

5. Hypothesentests - metrische Variable

5.1. Einstichproben t-test

Es wird getestet, ob sich der Mittelwert \bar{x} einer Stichprobe von einem vorgegebenen Wert μ_0 unterscheidet. Bei kleinen Stichproben ($n < 100$) müssen die Daten aus normalverteilter Grundgesamtheit stammen. Ist dies nicht der Fall, sollte man auf den **Wilcoxon-Test** ausweichen.

Laden Sie die Datei geburtsgewicht.dta in STATA (*Datenquelle: Heinecke, Hultsch, Repges: Medizinische Biometrie. Springer 1992, S.162*)

Im Mittel beträgt das Geburtsgewicht gesunder Neugeborener 3500g. Es soll anhand einer Stichprobe von 20 Neugeborenen einer Klinik geprüft werden, ob das auch in deren Einzugsbereich zutrifft.

A. Prüfung auf Normalverteilung (da $n = 20$) mit Shapiro-Wilk Test:

swilk gebgew

```
. swilk gebgew
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
gebgew	20	0.95425	1.083	0.160	0.43626

Die Daten können als normalverteilt angenommen werden.

B. 1-Stichproben t-Test mit STATA:

ttest gebgew = 3500

```
. ttest gebgew = 3500
```

One-sample t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]
gebgew	20	3726.5	102.7088	459.3276	3511.528 3941.472

mean = mean(gebgew)

t = 2.2053

Ho: mean = 3500

degrees of freedom = 19

Ha: mean < 3500

Ha: mean != 3500

Ha: mean > 3500

Pr(T < t) = 0.9800

Pr(|T| > |t|) = 0.0400

Pr(T > t) = 0.0200

Ergebnis: Das mittlere Geburtsgewicht im Einzugsbereich der Klinik weicht deutlich ab von 3500g ($p = 0,04$), es liegt mit 3726,5g signifikant höher.

Bei kleinen Stichproben ($n < 50$, besser $n < 100$) ist die Teststatistik t-verteilt mit $(n-1)$ Freiheitsgraden. Dies gilt aber nur, wenn die Stichprobendaten aus einer normalverteilten Grundgesamtheit stammen.

5.2. Wilcoxon - Test

Ist bei kleinen Stichproben keine Normalverteilung gegeben, kann der Wilcoxon - Test weiterhelfen. Zur Demonstration sollen die 20 Geburtsgewichte aus dem obigen Beispiel dienen, obwohl hier Normalverteilung angenommen werden kann. Die Nullhypothese, dass das mittlere Geburtsgewicht im Einzugsbereich der Klinik 3500g beträgt, kann mit diesem Test auf dem 5% Niveau nicht verworfen werden ($p = 0,0619$). Durch die geringeren Voraussetzungen ist auch die Leistungsfähigkeit dieses Tests geringer.

signrank gebgew = 3500

```
. signrank gebgew = 3500
```

```
Wilcoxon signed-rank test
```

sign	obs	sum ranks	expected
positive	13	155	105
negative	7	55	105
zero	0	0	0
all	20	210	210

```
unadjusted variance      717.50
adjustment for ties      -0.13
adjustment for zeros      0.00
-----
adjusted variance        717.38
```

```
Ho: gebgew = 3500
```

```
z = 1.867
Prob > |z| = 0.0619
```

5.3. Vorzeichentest

Die geringsten Anforderungen an die Daten stellt der Vorzeichentest, bei dem die Vorzeichen der Differenzen zwischen den einzelnen Daten und dem vorgegebenen Wert betrachtet werden. Unter der Nullhypothese müssen gleich viele positive wie negative Abweichungen auftreten, d.h., der Median der Geburtsgewichte wäre gleich 3500g oder anders ausgedrückt: $H_0 : \text{Median}(\text{gebgew} - 3500) = 0$

signtest gebgew = 3500

```
. signtest gebgew = 3500
```

```
Sign test
```

sign	observed	expected
positive	13	10
negative	7	10
zero	0	0
all	20	20

```
One-sided tests:
```

```
Ho: median of gebgew - 3500 = 0 vs.
Ha: median of gebgew - 3500 > 0
Pr(#positive >= 13) =
  Binomial(n = 20, x >= 13, p = 0.5) = 0.1316
```

