

# Hypothesentest Anova

Code ▼

## 1) Grundlegende Konzepte: Was ist die einfaktoriellen Varianzanalyse ohne Messwiederholung

Die einfaktorielle Varianzanalyse – auch “einfaktorielle ANOVA”, da in Englisch “Analysis of Variance” – testet, ob sich die Mittelwerte mehrerer unabhängiger Gruppen (oder Stichproben) unterscheiden, die durch eine kategoriale unabhängige Variable definiert werden. Diese kategoriale unabhängige Variable wird im Kontext der Varianzanalyse als “Faktor” bezeichnet. Entsprechend werden die Ausprägungen der unabhängigen Variable “Faktorstufen” genannt, wobei auch der Begriff der “Treatments” gebräuchlich ist. Als “einfaktoriell” wird eine Varianzanalyse bezeichnet, wenn sie lediglich einen Faktor, also eine Gruppierungsvariable, verwendet (mehrfaktorielle Varianzanalyse).

Das Prinzip der Varianzanalyse besteht in der Zerlegung der Varianz der abhängigen Variable. Die Gesamtvarianz setzt sich aus der sogenannten “Varianz innerhalb der Gruppen” und der “Varianz zwischen den Gruppen” zusammen. Diese beiden Anteile werden im Rahmen einer Varianzanalyse miteinander verglichen. Die einfaktorielle ANOVA stellt eine Verallgemeinerung des t-Tests für unabhängige Stichproben für Vergleich von mehr als zwei Gruppen (oder Stichproben) dar. Die Fragestellung der einfaktoriellen Varianzanalyse wird oft so verkürzt: “Unterscheiden sich die Mittelwerte einer unabhängigen Variable zwischen mehreren Gruppen? Welche Faktorstufen unterscheiden sich?”

## 2) Aufgabenstellung - Beschreibung

500 Personen wurden in drei Gruppen eingeteilt, abhängig von deren Körpergestaltung (Normalgewicht, Übergewicht, Fettleibig). Es wurde auch die Blutdruckwerten jeder Einzelnen gemessen. Der Blutdruck (gemessen in systolische Druck) liegt normalerweise im Bereich von 110-130 mmHg. Es soll nun beantwortet werden, welche der drei Gewichtstypen die besten Blutdruckergebnisse zeigen.

## 3) Hypothese

H1: Es gibt einen Mittelwertsunterschied zwischen der Gewichtstypen (Normalgewicht, Übergewicht, Fettleibig) und der systolischen Blutdruck für mindestens einen Vergleich.

$$M_{normal} \neq M_{übergewicht} \neq M_{fettleibig}$$

H0: Es gibt keinen Mittelwertsunterschied zwischen der Gewichtstypen (Normalgewicht, Übergewicht, Fettleibig) und der systolischen Blutdruck.

$$M_{normal} = M_{übergewicht} = M_{fettleibig}$$

**Variable 1 = SystolicBP -> Blutdruck (mmHg) AV**

**Variable 2 = Overwt -> Gewichtstyp (Normalgewicht, Übergewicht, Fettleibig) UV**

## 4) Voraussetzungen für die einfaktoriellen Varianzanalyse ohne Messwiederholung

Die abhängige Variable ist intervallskaliert -> Erfüllt -> Blutdruck ist metrisch.

Die unabhängige Variable (Faktor) ist kategorial (nominal- oder ordinalskaliert) -> Erfüllt -> Die Gewichtstypen sind ordinal-skaliert.

Die durch den Faktor gebildeten Gruppen sind unabhängig -> Erfüllt -> Jeder Teilnehmer/In wurde einzeln gewogen.

Die abhängige Variablen ist normalverteilt innerhalb jeder der Gruppen -> Erfüllt -> siehe Histogramm.

Homogenität der Varianzen: Die Gruppen stammen aus Grundgesamtheiten mit annähernd identischen Varianzen der abhängigen Variablen -> Erfüllt -> siehe Levene-Test.

