

# Präsentation der statistischen Berichte

Anh Thu Bui Pham

# Zusammenfassung der Ergebnisse - Eine Variable

- Ausgangssituation: Es soll überprüft werden ob, CEOs von gewinn- und verlustgenerierenden Unternehmen unterschiedlich viel verdienen.
- Untersucht wurden zwei unabhängigen Gruppen mit metrischen Variable  
→ Gruppenvergleich
- Es sind 19 Beobachtung jeweils von Gewinn und Verlust Unternehmen (insgesamt 38)
- Variablen wird in Tsd \$ von den CEO Jahresgehalt gemessen.
- Es gab keine fehlenden Werte

- In der Gruppe von CEO die Gewinne erzielen ist der Betrag weitaus höher als in der Gruppe Verlust (siehe Boxplot)
- Es gibt sehr viele Ausreißer, besonders in der Gewinn Gruppe gibt es extreme Ausreißer (siehe Boxplot)
- wilcoxon-test wurde für die Hypothese rangezogen wegen Aureißer, nicht perfekter normalverteilung und ungleiche varianz
- Als Signifikanzniveau wurde  $\sigma = 0.001$  gewählt  $\rightarrow$  sehr strenge Niveau
- bei 0,001 Signifsnivau ist **der Unterschied nicht signifikant**

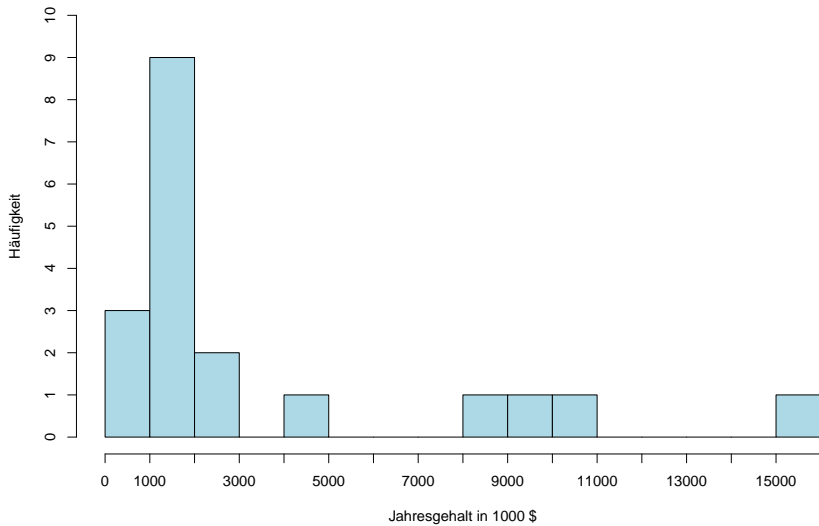
# R-Output - Eine Variable

## Anzeige der Daten

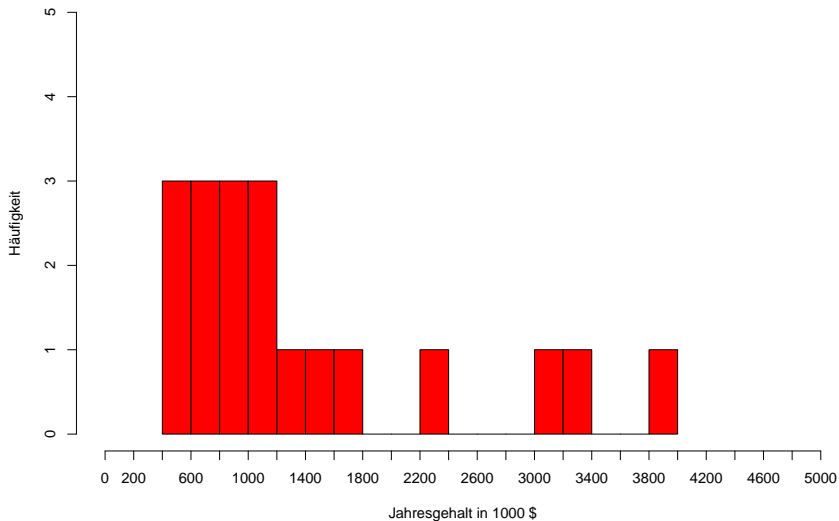
```
gewinne verluste
1      670      1238
2    15250     1088
3     1546       500
4     1700     4000
5     9525     1134
6     8888       900
```

