# Grundlagen des statistischen Testens

Dr. Mariana Nold

Institut für Soziologie,
Fakultät für Sozial- und Verhaltenswissenschaften,
Lehrstuhl für empirische Sozialforschung und Sozialstrukturanalyse

13. November 2017

# Übersicht

# Ziel der heutigen Veranstaltung ...

ist es die folgenden Fragen beantworten zu können:

## Zielfragen für heute

1. Was bedeuten die Begriffe Nullhypothese und Forschungshypothese?
2. Welche der beiden Hypothesen kann man nachweisen?
3. Wie sind der $\alpha$-Fehler und der $\beta$-Fehler definiert?
4. Was ist das Signifikanzniveau bzw. die Irrtumswahrscheinlichkeit?
5. Was versteht man unter einer Teststatistik?
6. Welche Rolle spielt der kritische Wert?
7. Was versteht man unter einem p-Wert?

# Die Grundidee des Signifikanztests

- Wir werden uns heute an einem Beispiel ansehen, welcher Logik der klassische Signifikanztest folgt.
- Unser Fokus liegt hierbei weniger auf die mathematischen Zusammenhänge, als vielmehr auf die wissenschaftstheoretische Legitimation dieser Vorgehensweise.
- Die Darstellung orientiert sich an den lesenswerten Kapiteln
    - *4.3.1: Die Grundidee von Signifikanztests (+Vorwort, ab S. 136)*
    - *4.3.2: Die Praxis von Signifikanztests am Beispiel des Testens von Mittelwertunterschieden (nur bis S. 151)*

  aus dem Buch “Statistik-Eine Einführung für Sozialwissenschaftler” von Ludwig-Mayerhofer, Liebeskind, Geißler, (im Folgenden mit LMLG abgekürzt.)

# Gibt es einen Unterschied hinsichtlich eines Merkmals in zwei Gruppen?

Ich werde heute anhand eines Beispiels darstellen, was die Behauptung: " Der Unterschied hinsichtlich eines Merkmals in zwei Gruppen ist signifikant" bedeutet. Zu dem Beispiel habe ich keine Daten.

## Reales Beispiel: Kann das Zufall sein?

In einer Einrichtung zur stationären Drogenfreien Therapie, hat ein Klient den Eindruck zu wenig Gulasch auf seinem Teller zu haben. Er fragt jemand vom Personal, ob es Zufall sein kann, dass er und die anderen Klientinnen und Klienten, wenn es Gulasch gibt, immer nur zwei oder drei Stücke Fleisch auf dem Teller haben und die Mitarbeiterinnen und Mitarbeiter deutlich mehr.

# Die Vermutung des Klienten

- Er hat die Vermutung, dass der Koch, der das Gulasch ausgibt tiefer schöpft, wenn ein Mitarbeiter oder eine Mitarbeiterin vor ihm steht.
- Wenn seine Vermutung zutrifft, dann sind es zwei Unterschiedliche Zufallsmechanismen.
- In diesem Fall, kann es vorkommen, dass eine Klientin bzw. ein Klient mehr Fleisch auf dem Teller hat als ein Mitarbeiter oder eine Mitarbeiterin, **aber nach Wahrscheinlichkeit, haben die Angestellten mehr Fleisch auf dem Teller, als die die dort eine Therapie machen?**
- Die Vermutung kann daher überprüft werden, indem man die Mengen an Fleisch vergleicht.
- Dabei sind verschieden Vorgehensweisen denkbar.

# Kann der Sachverhalt mit statistischen Methoden geklärt werden?

- Nur ein Stück weit, denn es bleibt immer eine Unsicherheit. Man kann allerdings die Unsicherheit quantifizieren.
- Desto mehr empirische Information uns vorliegt, desto sichere wird unsere Aussage. Mit anderen Worten: Desto länger wird diesen Ausgabeprozess beobachten, desto werden wir in der Beurteilung der Vermutung
- Dabei gilt die Annahme, dass der Prozess sich nicht verändert. Das bedeute, wir gehen davon aus, dass das Modell mit dem wir arbeiten, die ganze Zeit über gleicher Maßen gut passt.

# Ein geeignetes Modell I

Wir brauchen ein Modell, um auf der Modellebene arbeiten zu können.

