

3 Inferenz



Bayes:Start!

3.1 Lernsteuerung

3.1.1 Lernziele

Nach Absolvieren des jeweiligen Kapitels sollen folgende Lernziele erreicht sein.

Sie können ...

- die Definition von Inferenzstatistik sowie Beispiele für inferenzstatistische Fragestellungen nennen
- zentrale Begriffe nennen und in Grundzügen erklären
- den Nutzen von Inferenzstatistik nennen
- erläutern, in welchem Zusammenhang Ungewissheit zur Inferenzstatistik steht
- auch anhand von Beispielen erklären, was ein statistisches Modell ist
- die Grundkonzepte der Regression angeben
- Unterschiede zwischen klassischer und Bayes-Inferenz benennen
- Vor- und Nachteile der klassischen vs. Bayes-Inferenz diskutieren
- Die grundlegende Herangehensweise zur Berechnung des p-Werts informell erklären können

3.1.2 Begleitvideos

- [Video zur Inferenz, Teil 1](#)
- [Video zur Inferenz, Teil 2](#)

3.2 Wozu ist Statistik überhaupt da?

Ja, diese Frage haben Sie sich auch schon mal gestellt?

Abb. [Abbildung 3.1](#) gibt einen Überblick über die Ziele der Statistik.

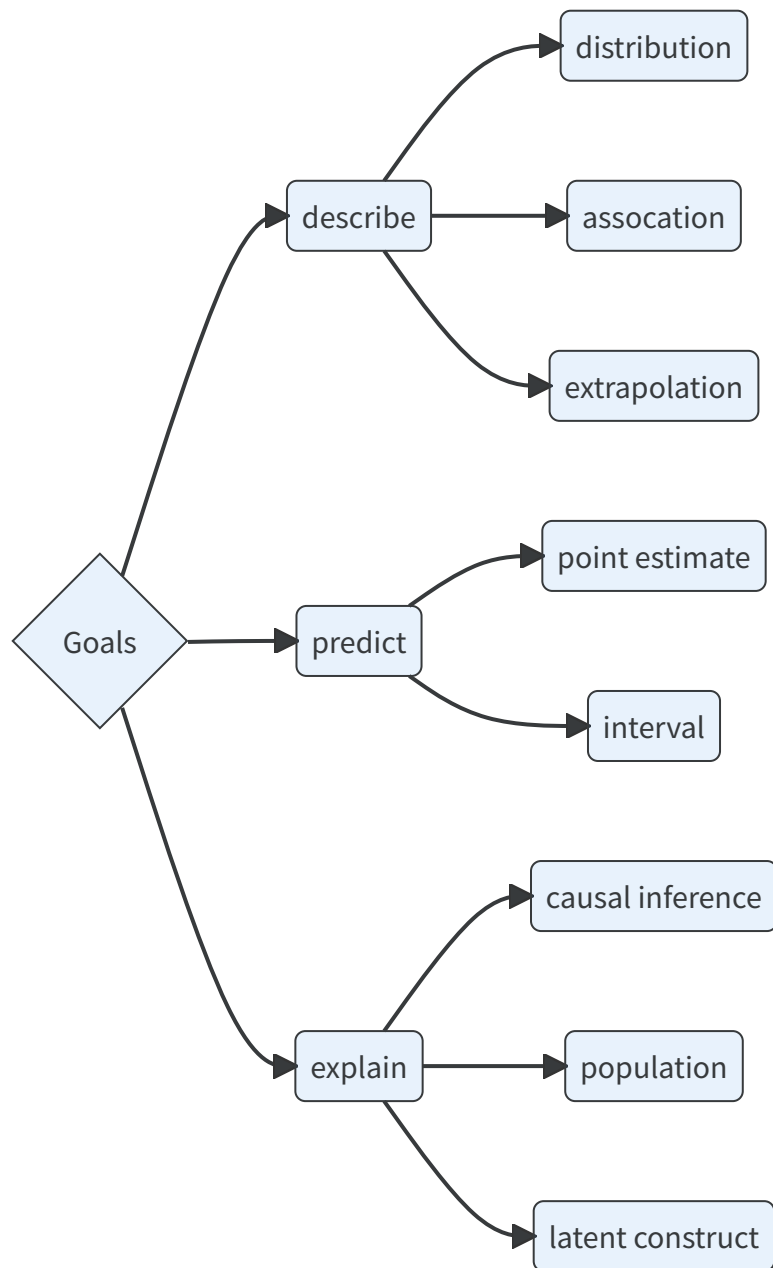


Abbildung 3.1: A taxonomy of statistical goals

***i* Hinweis**

Ziele existieren nicht “in echt” in der Welt. Wir denken sie uns aus. Ziele haben also keine ontologische Wirklichkeit, sie sind epistemologische Dinge (existieren nur in unserem Kopf). Das heißt, dass man sich nach Belieben Ziele ausdenken kann. Allerdings hülfe es, wenn man andere Menschen vom Nutzen der eigenen Ideen überzeugen kann.

3.3 Was ist Inferenz?

3.3.1 Inferenz als Generalisieren

Statistische Inferenz sieht sich drei “Herausforderungen” gegenüber, laut Gelman, Hill, und Vehtari (2021), Kap.

1.1. Diese betreffen das Schließen (oder Generalisieren) vom Einzelfall auf das Allgemeine:

1. Von der Stichprobe aus die Grundgesamtheit (Population)
2. Von der Experimental- auf die Kontrollgruppe (Kausalinferenz)

3. Von einem Messwert auf das zugrundeliegende Konstrukt

In diesem Kurs beschäftigen wir uns mit den ersten beiden Herausforderungen.

! Wichtig

Statistische Inferenz hat zum Ziel, vom Teil aufs Ganze zu schließen, bzw. vom Konkreten auf das Abstrakte.

3.4 Stichprobe vs. Population

Nehmen wir an, wir möchten herausfinden, wie groß der Anteil der R-Fans an der Population der Studierenden ist. Den Anteil der F-Fans bezeichnen wir der Einfachheit halber hier mit A ¹.

Das *Grundproblem der Inferenzstatistik* ist, dass wir an Aussagen zur Grundgesamtheit interessiert sind, aber nur eine Stichprobe, also einen Ausschnitt oder eine Teilmenge der Grundgesamtheit vorliegen haben.

Wir müssen also den Anteil der R-Fans auf Basis des Anteils in der Stichprobe für die Grundgesamtheit schließen: Wir verallgemeinern oder generalisieren von der Stichprobe auf die Grundgesamtheit, s. Abb.

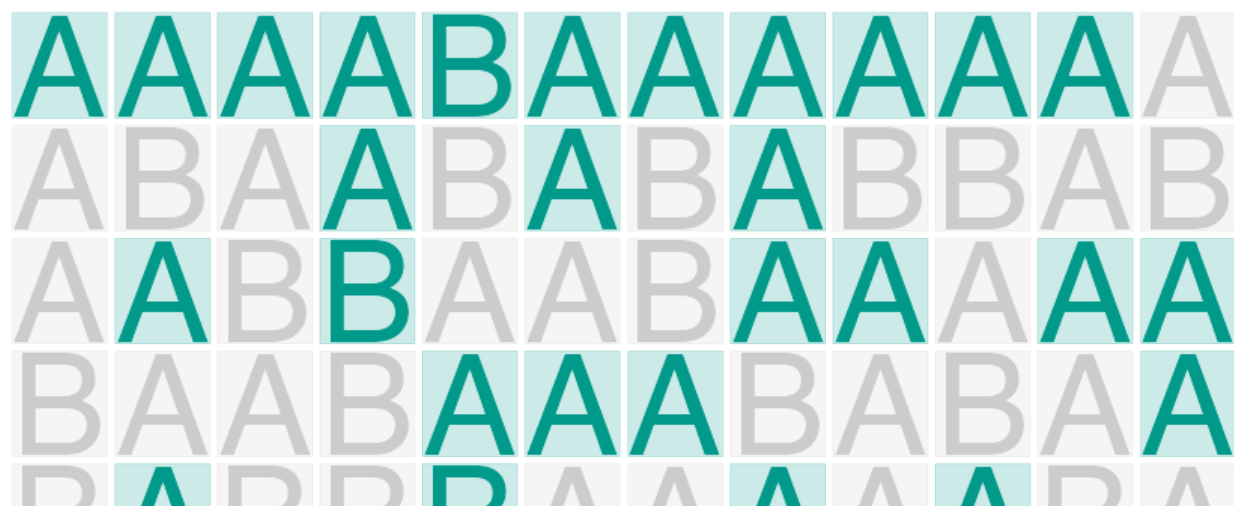
[Abbildung 3.2](#).

