# Inferenzstatistik

Population oder Grundgesamtheit, über die Aussage getroffen werden soll

# Inferenzstatistik

Population oder Grundgesamtheit, über die Aussage getroffen werden soll

Stichprobe statt Vollerhebung

# Inferenzstatistik

Population oder Grundgesamtheit, über die Aussage getroffen werden soll

Stichprobe statt Vollerhebung

Ziel: Auf der Basis der Stichprobendaten Aussagen über Grundgesamtheit treffen

# Inferenzstatistik

Population oder Grundgesamtheit, über die Aussage getroffen werden soll

Stichprobe statt Vollerhebung

Ziel: Auf der Basis der Stichprobendaten Aussagen über Grundgesamtheit treffen

Verfahren:

# Inferenzstatistik

Population oder Grundgesamtheit, über die Aussage getroffen werden soll

Stichprobe statt Vollerhebung

Ziel: Auf der Basis der Stichprobendaten Aussagen über Grundgesamtheit treffen

Verfahren:

- Schätzen (Punkt- und Intervall-Schätzer)

# Inferenzstatistik

Population oder Grundgesamtheit, über die Aussage getroffen werden soll

Stichprobe statt Vollerhebung

Ziel: Auf der Basis der Stichprobendaten Aussagen über Grundgesamtheit treffen

Verfahren:

- Schätzen (Punkt- und Intervall-Schätzer)
- Testen (Signifikanztest)

# Modell für Mittelwertvergleich

Grundgesamtheit kann durch ein Merkmal in zwei Teile (Subpopulationen) aufgeteilt werden

# Modell für Mittelwertvergleich

Grundgesamtheit kann durch ein Merkmal in zwei Teile (Subpopulationen) aufgeteilt werden

Fragestellung: Unterscheiden sich Verteilungen eines weiteren Merkmals der beiden Subpopulationen bezüglich der Lage?

# Modell für Mittelwertvergleich

Grundgesamtheit kann durch ein Merkmal in zwei Teile (Subpopulationen) aufgeteilt werden

Fragestellung: Unterscheiden sich Verteilungen eines weiteren Merkmals der beiden Subpopulationen bezüglich der Lage?

Ziehung von Stichproben aus diesen Subpopulationen

# Modell für Mittelwertvergleich

Grundgesamtheit kann durch ein Merkmal in zwei Teile (Subpopulationen) aufgeteilt werden

Fragestellung: Unterscheiden sich Verteilungen eines weiteren Merkmals der beiden Subpopulationen bezüglich der Lage?

Ziehung von Stichproben aus diesen Subpopulationen

Standardfall: Man unterstellt in der Grundgesamtheit eine Normalverteilung mit konstanter Varianz.
Lage der Verteilung bedeutet dann gleicher Mittelwert (richtiger: Erwartungswert)

# t-Test für zwei unabhängige Stichproben

# t-Test für zwei unabhängige Stichproben

$X_1, \ldots, X_m$ aus $\mathcal{N}(\mu_1, \sigma^2)$
$X_{m+1}, \ldots, X_n$ aus $\mathcal{N}(\mu_2, \sigma^2)$

# t-Test für zwei unabhängige Stichproben

$X_1, \ldots, X_m$ aus $\mathcal{N}(\mu_1, \sigma^2)$
$X_{m+1}, \ldots, X_n$ aus $\mathcal{N}(\mu_2, \sigma^2)$

$H_0$: $\mu_1 = \mu_2$ vs. $H_1$: $\mu_1 \neq \mu_2$

# t-Test für zwei unabhängige Stichproben

$X_1, \ldots, X_m$ aus $\mathcal{N}(\mu_1, \sigma^2)$
$X_{m+1}, \ldots, X_n$ aus $\mathcal{N}(\mu_2, \sigma^2)$

$H_0$: $\mu_1 = \mu_2$ vs. $H_1$: $\mu_1 \neq \mu_2$

Teststatistik:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_{\overline{X}}}$$

mit

$$\overline{X}_1 = \frac{1}{m} \sum_{i=1}^{m} X_i$$

$$\overline{X}_2 = \frac{1}{n-m} \sum_{i=m+1}^{n} X_i$$

$$S_{\overline{X}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2} \cdot \frac{1}{\sqrt{n}}$$