Gruppe	Min	Max	MW	Median	Varianz	StandAbw.	IQR
Gewinne	670	15250	3696	1782	17088864	4134	2150
Verluste	450	4000	1397	1088	1076459	1038	972.5

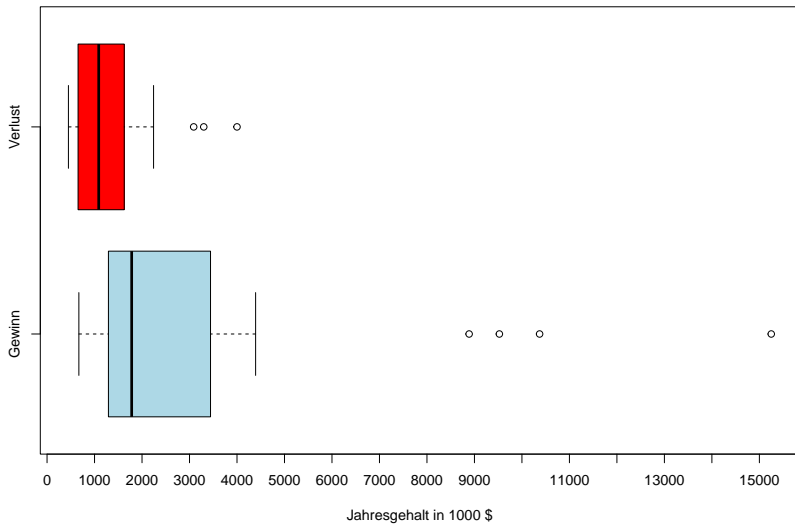
Histogramm der CEO-Gehälter (Gewinn-Unternehmen)



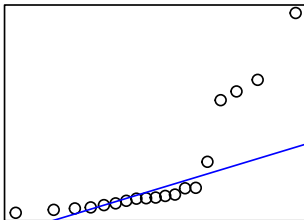
Histogramm der CEO-Gehälter (Verlust-Unternehmen)



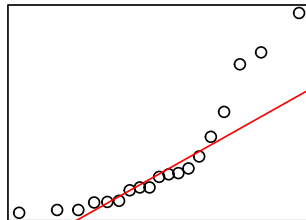
## Vergleich der CEO-Gehälter



## QQ: Gewinn



## QQ: Verlust





	F.Wert	Konfidenzintervall	p.Wert
<b>F</b>	15.88	[6.12, 41.21]	2.83e-07

## Hypothesentest

Wilcoxon rank sum test with continuity correction

data: daten\$gewinne and daten\$verluste

W = 269, p-value = 0.01019

alternative hypothesis: true location shift is not equal to 0

H0: Die Verteilung der CEO-Gehälter ist in beiden Gruppen gleich

H1: Die Verteilungen unterscheiden sich

Ergebnis: nicht signifikant bei 0.1% → H0 nicht abgelehnt

# Zusammenfassung der Ergebnisse - Zwei kategoriale Variablen

- Ausgangssituation: Es soll überprüft werden, ob Anzeigen wegen Drogendelikt auf freiem Fuß (ja) bzw. die Verhaftung (nein) der Personen durch job erklärt werden?
- Homogenitätsproblem: freie Fuß (A) soll auf Job (B) geprüft werden. Dh. prüfen ob Aussage A folgt aus B?
- Insgesamt 237 Beobachtungen, davon 211 gültig nach Bereinigung (26 fehlende Werte entfernt)
- Gruppenvariable  $\rightarrow$  Job; Zielvariable  $\rightarrow$  frei?