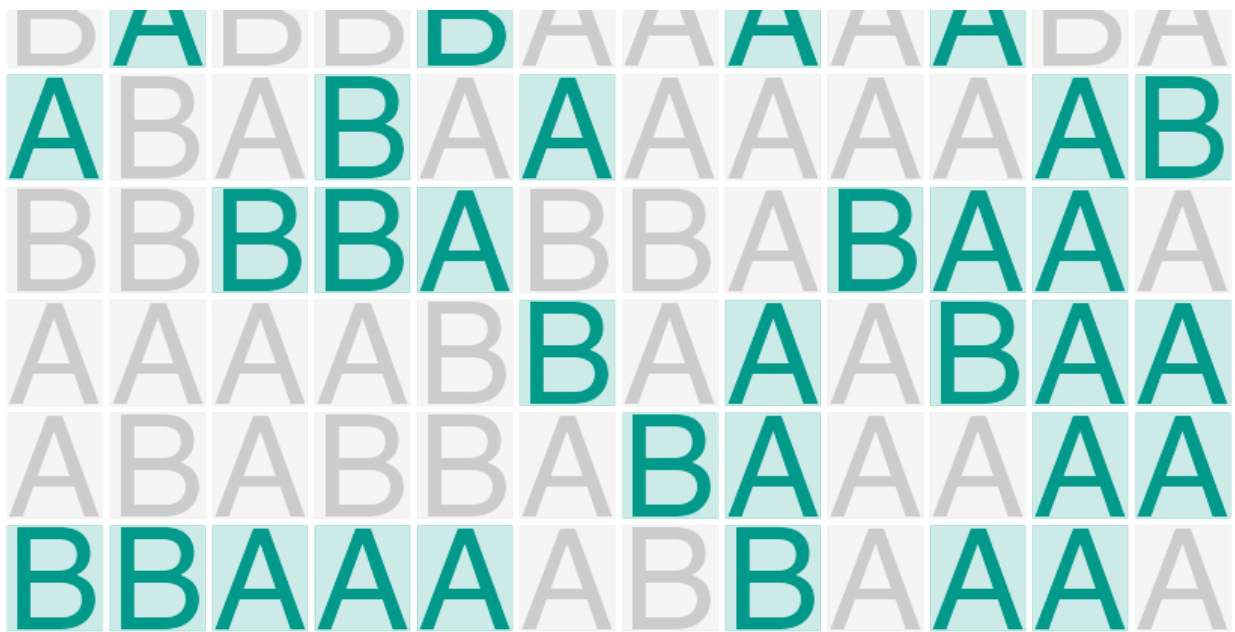
Population



(a) Population

Sample





(b) Sample

Abbildung 3.2: Population vs. sample (Image credit: Karsten Luebke)

Häufig ist das praktische Vorgehen recht simpel: Ah, in unserer Stichprobe sind 42% R-Fans!². Man schreibt: $p = 0.42$ (p wie *proportion*). Die Stichprobe sei repräsentativ für die Grundgesamtheit aller Studierender. Messerscharf schließen wir: In der Grundgesamtheit ist der Anteil der R-Fans auch 42%, $\pi = 0.42$.

3.4.1 Deskriptiv- vs. Inferenzstatistik

Statistik gibt es in zwei Geschmacksrichtungen, könnte man sagen: Deskriptiv- und Inferenzstatistik, s. Abb. [Abbildung 3.3](#). Einteilungen in Schubladen existieren nicht auf der Welt, sondern in unserem Kopf: Sie besitzen keine ontologische Realität, sondern eine epistemologische. Sie sind frei, sich andere Einteilungen der Statistik auszudenken. Es hilft allerdings, wenn man andere Menschen vom Wert seiner Idee überzeugen kann.

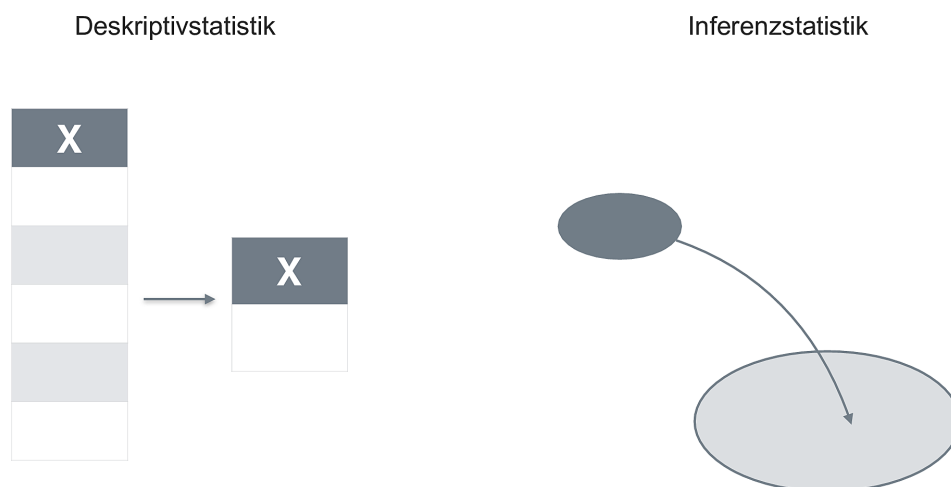


Abbildung 3.3: Deskriptiv- vs. Inferenzstatistik

Deskriptivstatistik fasst Stichprobenmerkmale zu Kennzahlen (Statistiken) zusammen.

Inferenzstatistik schließt von Statistiken auf Parameter (Kennzahlen von Grundgesamtheiten).

👁️ Schließen Sie die Augen und zeichnen Sie obiges Diagramm!

3.4.2 Wozu ist die Inferenzstatistik gut?

Definition 3.1 (Inferent) Inferenz bedeutet Schließen; auf Basis von vorliegenden Wissen wird neues Wissen generiert.

Inferenzstatistik ist ein Verfahren, das mathematische Modelle (oft aus der Stochastik) verwendet, um ausgehend von einer bestimmten Datenlage, die eine Stichprobe einer Grundgesamtheit darstellt, allgemeine Schlüsse zu ziehen.

🏆 Heute Nacht vor dem Schlafen wiederholen Sie die Definition. Üben Sie jetzt schon mal.

3.4.3 Deskriptiv- und Inferenzstatistik gehen Hand in Hand

Für jede beliebige Statistik (Kennzahl von Stichprobendaten) kann man die Methoden der Inferenzstatistik verwenden, s. Tabelle [Tabelle 3.1](#).

Tabelle 3.1: Bezeichnungen für Kennwerte

Kennwert	Stichprobe	Grundgesamtheit
Mittelwert	\bar{X}	μ
Streuung	sd	σ
Anteil	p	π
Korrelation	r	ρ
Regression	b	β

Für Statistiken (Daten einer Stichprobe) verwendet man *lateinische* Buchstaben; für Parameter (Population) verwendet man *griechische* Buchstaben.

🏆 Geben Sie die griechischen Buchstaben für typische Statistiken an!

