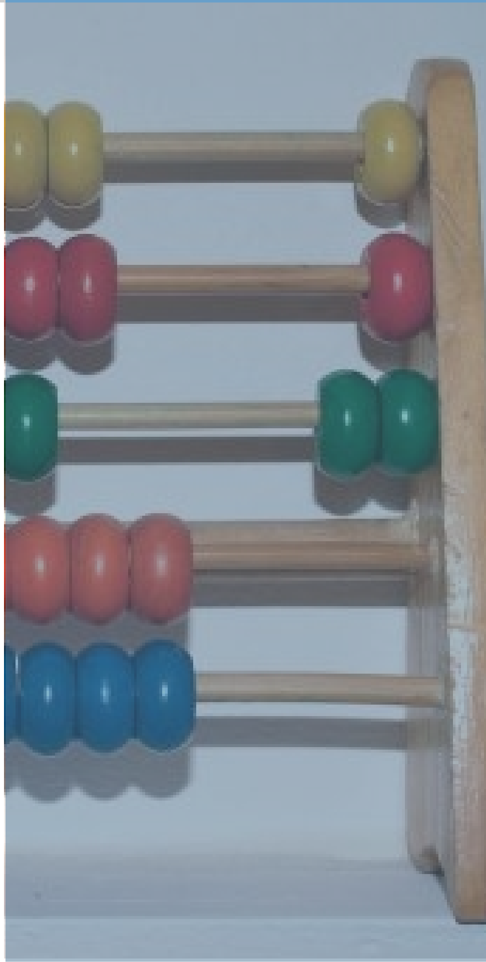
Umstellen der Kontrollgruppe (Referenz) auf mittelfeuchte Böden:

```
> wetness.new<-relevel(settlements$wetness, ref ="medium")
> summary.lm(aov(settlements$size~wetness.new))
…
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        26.268      2.267  11.587 1.15e-14 ***
wetness.newarid   -16.880      3.206  -5.265 4.48e-06 ***
wetness.newhumid    3.506      3.206   1.094     0.28
```

Ein signifikanter Unterschied besteht zwischen Kontrollgruppe (medium) und ariden Böden. Das ganze auf einen Blick mit multiplen t-Tests:

```
> pairwise.t.test(settlements$size,settlements$wetness,
p.adjust="bonferroni")
     Pairwise comparisons using t tests with pooled SD
data:  settlements$size and settlements$wetness
       arid    humid
humid  3.6e-07 -
medium 1.3e-05 0.84

P value adjustment method: bonferroni
```

ANOVA [4]



Zweifaktorielle ANOVA

Wenn der Einfluß zweier Gruppierungen auf die Größe untersucht werden soll. Hier: Siedlungsgröße und Region

Darstellung der Größe je Region:

```
> boxplot(size~regions, data=settlements)
```

Darstellung der Beziehungen von Region, Feuchtigkeit und Größe:

```
> interaction.plot(settlements$wetness, settlements$regions,
settlements$size)
```

