Statistik – Methoden zum Vergleich von zwei Gruppen

Trennschärfe und Poweranalyse

Trennschärfe / Power

- Bisheriger Fokus bei der Auswertung von Hypothesentests:
  - α-Fehler, Fehler 1.Art
  - Fehler: Wechsel zur Alternativhypothese, obwohl die Nullhypothese richtig ist
  - Wechsel zu Alternativhypothese erst, wenn die Wahrscheinlichkeit (1-α) des Fehlers klein genug ist
- Spezifität eines Hypothesentests

Trennschärfe / Power

Realität

Testergebnis		H_0 ist wahr	H_1 ist wahr
	Verbleib bei H_0	Korrekte Entscheidung $(1-\alpha)$ (Spezifität)	Fehler 2. Art (β)
	Wechsel zu H_1	Fehler 1.Art (α)	Korrekte Entscheidung $(1-\beta)$ (Trennschärfe)

Trennschärfe / Power

- Perspektivwechsel

- β-Fehler, Fehler 2.Art

- Fehler: Verbleib in der Nullhypothese, obwohl die Alternativhypothese richtig ist

- Die Wahrscheinlichkeit (1-β) mit dem Wechsel zur Alternativhypothese die korrekte Entscheidung zu treffen ist ein Gütekriterium für Hypothesentests

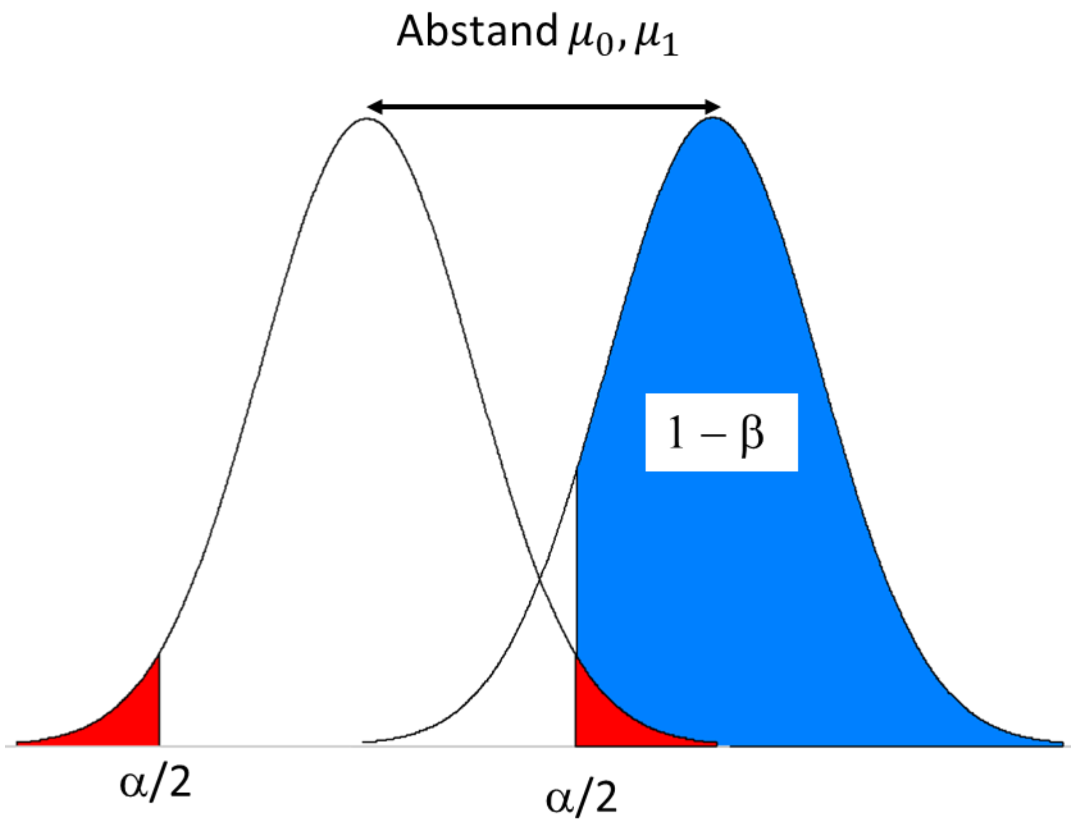
- Trennschärfe, Power eines Hypothesentests

Trennschärfe / Power

- Trennschärfe
- Unterschiede erkennen, wenn sie tatsächlich vorhanden sind
- Mit welcher Wahrscheinlichkeit wird die Nullhypothese H_0 korrekt abgelehnt
- Mit welcher Wahrscheinlichkeit wird die Alternativhypothese H_1 korrekt gewählt

Trennschärfe / Power

Zusammenhang Trennschärfe / Irrtumswahrscheinlichkeit



Trennschärfe / Power

Zusammenhang Trennschärfe / Irrtumswahrscheinlichkeit

- Wenn (1-β) klein ist, d.h. die Trennschärfe des Tests niedrig
  - Verbleibt man bei der Nullhypothese, ist diese nur eingeschränkter Wahrscheinlichkeit auch wahr
- Bei einem vorgegebenen Testszenario wirkt die Veränderung von α auch auf (1-β)
  - Absenken des α führt zu sinkender Teststärke (1-β)
  - Kompromiss erforderlich

Trennschärfe / Power

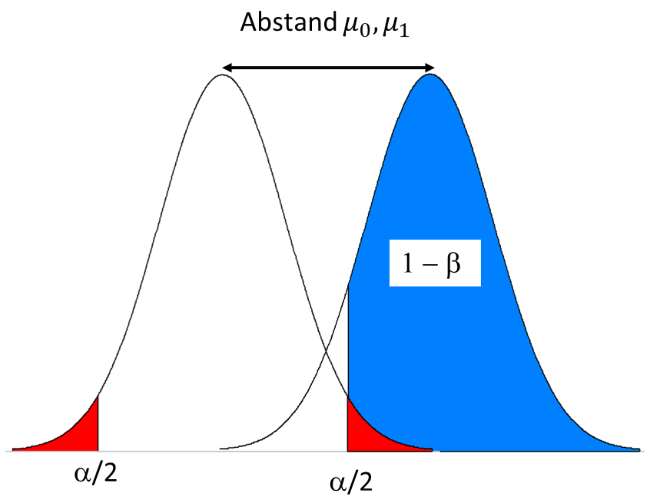
Trennschärfe-Analyse / Power-Analyse

- Bestimmung der erforderlichen Stichprobengröße, bei der mit der Wahrscheinlichkeit (1-β) ein Effekt einer bestimmten Größe erkannt werden kann