3.4.4 Schätzen von Parametern einer Grundgesamtheit

Meist begnügt man sich beim Analysieren von Daten nicht mit Aussagen für eine Stichprobe, sondern will auf eine Grundgesamtheit verallgemeinern.

Leider sind die Parameter einer Grundgesamtheit zumeist unbekannt, daher muss man sich mit *Schätzungen* begnügen.

Schätzwerte werden mit einem “Dach” über dem Kennwert gekennzeichnet, z.B.

Kennwert	Stichprobe	Grundgesamtheit	Schätzwert
Mittelwert	\bar{X}	μ	$\hat{\mu}$
Streuung	sd	σ	$\hat{\sigma}$
Anteil	p	π	$\hat{\pi}$
Korrelation	r	ρ	$\hat{\rho}$
Regression	b	β	$\hat{\beta}$

3.4.5 Beispiele für inferenzstatistische Fragestellungen

Sie testen zwei Varianten Ihres Webshops (V1 und V2), die sich im Farbschema unterscheiden und ansonsten identisch sind: Hat das Farbschema einen Einfluss auf den Umsatz?

- Dazu vergleichen Sie den mittleren Umsatz pro Tag von V1 vs. V2, \bar{X}_{V1} und \bar{X}_{V2} .
- Die Mittelwerte unterscheiden sich etwas, $\bar{X}_{V1} > \bar{X}_{V2}$
- Sind diese Unterschiede “zufällig” oder “substanziell”? Gilt also $\mu_{V1} > \mu_{V2}$ oder gilt $\mu_{V1} \leq \mu_{V2}$?
- Wie groß ist die Wahrscheinlichkeit³ $Pr(\mu_{V1} > \mu_{V2})$?

🔧 *Predictive Maintenance* ist ein Anwendungsfeld inferenzstatistischer Modellierung. Lesen Sie dazu S. 3 [dieses Berichts!](#)

3.5 Modellieren

3.5.1 Modellieren als Grundraster des Erkennens

In der Wissenschaft - wie auch oft in der Technik, Wirtschaft oder im Alltag - betrachtet man einen Teil der Welt näher, meist mit dem Ziel, eine Entscheidung zu treffen, was man tun wird oder mit dem Ziel, etwas zu lernen.

Nun ist die Welt ein weites Feld. Jedes Detail zu berücksichtigen ist nicht möglich. Wir müssen die Sache vereinfachen: Alle Informationen ausblenden, die nicht zwingend nötig sind. Aber gleichzeitig die Strukturelemente der wirklichen Welt, die für unsere Fragestellung zentral ist, beibehalten.

Dieses Tun nennt man *Modellieren*: Man erstellt sich ein Modell.

❗ Wichtig

Ein Modell ist ein vereinfachtes Abbild der Wirklichkeit.

Auf die Statistik bezogen heißt das, dass man einen Datensatz zusammenfasst, dass man das Wesentliche erkennt. Was ist das “Wesentliche”? Meist interessiert man sich für die Ursachen eines Phänomens? Etwa: “Wie kommt es bloß, dass ich ohne zu lernen die Klausur so gut bestanden habe?”⁴ Noch allgemeiner ist vom häufig am Zusammenhang von **X** und **Y** interessiert, s. [Abbildung 3.4](#), linker Teil, die ein Sinnbild eines statistischen Modells wiedergibt.

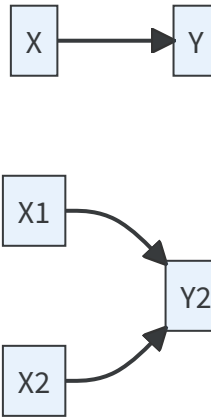


Abbildung 3.4: oben: Sinnbild eines statistischen Modells; unten: Sinnbild eines statistischen Modells, mit zwei Inputvariablen (Ursachen)

Das Diagramm hat Sie nicht so vom Hocker? Okay, ein statistisches Modell kann natürlich komplexer sein, z.B. wie in Abb. [Abbildung 3.4](#), rechter Teil, dargestellt.

Es hört sich zugspitzt an, aber eigentlich ist fast alles Modellieren: Wenn man den Anteil der R-Fans in einer Gruppe Studierender ausrechnet, macht man sich ein Modell: man vereinfacht diesen Ausschnitt der Wirklichkeit anhand einer statistischen Kennzahl, die das forschungsleitende Interesse zusammenfasst.

3.5.2 Vertiefung

Lesen Sie die Einführung zum Thema Modellieren bei Poldrack ([2022](#)) (Kap. 5.1).

Hinweis

Nutzen Sie die Übersetzungsfunktion Ihres Browsers, wenn Sie einen englischen Text lieber auf Deutsch lesen wollen. Oder einen deutschen lieber auf Englisch.

3.6 Regression

Einflussreiche Leute schwören auf die Regressionsanalyse ([Abbildung 3.5](#)).



Abbildung 3.5: One regression

3.6.1 Regression zum Modellieren

Die Regression ist eine Art Schweizer Taschenmesser: Für vieles gut einsetzbar.

Anstelle von vielen verschiedenen Verfahren des statistischen Modellierens kann man (fast) immer die Regression verwenden. Das ist nicht nur einfacher, sondern auch schöner. Wir werden im Folgenden stets die Regression zum Modellieren verwenden.

Dann wenden wir die Methoden der Inferenz auf die Kennzahlen der Regression an.

Hinweis

Regression + Inferenz = 💖

Alternativ zur Regression könnte man sich in den Wald der statistischen Verfahren begeben, [wie hier von der Uni Münster als Ausschnitt \(!\) aufgeführt](#).

Auf dieser Basis kann man meditieren, welches statistischen Verfahren man für eine bestimmte Fragestellung verwenden sollte, s. Abb. [Abbildung 3.6](#).

