

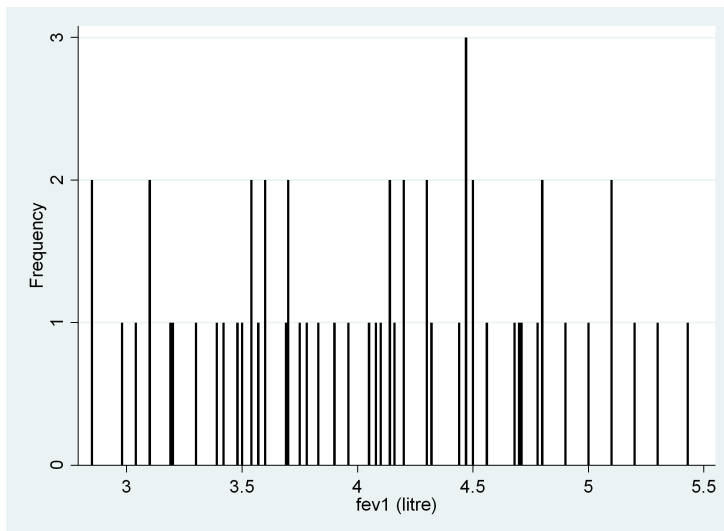
4. Grafiken

4.1. Metrische Daten - Histogramm

Zur Darstellung (Grafik oder Tabelle) der **Häufigkeitsverteilung** metrischer Daten ist oft eine Klassifizierung hilfreich. Als Demo laden wir die Datei fev.dta mit 57 Fällen ins Programm und stellen mit

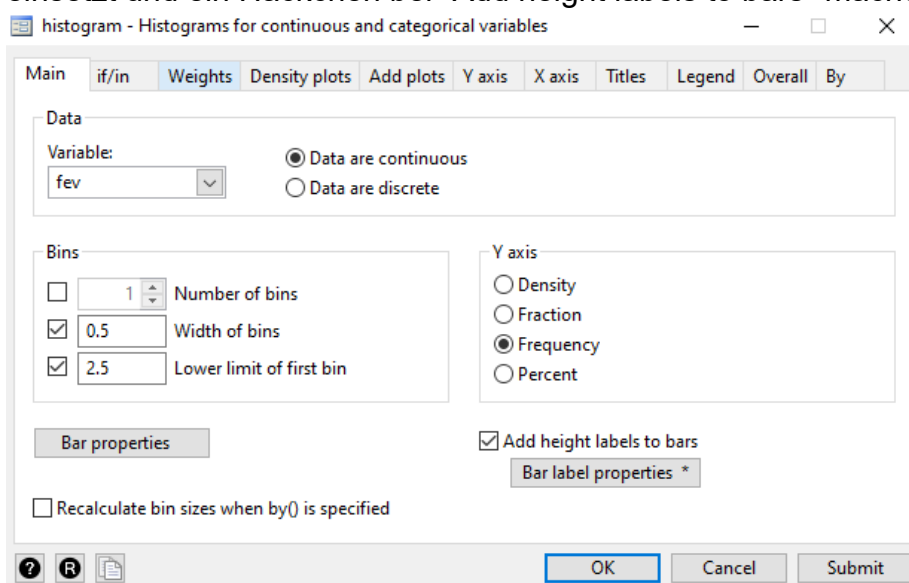
histogram fev, discrete frequency

die Häufigkeitsverteilung der einzelnen Meßpunkte dar. Diese Grafik ist wenig aussagekräftig. 32 Werte kommen 1x, 11 Werte 2x und 1 Wert kommt 3x vor.