Trennschärfe / Power

Einflussfaktoren der Trennschärfe

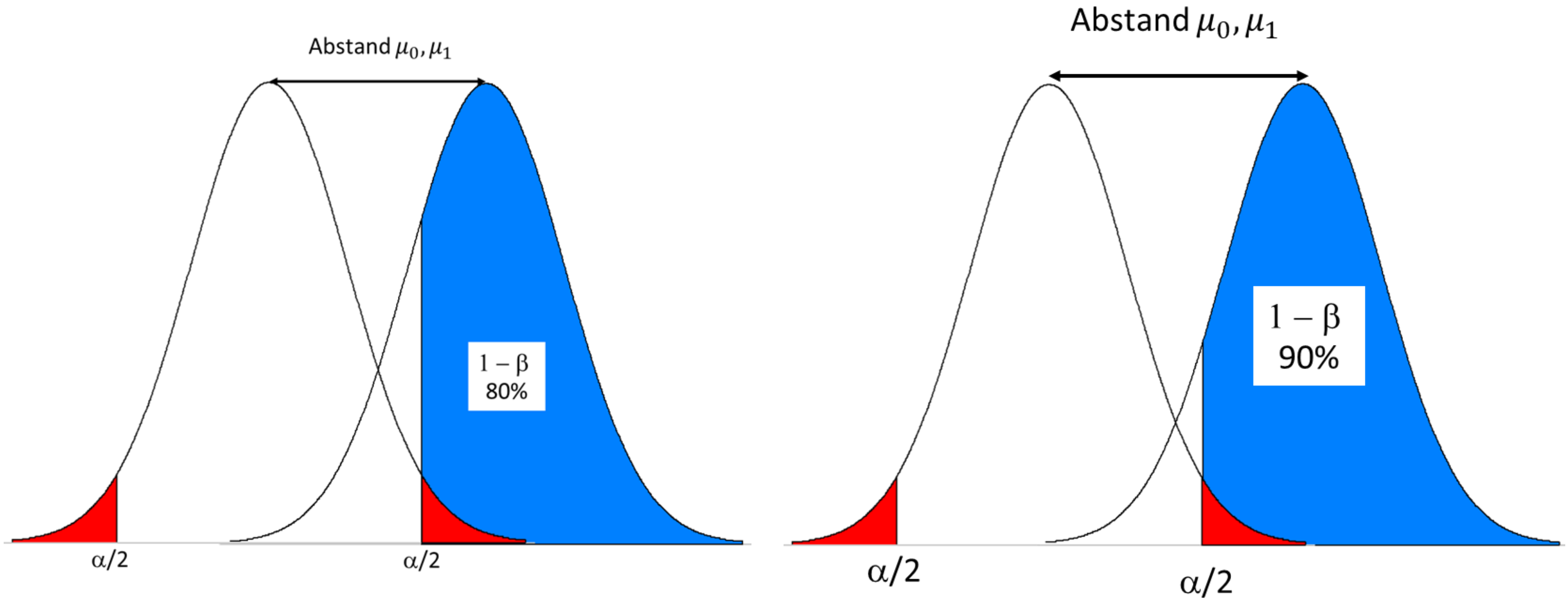
- Die Trennschärfe eines Tests kann auf verschiedene Arten verbessert werden
- Erkennen Sie Möglichkeiten?