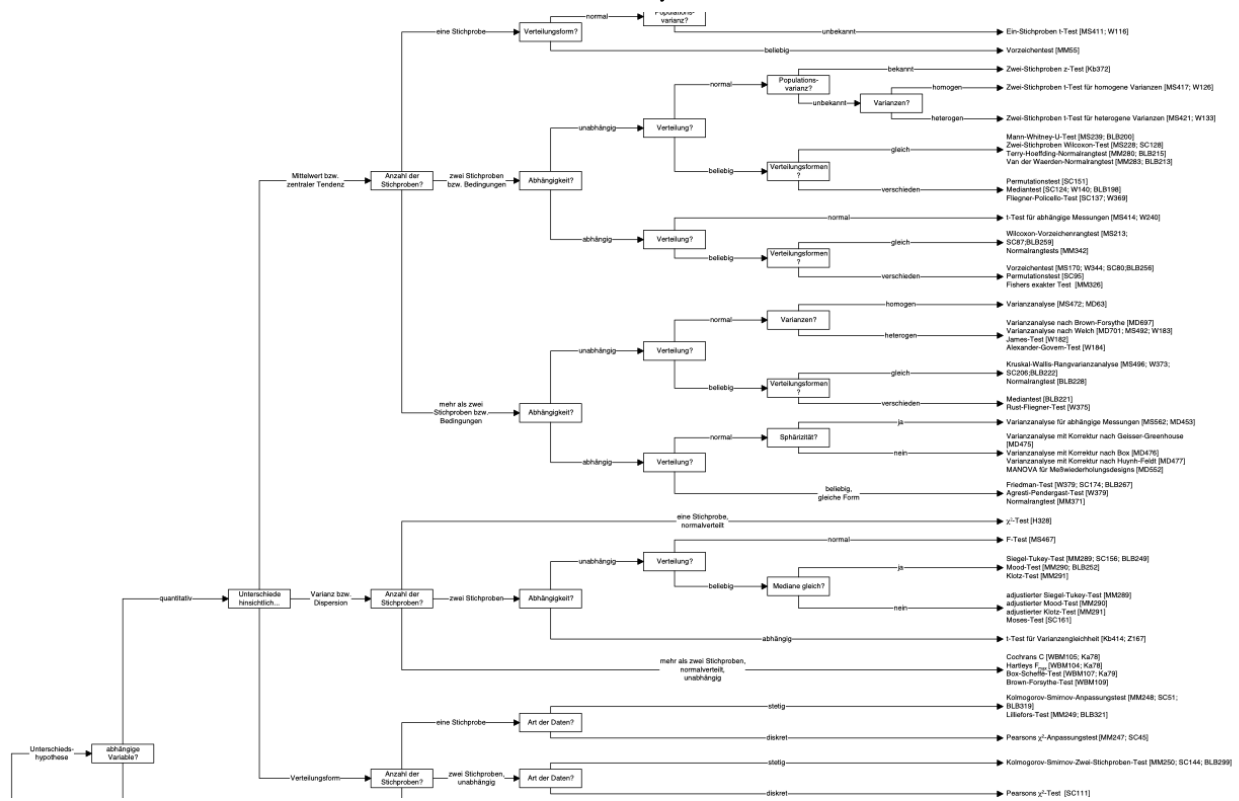


Abbildung 3.6: Wähle deine Statistik mit Bedacht

3.6.2 Viele statistische Verfahren sind Spezialfälle der Regression

Wie Jonas Kristoffer Lindeløv uns erklärt, sind viele statistische Verfahren, wie der sog. t-Test Spezialfälle der Regression, s. Abb. [Abbildung 3.7](#).

Common statistical tests are linear models

Last updated: 02 April, 2019

See worked examples and more details at the accompanying notebook: <https://lindelov.github.io/tests-as-linear>

	Common name	Built-in function in R	Equivalent linear model in R	Exact?	The linear model in words	Icon
Simple regression: $\text{lm}(y \sim 1 + x)$	y is independent of x P: One-sample t-test N: Wilcoxon signed-rank	t.test(y) wilcox.test(y)	$\text{lm}(y \sim 1)$ $\text{lm}(\text{signed_rank}(y) \sim 1)$	✓ for $N \geq 14$	One number (intercept, i.e., the mean) predicts y . - (Same, but it predicts the <i>signed rank</i> of y .)	
	P: Paired-sample t-test N: Wilcoxon matched pairs	t.test(y1, y2, paired=TRUE) wilcox.test(y1, y2, paired=TRUE)	$\text{lm}(y_2 - y_1 \sim 1)$ $\text{lm}(\text{signed_rank}(y_2 - y_1) \sim 1)$	✓ for $N \geq 14$	One intercept predicts the pairwise y - y differences. - (Same, but it predicts the <i>signed rank</i> of y - y .)	
	y ~ continuous x P: Pearson correlation N: Spearman correlation	cor.test(x, y, method='Pearson') cor.test(x, y, method='Spearman')	$\text{lm}(y \sim 1 + x)$ $\text{lm}(\text{rank}(y) \sim 1 + \text{rank}(x))$	✓ for $N \geq 10$	One intercept plus x multiplied by a number (slope) predicts y . - (Same, but with <i>ranked x</i> and y)	
	y ~ discrete x P: Two-sample t-test P: Welch's t-test N: Mann-Whitney U	t.test(y1, y2, var.equal=TRUE) t.test(y1, y2, var.equal=FALSE) wilcox.test(y1, y2)	$\text{lm}(y \sim 1 + G_2)^A$ $\text{glm}(y \sim 1 + G_2, \text{weights} = \dots)^A$ $\text{lm}(\text{signed_rank}(y) \sim 1 + G_2)^A$	✓ ✓ for $N \geq 11$	An intercept for group 1 (plus a difference if group 2) predicts y . - (Same, but with one variance <i>per group</i> instead of one common.) - (Same, but it predicts the <i>signed rank</i> of y .)	
	P: One-way ANOVA N: Kruskal-Wallis	aov(y ~ group) kruskal.test(y ~ group)	$\text{lm}(y \sim 1 + G_2 + G_3 + \dots + G_N)^A$ $\text{lm}(\text{rank}(y) \sim 1 + G_2 + G_3 + \dots + G_N)^A$	✓ for $N \geq 11$	An intercept for group 1 (plus a difference if group $\neq 1$) predicts y . - (Same, but it predicts the <i>rank</i> of y .)	
Multiple regression: $\text{lm}(y \sim 1 + x_1 + x_2 + \dots)$	P: One-way ANCOVA	aov(y ~ group + x)	$\text{lm}(y \sim 1 + G_2 + G_3 + \dots + G_N + x)^A$	✓	- (Same, but plus a slope on x .) Note: this is discrete AND continuous. ANCOVAs are ANOVAs with a continuous x .	
	P: Two-way ANOVA	aov(y ~ group * sex)	$\text{lm}(y \sim 1 + G_2 + G_3 + \dots + G_N + S_2 + S_3 + \dots + S_k + G_2 * S_2 + G_3 * S_3 + \dots + G_N * S_k)^A$	✓	Interaction term: changing sex changes the y ~ group parameters. Note: $G_{2:k,N}$ is an <i>indicator (0 or 1)</i> for each non-intercept levels of the <i>group</i> variable. Similarly for $S_{2:k}$ for <i>sex</i> . The first line (with G_2) is main effect of group, the second (with S_2) for <i>sex</i> and the third is the <i>group x sex</i> interaction. For two levels (e.g. male/female), line 2 would just be " S_2 " and line 3 would be S_2 multiplied with each G_i .	[Coming]
	Counts ~ discrete x N: Chi-square test	chisq.test(groupXsex_table)	Equivalent log-linear model $\text{glm}(y \sim 1 + G_2 + G_3 + \dots + G_N + S_2 + S_3 + \dots + S_k + G_2 * S_2 + G_3 * S_3 + \dots + G_N * S_k, \text{family} = \dots)^A$	✓	Interaction term: (Same as Two-way ANOVA.) Note: Run <i>glm</i> using the following arguments: <code>glm(model, family=poisson())</code> As linear-model, the Chi-square test is $\log(y) = \log(N) + \log(a) + \log(\beta) + \log(a\beta)$ where a_i and β_j are proportions. See more info in the <i>accompanying notebook</i> .	Same as Two-way ANOVA
	N: Goodness of fit	chisq.test(y)	$\text{glm}(y \sim 1 + G_2 + G_3 + \dots + G_N, \text{family} = \dots)^A$	✓	(Same as One-way ANOVA and see Chi-Square note.)	1W-ANOVA

List of common parametric (P) non-parametric (N) tests and equivalent linear models. The notation $y \sim 1 + x$ is R shorthand for $y = 1 \cdot b + a \cdot x$ which most of us learned in school. Models in similar colors are highly similar, but really, notice how similar they *all* are across colors! For non-parametric models, the linear models are reasonable approximations for non-small sample sizes (see "Exact" column and click links to see simulations). Other less accurate approximations exist, e.g., Wilcoxon for the sign test and Goodness-of-fit for the binomial test. The signed rank function is `signed_rank = function(x) sign(x) * rank(abs(x))`. The variables G_i and S_i are *dummy coded* *indicator variables* (either 0 or 1) exploiting the fact that when $\Delta x = 1$ between categories the difference equals the slope. Subscripts (e.g., G_2 or y_1) indicate different columns in data. Im requires long-format data for all non-continuous models. All of this is exposed in greater detail and worked examples at <https://lindelov.github.io/tests-as-linear>.

^A See the note to the two-way ANOVA for explanation of the notation.^B Same model, but with one variance per group: `glm(value ~ 1 + G, weights = varIdent(form = ~1|group), method="ML")`.Jonas Kristoffer Lindeløv
<https://lindelov.net>

Abbildung 3.7: Common statistical tests as linear models

3.6.3 In voller Pracht

Hier ist die Regressionsgleichung in voller Pracht; Abb. [Abbildung 3.8](#).