Datensatz:

Hide

```
Blood1 <- read.csv("C:/Data_Science_Projects/Repository_2/Aufgabe6/Blood1.csv")
View(Blood1)
```

Hide

```
library(dplyr) #-> Rename
Blood1 <- Blood1 %>% #pipe
  rename(Blutdruck = 'SystolicBP',
         Gewichtstyp = 'Overwt',
         Raucher = 'Smoke')
```

Hide

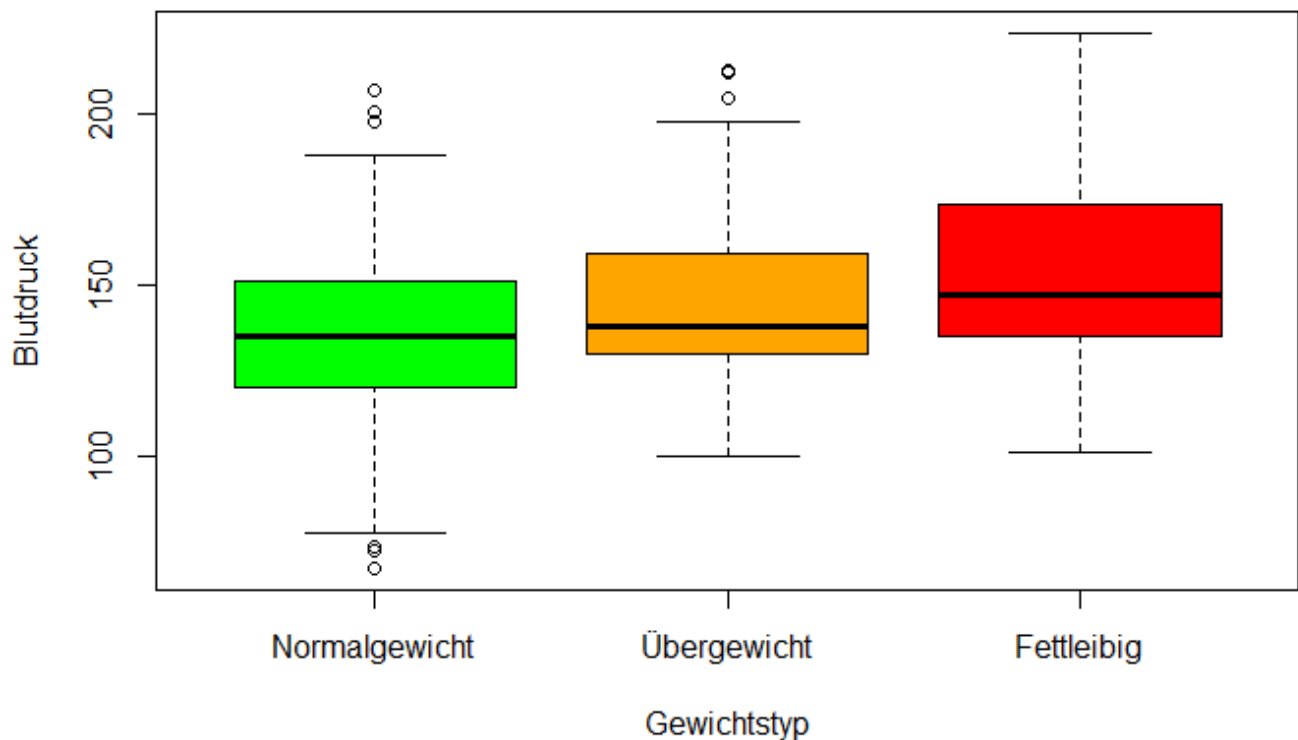
```
#library(dplyr) #-> Rekodieren
Blood1 <- Blood1 %>%
  mutate(Gewichtstyp = factor(Gewichtstyp, levels = c(0, 1, 2), labels = c("Normalgewicht",
"Übergewicht", "Fettleibig")))
```

## Boxplots zum Vergleich

Hide

```
boxplot(Blood1$Blutdruck ~ Blood1$Gewichtstyp, main = "Boxplots zum Vergleich", ylab = "Blutdruck", xlab= "Gewichtstyp" , col = c("green", "orange","red"))
```