Trennschärfe / Power

Einflussfaktor Abstand μ_0, μ_1

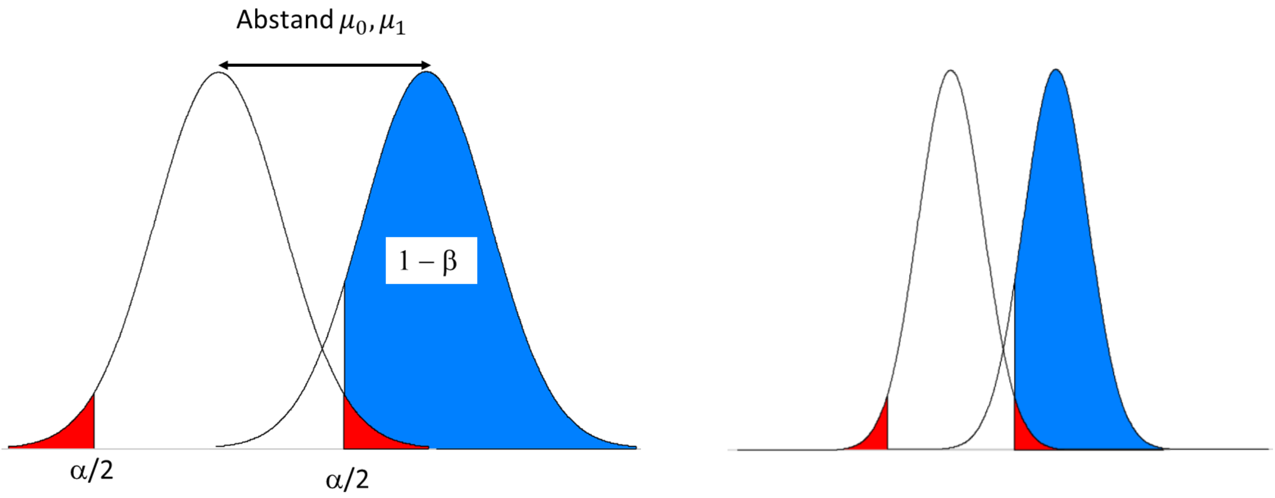
- Abstand Vergrößern 1-β wächst



Trennschärfe / Power

Einflussfaktor Varianz / Streuung

- Streuung verringern 1-β wächst

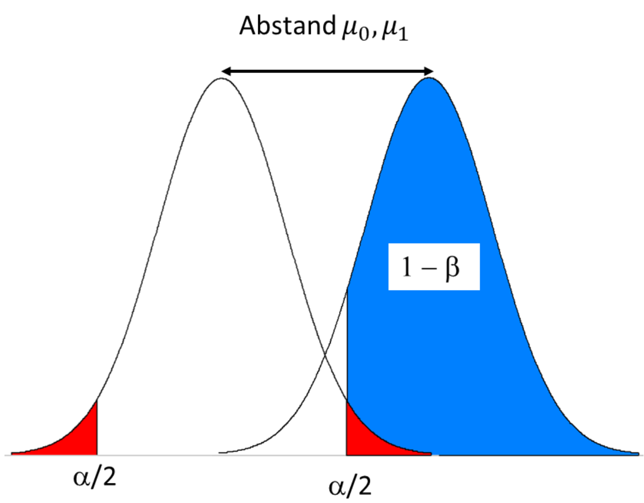


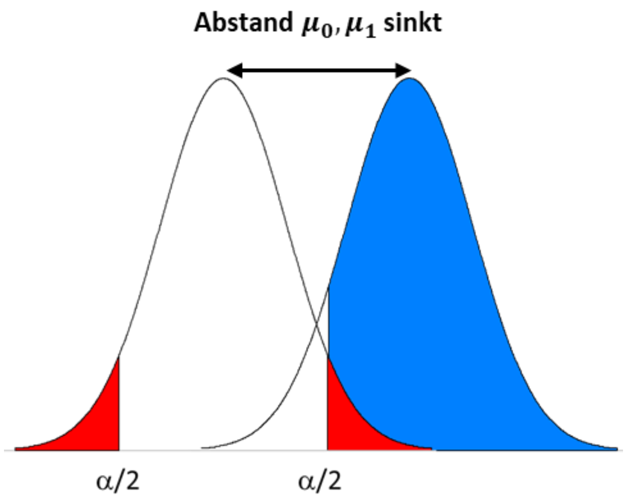
$\sigma = 1{,}0$ $\sigma = 0{,}5$

Trennschärfe / Power

Einflussfaktor Irrtumswahrscheinlichkeit α

- Irrtumswahrscheinlichkeit α erhöhen 1-β wächst





Trennschärfe / Power

Einflussfaktor Stichprobenumfang n

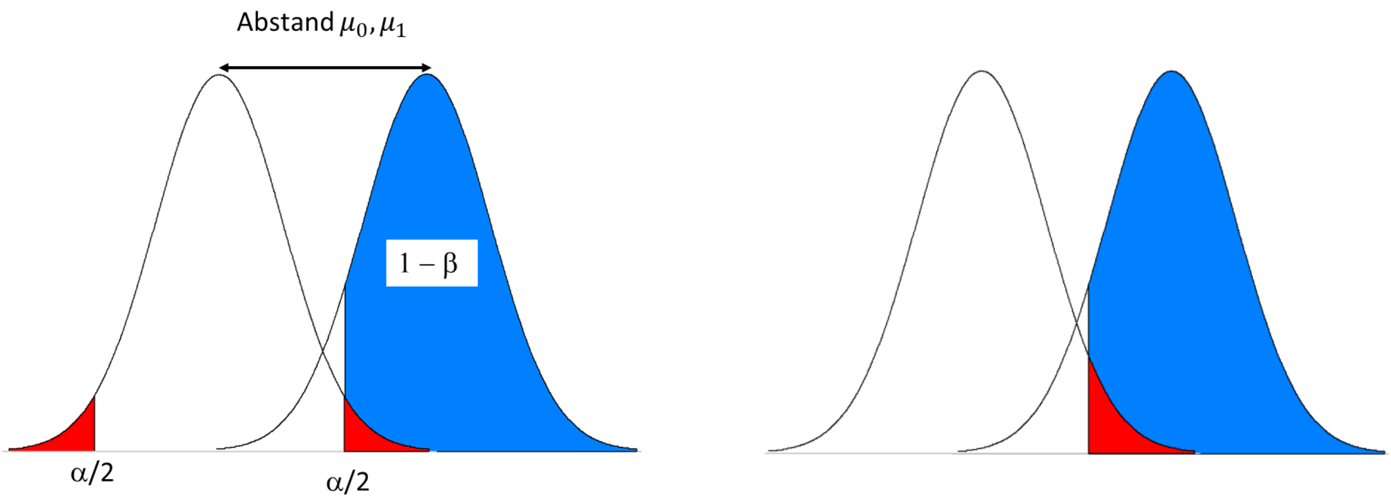
$$\boldsymbol{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}}$$

- Streuung sinkt 1-β wächst

Trennschärfe / Power

Einflussfaktor ein- / zweiseitiger Test

- Test einseitig (α wächst einseitig) 1-β wächst



Trennschärfe / Power

Einflussfaktor Testauswahl

- Für einzelne Fragestellungen gibt es unterschiedlich starke Tests (z.B. Prüfung auf Normalverteilung)
- Auswahl starker Tests
- Parametrische Tests sind i.a. stärker als nicht-parametrische Tests
- Wo möglich parametrische Tests wählen

Trennschärfe / Power

Grundlegendes Problem bei der Berechnung von Teststärke bzw. Stichprobengröße:

- Bestimmung der Effektstärke

- Für die t-Testfamilie kennen wir bereits den Cohen‘s D

- Für alle weitergehenden Berechnungen benötigen wir Definitionen der entsprechenden Effektstärken

Trennschärfe / Power

Beispiel Einfaktorielle ANOVA

- Grundlage der Berechnung von Trennschärfe bzw. Stichprobenumfang ist eine F-Verteilung, meistens um einen Nichtzentralitätsparameter ergänzt

