

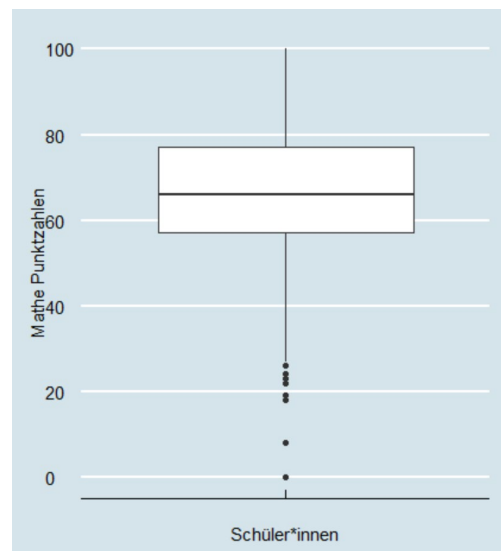
## Zusatzaufgabe Soziologie

### Aufgabe 1a

Variable	Skalenniveau	Ausprägungen
<i>gender</i>	Dichotom	male – female
<i>race_ethnicity</i>	Nominal	group A – group E
<i>parental_level_of_education</i>	Ordinal	high school – masters degree
<i>lunch</i>	Dichotom	standard – free/reduced
<i>test_preparation_course</i>	Dichotom	none - completed
<i>math_score</i>	Ratioskala	0 – 100
<i>reading_score</i>	Ratioskala	0 – 100
<i>writing_score</i>	Ratioskala	0 – 100

### Aufgabe 1b

Der Median liegt zwischen 60 und 80 Punkten. Der Interquartilsabstand beginnt ein wenig unter 60 Punkten und endet ein wenig unter 80 Punkten. In diesem Bereich sind 50% aller Schüler\*innen. Die Linie, die aus der Box oben und unten herausgeht, sind die Whisker. Diese zeigen das Maximum der Verteilung ohne Ausreisser ( $> 1.5$  Interquartilsabstände zur Box)



### Aufgabe 1c

Mittelwert = 67.77067

Median = 68

Der Median und der Mittelwert der Daten sind sehr nahe beieinander. Es ist also weder stark rechtsschief noch stark linksschief. Sehr grosse Ausreisser nach oben sind auch nicht vorhanden oder sogar möglich, da die Werte nach oben mit 100 limitiert sind.

**Aufgabe 2a**

Die Regressionsgerade hat eine Steigung von 0.99 und einen Intercept von -0.66. Es liegt ein sehr starker Zusammenhang vor, denn mit einem  $R^2$  value von 0.91 lassen sich die Werte zu 91% voraussagen

```
Call:
lm(formula = writing_score ~ reading_score, data = StudPerf)

Residuals:
    Min       1Q   Median       3Q      Max
-12.9573  -2.9573   0.0363   3.1026  15.0557

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.667554   0.693792  -0.962   0.336
reading_score  0.993531   0.009814  101.233 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.529 on 998 degrees of freedom
Multiple R-squared:  0.9113,    Adjusted R-squared:  0.9112
F-statistic: 1.025e+04 on 1 and 998 DF,  p-value: < 2.2e-16
```

**Aufgabe 2b**

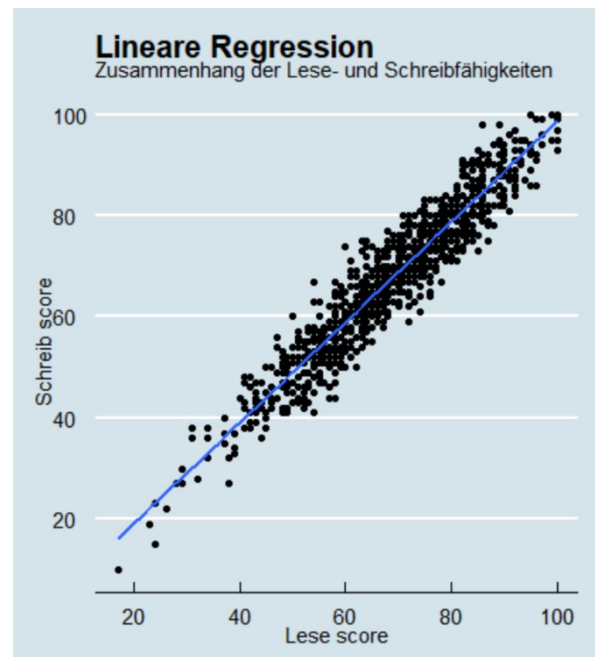
Es besteht ein positiv linearer Zusammenhang dessen Steigung und X-Achsenschnitt im vorherigen Regressionsmodell vorhergesagt wurden.