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon$$

Anhan der Gleichung erkennt man auch, warum man von einem *linearen Modell* spricht: Y wird als gewichteter Mittelwert mehrerer Summanden berechnet. Dabei wird X nicht mit “fortgeschrittenen” Transformationen wie Quadradieren oder Exponenzieren beglückt, sondern nur mit den Regressionsgewichten multipliziert.

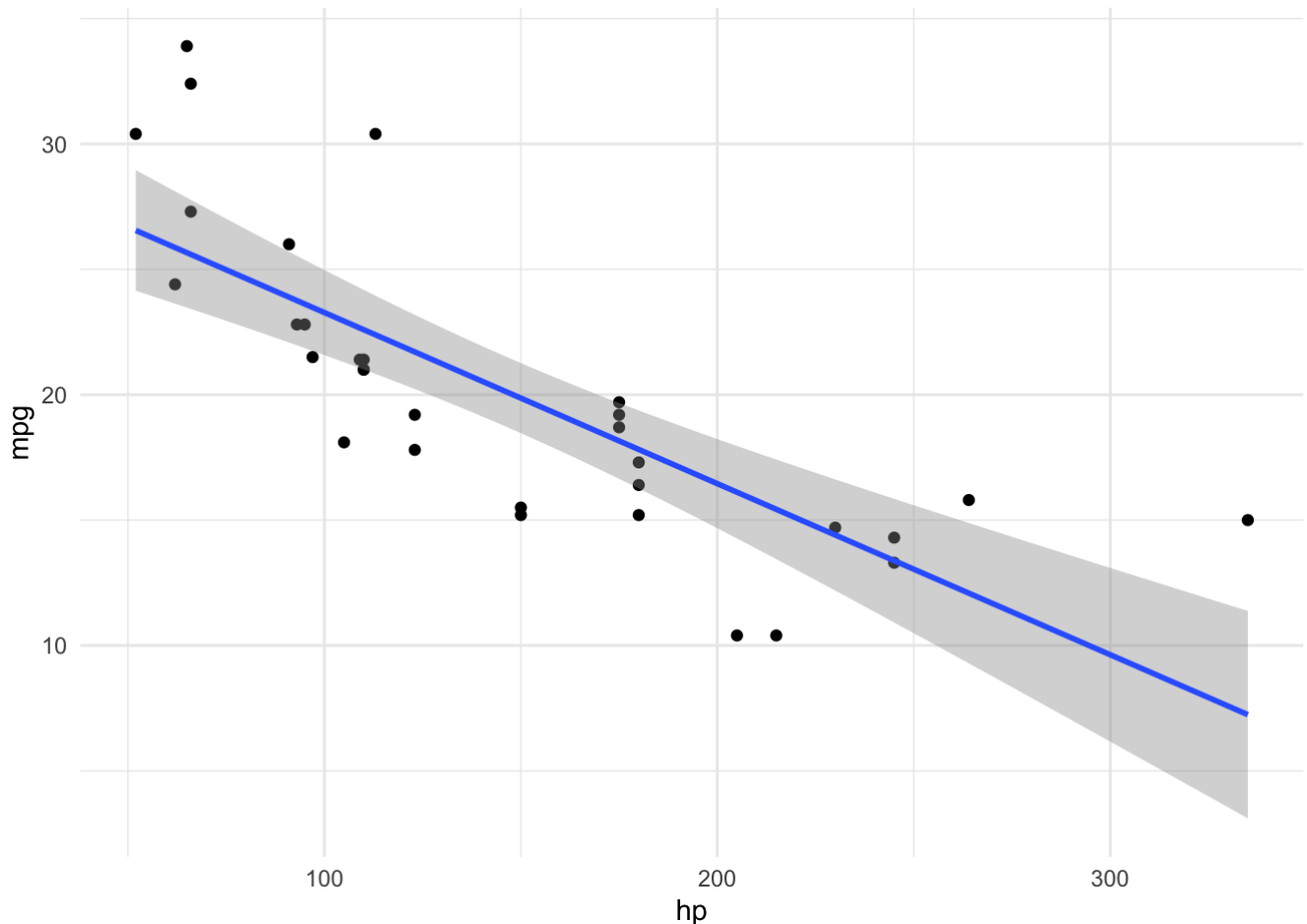


Abbildung 3.8: Die Regressionsgerade in voller Pracht

3.7 Unsicherheit

3.7.1 Inferenz beinhaltet Unsicherheit

Inferenzstatistische Schlüsse sind mit Unsicherheit behaftet: Schließlich kennt man nur einen Teil (die Stichprobe) eines Ganzen (die Population), möchte aber vom Teil auf das Ganze schließen.

! Wichtig

Nichts Genaues weiß man nicht: Schließt man von einem Teil auf das Ganze, so geschieht das unter Unsicherheit. Man spricht von Ungewissheit, da man die Unsicherheit das Wissen über das Ganze betrifft.

Schließt man etwa, dass in einer Grundgesamtheit der Anteil der R-Fans bei 42% liegt, so geschieht das unter Unsicherheit. Man ist sich nicht sicher, dass es wirklich 42% in der Population sind - und nicht etwa etwas mehr oder etwas weniger. Schließlich hat man *nicht* die ganze Population gesehen bzw. vermessen. *Sicher* ist man sich hingegen für die Stichprobe (Messfehler einmal ausgeblendet).

Zur Bemessung der Unsicherheit (Ungewissheit) bedient man sich der Wahrscheinlichkeitsrechnung (wo immer möglich).

Die Wahrscheinlichkeitstheorie bzw. -rechnung wird auch als die Mathematik des Zufalls bezeichnet.

Definition 3.2 (Zufälliges Ereignis) Unter einem zufälligen Ereignis (random) verstehen wir ein Ereignis, das nicht (komplett) vorherzusehen ist, wie etwa die Augenzahl Ihres nächsten Würfelwurfs. Zufällig bedeutet nicht (zwangsläufig), dass das Ereignisse keine Ursachen besitzt. So gehorchen die Bewegungen eines Würfels den Gesetzen der Physik, nur sind uns diese oder die genauen Randbedingungen nicht (ausreichend) bekannt.

🤖 Welche physikalischen Randbedingungen wirken wohl auf einen Münzwurf ein?

3.7.2 Beispiele zur Quantifizierung von Ungewissheit

Aussagen mit Unsicherheit können unterschiedlich präzise formuliert sein.

- Morgen regnet's \Leftrightarrow Morgen wird es hier mehr als 0 mm Niederschlag geben ($p = 97\%$).
- Methode A ist besser als Methode $B \Leftrightarrow$ Mit einer Wahrscheinlichkeit von 57% ist der Mittelwert für Methode A höher als für Methode B .
- Die Maschine fällt demnächst aus \Leftrightarrow Mit einer Wahrscheinlichkeit von 97% wird die Maschine in den nächsten 1-3 Tagen ausfallen, laut unserem Modell.
- Die Investition lohnt sich \Leftrightarrow Die Investition hat einen Erwartungswert von 42 Euro; mit 90% Wahrscheinlichkeit wird der Gewinn zwischen -10000 und 100 Euro.

🤖 Geben Sie weitere Beispiele an!

3.7.3 Zwei Arten von Ungewissheit

Im Modellieren im Allgemeinen und in Regressionsmodellen im Besonderen lassen sich (mindestens) zwei Arten von Ungewissheiten angeben, s. auch Abb. [Abbildung 3.9](#).

1. Wie (un)gewiss ist man sich über den Wert des Regressionsgewichts?
2. Wie (un)gewiss ist man sich über den Wert von Y ? Schließlich könnte es ja Einflüsse (X) geben, die man nicht berücksichtigt hat.