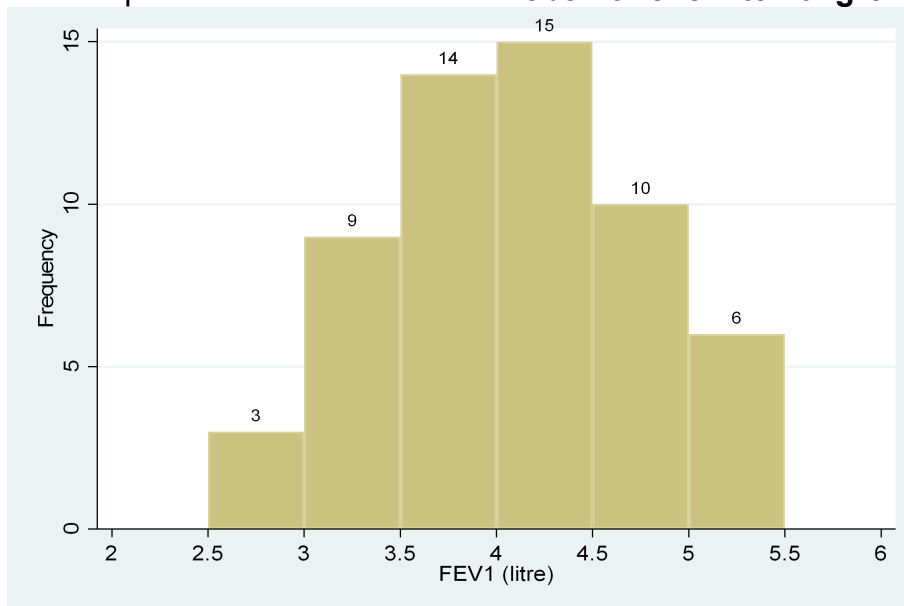
Daher wird eine Klassierung vorgenommen. Für 57 Werte erhält man als Anhaltspunkt $\sqrt{57} = 7,5 \approx 7$ Klassen. Mit **sum fev** erhält man ua.: Min = 2,85 und Max = 5,43 .

Die untere Grenze der ersten Klasse sei nun 2,5 , die obere Grenze der letzten Klasse 5,5. Die Differenz beträgt 3,0. Teilt man in 6 Klassen, so erhält man eine Klassenbreite von 0,5. Im Menü *Graphics* >> *Histogram* erhält man folgende Dialogbox, in die man diese Werte einsetzt und ein Häkchen bei "Add height labels to bars" machen kann.



Die Erzeugung der folgenden Grafik über das Menü entspricht dem Kommando **histogram fev, width(0.5) start(2.5) frequency addlabel**

Die entsprechenden Klassen haben **oben offene Intervallgrenzen**.



Die Annahme einer Normalverteilung kann z.B. mit dem Shapiro-Wilk-Test geprüft werden, der die Normalverteilungsannahme hier auch bestätigt.

```
. swilk fev
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
fev	57	0.97832	1.131	0.264	0.39576

Durch Erzeugung einer neuen Var **fevcat** mit den folgenden Kommandos gelingt die numerische (tabellarische) Darstellung der Klassen mit **oben offenen Intervallgrenzen**. Diese Kommandofolge kann auch in einem Do-File (siehe "help doedit") gespeichert werden.

```
generate fevcat=3 if fev<3
replace fevcat=3.5 if fev>=3 & fev<3.5
replace fevcat=4 if fev>=3.5 & fev<4
replace fevcat=4.5 if fev>=4 & fev<4.5
replace fevcat=5 if fev>=4.5 & fev<5
replace fevcat=5.5 if fev>=5 & fev<5.5
```

```
. tabulate fevcat
```

fevcat	Freq.	Percent	Cum.