## Boxplots zum Vergleich



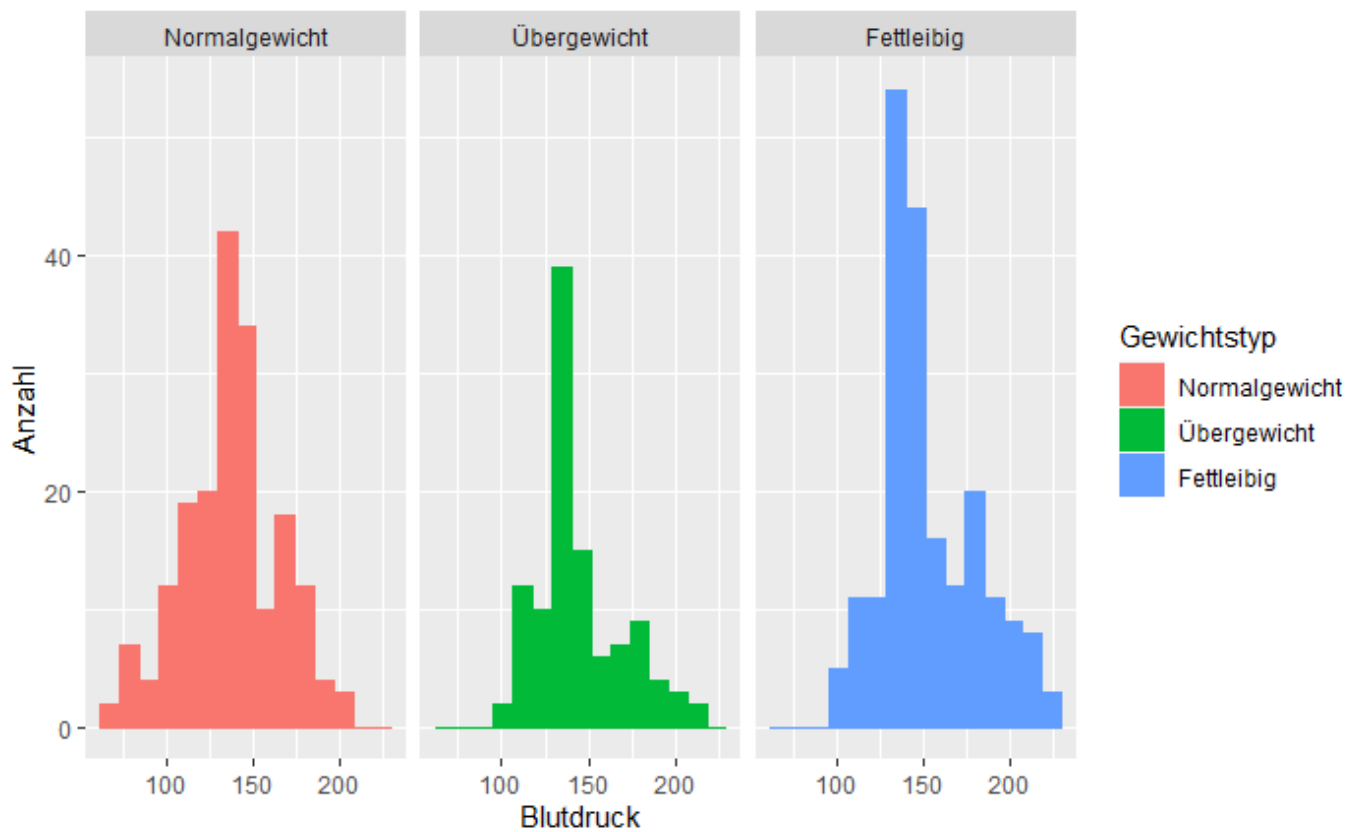
Die Boxplots zeigen ein paar Ausreisser. Die Verteilungen scheinen sich von einander zu unterscheiden, allerdings nicht so eindeutig bei Normalgewicht und Übergewicht. Darüber hinaus ist bemerkbar, dass der Median bei der unabhängigen Variable "Gewicht" für alle drei Ausprägungen (Normalgewicht, Übergewicht, Fettleibig) dicht beieinander liegen.

## Normalverteilung - Prüfung mittels Histogramm und QQPlot

Um einen ersten Überblick über die Daten zu gewinnen, empfiehlt es sich, ein Histogramm zu erstellen.