- Das Modell ist folgendes:
  Ich bezeichne mit der Zufallsvariable $X$, die Menge Fleisch auf pro Teller, in der Gruppe der Klientinnen und Klienten und mit $Y$ entsprechen die Menge Fleisch auf den Tellern der Mitarbeiterinnen und Mitarbeiter.
- In einem Gedankenexperiment, wird der Prozess einen Monat lang beobachtet.
- Immer, wenn es Gulasch gibt, wird aus jeder Gruppe eine einfache Zufallsstichprobe gezogen. Die am Ende erreichten Stichprobenumfänge werden mit $n_1$ (für $X$) und $n_2$ (für $Y$) bezeichnet.

# Die Daten

- In der List der Klienten und Klientinnen sehen die Daten formal so aus: $x_1, x_2, \ldots, x_{n_1}$, dabei entspricht jedes $x_i$ einer beobachteten Zahl.
- In der List der Mitarbeiter und Mitarbeiterinnen sehen die Daten formal so aus: $y_1, y_2, \ldots, y_{n_2}$, dabei entspricht jedes $y_i$ einer beobachteten Zahl.
- Welche Grafik würden sie erstellen, um einen ersten Eindruck zu gewinnen, ob sie die beiden Verteilungen unterscheiden?
- Gut sind z. B. zwei Boxplots nebeneinander.

# Ein geeignetes Modell II

- In formaler Schreibweise haben wir $n_1$ Realisation der Zufallsvariable $X_1$ und $n_2$ Realisation der Zufallsvariable $Y$.
- Ich mache jetzt die folgenden Modellnahmen $X_i$ iid wie

$$X \sim N(\mu_X, \sigma_X) \tag{1}$$

  und $Y_i$ iid wie

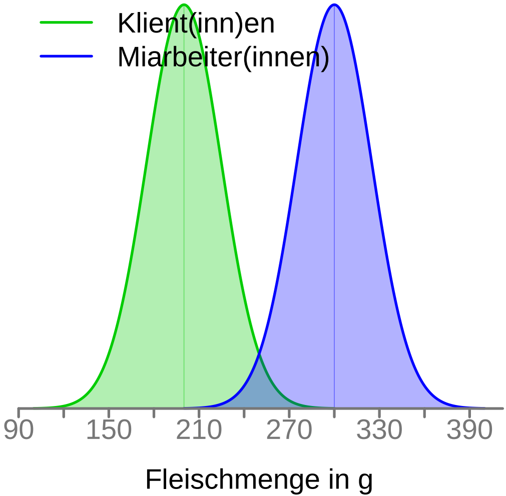
$$Y \sim N(\mu_Y, \sigma_Y). \tag{2}$$

- iid bedeutet unabhängig und identisch verteilt.

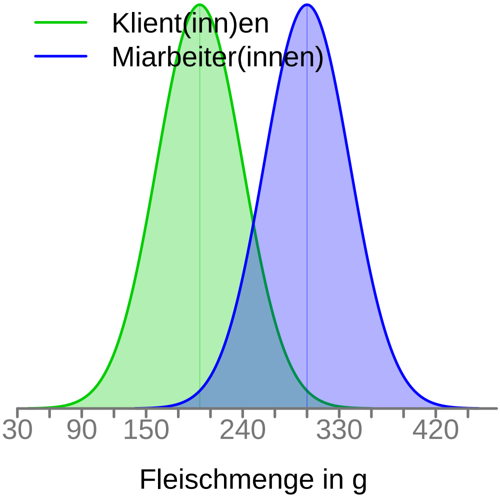
# Vergleich von Dichtkurven

- Die folgenden Bilder zeigen in grün die Modell-Dichte für $X$ und in blau die Modell-Dichte für $Y$.
- In den grafisch dargestellten Modellen gilt jeweils: $\sigma_X = \sigma_Y$
- Bitte beantworten Sie für jedes der Bilder die folgende Frage: Wenn die Modell so der Realität entsprechen würden, was würde das bedeuten?
- Informell formuliert: Wie fair bzw. unfair ist die Gulasch-Ausgabe jeweils?