3	3	5.26	5.26
3.5	9	15.79	21.05
4	14	24.56	45.61
4.5	15	26.32	71.93
5	10	17.54	89.47
5.5	6	10.53	100.00
Total	57	100.00	

Oder man verwendet das etwas längere Kommando

recode fev (2.5/2.99=3) (3.0/3.49=3.5) (3.5/3.99=4) (4/4.49=4.5) (4.5/4.99=5) (5/5.5=5.5),
gen(fevcat)

Dabei bleibt die Breite der Intervalle konstant bei 0,5

Möchte man nach **unten offene Intervalle** verwenden, so lautet die Kommandofolge für eine Klasseneinteilung:

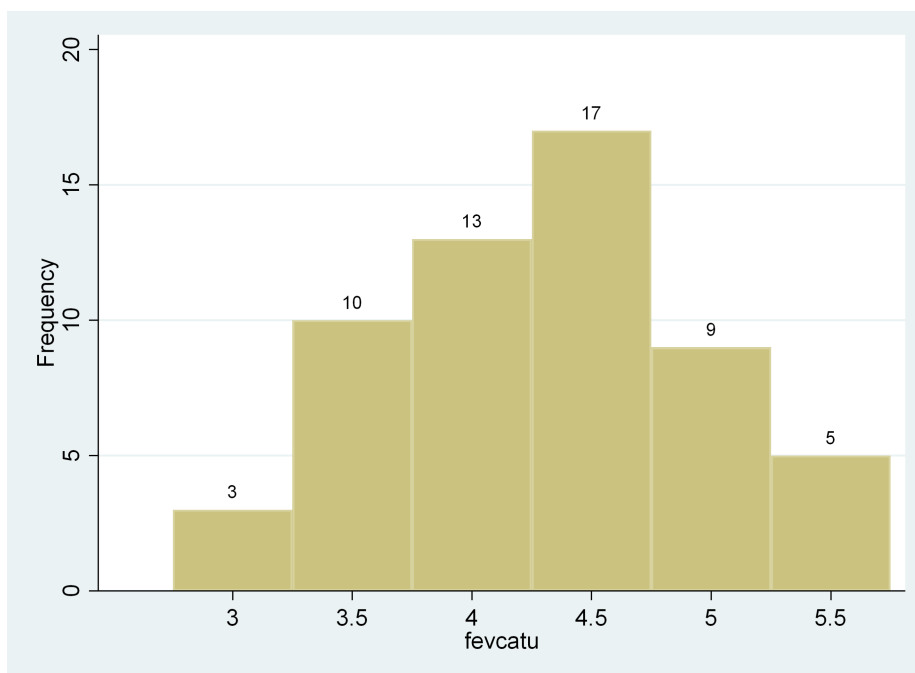
generate fevcatu=3 if fev<=3
replace fevcatu=3.5 if fev>3 & fev<=3.5
replace fevcatu=4 if fev>3.5 & fev<=4
replace fevcatu=4.5 if fev>4 & fev<=4.5
replace fevcatu=5 if fev>4.5 & fev<=5
replace fevcatu=5.5 if fev>5 & fev<=5.5

Oder man verwendet alternativ das Kommando:

generate fevcatu= **autocode**(fev,6, 2.5, 5.5)

mit dem man die Werte automatisch in 6 gleich große, nach unten offene Intervalle, einteilt zwischen dem Minimum 2,5 und dem Maximum 5,5. Dies führt allerdings zu einer leichten Veränderung der grafischen Darstellung im Histogramm mit dem Kommando:

histogram fevcatu, **discrete width**(0.5) **start**(3) **frequency** **addlabel**



tab fevcat

fevcat	Freq.	Percent	Cum.