```
Ho: median of gebgew - 3500 = 0 vs.
Ha: median of gebgew - 3500 < 0
Pr(#negative >= 7) =
  Binomial(n = 20, x >= 7, p = 0.5) = 0.9423
```

```
Two-sided test:
```

```
Ho: median of gebgew - 3500 = 0 vs.
Ha: median of gebgew - 3500 != 0
Pr(#positive >= 13 or #negative >= 13) =
  min(1, 2*Binomial(n = 20, x >= 13, p = 0.5)) = 0.2632
```

Der hier durchgeführte exakte Binomialtest ist nicht signifikant, d.h., 13 positive Vorzeichen sind noch durch den Zufall erklärbar auf dem 5% Niveau.


```
. sdtest ht, by(sex)
```

Variance ratio test

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
1	44	165.8068	.8834777	5.860328	164.0251	167.5885
2	58	177.3034	.8281751	6.307194	175.6451	178.9618
combined	102	172.3441	.8273131	8.355452	170.703	173.9853

ratio = sd(1) / sd(2) f = 0.8633
 Ho: ratio = 1 degrees of freedom = 43, 57

Ha: ratio < 1 Ha: ratio != 1 Ha: ratio > 1
 Pr(F < f) = 0.3097 2*Pr(F < f) = 0.6195 Pr(F > f) = 0.6903

Varianzgleichheit kann angenommen werden (p = 0,6195).

5.5. Varianzanalyse (ANOVA)

Datei lung1984_gr.dta

Für den Test auf Unterschied der Mittelwerte zwischen mehr als 2 Gruppen: ANOVA

oneway ht gr, bonferroni tabulate

```
. oneway ht gr, bonferroni tabulate
```

gr	Summary of Height (cm)		
	Mean	Std. Dev.	Freq.
1	166.452	6.0124265	25
2	169.06	9.1598768	25
3	177.008	6.7435726	25
4	176.52222	5.9672398	27
Total	172.34412	8.3554523	102

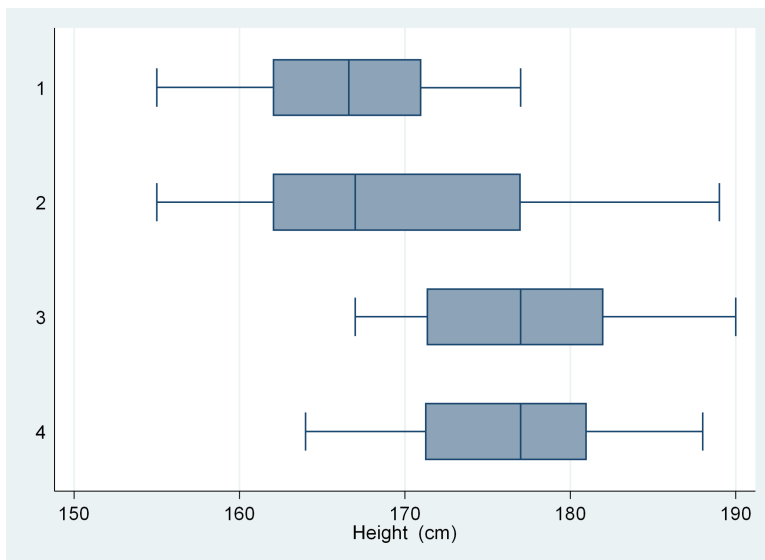
Source	Analysis of Variance				
	SS	df	MS	F	Prob > F
Between groups	2152.68394	3	717.561314	14.36	0.0000
Within groups	4898.48796	98	49.9845711		
Total	7051.17191	101	69.8135832		

Bartlett's test for equal variances: chi2(3) = 6.3255 Prob>chi2 = 0.097

Comparison of Height (cm) by gr (Bonferroni)			
Row Mean- Col Mean	1	2	3
2	2.608 1.000		
3	10.556 0.000	7.948 0.001	
4	10.0702 0.000	7.46222 0.001	-.485778 1.000

Die Voraussetzungen für Varianzgleichheit sind erfüllt. Bartlett's Test liefert: p = 0,097.
 Die ANOVA - Tabelle zeigt einen signifikanten Unterschied zwischen den 4 Gruppen, aber

nicht, zwischen welchen Gruppen. Die Tabelle der Mehrfachvergleiche (adjustiert nach Bonferroni) zeigt signifikante Unterschiede zwischen den Gruppen 1-3 , 1-4 , 2-3 und 2-4, wie man es aufgrund der Mittelwerte vermuten könnte. Mit dem Kommando **graph hbox ht, over(gr)** erhält man folgende Box-Plots:



6. Hypothesentests - dichotome Variable

6.1. Chi² - Unabhängigkeitstest

Mit einer Zufallsstichprobe werden im Rahmen einer Studie u.a. zwei dichotome Merkmale simultan erfasst, z.B. Geschlecht (w[1],m[2]) und Karies (ja = 1, nein = 0). Bei solchen Studien kann nur die Fallzahl n frei gewählt werden (Fallzahlplanung), die Randsummen der Vierfeldertafel $(a+b)$, $(a+c)$, $(c+d)$ und $(b+d)$ ergeben sich zufällig. Man möchte die Frage beantworten, ob Karies mit dem Geschlecht assoziiert ist oder ob beide Merkmale voneinander unabhängig sind (geschl_karies.dta). Hierzu dient der Chi²-Unabhängigkeitstest.

tabulate geschl karies, chi2 exact

a)

