

Analyse der Herzgesundheitsdaten

Kirill Khvastow, Inga Stakhurska, Stefan Shushpanov

WS 2023/2024

Einführung

- ▶ **Ziel des Projekts:** Untersuchung der Zusammenhänge zwischen verschiedenen Gesundheitsaspekten und Herzkrankheiten.
- ▶ **Datenquelle:** Kaggle-Dataset von Amir Mahdiabbootalebi.
- ▶ **Schlüsselfragen:** Gibt es einen signifikanten Unterschied im Ruheblutdruck je nach Brustschmerztyp?

Datensatz

- ▶ **Umfasst:** 748 Messungen zu Alter, Geschlecht, Brustschmerztyp, Ruheblutdruck, Cholesterinspiegel, maximale Herzrate und Vorhandensein von Herzkrankheiten.
- ▶ **Zielvariablen:** Ruheblutdruck und Brustschmerztyp.

Explorative Analyse

Numerische Zusammenfassungen

- ▶ Die Tabelle zeigt Minimum, Maximum, Durchschnitt und Varianz für Alter, RestingBP, Cholesterin und MaxHR.

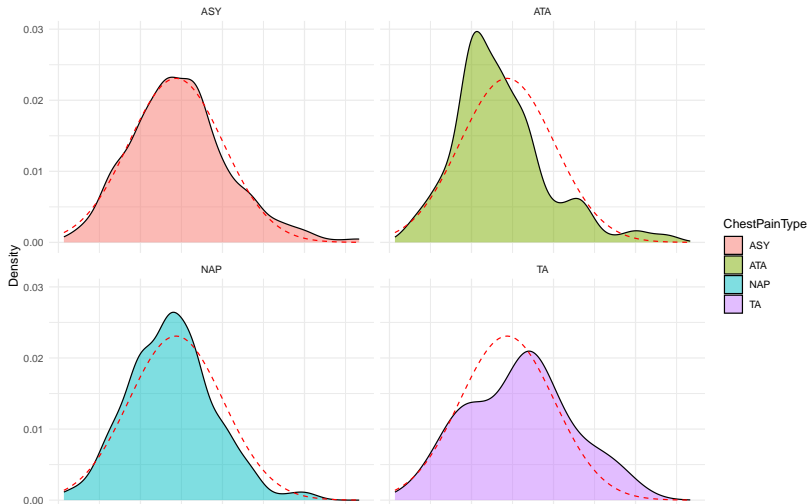
##	Minimum	Maximum	Mean	Variance
## Age	28	77	52.88204	90.36191
## RestingBP	92	200	133.02279	298.69344
## Cholesterol	85	603	244.63539	3499.13936
## MaxHR	69	202	140.22654	601.43183

Explorative Analyse

Verteilungen

- Unterschiede im Ruheblutdruck bei verschiedenen Brustschmerztypen in man.

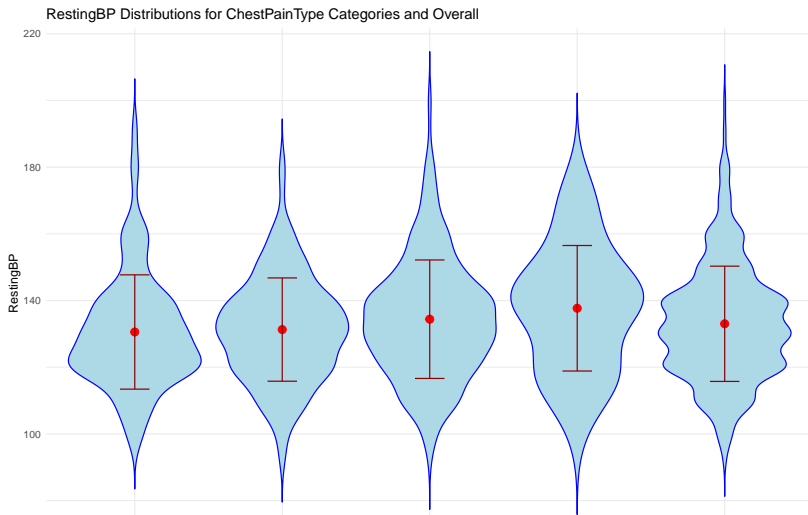
Probability Distribution of RestingBP for ChestPainType