3	3	5.26	5.26
3.5	10	17.54	22.81
4	13	22.81	45.61
4.5	17	29.82	75.44
5	9	15.79	91.23
5.5	5	8.77	100.00
Total	57	100.00	

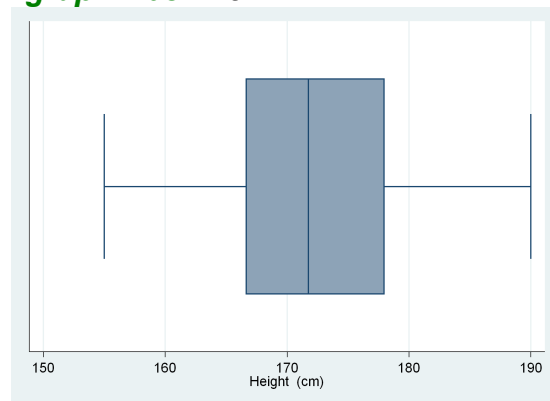
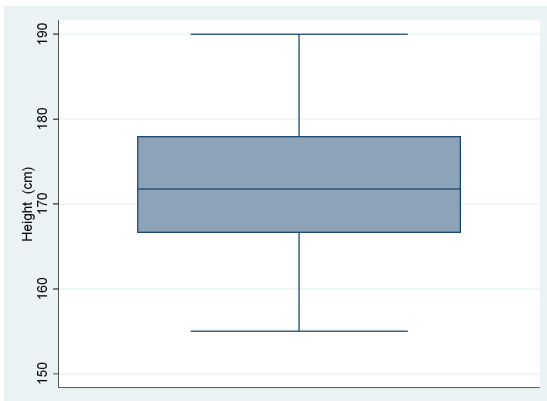
4.2. Metrische Daten - Boxplot

Zur Demonstration laden wir die Datei lung1984.dta in Stata. Die Variable **ht** beinhaltet die Größe von 102 Personen. Mit

graph box ht

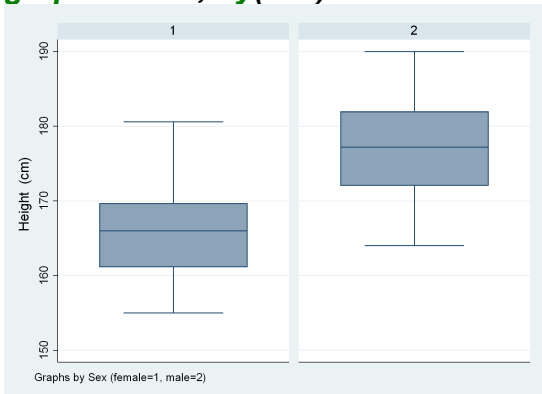
erhält man einen vertikalen Boxplot und mit

graph hbox ht einen horizontalen Boxplot.

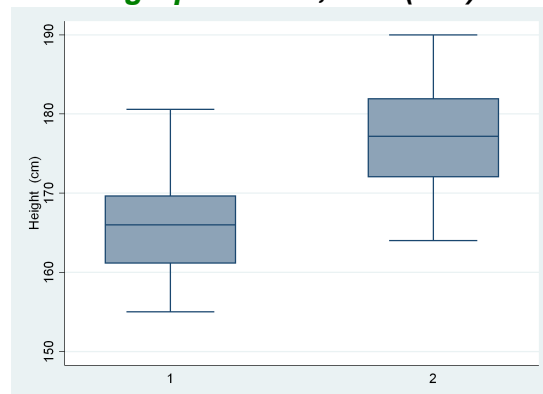


Die Option **by(sex)** dient zur Visualisierung von Gruppenvergleichen, hier dem Geschlecht.

graph box ht, by(sex)



graph box ht, over(sex)



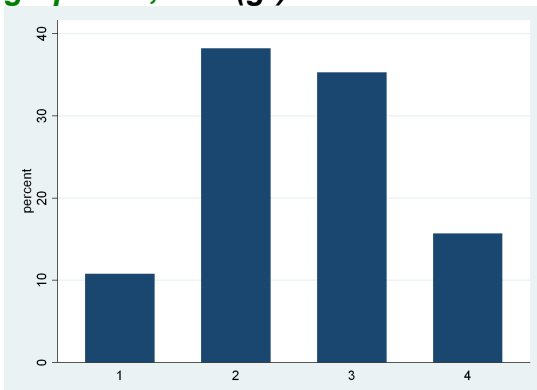
Die Option **over(sex)** liefert ein analoges Ergebnis in optischer Variante.

Viele weitere Möglichkeiten zur Darstellung von Boxplots und Histogrammen findet man unter **help graph box** und **help histogram** in der Hilfe von Stata.

4.3. Kategoriale Daten - Balken- und Tortendiagramm

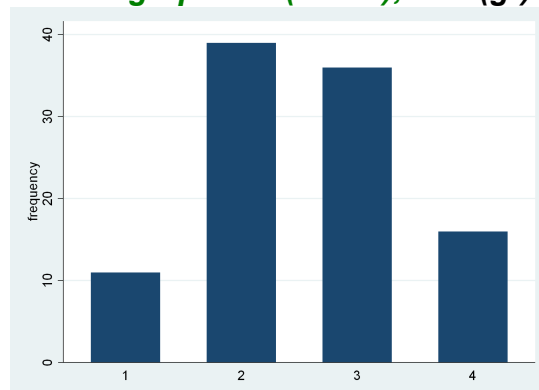
Wir öffnen die Datei lung1984_gr.dta und betrachten hier die Häufigkeit der 4 Ausprägungen 1, 2, 3, 4 der Variablen **gr**.