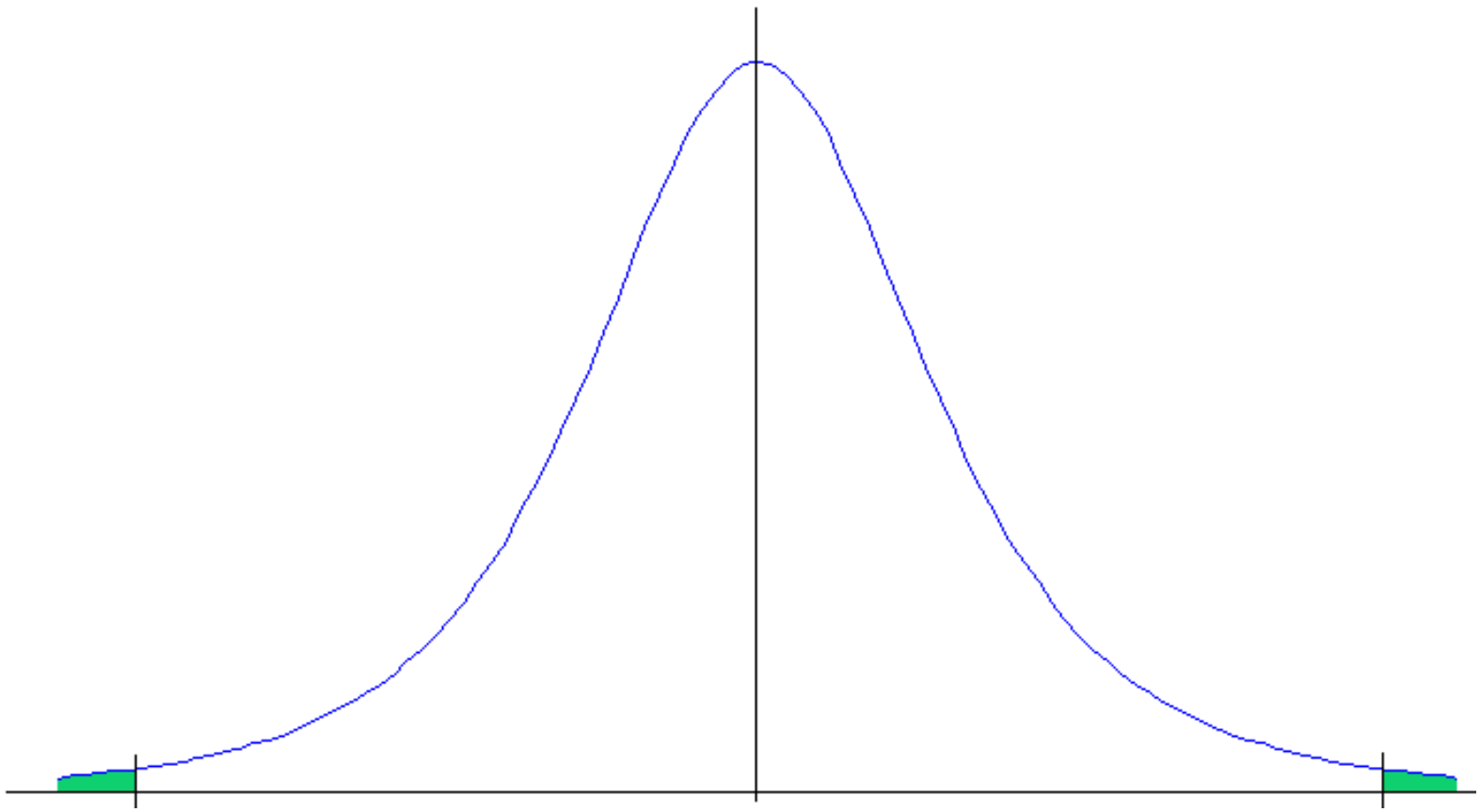
# t-Test für zwei unabhängige Stichproben

$t$ unter $H_0$ t-verteilt mit $n - 2$ Freiheitsgraden

# t-Test für zwei unabhängige Stichproben

$t$ unter $H_0$ t-verteilt mit $n - 2$ Freiheitsgraden

Entscheidung für $H_1$, falls $|t|$ zu groß



# T-Test mit Stata

```
cd p:/statistik/stata/daten
use data1
generate east = state >= 11 if !missing(state)
tabulate east, summarize(ybirth)
```

# T-Test mit Stata

```
cd p:/statistik/stata/daten
use data1
generate east = state >= 11 if !missing(state)
tabulate east, summarize(ybirth)
```

Aufteilung in West- und Ostdeutschland nach Bundesländern
Ausgabe von Mittelwert und Standardabweichung vom Geburtsjahr neben Häufigkeiten

# T-Test mit Stata

```
cd p:/statistik/stata/daten
use data1
generate east = state >= 11 if !missing(state)
tabulate east, summarize(ybirth)
```

Aufteilung in West- und Ostdeutschland nach Bundesländern
Ausgabe von Mittelwert und Standardabweichung vom Geburtsjahr neben Häufigkeiten

Ist der Mittelwert des Alters in beiden Subpopulationen gleich groß?