Explorative Analyse

Verteilungen

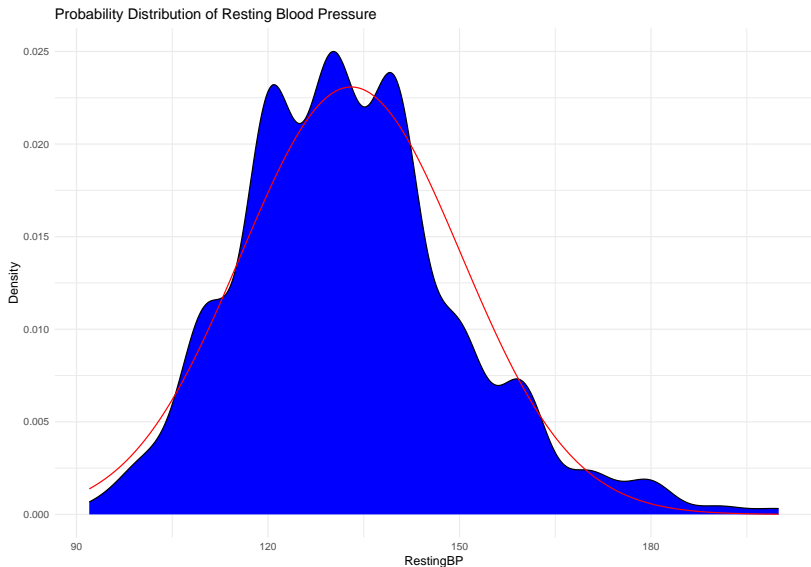
- Unterschiede im Ruheblutdruck bei verschiedenen Brustschmerztypen.



Explorative Analyse

Verteilungen

► Normalverteilung von Ruheblutdruck.



Statistische Methode

- ▶ **Ziel:** Überprüfung der Hypothese, ob Brustschmerztypen den Ruheblutdruck beeinflussen.
- ▶ **Methode:** Varianzanalyse (ANOVA) gefolgt von einem Tukey-HSD-Test.

Statistische Methode

Nullhypothese H_0 : Erwartungswerte des Ruheblutdrucks der vier Kategorien sind gleich.

Alternativhypothese H_1 : Es gibt mindestens zwei Gruppen, deren Erwartungswerte sich statistisch signifikant unterscheiden.

Annahmen für ANOVA:

1. Die Beobachtungen sind unabhängig.
2. Die Störterme sind $N(0, \sigma^2)$ verteilt.
3. Die Varianz der verschiedenen Gruppen soll gleich sein.

$$F \equiv \frac{MQA}{MQR} = \frac{\sum_{i=1}^k n_i (\bar{y}_{i.} - \bar{y}_{..})^2 / (k - 1)}{\sum_{i=1}^k (n_i - 1) s_i^2 / (N - k)}$$

Ergebnisse

- **ANOVA-Ergebnisse:** Einige Gruppen zeigen statistisch signifikante Unterschiede im Erwartungswert des Ruheblutdrucks.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ChestPainType	3	3101	1033.8	3.496	0.0153 *
Residuals	742	219425	295.7		

Significance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ergebnisse

- **Tukey-HSD-Test:** Signifikante Unterschiede im Ruheblutdruck zwischen einigen Brustschmerztypen, aber nicht alle Unterschiede sind statistisch signifikant.

Comparison	Difference	Lower Bound	Upper Bound	Adjusted p
ATA-ASY	-3.833735	-7.970138	0.302668	0.0805519
NAP-ASY	-3.110059	-7.221041	1.000922	0.2090459
TA-ASY	3.282927	-4.005327	10.57118	0.6524202
NAP-ATA	0.723676	-4.114931	5.562283	0.9805744
TA-ATA	7.116662	-0.605423	14.83875	0.0832181
TA-NAP	6.392986	-1.315511	14.10148	0.1428788

Schlussfolgerung

- ▶ Die Analyse deutet auf eine komplexe Beziehung zwischen Brustschmerztyp und Ruheblutdruck hin.
- ▶ Weitere Forschung ist notwendig, um diese Beziehungen vollständig zu verstehen und die Ursachen von Herzkrankheiten besser zu identifizieren.

Repository-Link

- ▶ Besuchen Sie unser GitHub-Repository für den vollständigen Code und weitere Details: **GitHub-Link**