Hide

```
library(ggplot2)
Blood1 %>%
  group_by(Gewichtstyp) %>%
  ggplot(aes(Blutdruck, color=Gewichtstyp)) +
  geom_histogram(aes(fill = Gewichtstyp), bins = 15) +
  facet_wrap(~Gewichtstyp) +
  theme_grey()+
  labs(x= "Blutdruck", y = "Anzahl" )
```



QQPlot

Hide

```
library(car)
```

```
Lade nütziges Paket: carData
Registered S3 method overwritten by 'data.table':
  method      from
print.data.table

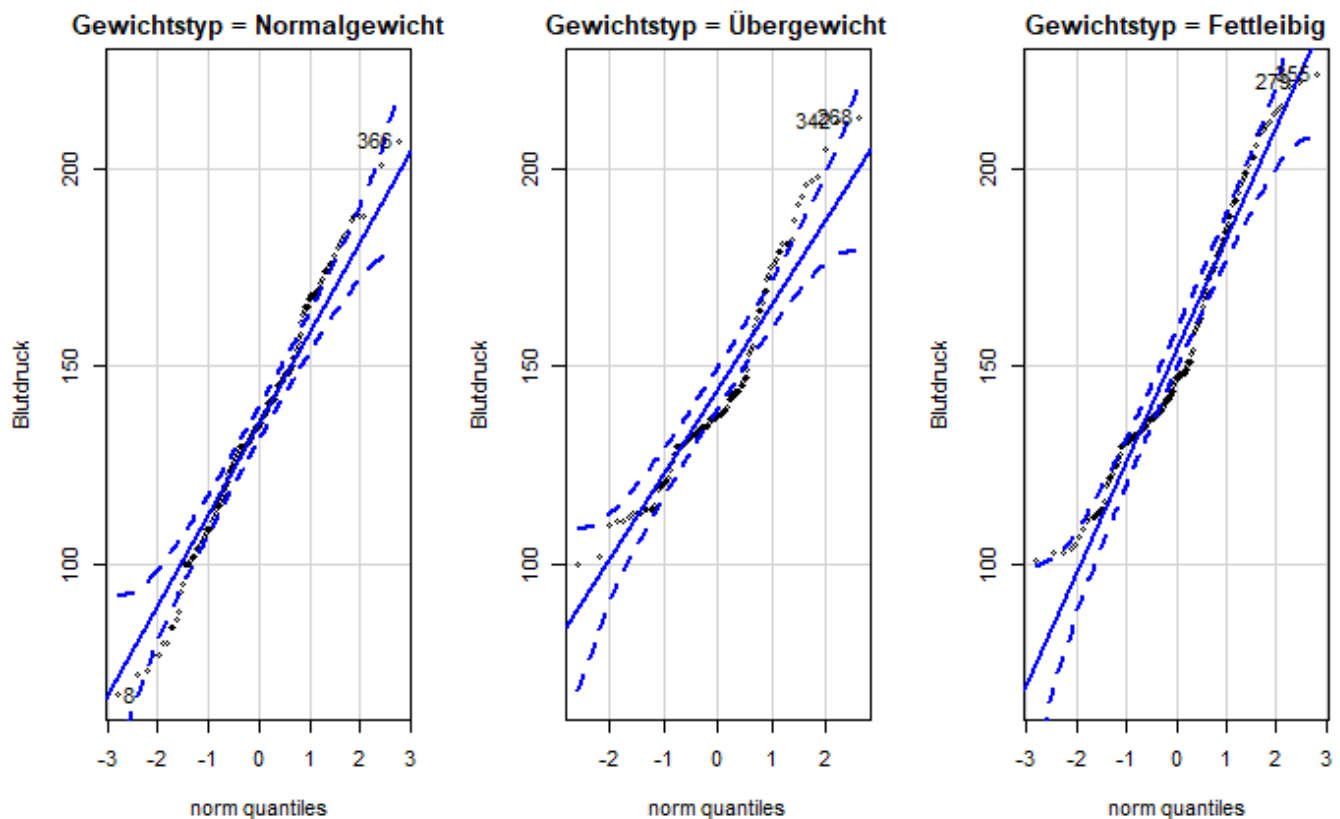
Attache Paket: 拖车car拖车

The following object is masked from 拖车package:dplyr拖车:

  recode
```