- Auffällig sind die ungleiche Gruppenverteilung Job mit 76% vs Arbeitslos mit 24% (siehe Kacheldiagramm)
- Als Signifikanzniveau wurde  $\alpha = 0.01$  gewählt
- Mit einem  $\chi^2$ -Test auf Gleichverteilung konnte nachgewiesen werden, dass Beschäftigte und Arbeitslose werden nicht gleich häufig verhaftet/freigelassen.
  - ▶ Wert der Teststatistik: 8.0499
  - ▶ p-Wert (gerundet): 0.004551

# R-Output - Zwei kategoriale Variablen

## Rohdaten und bereinigte Daten

	frei	job
1	ja	ja
2	ja	ja
3	<NA>	ja
4	ja	nein
5	ja	ja
6	ja	ja

	frei	job
1	ja	ja
2	ja	ja
4	ja	nein
5	ja	ja
6	ja	ja

## Kontingenztafel

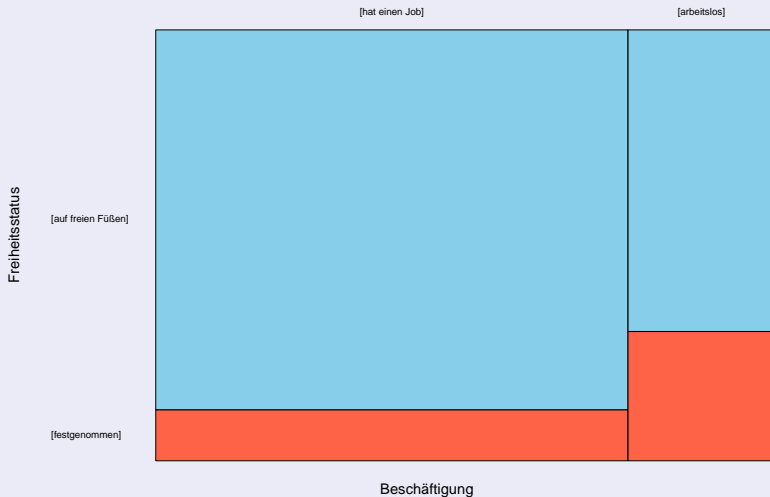
Absolute Häufigkeiten:

Beschäftigung	Freiheitsstatus		Gesamt
	[auf freien Füßen]	[festgenommen]	
[hat einen Job]	142	19	161
[arbeitslos]	35	15	50
Gesamt	177	34	211

Relative Häufigkeiten innerhalb der Job-Gruppen (in %):

Beschäftigung	Freiheitsstatus	
	[auf freien Füßen]	[festgenommen]
[hat einen Job]	88.2	11.8
[arbeitslos]	70.0	30.0

**Kacheldiagramm: Beschäftigung und Freiheitsstatus**



## Hypothese

Pearson's Chi-squared test with Yates' continuity correction

data: kontingenz

X-squared = 8.0499, df = 1, p-value = 0.004551

H0 → Der Anteil der Personen auf freiem Fuß ist gleich bei Beschäftigten und Arbeitslosen  
H1 → Der Anteil unterscheidet sich zwischen den beiden Gruppen.

Ergebnis: H0 wird abgelehnt, es ist ein signifikanter Unterschied bei Signifikantsniveau von 1%. Die Verteilung des Freiheitsstatus unterscheidet sich zwischen den Job-Gruppen!

## Zusammenfassung der Ergebnisse - Korrelation

- Ausgangssituation: Wir sollen untersuchen, ob ein Zusammenhang zwischen den beiden Verbrauchsangaben – dem WLTP-Wert und dem realen Verbrauchswert – besteht und wie stark es ist.
- keine fehlenden Werte
- jedoch ein einzelne extrem Wert der unplausible ist (51l/100km Verbrauch) wird aus den Daten entfernt und für diese Analyse nicht herangezogen
- Ursprünglich waren 67 Fahrzeuge angekündigt, aber nur 52 wurden bereitgestellt
  - ▶ ohne unplausible Außreißer insgesamt 51 Beobachtungen

Variable	Beschreibung
modell	Name des Automodells
wltp	Verbrauchswerte laut WLTP (in l/100km)
test	Reale Verbrauchswerte (in l/100km)