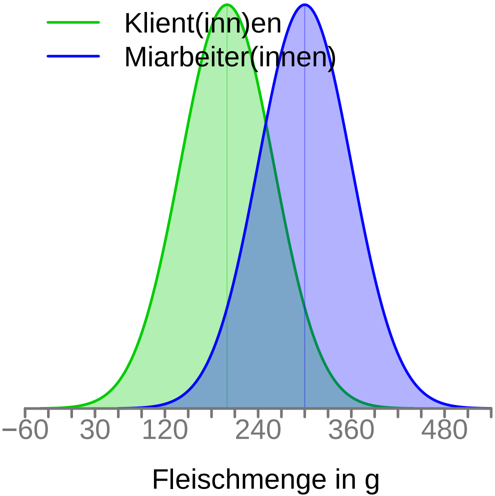
## $\mu_X = 200$, $\mu_Y = 300$, $\sigma_X = \sigma_Y = 25$



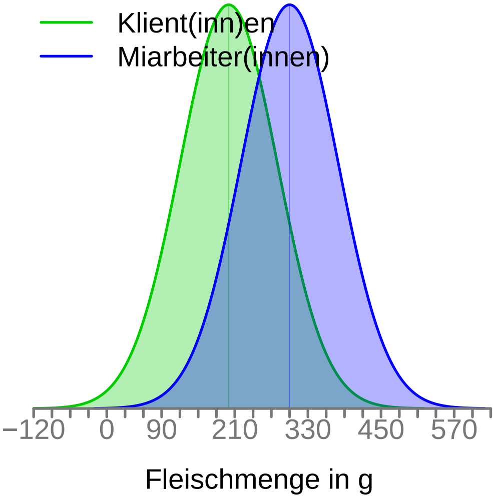
$\mu_X = 200,\ \mu_Y = 300,\ \sigma_X = \sigma_Y = 40$



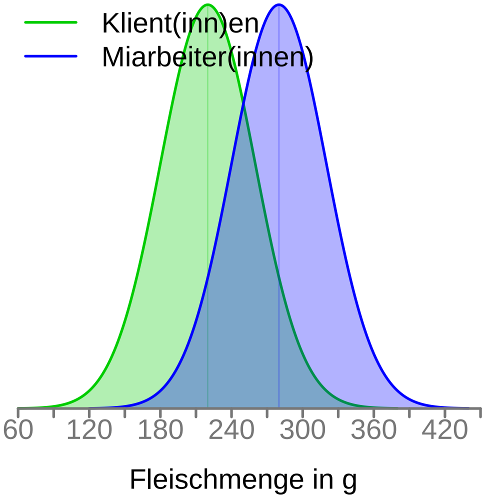
## $\mu_X = 200,\ \mu_Y = 300,\ \sigma_X = \sigma_Y = 60$



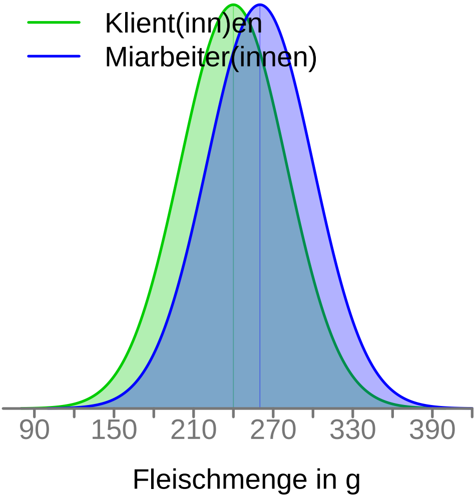
$\mu_X = 200$, $\mu_Y = 300$, $\sigma_X = \sigma_Y = 80$



$\mu_X = 220$, $\mu_Y = 280$, $\sigma_X = \sigma_Y = 40$



## $\mu_X = 240$, $\mu_Y = 260$, $\sigma_X = \sigma_Y = 40$



# Die Rolle der Stichprobenumfänge $n_1$ und $n_2$.

- Desto mehr Informationen man hat, desto kleiner sind die Unterschiede die der Test aufdecken kann.
- Nehmen wir an, die Differenz $\delta := \mu_x - \mu_y$ beträgt in Wirklichkeit nur $10g$, dann braucht man einen große Stichprobenumfänge $n_1$ und $n_2$ um diese kleine Differenz aufdecken zu können.
- Wenn $\delta$ einen Wert von 100 g hat wird bereits ein kleiner Stichprobenumfang von z. B. 10 Personen pro Gruppe zu einer Ablehnung der Nullhypothese führen.
- Merksatz: Desto stärker die Nullhypothese verletzt ist, desto weniger empirische Information braucht man, um die Forschungshypothese nachzuweisen.