- Die Berechnung wird dabei iterativ ausgeführt

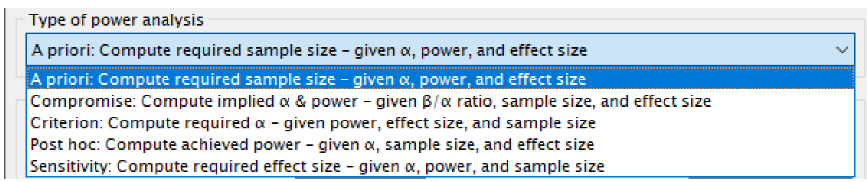
- Auf die entsprechenden Formulierungen wird an dieser Stelle verzichtet

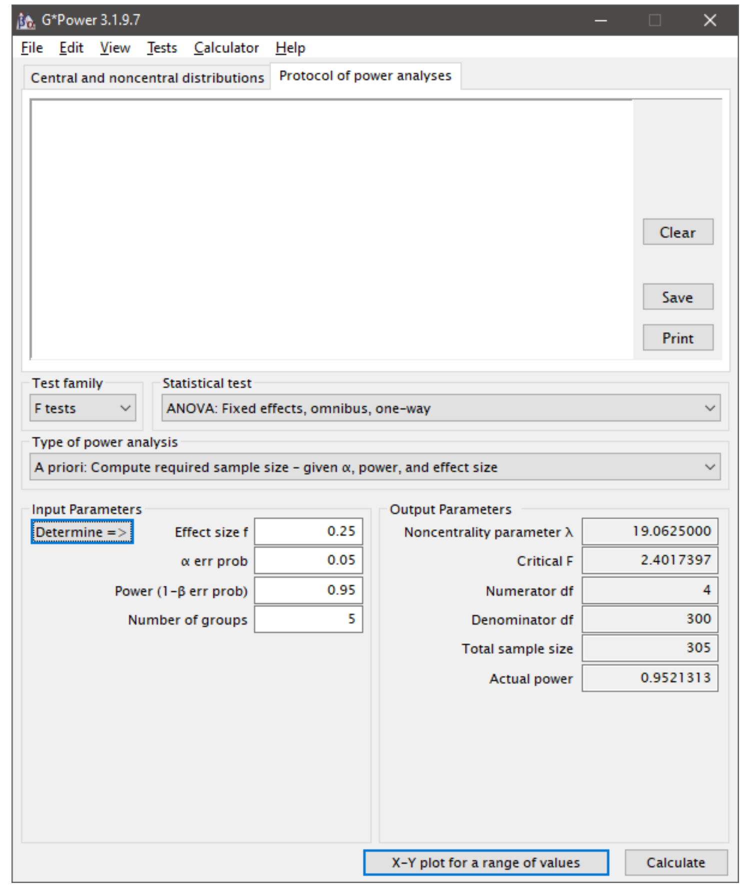
- Berechnung über RStudio und gpower möglich, keine Berechnung über den RCommander

Trennschärfe / Power

- Eingabefenster gPower für die einfaktorielle Varianzanalyse (Beispiel)