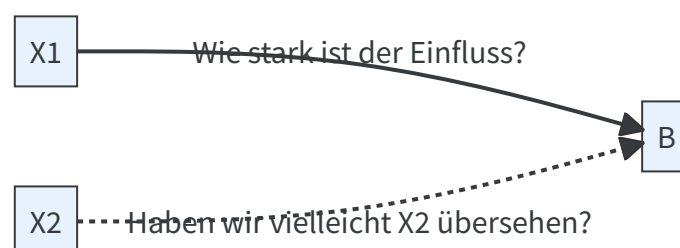


Abbildung 3.9: Zwei Arten der Ungewissheit beim Modellieren

3.7.4 Ich weiß, was ich nicht weiß: Ungewissheit angeben

Streng genommen ist eine Inferenz aus Angabe der Ungewissheit (Genuaigkeit der Schätzung) wertlos.

Angenommen, jemand sagt, dass sie den Anteil der R-Fans (in der Population) auf 42% schätzt, lässt aber offen wie *sicher* (präzise) die Schätzung ist. Wir wissen also nicht, ob z.B. 2% oder 82% noch erwartbar sind. Oder ob man im Gegenteil mit hoher Sicherheit sagen kann, die Schätzung schließt sogar 41% oder 43% aus.

! Wichtig

Eine Inferenz nennt man auch Schätzung. Es sollte immer die Genauigkeit (Ungewissheit) der Schätzung angegeben werden.

Im Rahmen der Regressionsanalyse schlägt sich die Ungewissheit an zwei Stellen nieder:

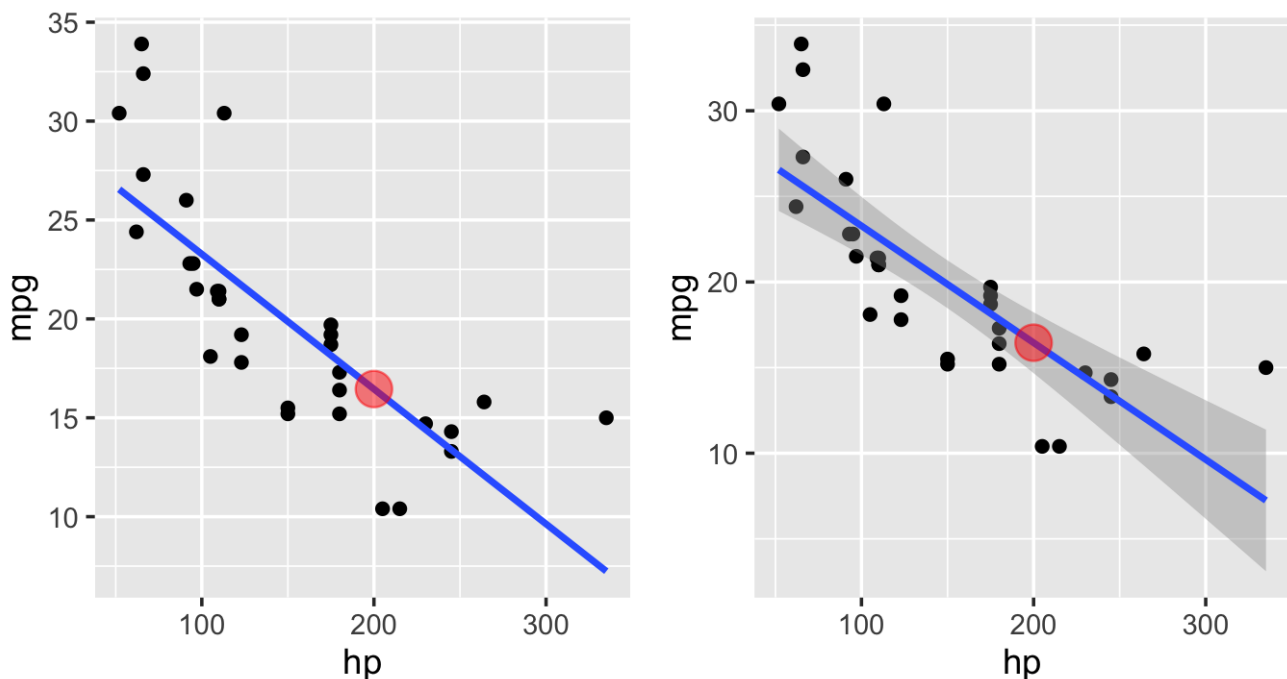
1. zur Lage der Regressionsgeraden (β_0, β_1)
2. zu Einflüssen (X), die unser Modell nicht kennt (ϵ, σ)

3.7.5 Visualisierung von Ungewissheit

Definition 3.3 (Punktschätzer) Gibt man nur einen Punktwert an, wie 42%, als Ergebnis einer Inferenz, spricht man von einem *Punktschätzer*.

Punktschätzer beinhalten *keine Angabe* der Schätz(un)genauigkeit, s. Abb. **?fig-punktschaetzer2**, links. Rot markiert: Die Punktschätzung von `mpg` für `hp=200`.

Eine Punktschätzung mittels einer Regressionsanalyse ohne (links) bzw. mit Ungewissheitsintervall (rechts, in grau)

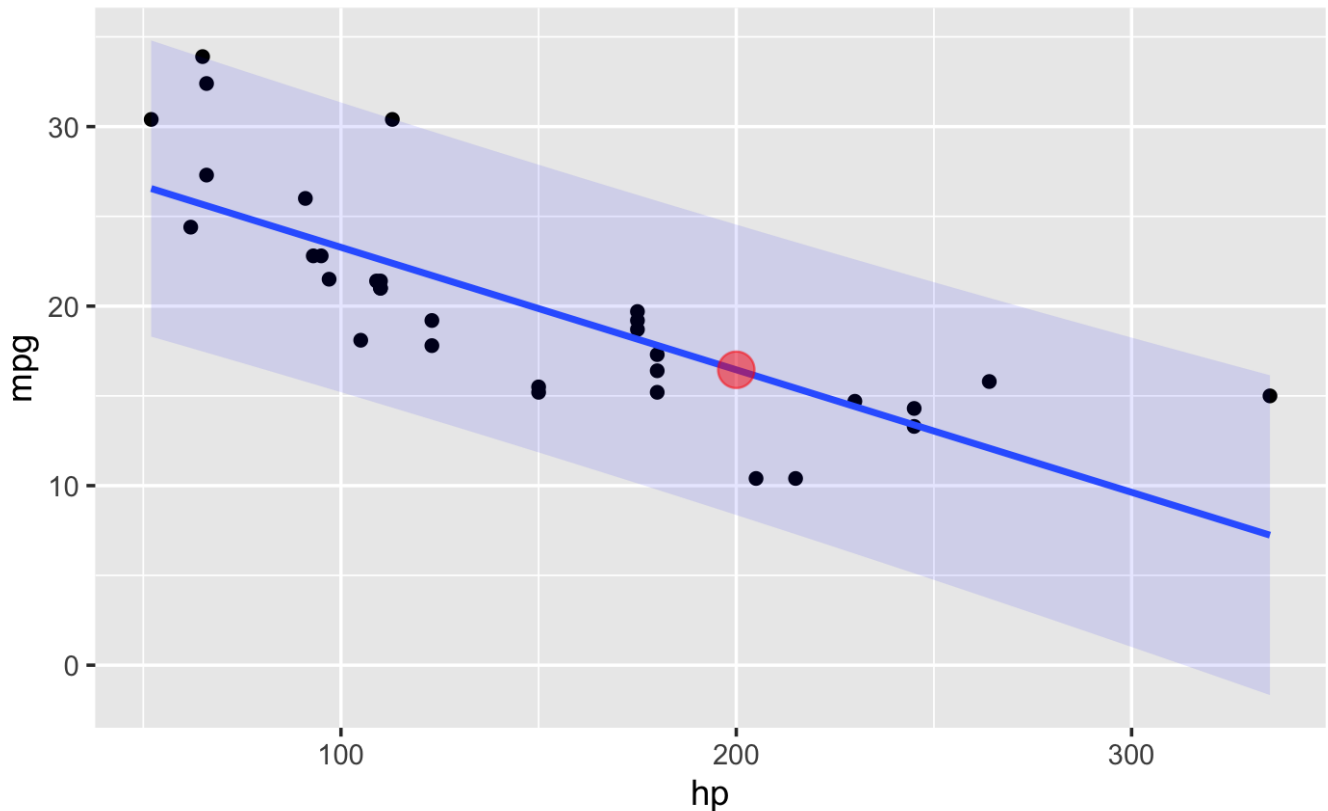


Eine Punktschätzung und ihre Ungewissheit

In Abb. **?fig-punktschaetzer2**, rechts, ist die Ungewissheit in den Regressionskoeffizienten visualisiert: Wie sicher sind wir uns zur Stärke des Zusammenhangs von X und Y, vgl. [Definition 3.3](#).