Hide

```
qqPlot(Blutdruck ~ Gewichtstyp, data=Blood1,
       layout=c(1, 3))
```



Alle drei Ausprägungen des Gewichts sind relativ gut normalverteilt. Aus dem Histogramm und dem QQPlot ist ersichtlich, dass die Ausprägung "Normalgewicht" am nächsten an der Normalverteilung dran ist, da sich die Datenpunkte nicht weit von der Geradenlinien befinden.

## 5) Deskriptive Statistiken

Die Tabelle in Abbildung gibt die Mittelwerte, Standardabweichungen und Grössen aller drei Gruppen wieder. Diese Informationen werden für die Berichterstattung verwendet.

Hide

```
Blood1 %>%
group_by(Gewichtstyp) %>%
  summarize(Anzahl = n(), Mittelwert = mean(Blutdruck), Median = median(Blutdruck), Standarda
bweichung = sd(Blutdruck)) %>%
  mutate_if(is.numeric, round, 2)
```

Gewichtstyp <fctr>	Anzahl <dbl>	Mittelwert <dbl>	Median <dbl>	Standardabweichung <dbl>
1 Normalgewicht	187	136.32	135	27.27
2 Übergewicht	109	144.37	138	25.08
3 Fettleibig	204	153.18	147	27.81
3 rows				

Es gibt einen Mittelwertsunterschied zwischen den Gruppen. Normalgewicht ( $M = 136.32$ ,  $SD = 27.27$ ,  $n=187$ ) zeigt die besten Blutdruckergebnisse, da der Mittelwert der Normalgewichtigen nahe an dem Normalbereich von 110-130 mmHg liegt, gefolgt von Übergewicht ( $M = 144.37$ ,  $SD = 25.08$ ,  $n=109$ ). Den höchsten Blutdruckwert sehen wir bei Fettleibig ( $M = 153.18$ ,  $SD = 27.81$ ,  $n=204$ ).

## 6) Prüfung der Varianzhomogenität (Levene-Test)

Der Levene-Test prüft die Nullhypothese, dass die Varianzen der Gruppen sich nicht unterscheiden. Ist der Levene-Test nichtsignifikant, so kann von homogenen Varianzen ausgegangen. Wäre der Levene-Test jedoch signifikant, so wäre eine der Grundvoraussetzungen der Varianzanalyse verletzt. Gegen leichte Verletzungen gilt die Varianzanalyse als robust; vor allem bei genügend grossen und etwa gleich grossen Gruppen sind Verletzungen nicht problematisch. Bei ungleich grossen Gruppen führt eine starke Verletzung der Varianzhomogenität zu einer Verzerrung des F-Tests. Alternativ können dann auf den Welch-Test zurückgegriffen werden. Dabei handelt es sich um adjustierte F-Tests.

Hide

```
leveneTest(Blood1$Blutdruck ~ Blood1$Gewichtstyp, center="mean")
```

```
Levene's Test for Homogeneity of Variance (center = "mean")
      Df F value Pr(>F)
group  2  1.4169 0.2434
      497
```

Im vorliegenden Beispiel ist der Levene-Test nicht signifikant ( $F(2, 497) = 1.4169$ ,  $p = 0.2434$ ), da der P-Value mit 0.2434 größer ist als 0.05 so dass von Varianzhomogenität ausgegangen werden kann. Das heisst es muss keine Welch-Korrektur durchgeführt werden.