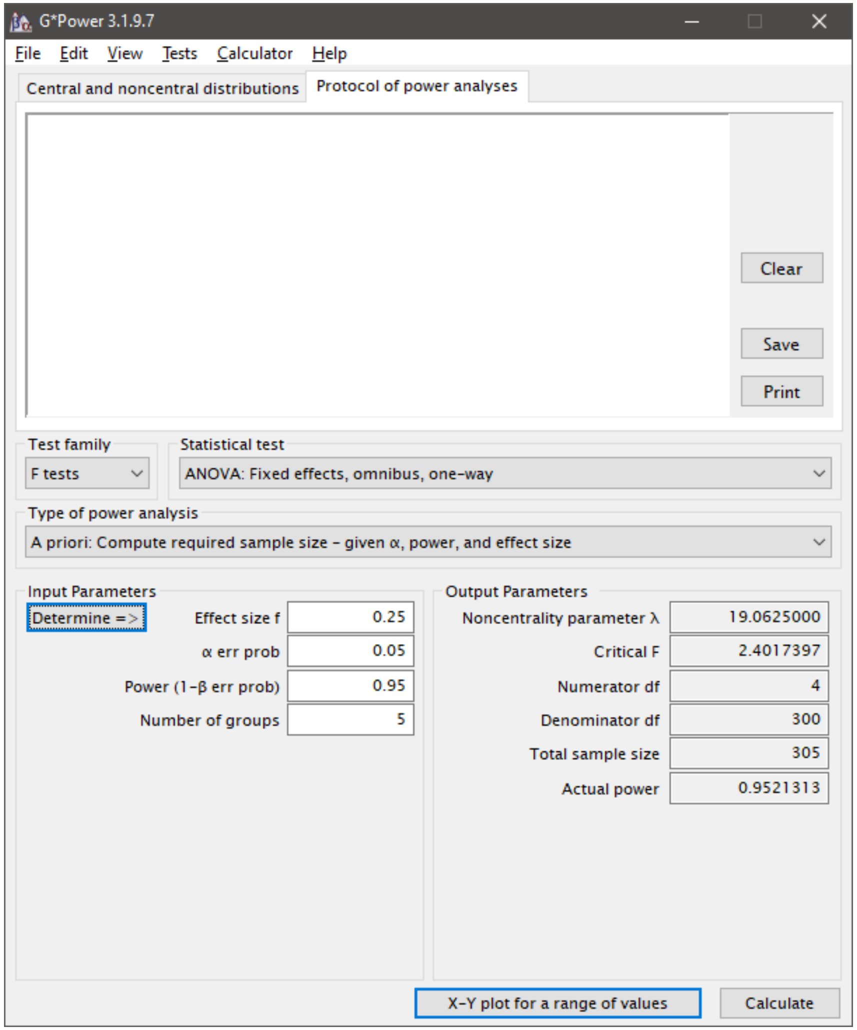
- Type of power analysis



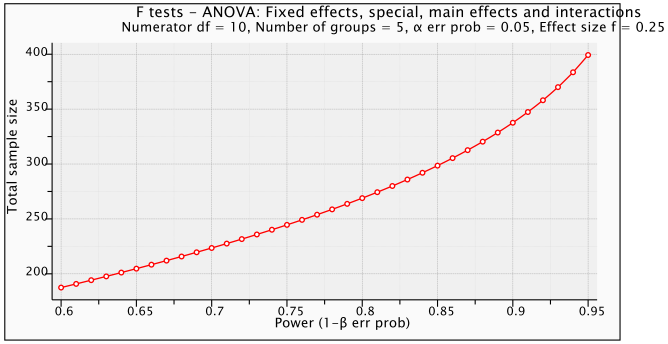


Trennschärfe / Power

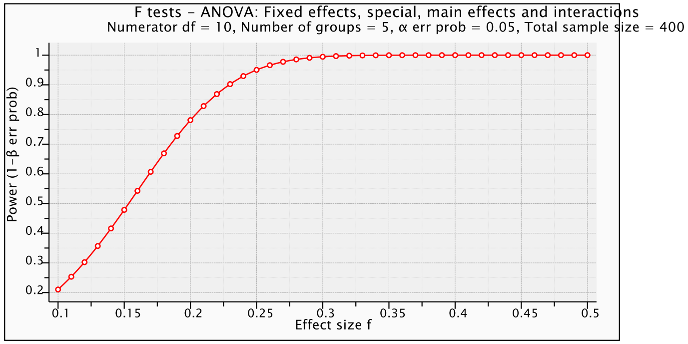
- Ausgabe
- Sample Size
- Power



Trennschärfe / Power



- Verschiedene Möglichkeiten die Operationscharakteristik (OC) grafisch darzustellen



Trennschärfe / Power

- Zur Berechnung von Stichprobengrößen kann man neben gpower auch ein Paket in r nutzen

- Die Berechnung erfolgt über RStudio, nicht über den RCommander

```
> install.packages(„pwr“)
> library(pwr)
```

bzw.

```
> install.packages(„WebPower“)
> library(WebPower)
```

- Die enthaltenen Analysen kann man sich in der r-Hilfe anschauen

Trennschärfe / Power

Für Aussagen über Stichprobengröße benötigen wir folgende Informationen:

- Effektgröße
  - Wie groß ist der Effekt, den man unterscheiden möchte
  - Je größer der Effekt ist, um so leichter ist er festzustellen bzw. um so kleiner kann die Stichprobe sein

Trennschärfe / Power

- Power bzw. Teststärke (1-β)
  - Wahrscheinlichkeit die Nullhypothese korrekt abzulehnen bzw. die Alternativhypothese zu wählen, wenn sie richtig ist
  - Je größer (1-β), um so leichter kann man einen Effekt identifizieren, aber um so größer müssen Stichproben sein

Trennschärfe / Power

- Signifikanzniveau α

    - Wahrscheinlichkeit, die Nullhypothese abzulehnen, obwohl sie richtig ist

    - Je niedriger das Signifikanzniveau gewählt wird, um so besser lehnt man eine Alternativhypothese ab, wenn sie falsch ist, um so größere Stichproben benötigt man aber auch

Trennschärfe / Power

- Wir haben bereits im Zusammenhang mit der t-Testfamilie die Effektgröße Cohen‘s D kennengelernt

- Für andere Testverfahren benötigen wir aber andere Effektdefinitionen

- Diese folgen nun mit einer Darstellung ihrer Nutzung in r und gpower

Effektdefinition für verschiedene Tests

t-Test, 1 Stichprobe

- Effekt: Cohen‘s D

- $D = \frac{\bar{x} - \mu_0}{s}$

pwr.t.test(n =, d = , sig.level = , power = , type = "one.sample“, alternative=)

- **n** Stichprobengröße
- **d** Effektgröße
- **sig.level** α
- **power** $1 - \beta$
- **type** Test
- **alternative** “less“, “two.sided“, “greater“