ANOVA:

```
> summary(aov(size~wetness*regions, data=settlements))
                Df Sum Sq Mean Sq F value   Pr(>F)
wetness          2   3564  1782.1  29.332 3.08e-06 ***
regions         13   1372   105.6   1.738    0.142
wetness:regions 12    832    69.4   1.142    0.391
Residuals       17   1033    60.8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Schreibweise in der Formel:

*: Die Beziehung der Variablen untereinander wird berücksichtigt

+: Die Beziehung der Variablen untereinander wird nicht berücksichtigt

ANOVA [5]

Zweifaktorielle ANOVA

Ergebnis: nur Feuchtigkeit hat signifikanten Einfluß auf Siedlungsgröße, nicht Region und auch nicht kombinierte Effekte von Region und Feuchtigkeit sind signifikant

Detaillierte Ausgabe:
> summary.lm(aov(size~wetness*regions, data=settlements))

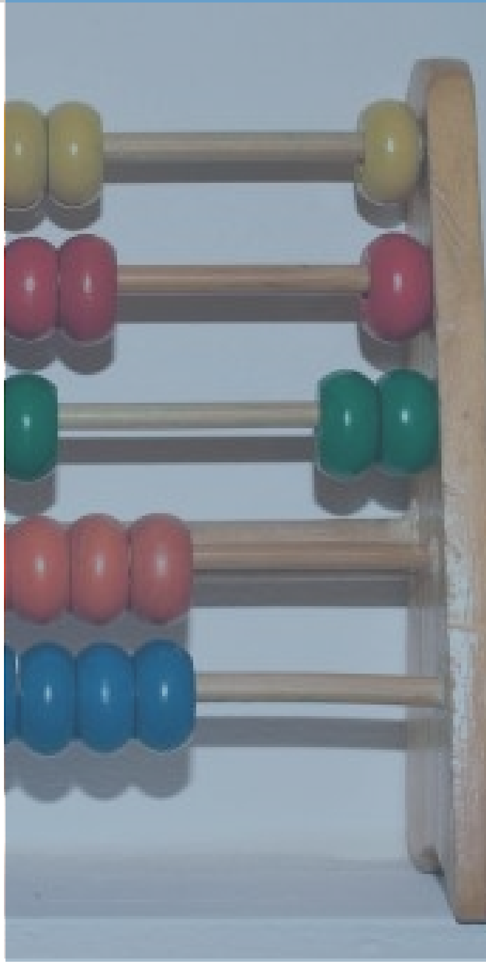
Ändern der Kontrollgruppe für die Regionen:

```
> regions.new<-relevel(settlements$region, ref ="Herzogtum
Lauenburg")
```

Detailierte Ausgabe:

```
> summary.lm(aov(size~wetness.new*regions.new, data=settlements))
```

Ergebnis: Von unserer Kontrollgruppe (mittelfeuchte Böden im Herzogtum Lauenburg) weichen keine anderen Gruppen signifikant ab. Nur die Feuchtigkeit hat einen signifikanten Einfluß auf die Siedlungsgröße, dabei sind nur Siedlungen auf feuchten Böden signifikant unterschiedlich von unserer Kontrollgruppe (mittelfeuchte Böden).



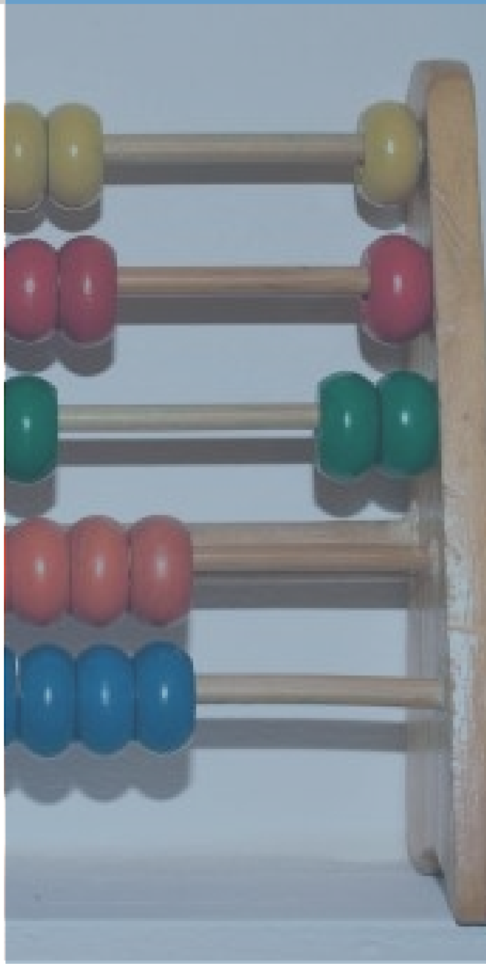
ANOVA [6]

Varianten der ANOVA

Einfaktorielle Varianzanalyse
Ein unabhängiger Faktor (Gruppierungsvariable), eine abhängige Variable

Zwei- / mehrfaktorielle Varianzanalyse
Zwei / mehrere unabhängige Faktoren (Gruppierungsvariable), eine abhängige Variable

multivariate Varianzanalyse (MANOVA)
Ein/Zwei/Viele unabhängige Faktoren (Gruppierungsvariable), mehrere abhängige Variable



ende_09_parametrische_tests

Nächste Sitzung 20. Januar 2015:
Regression und Korrelation

Bitte lesen Sie Shennan 6
„Estimation and Testing with the Normal Distribution“