## 7) Ergebnisse der einfaktoriellen Varianzanalyse ohne Messwiederholung ohne Welch-Korrektur

Hide

```
Blood1Anova <- aov(data=Blood1, Blood1$Blutdruck~Blood1$Gewichtstyp)
summary(Blood1Anova)
```

```
              Df Sum Sq Mean Sq F value    Pr(>F)
Blood1$Gewichtstyp  2   27801     13900    19.02 1.1e-08 ***
Residuals          497  363274         731
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Das Gesamtmodel ist signifikant geworden ( $F(2, 497) = 19.02$ ,  $p = .000$ ). Allerdings lässt sich aufgrund dieses Tests nicht bestimmen, welche der drei Gruppen sich signifikant voneinander unterscheiden. Es ist denkbar, dass sich lediglich ein Paar signifikant unterscheidet und die übrige kein signifikanten Unterschied zeigt. Daher wird ein Post-hoc-Test durchgeführt.

## 8) Post-hoc-Tests

Obwohl der F -Test zeigt, dass ein Haupteffekt von Gewichtstypen auf Blutdruck besteht, muss anhand von Post-hoc-Tests geklärt werden, zwischen welchen Faktorstufen (Gewichtstypen) signifikante Unterschiede bezüglich der Blutdruck bestehen.

$$\frac{k \cdot (k - 1)}{2} = \frac{3 \cdot (3 - 1)}{2} = \frac{6}{2} = 3$$

Bei der Berechnung von Post-hoc-Tests wird im Prinzip für jede Kombination von zwei Mittelwerten ein t -Test durchgeführt. Im aktuellen Beispiel mit drei Gruppen sind dies 3 Tests. Multiple Tests sind jedoch problematisch, da der Alpha-Fehler (die fälschliche Ablehnung der Nullhypothese) mit der Anzahl der

Vergleiche steigt.

Wird nur ein t-Test mit einem Signifikanzlevel von .05 durchgeführt, so beträgt die Wahrscheinlichkeit des Nicht-Eintreffens des Alpha-Fehlers 95 Prozent. Werden jedoch sechs solcher Paarvergleiche vorgenommen, so beträgt die Nicht-Eintreffens-Wahrscheinlichkeit des Alpha-Fehlers  $(.95)^3 = .857$ . Um die Wahrscheinlichkeit des Eintreffens des Alpha-Fehlers zu bestimmen, wird  $1 - .857 = .143$  gerechnet. Die Wahrscheinlichkeit des Eintreffens des Alpha-Fehlers liegt somit bei 14.3 Prozent. Diese Fehlerwahrscheinlichkeit wird als "Familywise Error Rate" bezeichnet.

Um dieses Problem zu beheben kann zum Beispiel die Turkey angewendet werden. RStudio rechnet das neue Niveau ein, daher können wir weiter auf 0.05 testen.

Hide