```
. tabulate geschl karies, chi2 exact colsort
```

geschl	karies		Total
	1	0	
w	27	14	41
m	26	12	38
Total	53	26	79

```

Pearson chi2(1) = 0.0589 Pr = 0.808
Fisher's exact = 1.000
1-sided Fisher's exact = 0.499

```

Einzeldaten im Editor

geschl	karies
2	0
1	0
2	1
2	0
2	1
2	0
1	1
-	-
-	-
-	-

Die beiden Merkmale sind voneinander unabhängig. Der p-Wert ist nicht signifikant. STATA bietet dazu noch den exakten Test nach Fisher für kleine Fallzahlen.

Neben **tabulate geschl karies** für die Einzeldaten gibt es zwei weitere Berechnungsmöglichkeiten:

b) Eingabe der Daten mit GewichtsvARIABLE n:

tabulate geschl karies [fweight = n], chi2 exact

mit

fweight = frequency weights

liefert gleiches Ergebnis wie a)

geschl	karies	n
1	1	27
1	0	14
2	1	26
2	0	12

c) Eingabe der Häufigkeiten der 2x2 - Tafel über den

Table calculator:

tabi 27 14 \ 26 12, chi2 exact

liefert gleiches Ergebnis wie a)

6.2. Chi² - Homogenitätstest

Vergleicht man zwei Zufallsstichproben, z.B. zwei Therapiegruppen hinsichtlich einer dichotomen Charakteristik (z.B. gestorben, überlebt), so ergibt sich die Fragestellung, ob die Anteile der z.B. Gestorbenen in beiden Gruppen gleich sind.

Als Beispiel soll die bekannte APSAC - Studie dienen, bei der für Patienten mit akutem Herzinfarkt ein neuer Blutgerinnungshemmer APSAC (*Anisoylated Plasminogen Streptokinase Activator Complex*) mit der damaligen Heparin - Standardtherapie verglichen wurde. Die Anzahl der Teilnehmer in den zwei Gruppen (Zeilensummen) wird vorgegeben. Hauptzielkriterium war die Verminderung der Hospitalmortalität (Sterblichkeit) innerhalb der ersten 28 Tage nach Infarkt. Die Studiendaten sind in der folgenden 2x2 - Tafel zusammengefasst. Unter APSAC versterben 9 von 162 Patienten (5,6%), unter Heparin 19 von 151 (12,6%). Es kann wieder die Chi²-Teststatistik (wie bei Punkt 6.1.b) angewendet werden, die den Wert Chi² = 4,7381 liefert. Der Unterschied ist gerade signifikant (p = 0,03) und deutet darauf hin, dass die Mortalität unter APSAC geringer ist, als unter Heparin. Der exakte Test liefert ein übereinstimmendes, wenn auch knapperes Ergebnis.

	gestorben	überlebt	
APSAC	9	153	162
Heparin	19	132	151
	28	285	313

Datenquelle: Schumacher, Schulgen:
Methodik klinischer Studien. Springer 2002
(APSAC_gewichtet.dta)

a)

tabulate Medikament Ereignis [fweight = n], chi2 exact

`. tabulate Medikament Ereignis [fweight = n], chi2 exact`

Medikament	Ereignis		Total
	gestorben	überlebt	
APSAC	9	153	162
Heparin	19	132	151
Total	28	285	313