- realltive symmetrische Verteilung (siehe summary(daten) und boxplot)
- Beim Test fällt der real Verbrauch höher aus als der WLTP Wert (siehe Boxplot und Streudiagramm)
- Autos mit hohen WLTP haben tendenzielle hohe real Verbrauch (siehe positive Regressionlinie im Streudiagramm)
- Als Signifikanzniveau wurde  $\sigma = 0.01$  gewählt
- Mit einem Pearson-Korrelation konnte nachgewiesen werden, dass es ein mäßige positive und eine statistisch signifikanter linearer Zusammenhang gibt.
  - ▶ Korrelationskoeffizient: 0.4452521
  - ▶ p-Wert (gerundet): 0.00106

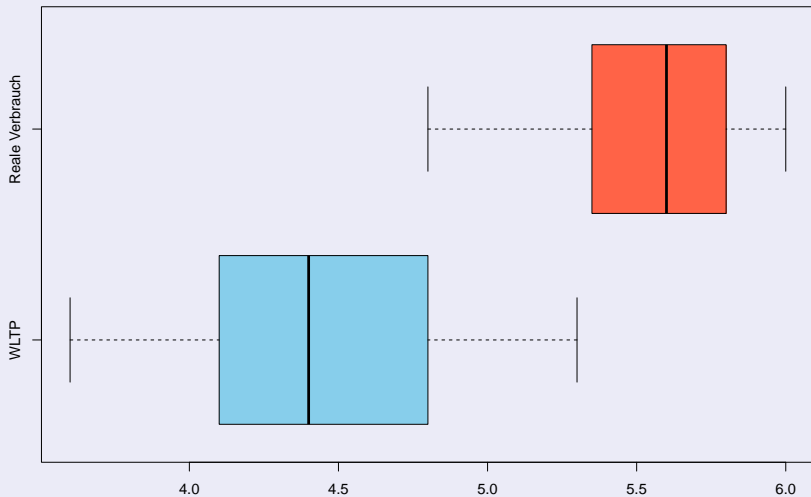
# R-Output - Korrelation

## Daten anzeigen

	modell	wltp	test
1	VW Golf 1 6 TDI	4.5	5.2
2	Mercedes SLK 250 CDI BlueEFFICIENCY	4.9	6.0
3	Peugeot 308 eHDI FAP 110 EGS6 Stop+Start	3.8	6.0
4	Chevrolet Aveo 1 3	4.1	5.4
5	BMW X1 sDrive18d	4.9	5.4
6	Nissan Qashqai 4x4 1 6 dCi	5.1	5.9

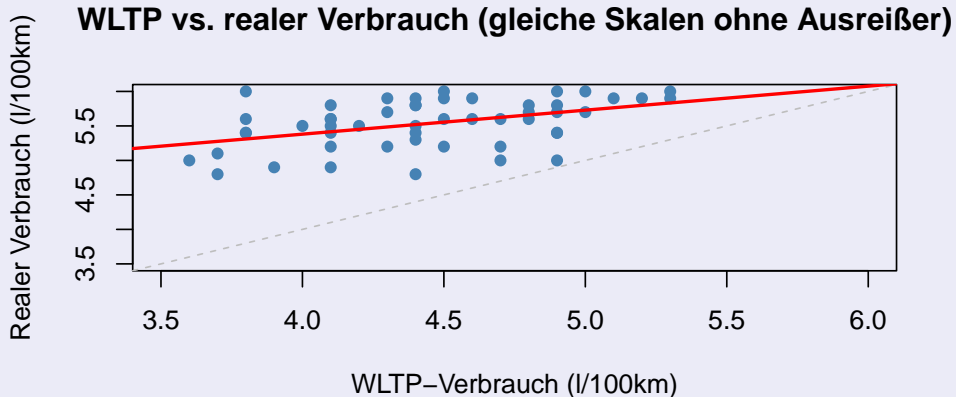
modell	wltp	test
Length:51	Min. :3.600	Min. :4.800
Class :character	1st Qu.:4.100	1st Qu.:5.350
Mode :character	Median :4.400	Median :5.600
	Mean :4.445	Mean :5.535
	3rd Qu.:4.800	3rd Qu.:5.800