```
TukeyHSD(aov(data=Blood1, Blood1$Blutdruck ~ Blood1$Gewichtstyp))
```

```
Tukey multiple comparisons of means
 95% family-wise confidence level
```

```
Fit: aov(formula = Blood1$Blutdruck ~ Blood1$Gewichtstyp, data = Blood1)
```

```
$`Blood1$Gewichtstyp`
```

	diff	lwr	upr	p adj
Übergewicht-Normalgewicht	8.051464	0.3927115	15.71022	0.0366867
Fettleibig-Normalgewicht	16.865865	10.4316024	23.30013	0.0000000
Fettleibig-Übergewicht	8.814400	1.2740746	16.35473	0.0170703

Es wird ersichtlich, dass sich die Gewichtstypen bezüglich der Blutdruck signifikant unterscheiden. Die p adj Values sind alle unter 0.05.

Es können also drei unabhängige/ generalisierbare Gruppen von Gewichtstypen gebildet werden.

Es werden drei Gruppen gebildet - keine Veränderung.

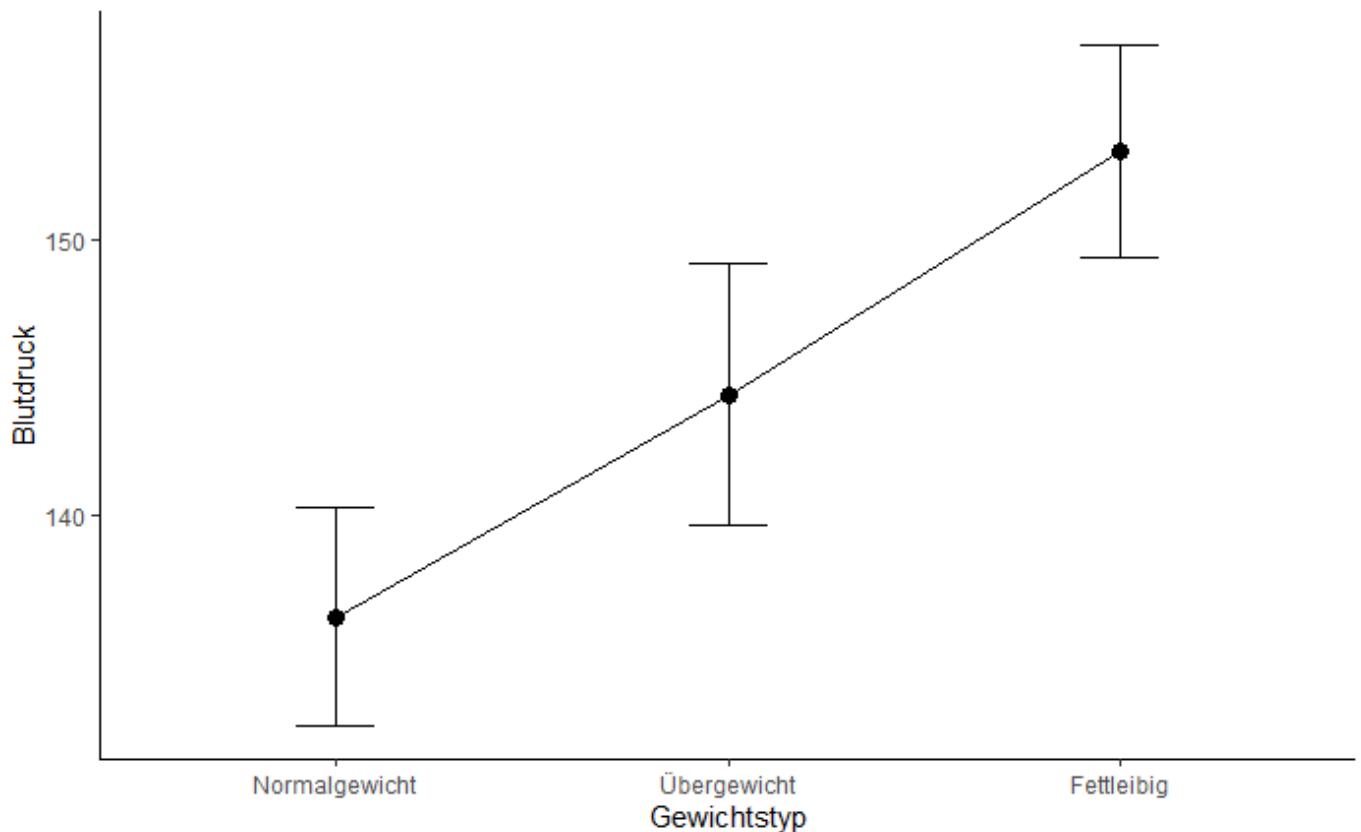
## 9) Profildiagramm

Spannend ist auch sich die Mittelwerte Hilfe dieses Plots anzeigen zu lassen.

Hide

```
ggplot(Blood1, aes(x=Gewichtstyp, y=Blutdruck, group=1))+
  stat_summary(fun.y = mean, geom="point", size=3)+
  stat_summary(fun.y = mean, geom="line")+
  stat_summary(fun.data = mean_cl_normal, geom="errorbar", width=.2, size=.25)+
  labs(x="Gewichtstyp", y="Blutdruck")+
  theme_classic()
```

```
`fun.y` is deprecated. Use `fun` instead.`fun.y` is deprecated. Use `fun` instead.
```



Wie der Plot in Abbildung erkennen lassen, bestehen bezüglich der drei Gewichtstypen Unterschiede im Mittelwert.

## 10) Berechnung der Effektstärke

### Das partielle Eta-Quadrat

Das partielle Eta-Quadrat (partielles  $\eta^2$ ) ist ein Mass für die Effektgrösse: Es setzt die Variation, die durch einen Faktor erklärt wird, in Bezug mit jener Variation, die nicht durch andere Faktoren im Modell erklärt wird. Das heisst, es wird ausschliesslich jene Variation betrachtet, welche nicht durch die anderen Faktoren im Modell erklärt wird. Das partielle Eta-Quadrat zeigt, welchen Anteil davon ein Faktor erklärt. Im Falle der einfaktoriellen Varianzanalyse ist das partielle Eta-Quadrat jener Anteil der korrigierten Gesamtvariation, der durch das Modell erklärt wird.

$$\eta^2 = \frac{QS_{Zwischen}}{QS_{total}}$$

$$\eta_{par.}^2 = \frac{QS_{Zwischen}}{QS_{zwischen} + QS_{innerhalb}}$$

Hide