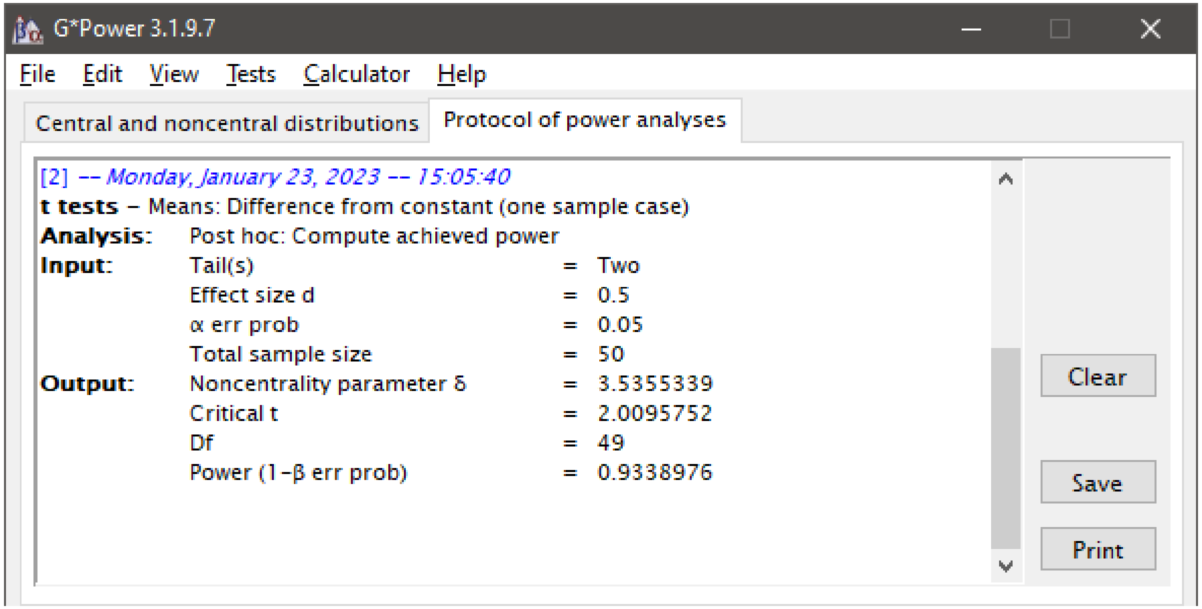
Effektdefinition für verschiedene Tests

t-Test, 1 Stichprobe

```
> pwr.t.test(d=0.50,n=50, sig.level =0.05, type="one.sample", alternative="two.sided")

One-sample t test power calculation

              n = 50
              d = 0.5
      sig.level = 0.05
          power = 0.9338976
    alternative = two.sided



> pwr.t.test(d=0.50, sig.level =0.05,power=0.9, type="one.sample", alternative="two.sided")


     One-sample t test power calculation

              n = 43.99548
              d = 0.5
      sig.level = 0.05
          power = 0.9
    alternative = two.sided
```

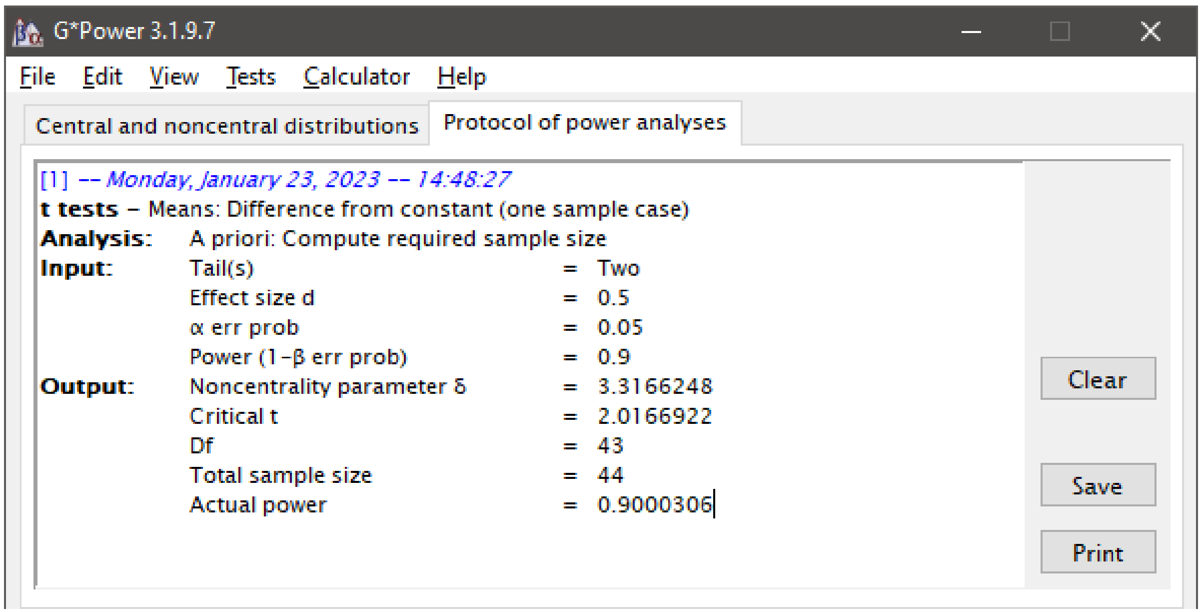
Effektdefinition für verschiedene Tests

t-Test, 1 Stichprobe



Effektdefinition für verschiedene Tests

t-Test, 1 Stichprobe



Effektdefinition für verschiedene Tests

t-Test, 2 unabhängige Stichproben

- Effekt: Cohen‘s D

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

pwr.t.test(n =, d = , sig.level= , power = , type = "two.sample“, alternative=)

- **n** Stichprobengröße
- **d** Effektgröße
- **sig.level** α
- **power** $1 - \beta$
- **type** Test
- **alternative** “less“, “two.sided“, “greater“

Effektdefinition für verschiedene Tests

t-Test, gepaarte Stichproben

- Effekt: Cohen‘s D
- $D = \frac{\bar{x}_2 - \bar{x}_1}{s_{pooled}}$

pwr.t.test(n =, d = , sig.level= , power = , type = “paired“, alternative=)

- **n** Stichprobengröße
- **d** Effektgröße
- **sig.level** α
- **power** $1 - \beta$
- **type** Test
- **alternative** “less“, “two.sided“, “greater“

Effektdefinition für verschiedene Tests

Einfaktorielle Varianzanalyse (ANOVA)

- Effekt: f

- $f = \sqrt{\frac{\eta^2}{1-\eta^2}}$ mit $\eta^2 = \frac{SS_{Zwischen}}{SS_{Gesamt}}$

pwr.anova.test(k =, n = , f = , sig.level = , power =)

- **k** Anzahl der Gruppen
- **n** Stichprobengröße (pro Gruppe)
- **f** Effektgröße
- **sig.level** α
- **power** $1 - \beta$