Auch wenn wir uns *sicher* im Hinblick auf die Regressionsgewichte in Abb. [Abbildung 3.10](#) bliebe eine *Restungewissheit*. Unsere Schätzungen wären auch dann nicht sicher, nicht fehlerfrei. Das liegt daran, da das Modell nicht alle Einflüsse auf Y berücksichtigt, sondern nur einen, hier als X bezeichnet.

In Abb. [Abbildung 3.10](#) ist nicht nur die Ungewissheit durch die Regressionsgewichte, sondern auch die “Restungewissheit” dargestellt. In diesem Fall spricht man von einem “Vorhersageintervall”, da man nicht nur von “typischen Fällen” auf der Regressiongeraden spricht, sondern für echte Fälle Vorhersagen (Schätzungen) tätigt, wo auch die zweite Art von Ungewissheit relevant ist.



Das Vorhersageintervall zeigt eine Punktschätzung und ihre Ungewissheit

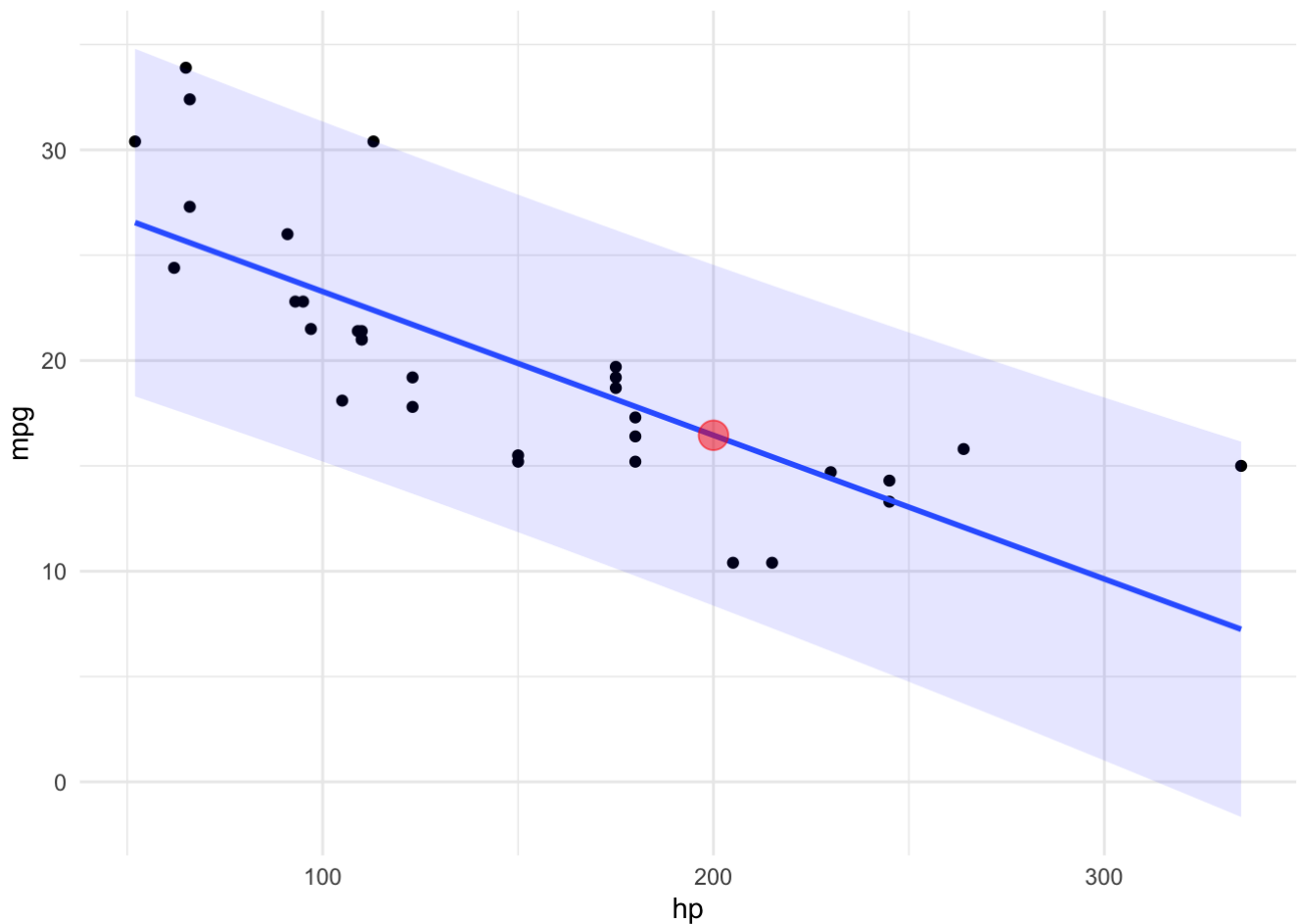


Abbildung 3.10: Zweifache Ungewissheit in den Regressionskoeffizienten - Vorhersageintervall

Wie man sieht, wird die Ungewissheit *größer*, wenn man beide Arten der Ungewissheit berücksichtigt. Das Vorhersage-Intervall berücksichtigt Ungewissheit in $\beta_0, \beta_1, \epsilon$ bei der Vorhersage von \hat{y}_i .

👉 Geben Sie ein vergleichbares Beispiel an!

3.7.6 Konfidenzintervall

Wir sehen hier, dass ein “Ungewissheitskorridor” angegeben wird. Entsprechend wird nicht ein *Punktschätzer*, sondern ein *Schätzbereich* angegeben. Man spricht auch von einem *Konfidenzintervall* oder *Unsicherheitsbereich*⁵

Definition 3.4 (Konfidenzintervall) Ein Konfidenzintervall (confidence intervall, CI) ist ein Oberbegriff für Schätzbereiche. Die Grenzen eines Konfidenzintervall markieren einen Bereich plausibler Werte für einen Parameter.

Ein Konfidenzintervall wird häufig mit 90% oder 95% Genauigkeit angegeben. Im Kontext der Bayes-Analyse ist das einfach zu interpretieren. Sagen wir, wir finden, dass in einem Modell ein 95%-Konfidenzintervall für den Anteil der R-Fans angegeben wird, dass sich von 40 bis 44 Prozent erstreckt. Dieser Befund lässt sich so interpretieren: “Laut Modell liegt der gesuchte Anteil mit einer Wahrscheinlichkeit von 95% im Bereich von 40 bis 44 Prozentpunkten.”

👉 Interpretieren Sie den Ungewissheitskorridor!

3.8 Klassische vs. Bayes-Inferenz

3.8.1 Klassische Inferenz: Frequentismus

- Die Berücksichtigung von Vorwissen zum Sachgegenstand wird vom Frequentismus als subjektiv zurückgewiesen.
- Nur die Daten selber fließen in die Ergebnisse ein
- Wahrscheinlichkeit wird über relative Häufigkeiten definiert.
- Es ist nicht möglich, die Wahrscheinlichkeit einer Hypothese anzugeben.
- Stattdessen wird angegeben, wie häufig eine vergleichbare Datenlage zu erwarten ist, wenn die Hypothese gilt und der Versuch sehr häufig wiederholt ist.
- Ein Großteil der Forschung (in den Sozialwissenschaften) verwendet diesen Ansatz.