Vergleich der Verbrauchswerte (ohne Ausreißer)



## Streudiagramm

Die Identitätslinie ( $y = x$ ) zeigt, wo WLTP und realer Verbrauch gleich wären. Die meisten Punkte liegen oberhalb dieser Linie → tatsächlicher Verbrauch ist immer höher als die WLTP-Angabe.“



## Hypothesentest

Pearson's product-moment correlation

```
data:  daten$wltp and daten$test
t = 3.4808, df = 49, p-value = 0.00106
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1933994 0.6420532
sample estimates:
      cor
0.4452521
```

# Zusammenfassung der Ergebnisse - Regression

- Ausgangssituation: Es soll untersucht werden, ob und wie die Kosten pro Kilo Druckmaterial die Druckqualität beeinflussen (geringere Schichtdicke → bessere Qualität).
- Es liegen keine fehlenden Werte vor.
- 15 Beobachtungen insgesamt → sehr kleine Stichprobe.
- Auffällig: Ein Gerät mit sehr hoher Kostenangabe (135€) und feiner Schichtdicke (0.025mm) sticht als potenzieller Ausreißer hervor, wurde aber berücksichtigt.

- Tendenz: Je höher die Materialkosten, desto geringer die Schichtdicke.
- Als Signifikanzniveau wurde  $\alpha = 0.01$  gewählt
- Residuenplot zeigt keine auffällige Struktur – Annahme der Homoskedastizität erfüllt.
- QQ-Plot zeigt größtenteils normalverteilte Residuen, nur leichte Abweichungen am Rand – akzeptabel bei kleiner Stichprobe
- Das Regressionsmodell zeigt eine negative Steigung ( $\beta = -0.000576$ ),  $\rightarrow$  der Zusammenhang ist jedoch nicht signifikant ( $p = 0.311$ ).

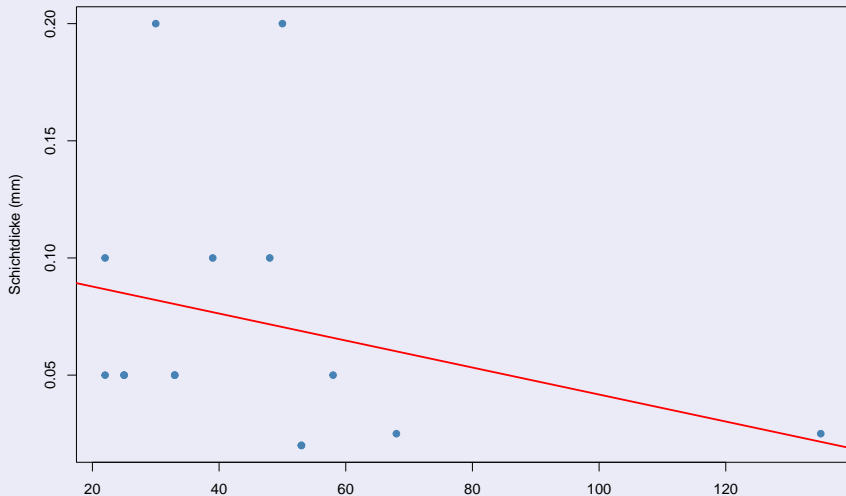
## R-Output - Regression

	schichtdicke	kosten
XYZ-Printing-Da-Vinci-Jr-2-0-Mix	0.200	50
XYZ-Printing-Da-Vinci-1-0-Pro	0.100	39
Anycubic-Vyper	0.100	22
Velleman-K8200	0.200	30
Formlabs-Form-3	0.025	135
Ultimaker-3	0.020	53