# Die Hypothese

Wenn man die Vermutung als statistische Hypothese formuliert, lautet sie

$$\mu_X < \mu_Y.$$

Die Gegenvermutung lautet:

$$\mu_X \geq \mu_Y.$$

Man nennt die Gegenvermutung im Nullhypothese und die Vermutung Forschungshypothese (bzw. Alternativhypothese). Die beiden Hypothesen sie so formuliert, die eine das logische Gegenteil der anderen ist.

# Die Grundidee des Signifikanztest I

- Beim klassischen statistischen Testen, gehen wir davon aus, dass die Nullhypothese gilt.
- Die Nullhypothese ist das Gegenteil von der Vermutung die wir nachweisen möchten. Wir gehen also davon aus, dass die Hypothese die wir nachweisen möchten nicht gilt.
- Die Vorgehensweise erinnert an die Unschuldsvermutung im deutschen Strafrecht aus Perspektive der Staatsanalschaft.
- Die Staatsanwaltschaft möchte die Schuld nachweisen, die Schuld entspricht dann der Forschungshypothese. Sie geht von der Unschuld aus. Die Unschuld ist die Nullhypothese.

# Die Grundidee des Signifikanztest II (vgl. LMLG S. 139)

- Unsere Nullhypothese (Unschuldsvermutung) ist $\mu_X \geq \mu_Y$. Wir gehen davon aus, dass sie die Realität beschreibt.
- Wenn wir nun in der Stichprobe tatsächlich ein Ergebnis beobachten, das im Lichte der Nullhypothese unwahrscheinlich ist, so spricht das *gegen* diese hypothetische Annahme.
- Wir schließen dann, dass die Nullhypothese nicht haltbar ist, mit Bezug auf die Daten.
- Wenn die Nullhypothese nicht haltbar ist, dann wird sie abgelehnt.
- Wenn wir die Hypothese $\mu_X \geq \mu_Y$ ablehnen, dann folgt logische das wir $\mu_X < \mu_Y$ nachgewiesen haben.

# Ein Irrtum ist möglich, ...

genauer zwei Irrtümer sind möglich:

1) Wir lehnen die Nullhypothese ab, obwohl sie wahr ist. Man nennt diesen Fehler $\alpha$-Fehler oder Fehler 1. Art.
2) Wir behalten die Nullhypothese bei, obwohl sie falsch ist. Man nennt diesen Fehler $\beta$-Fehler oder Fehler 2. Art.

**Vorsicht!**

Es ist nicht möglich die Nullhypothese nachzuweisen. Man kann sie nur beibehalten, weil nicht genug empirische Evidenz dagegen spricht.

Eben darin liegt der Grund dafür, dass man das was man Nachweisen möchte, nicht als Nullhypothese formuliert.

# Jede Entscheidung ist mit einem möglichen Fehler verbunden

- Wie gehen davon aus, dass der Koch den Klientinnen und Klienten mindestens so viel Fleisch auf den Teller schöpft, wie den Mitarbeiterinnen und Mitarbeitern. In Formalsprache:

$$H_0 : \mu_X \geq \mu_Y$$

- Wenn wir die Nullhypothese beibehalten, besteht die Möglichkeit einen Fehler 2. Art zu machen.
- Wenn wir sie ablehnen, besteht die Möglichkeit den Fehler 1. Art zu machen.

# Die Bedeutung des Signifikanzniveaus

- Der Begriff Signifikanzniveau ist gleichbedeutend mit dem Begriff Irrtumswahrscheinlichkeit.
- Desto niedriger das Signifikanzniveau (bzw. die Irrtumswahrscheinlichkeit), desto stärker müssen die Daten gegen die Nullhypothese sprechen damit sie verworfen wird.
- Sehr häufig wählt man das Signifikanzniveau 5%.
- Wenn wir noch mehr Sicherheit möchten, um möglichst sicher zu sein, dass wir nicht $H_0$ verwerfen, obwohl es in Wirklichkeit gilt, dann können wir auch ein Signifikanzniveau von 1% wählen.

# Nullhypothese verwerfen

Wenn die Analyse der Daten, also signifikant gegen die Nullhypothese spricht mit einem Signifikanzniveau von 5%, dann können wir dem Koch sagen:

## Inhaltliche Bedeutung

Lieber Koch, es kann sein, dass $H_0 : \mu_X \geq \mu_Y$ gilt, aber es ist sehr unwahrscheinlich, dass wir unter Gültigkeit von $H_0$ die Daten so wie sind beobachten können. Wir gehen davon aus, dass es kein Zufall ist und verwerfen die Nullhypothese. Die Wahrscheinlichkeit für einen Irrtum beträgt 5%.