# T-Test mit Stata

```
cd p:/statistik/stata/daten
use data1
generate east = state >= 11 if !missing(state)
tabulate east, summarize(ybirth)
```

Aufteilung in West- und Ostdeutschland nach Bundesländern
Ausgabe von Mittelwert und Standardabweichung vom Geburtsjahr neben Häufigkeiten

Ist der Mittelwert des Alters in beiden Subpopulationen gleich groß?

Durchführung mit Kommando `ttest`

# T-Test mit Stata

```
ttest ybirth, by(east)
```

# T-Test mit Stata

ttest ybirth, by(east)

```
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |    4068    1959.828    .2844148    18.14023    1959.271    1960.386
       1 |    1343    1958.477    .4926276     18.0533    1957.511    1959.444
---------+--------------------------------------------------------------------
combined |    5411    1959.493    .2464185    18.12642     1959.01    1959.976
---------+--------------------------------------------------------------------
    diff |            1.350881    .5702133                .2330338    2.468729
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   2.3691
Ho: diff = 0                     degrees of freedom =     5409

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9911         Pr(|T| > |t|) = 0.0179          Pr(T > t) = 0.0089
```

# T-Test mit Stata

P-Wert= 0.0179

# T-Test mit Stata

P-Wert= 0.0179

$\Rightarrow H_0$: $\mu_1 = \mu_2$ kann zum 5%-Niveau verworfen werden

# T-Test mit Stata

P-Wert= 0.0179

$\Rightarrow H_0$: $\mu_1 = \mu_2$ kann zum 5%-Niveau verworfen werden

bei Problem mit ungleichen Standardabweichungen oder Varianzen in Subpopulationen

# T-Test mit Stata

P-Wert= 0.0179

$\Rightarrow H_0$: $\mu_1 = \mu_2$ kann zum 5%-Niveau verworfen werden

bei Problem mit ungleichen Standardabweichungen oder Varianzen in Subpopulationen

```
ttest ybirth, by(east) unequal
```

# Daten lesen und schreiben

```
cd p:/statistik/stata/daten
use data1
describe
browse
```

# Daten lesen und schreiben

```
cd p:/statistik/stata/daten
use data1
describe
browse
```

Datenmatrix, mit `browse` Darstellung ähnlich der in SPSS aber nicht bearbeitbar

# Daten lesen und schreiben

```
cd p:/statistik/stata/daten
use data1
describe
browse
```

Datenmatrix, mit `browse` Darstellung ähnlich der in SPSS aber nicht bearbeitbar

mit `edit` ist auch Bearbeiten möglich

# Daten lesen und schreiben

```
cd p:/statistik/stata/daten
use data1
describe
browse
```

Datenmatrix, mit `browse` Darstellung ähnlich der in SPSS aber nicht bearbeitbar

mit `edit` ist auch Bearbeiten möglich

Alt + F4 Schließen des browse-Fensters

# Import maschinenlesbarer Daten

Typ kann i.d.R. anhand Extension festgestellt werden

# Import maschinenlesbarer Daten

Typ kann i.d.R. anhand Extension festgestellt werden

| | |
|---|---|
| .dta | Stata |
| .odf | Open Office Calc |
| .rda, .rdata | R |
| .sav | SPSS |
| .xls, .xlsx | Excel |
| .xpt | SAS (portable) |

# Import maschinenlesbarer Daten

Typ kann i.d.R. anhand Extension festgestellt werden

| | |
|---|---|
| .dta | Stata |
| .odf | Open Office Calc |
| .rda, .rdata | R |
| .sav | SPSS |
| .xls, .xlsx | Excel |
| .xpt | SAS (portable) |

`type` *dateiname*

# Import maschinenlesbarer Daten

Typ kann i.d.R. anhand Extension festgestellt werden

| | |
|---|---|
| .dta | Stata |
| .odf | Open Office Calc |
| .rda, .rdata | R |
| .sav | SPSS |
| .xls, .xlsx | Excel |
| .xpt | SAS (portable) |

```
type dateiname
```

```
type popst1.dta
type popst1.xls
type popst1.raw
```

# Lesen von Excel-Dateien

Struktur in Excel-Datei kann vielfältig und nicht unmittelbar passend sein!

# Lesen von Excel-Dateien

Struktur in Excel-Datei kann vielfältig und nicht unmittelbar passend sein!