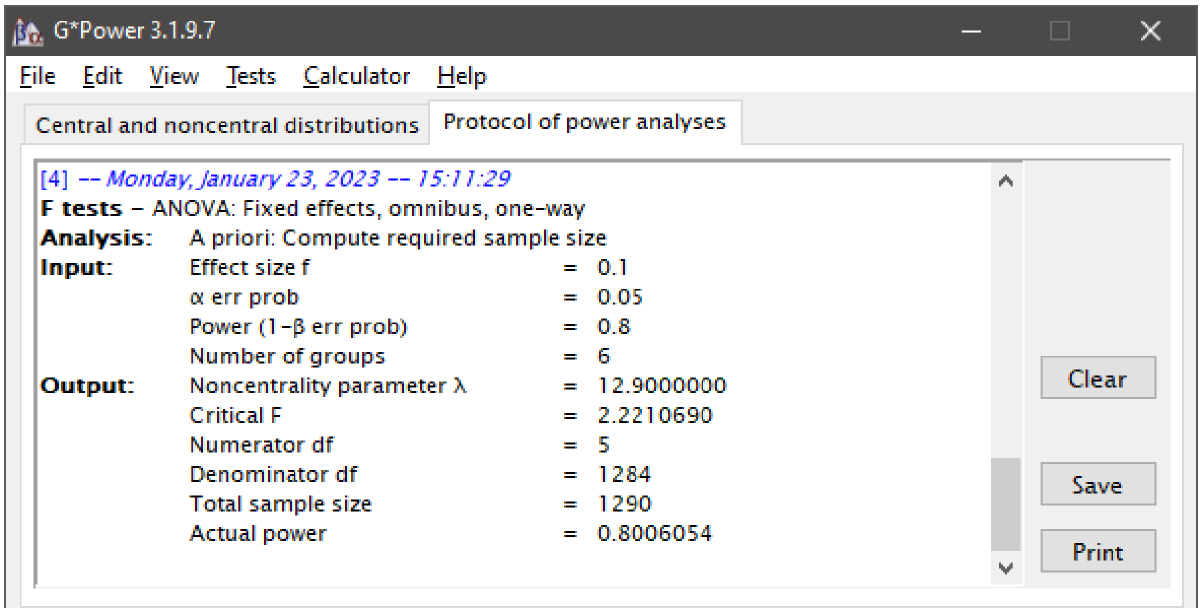
Effektdefinition für verschiedene Tests

Einfaktorielle Varianzanalyse (ANOVA)

```
> pwr.anova.test (k =6 , f =0.1 , sig.level =0.05 , power =0.80)

 Balanced one-way analysis of variance power calculation

              k = 6
              n = 214.7178
              f = 0.1
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

Effektdefinition für verschiedene Tests

Einfaktorielle Varianzanalyse (ANOVA)



Effektdefinition für verschiedene Tests

Single Proportion

- Effekt: h

- $h = 2\ arc\ sin\sqrt{p_1} - 2\ arc\ sin\ \sqrt{p_0}$

pwr.p.test(h = , n = , sig.level =, power =, alternative =)

- **h** Effektgröße
- **n** Anzahl der Beobachtungen
- **sig.level** α
- **power** $1 - \beta$
- **alternative** “less“, “two.sided“, “greater“

Effektdefinition für verschiedene Tests

Two Proportion

- Effekt: h

- $h = 2\ arc\ sin\sqrt{p_2} - 2\ arc\ sin\ \sqrt{p_1}$

pwr.2p.test(h = , n = , sig.level =, power =, alternative =)

- **h** Effektgröße
- **n** Anzahl der Beobachtungen (pro Stichprobe)
- **sig.level** α
- **power** $1 - \beta$
- **alternative** “less“, “two.sided“, “greater“

Effektdefinition für verschiedene Tests

χ^2-Test (Anpassung)

- Effekt: w

- $w = \sqrt{\frac{\chi^2}{n * df}}$

pwr.chisq.test(w =, N = , df = , sig.level=, power =)

- **w** Effektgröße
- **N** Gesamtzahl der Beobachtungen
- **sig.level** α
- **power** $1 - \beta$

Effektdefinition für verschiedene Tests

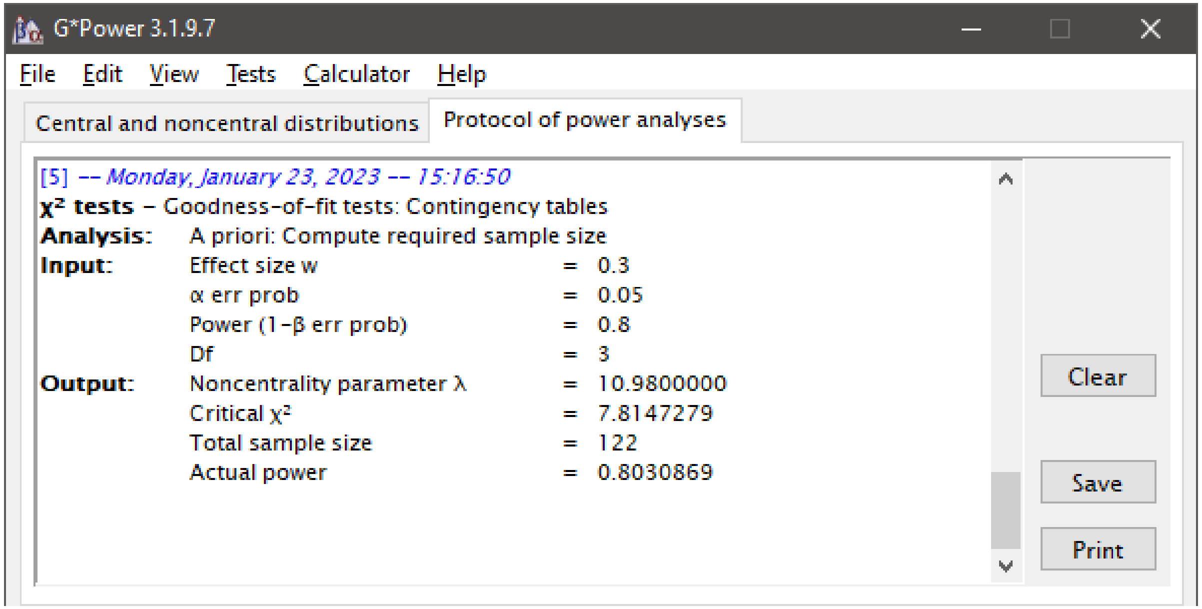
χ^2-Test (Anpassung)

\> pwr.chisq.test (w=0.3, df=3, sig.level =0.05, power=0.8)

```
    Chi squared power calculation

              w = 0.3
              N = 121.1396
             df = 3
      sig.level = 0.05
          power = 0.8