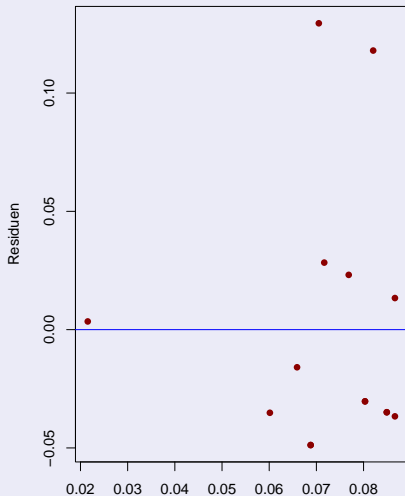
# Streudiagramm

Streudiagramm: Kosten vs. Schichtdicke

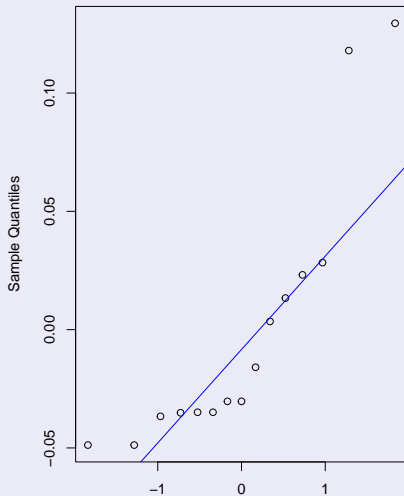


## Residueplot & Resdne QQ-Plot

Residuenplot



QQ-Plot der Residuen



## Regressionsmodell

```
Call: lm(formula = schichtdicke ~ kosten, data = daten) Residuals: Min 1Q  
Median 3Q Max -0.04879 -0.03503 -0.03031 0.01825 0.12948 Coefficients:  
Estimate Std. Error t value Pr(>|t|)  
(Intercept) 0.0993250 0.0294262 3.375 0.00497 ** kosten -0.0005762  
0.0005467 -1.054 0.31113  
Signif. codes: 0 ' ' 0.001 ' ' 0.01 ' ' 0.05 ' ' 0.1 ' ' 1  
Residual standard error:  
0.05823 on 13 degrees of freedom Multiple R-squared: 0.07871, Adjusted  
R-squared: 0.007842 F-statistic: 1.111 on 1 and 13 DF, p-value: 0.3111
```

# Zusammenfassung der Ergebnisse - Zeitreihenanalyse

- Ausgangssituation: Wir sollen den Börsenkurstrend von CICI Bank Limited (NYSE:IBN) ermitteln.
- Zeitraum: 18. Februar – 07. April 2025 (35 Handelstage)
- Verlauf des Kurses:
- Anfangs: leichter Abwärtstrend
- Ab Mitte März: deutlicher Kursanstieg
- Ende: Konsolidierung mit leichtem Rückgang

- Trend: Bestätigt klaren Aufwärtstrend ab Woche 11 bis etwa Woche 13, danach flach
- Saisonale Effekte klar erkennbar (besonders mittwochs)
  - ▶ Wiederkehrendes 5-Tage-Muster (Handelswoche)
  - ▶ Besonders Mittwoch ( $s_3$ ) zeigt positiven Effekt (+0.117), Montag/Freitag negativ
- Zufällige Komponente: Schwache Restschwankungen → gute Modellstruktur
- Glättungsparameter:
  - ▶ Alpha ( ) = 0.398 → moderate Berücksichtigung neuer Daten
  - ▶ Beta ( ) = 0.901 → hohe Sensitivität auf Trendänderungen
  - ▶ Gamma ( ) = 0.050 → stabile Saisonalität wird nur langsam angepasst
- Trendkoeffizient:  $b = -0.266$  → Modell erkennt rückläufige Tendenz zum Ende hin
- Prognose weist auf potenzielle Kursabschwächung hin – jedoch mit Unsicherheit