3.8.2 Bayesianische Inferenz

- Vorwissen (Priori-Wissen) fließt explizit in die Analyse ein (zusammen mit den Daten).
- *Wenn* das Vorwissen gut ist, wird die Vorhersage genauer, ansonsten ungenauer.
- Die Wahl des Vorwissens muss explizit (kritisierbar) sein.
- In der Bayes-Inferenz sind Wahrscheinlichkeitsaussagen für Hypothesen möglich.
- Die Bayes-Inferenz erfordert mitunter viel Rechenzeit und ist daher erst in den letzten Jahren (für gängige Computer) komfortabel geworden.

3.8.3 Vergleich von Wahrscheinlichkeitsaussagen

3.8.3.1 Frequentismus

Die zentrale Statistik heißt der *p-Wert*

Der p-Wert ist so definiert: “Wie wahrscheinlich ist der Wert der Teststatistik (oder noch extremere Werte), vorausgesetzt die Nullhypothese gilt und man wiederholt den Versuch unendlich oft (mit gleichen Bedingungen, aber zufällig verschieden und auf Basis unseres Modells)?”

Findet man $p < .05$ (oder einen anderen Prozentwert, aber meistens wird 5% hergenommen), so spricht man von “(statistischer) Signifikanz” und nimmt dies als Beleg, dass man einen Effekt gefunden hat, die Hypothese eines Nulleffekts (z.B. kein Zusammenhang von X und Y) also verwerfen kann.

3.8.3.2 Bayes-Statistik

Die zentrale Statistik ist die *Posteriori-Verteilung*.

Die Posteriori-Verteilung beantwortet uns die Frage: “Wie wahrscheinlich ist die Forschungshypothese, jetzt, nachdem wir die Daten kennen, auf Basis unseres Modells?”

🔍 Recherchieren Sie eine Definition des p-Werts und lesen Sie sie genau.

[In diesem Post](#) wird für Bayes geworben und (vielleicht einseitig) Stellung pro Bayes bezogen.

3.8.4 Frequentist und Bayesianer

Im Cartoon 1132 [von xkcd](#) wird sich über das Nicht-Berücksichtigen von Vorab-Informationen (Prior-Verteilung) lustig gemacht, s. [Abbildung 3.11](#).

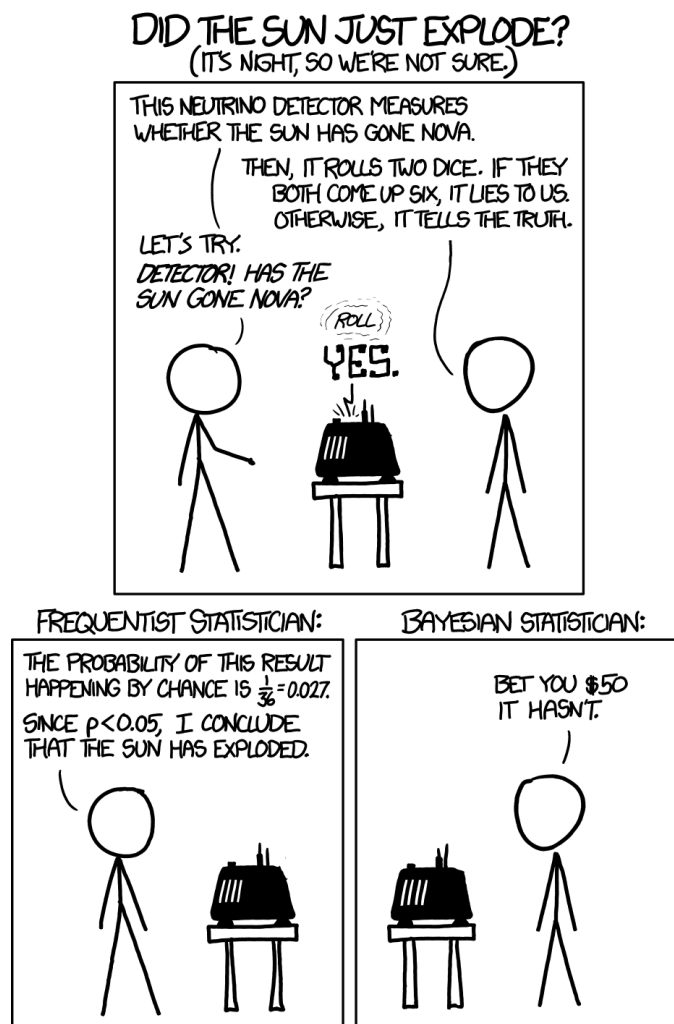


Abbildung 3.11: Frequentist wettet mit Bayesianer

[Quelle](#)

3.8.5 Der p-Wert ist wenig intuitiv



[from Imgflip Meme Generator](#)

3.8.6 Beispiel zum Nutzen von Apriori-Wissen 1

Ein Betrunkener behauptet, er könne hellsehen. Er wirft eine Münze 10 Mal und sagt jedes Mal korrekt vorher, welche Seite oben landen wird.

Die Wahrscheinlichkeit dieses Ergebnisses ist sehr gering (2^{-10}) unter der Hypothese, dass die Münze fair ist, dass Ergebnis also “zufällig” ist.

Unser Vorwissen lässt uns allerdings trotzdem an der Hellsichtigkeit des Betrunkenen zweifeln, so dass die meisten von uns die Hypothese von der Zufälligkeit des Ergebnisses wohl nicht verwerfen.

3.8.7 Beispiel zum Nutzen von Apriori-Wissen 2

Eine Studie (vgl. Gelman, Hill, und Vehtari (2021)) fand einen “großen Effekt” auf das Einkommen von Babies, eine Stunde pro Woche während zwei Jahren an einem psychosozialen Entwicklungsprogramm teilnahmen (im Vergleich zu einer Kontrollgruppe), $n = 127$.

Nach 20 Jahren war das mittlere Einkommen der Experimentalgruppe um 42% höher (als in der Kontrollgruppe) mit einem Konfidenzintervall von $[+2\%, +98\%]$.

Allerdings lässt uns unser Vorwissen vermuten, dass so ein Treatment das Einkommen nach 20 Jahren kaum verdoppeln lässt. Wir würden den Effekt lieber in einem konservativeren Intervall schätzen (enger um Null).

3.9 Literatur

Bei Gelman, Hill, und Vehtari (2021), Kap. 1 findet sich eine Darstellung ähnlich zu der in diesem Kapitel.

3.10 Fazit

! Wichtig

[Kontinuierliches Lernen](#) ist der Schlüssel zum Erfolg.

3.11 Aufgaben

1. [Griech-Buchstaben-Inferenz](#)
2. [korr-als-regr](#)
3. [ttest-als-regr](#)
4. [ttest-skalenniveau](#)
5. [adjustieren2](#)
6. [inferenz-fuer-alle](#)
7. [adjustieren1](#)
8. [ungewiss-arten-regr](#)
9. [vorhersageintervall1](#)
10. [lm-standardfehler](#)
11. [punktschaetzer-reicht-nicht](#)

3.12 —



-
1. ~~Meistens~~ Manchmal darf man bei der Statistik nicht nach einem tieferen Sinn suchen. Ist Statistik eine Art moderne Kunst? [↗](#)
 2. Mancheiner hätte mit mehr gerechnet [↗](#)
 3. oft mit *Pro* oder *p* abgekürzt, für *probability* [↗](#)
 4. Das ist natürlich nur ein fiktives, komplett unrealistisches Beispiel, das auch unklaren Ursachen den Weg auf diese Seite gefunden hat. [↗](#)
 5. Tatsächlich gibt es mehrere Synonyme oder ähnliche Begriffe für Konfidenzintervall. Wir kommen später darauf detaillierter zu sprechen. [↗](#)