```

Pearson chi2(1) = 4.7381 Pr = 0.030
Fisher's exact = 0.046
1-sided Fisher's exact = 0.023
    
```

b) Eine weitere Möglichkeit bietet der 2-Stichprobentest für Anteilswerte der Gestorbenen. Verglichen werden APSAC (x) mit 9 von 162 Patienten (5,6%) und Heparin (y) mit 19 von 151 Patienten (12,6%).

prtesti 162 9 151 19, count

```
. prtesti 162 9 151 19, count
```

Two-sample test of proportions

x: Number of obs = 162
y: Number of obs = 151

	Mean	Std. Err.	z	P> z	[95% Conf. Interval]
x	.0555556	.0179968			.0202825 .0908286
y	.1258278	.0269897			.0729289 .1787267
diff	-.0702723	.0324396			-.1338527 -.0066918
	under Ho:	.0322837	-2.18	0.030	

diff = prop(x) - prop(y) z = -2.1767
Ho: diff = 0

Ha: diff < 0 Pr(Z < z) = 0.0148
Ha: diff != 0 Pr(|Z| > |z|) = 0.0295
Ha: diff > 0 Pr(Z > z) = 0.9852

Beide Methoden ergeben ein signifikantes Ergebnis und man erkennt außerdem wegen $\chi^2 = z^2$, hier: $(4,738) = (-2.1767)^2$ die Äquivalenz der beiden Teststatistiken.

6.3. McNemar Test

Werden zwei Beobachtungen eines dichotomen Merkmals (1,0) an den gleichen Individuen oder Objekten durchgeführt, so sind sie nicht unabhängig (sie sind **verbunden**). Es wurden z.B. 40 Personen hinsichtlich einer Eigenschaft (vorhanden = 1, nicht vorhanden = 0) von 2 Personen (A und B) beurteilt. Es ergaben sich die in der Tabelle angegebenen Häufigkeiten (mc_nemar.dta).

Der Anteil „1“ beträgt bei A 17/40 und bei B 19/40. Wegen $b+c = 8$ wird der exakte Test verwendet.

Ergebnis: In der Grundgesamtheit sind die Beurteilungen gleich ($p = 0,7266$)

```
. symmetry A B [fweight=n], exact
```

A	B		Total
	0	1	
0	18	5	23
1	3	14	17
Total	21	19	40

	chi2	df	Prob>chi2
Symmetry (asymptotic)	0.50	1	0.4795
Marginal homogeneity (Stuart-Maxwell)	0.50	1	0.4795
Symmetry (exact significance probability)			0.7266

Eine weitere Möglichkeit bietet das Kommando

mcc A B [fweight = n]

aus dem Stata-Menü: *Statistics > Epidemiology and related > Tables for epidemiologists > Matched case-control studies*

mit den Bezeichnungen: A = Cases, B = Controls und 1 = Exposed, 0 = Unexposed

```
. mcc A B [fweight = n]
```

Cases	Controls		Total
	Exposed	Unexposed	
Exposed	14	3	17
Unexposed	5	18	23
Total	19	21	40

```
McNemar's chi2(1) =      0.50    Prob > chi2 = 0.4795
Exact McNemar significance probability = 0.7266
```

oder nach Eingabe der aggregierten Daten in den "Matched case-control calculator" des gleichen Stata-Menüs

mcci 14 3 5 18

```
. mcci 14 3 5 18
```

Cases	Controls		Total
	Exposed	Unexposed	
Exposed	14	3	17
Unexposed	5	18	23
Total	19	21	40

```
McNemar's chi2(1) =      0.50    Prob > chi2 = 0.4795
Exact McNemar significance probability = 0.7266
```

In allen Fällen erhält man das gleiche Ergebnis.