graph bar, over(gr) liefert:



jeweils mit Prozentangabe

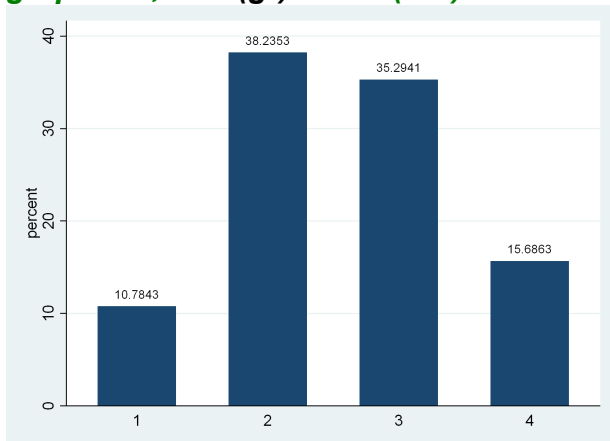
graph bar (count), over(gr)



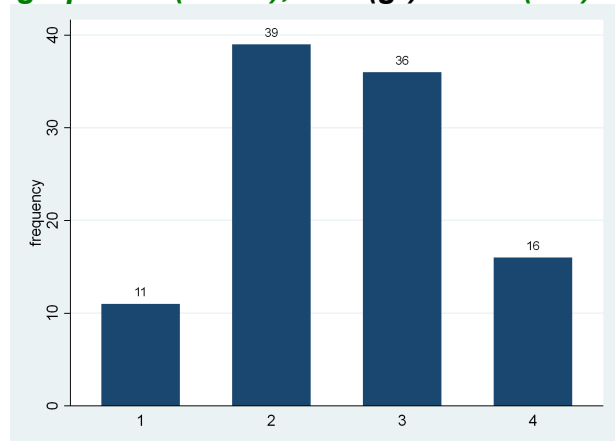
oder absoluter Häufigkeit

Eine Beschriftung der Balken liefert die Option ***blabel(bar)***

graph bar, over(gr) blabel(bar)



graph bar (count), over(gr) blabel(bar)

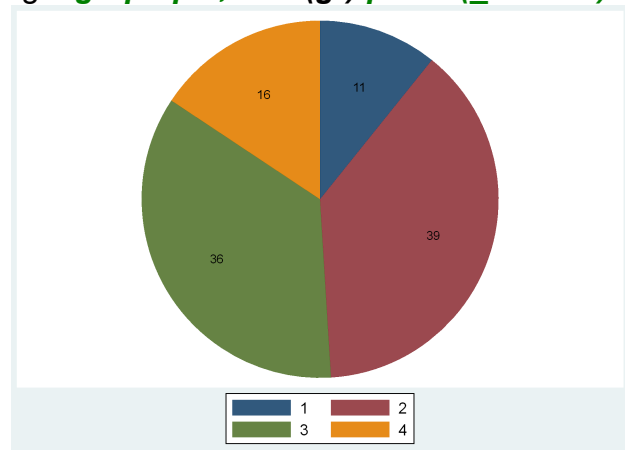
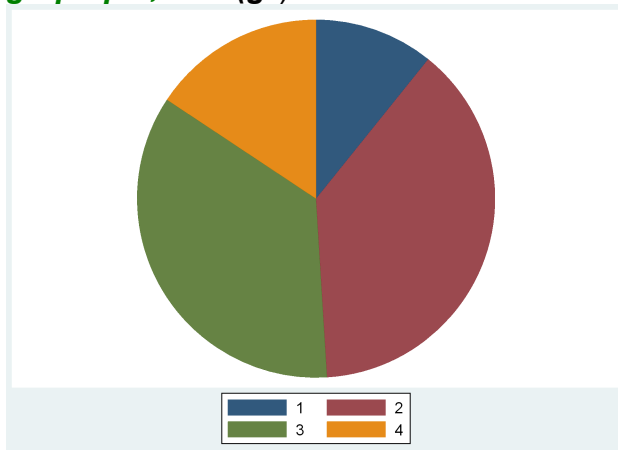


Tortendiagramme werden mit dem Kommando ***graph pie, over(gr)*** aufgerufen.