**Aufgabe 2c**

Ein Schreibscore von 75.53981 wird bei einem Lesescore von 75 prognostiziert. Dies entspricht einem fast perfekten Zusammenhang.

**Aufgabe 2d**

Der Intercept wird beim hinzunehmen der Variable *gender* grösser. Mit einem Wert von 0.98 wird der Zusammenhang kleiner als ohne die Zunahme von *gendermale*. Männer haben somit einen negativen Einfluss auf den *writing\_score*.



```
Call:
lm(formula = writing_score ~ reading_score + gender, data = StudPerf)

Residuals:
    Min       1Q   Median       3Q      Max
-12.0239  -3.0482   0.0261   3.0062  13.8275

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.664004   0.740310   2.248   0.0248 *
reading_score  0.975141   0.009842  99.084 < 2e-16 ***
gendermale   -2.198267   0.287419  -7.648 4.78e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.404 on 997 degrees of freedom
Multiple R-squared:  0.9162,    Adjusted R-squared:  0.916
F-statistic: 5448 on 2 and 997 DF,  p-value: < 2.2e-16
```

**Aufgabe 3a**

0. Zunächst ziehen wir jeweils eine Stichprobe von je 300 Studierenden.
1. Null- und Alternativhypothese formulieren und Signifikanzniveau festlegen  
**H<sub>0</sub>**: Studierende, die den Vorbereitungskurs absolviert haben, haben keine andere Leistung in der Mathematik, wie Studierende ohne Vorbereitungskurs.  
 $\rightarrow \text{Prep} - \text{NoPrep} = 0$   
**H<sub>A</sub>**: Studierende, die den Vorbereitungskurs absolviert haben, haben eine andere Leistung in der Mathematik, wie Studierende ohne Vorbereitungskurs  
 $\rightarrow \text{Prep} - \text{NoPrep} \neq 0$   
 Signifikanzniveau: 5%
2. Prüfgrösse (z- oder t-Wert) und deren Verteilung bestimmen  
 Varianz der Grundgesamtheit ist nicht bekannt, wir verwenden die geschätzte Stichprobenvarianz  $\rightarrow$  **t-Test**
3. Ablehnungsbereich der Nullhypothese kennzeichnen  
 ungerichtete Hypothese = zweiseitiger Ablehnungsbereich (je 2.5%)  
 H<sub>0</sub> verwerfen:  $t > t(0.975; 597) = 2.024$  oder  $t < t(0.025; 597) = -2.024$
4. Prüfgrösse berechnen und Entscheid über Nullhypothese treffen  
 t-Wert = 3.78  
 t-Wert ist grösser als der kritische Wert 2.024, Nullhypothese kann somit verworfen werden.

Welch Two Sample t-test

```
data: Prep and NoPrep
t = 3.7752, df = 596.49, p-value = 0.0001759
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.175002 6.891664
sample estimates:
mean of x mean of y
69.44333 64.91000
```

**Aufgabe 3b**

Um das Risiko eines  $\beta$ -Fehlers zu minimieren, maximieren wir die Stichprobengrösse.  
 Der t-Wert wird grösser und deutet weiter darauf hin, die Nullhypothese abzulehnen

Welch Two Sample t-test

```
data: PrepFull and NoPrepFull
t = 5.787, df = 770.08, p-value = 1.043e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.712041 7.523257
sample estimates:
mean of x mean of y
69.69553 64.07788
```