```
library(effectsize)
Blood1eta <- effectsize::eta_squared(Blood1Anova, partial = TRUE)
```

For one-way between subjects designs, partial eta squared is equivalent to eta squared. Returning eta squared.

Hide

Blood1eta



Parameter	Eta2	90% CI
Blood1\$Gewichtstyp	0.07	[0.04, 0.11]

Im vorliegenden Beispiel beträgt das partielle Eta-Quadrat .07. Das heisst, es wird 7% der Variation in Blutdruck durch Gewichtstypen aufgeklärt. Das partielle Eta<sup>2</sup> wird gerundet. "90% CI" beschreibt das Konfidenzintervall für 90 %. Dieses liegt hier zwischen 4% und 11%.

## Effektstärke

Um die Bedeutsamkeit eines Ergebnisses zu beurteilen, werden Effektstärken berechnet. Im Beispiel sind zwar einige der Mittelwertsunterschiede zwar signifikant, doch es stellt sich die Frage, ob sie gross genug sind, um als bedeutend eingestuft zu werden.

Es gibt verschiedene Arten die Effektstärke zu messen. Zu den bekanntesten zählen die Effektstärke von Cohen (d) und der Korrelationskoeffizient (r) von Pearson.

Da R das partielle Eta-Quadrat ausgibt, wird dieses hier in die Effektstärke nach Cohen (1988) umgerechnet. In diesem Fall befindet sich die Effektstärke immer zwischen 0 und unendlich.

$$f = \sqrt{\frac{\eta^2}{1 - \eta^2}}$$

Hide

```
Blood1eff<- sqrt(Blood1eta$Eta2/(1-Blood1eta$Eta2))
Blood1eff
```

```
[1] 0.276638
```

Um zu beurteilen, wie gross dieser Effekt ist, kann man sich an der Einteilung von Cohen (1988) orientieren:

Schwacher Effekt:  $0.10 < ||f|| < 0.25$   
 Schwacher bis mittlerer Effekt:  $0.25 = ||f||$   
 Mittlerer Effekt:  $0.25 < ||f|| < 0.40$   
 Mittlerer bis starker Effekt:  $0.40 = ||f||$   
 Starker Effekt:  $0.40 < ||f||$

Damit entspricht eine Effektstärke von 0.276 einem mittleren Effekt.

## 11) Eine Aussage

Die Gewichtstypen haben einen signifikanten Einfluss auf den Blutdruck ( $F(2, 497) = 19.02, p = .000$ ). 7% der Streuung der Blutdruck-Werte um den Gesamtmittelwert kann durch die Gewichtstypen erklärt werden. Die Effektstärke nach Cohen (1988) liegt bei  $f = 0.276$  und entspricht einem mittleren Effekt.  $H_0$  wird abgelehnt,  $H_1$  angenommen.

Post-hoc-Tests mit Tukey zeigen, dass sich drei Gruppen von Gewichtstypen bilden lassen (alle  $p < .05$ ): Normalgewicht ( $M = 136.32, SD = 27.27, n=187$ ), Übergewicht ( $M = 144.37, SD = 25.08, n=109$ ) und Fettleibig ( $M = 153.18, SD = 27.81, n=204$ ) bilden jede für sich eine eigene Gruppe.

Damit kann festgehalten werden, dass alle drei Gruppen unabhängige Gruppen bilden und sich signifikant unterscheiden. Normalgewicht ist am nächsten an dem Normalbereich Blutdrucks von 110-130 mmHg und Fettleibig die am weitesten entfernt.