graph pie, over(gr)

oder mit Beschriftung

graph pie, over(gr) plabel(_all sum)

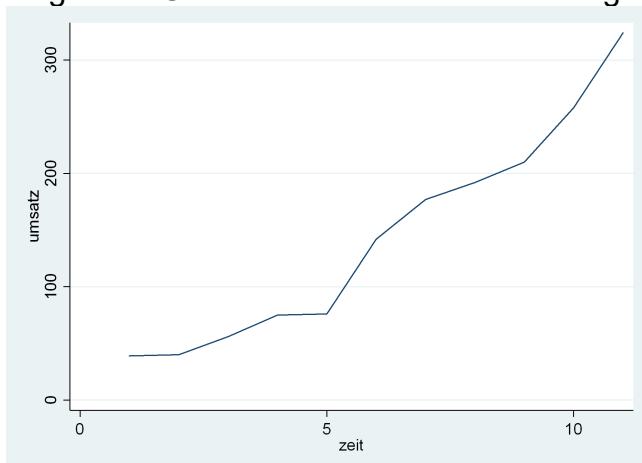


Weitere umfangreiche Bearbeitungen von Grafiken ermöglicht der **Grafik Editor**. Einige Beispiele sind im Folgenden unter 4.4. aufgeführt für die Var ***gr*** aus `lung1984_gr.dta`. Mit ***graph bar, over(gr)*** erscheint die Balkengrafik im Grafikfenster (oben links).

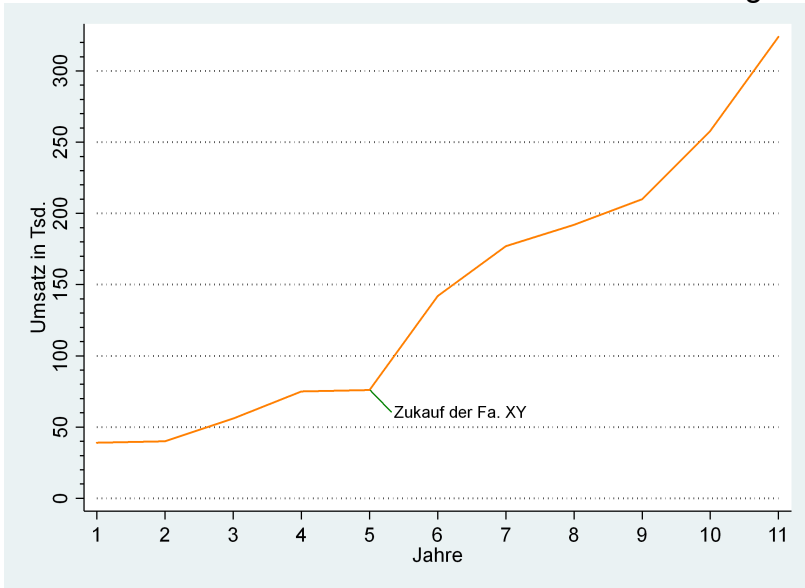
4.4 Liniendiagramm

Zur Demonstration laden wir die Datei `umsatz_zeit.dta` in Stata. Mit


graph twoway line umsatz zeit erhält man ein einfaches Liniendiagramm für die Entwicklung eines Umsatzes mit der Zeit. Das Diagramm erhält man auch mit ***line umsatz zeit***.



Mit dem Grafikeditor lassen sich diverse Veränderungen einfügen.