NOTE: N is the number of observations
```

Effektdefinition für verschiedene Tests

χ^2-Test (Anpassung)



Effektdefinition für verschiedene Tests

Einfache lineare Regression

- Effekt: f2

- $f2 = r = \sqrt{R^2}$ (hier sollte der angepasste (adjusted) Wert genutzt werden)

pwr.f2.test(u =, v= , f2=, sig.level=, power =)

- **u** Freiheitsgrade Zähler
- **v** Freiheitsgrade Nenner
- **f2** Effektstärke
- **sig.level** α
- **power** $1 - \beta$

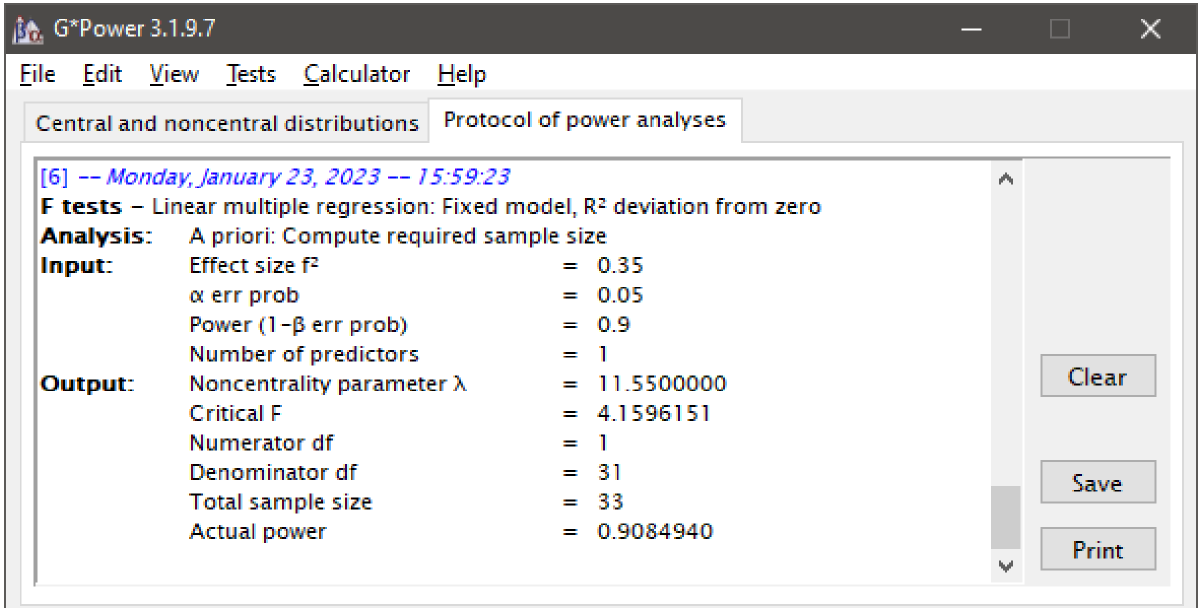
Effektdefinition für verschiedene Tests

Einfache lineare Regression

```
> pwr.f2.test(u=1, f2=0.35, sig.level =0.05, power=0.9)

     Multiple regression power calculation

              u = 1
              v = 30.06969
             f2 = 0.35
      sig.level = 0.05
          power = 0.9
```

Effektdefinition für verschiedene Tests

<u>Einfache lineare Regression</u>



Effektdefinition für verschiedene Tests

Multiple lineare Regression

- Effekt: f2

- $f2 = r = \sqrt{R^2}$ (hier sollte der angepasste (adjusted) Wert genutzt werden)

pwr.f2.test(u =, v= , f2=, sig.level=, power =)

- **u** Freiheitsgrade Zähler
- **v** Freiheitsgrade Nenner
- **f2** Effektstärke
- **sig.level** α
- **power** $1 - \beta$

Effektdefinition für verschiedene Tests

Korrelation

- Effekt: r

- $r = Korrelationskoeffizient$

pwr.r.test(n = , r = , sig.level = , power =)

- **n** Anzahl der Beobachtungen
- **r** Effektstärke
- **sig.level** α
- **power** $1 - \beta$

Effektdefinition für verschiedene Tests

Nicht-parametrische Tests (Wilcoxon)

Effekt: Cohen‘s D

pwr.t.test(n = , d = , sig.level= , power = , type = , alternative=)

- **n** Stichprobengröße
- **d** Effektstärke
- **sig.level** Signifikanzniveau
- **power** Teststärke
- **type** „one.sample“, „two.sample“, „paired“
- **alternative** “less“, “two.sided“, “greater“

(Gleiche Berechnung wie parametrisch)

Effektdefinition für verschiedene Tests

Nicht-parametrische Tests (Kruskal-Wallis)

- Effekt: f

- $f = \sqrt{\frac{\eta^2}{1-\eta^2}}$ mit $\eta^2 = \frac{SS_{zwischen}}{SS_{Gesamt}}$

pwr.anova.test(k = , n = , f = , sig.level = , power =)

- **k** Anzahl der Gruppen
- **n** Stichprobengröße (pro Gruppe)
- **f** Effektgröße
- **sig.level** α
- **power** $1 - \beta$

(Gleiche Berechnung wie parametrisch)

Effektdefinition für verschiedene Tests

Mehrfaktorielle Varianzanalyse (ANOVA)

- Effekt: f
- $f = \sqrt{\frac{\eta^2}{1-\eta^2}}$ mit $\eta^2 = \frac{SS_{zwischen}}{SS_{Gesamt}}$

wp.kanova(n = , ndf = , f = , ng = , alpha = , power =)

- **n** Stichprobengröße
- **ndf** Zählerfreiheitsgrade
- **f** Effektgröße
- **ng** Anzahl der Gruppen
- **alpha** α
- **power** $1 - \beta$

Effektdefinition für verschiedene Tests

Mehrfaktorielle Varianzanalyse (ANOVA)

```
> wp.kanova(ndf =2, f=0.25, ng=12, alpha=0.05, power=0.8)

Multiple way ANOVA analysis

           n ndf      ddf    f ng alpha power
    157.3764   2 145.3764 0.25 12  0.05   0.8

NOTE: Sample size is the total sample size
```

Effektdefinition für verschiedene Tests

Mehrfaktorielle Varianzanalyse (ANOVA)