# R-Output - Zeitreihenanalyse

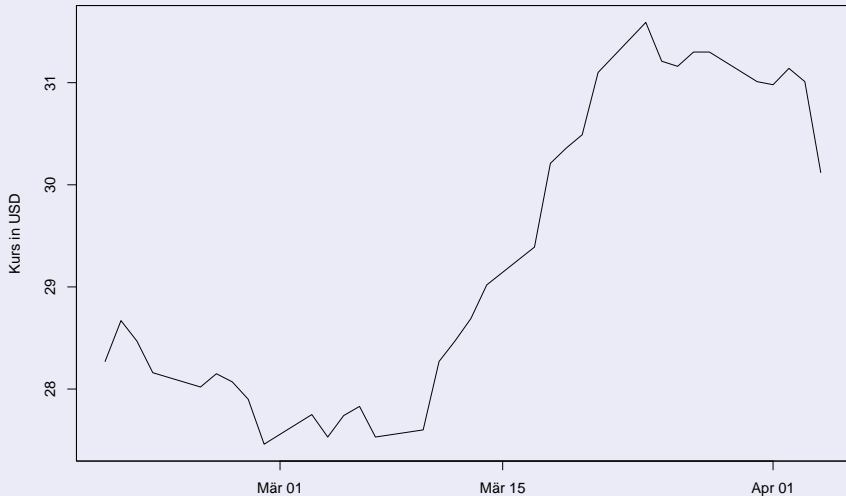
## Daten importieren und anzeigen

```
time series ends    2025-04-04
```

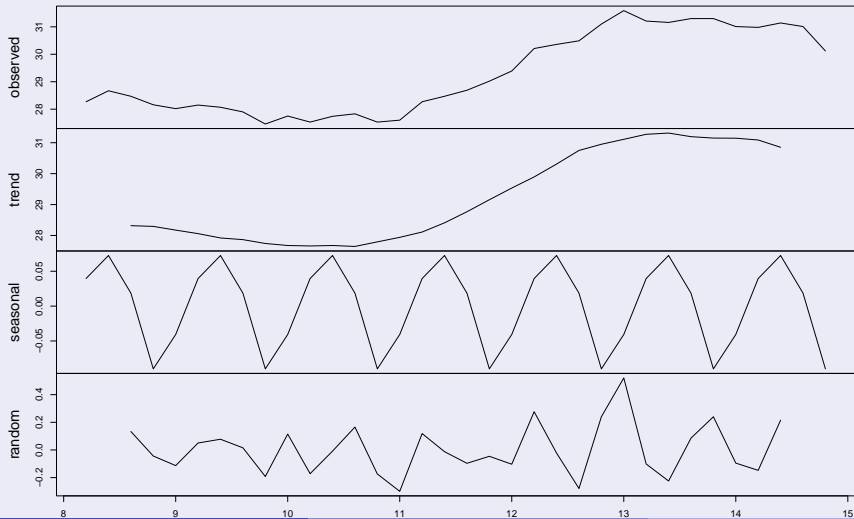
Index	Low
Min. :2025-02-18	Min. :27.46
1st Qu.:2025-02-28	1st Qu.:28.03
Median :2025-03-12	Median :28.68
Mean :2025-03-12	Mean :29.29
3rd Qu.:2025-03-24	3rd Qu.:31.00
Max. :2025-04-04	Max. :31.59

	Low
2025-02-18	28.27
2025-02-19	28.67
2025-02-20	28.47
2025-02-21	28.16

Tägliche Tiefstkurse von ICICI Bank



Decomposition of additive time series





## Holt Winters Modell

Holt-Winters exponential smoothing with trend and additive season

Call:

```
HoltWinters(x = data_ts)
```

Smoothing parameters:

alpha: 0.3977184

beta : 0.9011878

gamma: 0.04984415

Coefficients:

[,1]

a 30.59224317

b -0.26593339

s1 -0.13350313

s2 0.05519932

Time Series:

Start = c(15, 1)

End = c(17, 1)

Frequency = 5

	fit	upr	lwr
15.0	30.19281	30.86755	29.51806
15.2	30.11558	30.96150	29.26966
15.4	29.91112	31.04299	28.77924
15.6	29.60834	31.11464	28.10204
15.8	29.13046	31.07878	27.18214
16.0	28.86314	31.32066	26.40562
16.2	28.78591	31.78520	25.78662
16.4	28.58145	32.16506	24.99785
16.6	28.27868	32.48532	24.07204
16.8	27.80080	32.66634	22.93526
17.0	27.53347	33.10137	21.96558

## Prognose der Börsenkurse mit Holt-Winters