4.5. Der Grafikeditor

Öffnen des Grafikeditors mit *File >> Start Graph Editor* oder Klick auf 

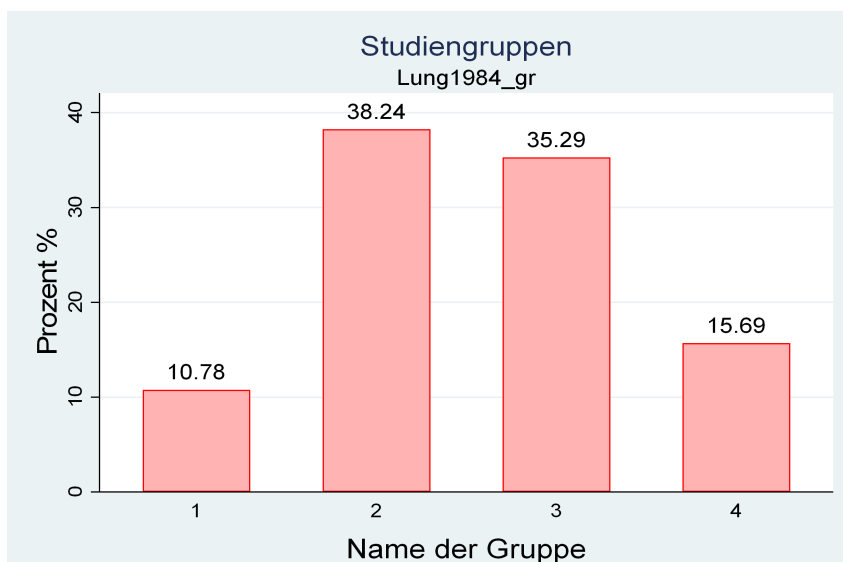
5

Bezeichnungen:

- 1 Menüleiste
- 2 Symbolleiste, Toolbar (2)
- 3 Kontextabhängige Toolbar (3)
- 4 Werkzeuge
- 5 Grafik
- 6 Objektabhängige Einstellungen

Das Menü Graph bietet ebenso wie die Toolbar (3) und Objects (6) vielfältige Möglichkeiten, das Aussehen der Grafik zu verändern. Einige Beispiele:

- zur Änderung der Beschriftung der y-Achse klicken wir auf "percent" und ändern im Textfeld von Toolbar (3) den Eintrag in "Prozent %". Danach die Eingabetaste.
- zum Hinzufügen von Überschriften klicken wir bei Objekt (6) auf "title" und schreiben in das Textfeld von Toolbar (3) "Studiengruppen" und unter "subtitle" "lung1984_gr". Danach drücken wir wieder die Eingabetaste.
- zur Beschriftung der einzelnen Säulen klicken wir auf "bar region" in (6) oder in das weiße Feld der Balken. In Toolbar (3) klicken wir auf "More.." und in der folgenden Dialogbox auf "Labels". Unter Label: wählen wir "bar". Bei "Format" klicken wir auf das kleine rechte Kästchen, wählen dort "Fixed numeric", "Total digits": 3 und "Digits right of decimal": 2. Jeweils mit Eingabe schließen wir ab und erhalten die Prozentangaben über den Balken mit 2 Dezimalstellen.
- zur Beschriftung der x-Achse gehen wir unter (6) auf "grpaxis" >> "title" und tragen dort "Name der Gruppe" ein.
- Durch Anklicken der Texte kann noch die Schriftgröße geändert werden.



Praktisch ist der kleine rote Punkt in der Toolbar (2), rechts. Er ermöglicht die Aufzeichnung und Speicherung aller Veränderungen an der ursprünglichen Grafik, die bei ähnlichen Grafiken mit dem grünen Pfeil aufgerufen und schnell wiederholt werden können.

Eine Übersicht über die vielen Möglichkeiten zur Formatierung unterschiedlicher Grafiken bietet das STATA GRAPHICS REFERENCE MANUAL (g.pdf).

STATA - Kommandos für Grafiken

histogram var, discrete frequency

histogram var, width(#) start(#) frequency addlabel

recode var, gen(var2)

histogram var, discrete width(#) start(#) frequency addlabel

graph box var, by(var2)

graph bar, over(var)

graph bar, over(var) blabel(bar)

graph pie, over(var) plabel(_all sum)

graph twoway line