Importfilter für Excel

# Lesen von Excel-Dateien

Struktur in Excel-Datei kann vielfältig und nicht unmittelbar passend sein!

Importfilter für Excel

```
import excel popst1, clear
describe
```

# Lesen von Excel-Dateien

Struktur in Excel-Datei kann vielfältig und nicht unmittelbar passend sein!

Importfilter für Excel

```
import excel popst1, clear
describe
```

Variablen sind Strings wegen Beschreibung in erster Zeile

# Lesen von Excel-Dateien

Struktur in Excel-Datei kann vielfältig und nicht unmittelbar passend sein!

Importfilter für Excel

```
import excel popst1, clear
describe
```

Variablen sind Strings wegen Beschreibung in erster Zeile

```
import excel popst1, clear firstrow
```

# Lesen von Excel-Dateien

Struktur in Excel-Datei kann vielfältig und nicht unmittelbar passend sein!

Importfilter für Excel

```
import excel popst1, clear
describe
```

Variablen sind Strings wegen Beschreibung in erster Zeile

```
import excel popst1, clear firstrow
```

`rename` zum Umbenennen der Variablen verwendebar

# Lesen von Excel-Dateien

Variablen- und Bereichsangabe möglich:

# Lesen von Excel-Dateien

Variablen- und Bereichsangabe möglich:

```
import excel state area pop_total pop_male
   pop_female using popst1, clear cellrange(A2:E17)
```

# Lesen von Excel-Dateien

Variablen- und Bereichsangabe möglich:

```
import excel state area pop_total pop_male
   pop_female using popst1, clear cellrange(A2:E17)
```

Stata kann ohne Zusatztools keine SPSS-Datendatei (*.sav*) lesen!

# Lesen von Excel-Dateien

Variablen- und Bereichsangabe möglich:

```
import excel state area pop_total pop_male
   pop_female using popst1, clear cellrange(A2:E17)
```

Stata kann ohne Zusatztools keine SPSS-Datendatei (*.sav*) lesen!

Aber SPSS kann Daten in Stata-Format (*.dta*) speichern.

# Import von ASCII-Daten

exemplarisch mit csv-Datei

# Import von ASCII-Daten

exemplarisch mit csv-Datei

```
type popst1.raw
```

# Import von ASCII-Daten

exemplarisch mit csv-Datei

```
type popst1.raw
```

Komma als Trennzeichen zwischen Variablenwerten (entspricht Voreinstellung)

# Import von ASCII-Daten

exemplarisch mit csv-Datei

```
type popst1.raw
```

Komma als Trennzeichen zwischen Variablenwerten (entspricht Voreinstellung)

```
import delimited using popst1.raw, clear
```

# Import von ASCII-Daten

exemplarisch mit csv-Datei

```
type popst1.raw
```

Komma als Trennzeichen zwischen Variablenwerten (entspricht Voreinstellung)

```
import delimited using popst1.raw, clear
```

Angabe der Variablennamen vor `using` möglich

# Import von ASCII-Daten

exemplarisch mit csv-Datei

```
type popst1.raw
```

Komma als Trennzeichen zwischen Variablenwerten (entspricht Voreinstellung)

```
import delimited using popst1.raw, clear
```

Angabe der Variablennamen vor `using` möglich

oder mit `rename`

```
rename (v1 v2) (state area)
```

# Direkte Dateneingabe

Zunächst mit `clear` Arbeitsspeicher leeren

# Direkte Dateneingabe

Zunächst mit `clear` Arbeitsspeicher leeren

Methode 1:

# Direkte Dateneingabe

Zunächst mit `clear` Arbeitsspeicher leeren

Methode 1:mit

```
edit
```

Editorfenster öffnen und hier Eingabe vornehmen

# Direkte Dateneingabe

Zunächst mit `clear` Arbeitsspeicher leeren

Methode 1:mit

```
edit
```

Editorfenster öffnen und hier Eingabe vornehmen

Methode 2:

# Direkte Dateneingabe

Zunächst mit `clear` Arbeitsspeicher leeren

Methode 1:mit

```
edit
```

Editorfenster öffnen und hier Eingabe vornehmen

Methode 2: `input`-Befehl verwenden

# Direkte Dateneingabe

Zunächst mit `clear` Arbeitsspeicher leeren

Methode 1:mit

```
edit
```

Editorfenster öffnen und hier Eingabe vornehmen

Methode 2: `input`-Befehl verwenden

```
input year temp
 1. 1984 10.3
 2. 1985 9.7
 3. end
```