# Wie kann man entscheiden, ab wann man $H_0$ verwirft?

- Für einen (klassischen) parametrischen statistischen Test, braucht man ein Teststatistik.
- Für die Herleitung unserer Teststatistik brauchen wir die Modellannahme der Normalverteilung, also $X \sim N(\mu_X, \sigma_X)$ und $Y \sim N(\mu_Y, \sigma_Y)$. Wir nehmen an $X_i$ iid wie $X$ und $Y_i$ iid wie Y.
- Im Folgenden mache ich die zusätzliche Annahme $\sigma_X = \sigma_Y$.
- Die zeichnet sich durch zwei Eigenschaften aus:
  1) Sie ist sensibel für das Testproblem.
  2) Ihre Verteilung unter $H_0$ ist bekannt.

# Die zusätzliche Annahme $\sigma_X = \sigma_Y$.

- Je nachdem ob man diese Annahme macht oder nicht, gibt es zwei unterschiedliche Tests, die auch auf unterschiedlichen Teststatistiken beruhen.
- Ich mache diese Annahme jetzt, um anhand der entsprechenden Teststatistik, zu zeigen, wie man darüber entscheidet, ob man die Nullhypothese beibehält oder nicht.
- Es gibt auch einen stat. Test mit der Nullhypothese $H_0 : \sigma_X = \sigma_Y$.
- Manchmal rechnet man zuerst diesen Test um dann zu entscheiden, ob man von der Gleichheit der Standardabweichungen ausgehen kann oder nicht.
- Natürlich ist auch hier ein Irrtum möglich.

# Unsere Teststatistik

Leseempfehlung: (vglm LMLG S. 141- 152)

- Die Nullhypothese $H_0$ bzw. die Forschungshypothese $H_1$ beinhalten eine Beziehung zwischen den Erwartungswerten der Normalverteilung $\mu_X$ und $\mu_Y$.
- Wir hatten schon im letzten Semester gesehen, dass das arithmetische Mittel $\bar{X}$ bzw. $\bar{Y}$ ein guter Schätzer für $\mu_X$ bzw $\mu_Y$ ist.
- Mit den Gesetzen der Wahrscheinlichkeitstheorie lässt sich auch nachweisen, dass das arithmetische Mittel als Schätzer für den Erwartungswert bei normalverteilten Daten eine hohe Güte hat.

# Die Teststatistik entwickeln I

Die Nullhypothese lautet:

$$H_0 : \mu_X \geq \mu_Y$$

Gleichbedeutend damit ist:

$$H_0 : \mu_X - \mu_Y \geq 0$$

Aus den gerade genannten Gründen ist die Differenz der Mittelwerte

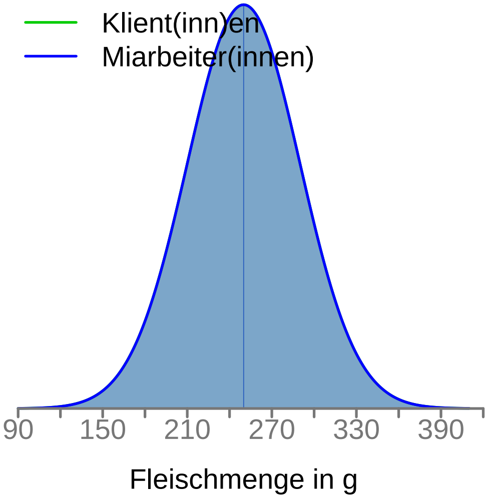
$$\bar{X} - \bar{Y}$$

ein plausibler Schätzer für die Differenz der Modellparameter $\mu_X$ und $\mu_Y$.

# Die Teststatistik entwickeln II

- Wir werden also aus den Daten $Z := \bar{X} - \bar{Y}$ berechnen und auf das Ergebnis dieser Rechnung unsere Entscheidung über die Ablehnung bzw. Beibehaltung der Nullhypothese stützen.
- Wir müssen dafür wissen, wie sich die Zufallsvariable $Z$ verhält, denn wir müssen wissen, welche Werte von $Z$ gegen die Nullhypothese sprechen und auch wie stark sie dagegen sprechen.
- Ein Problem dabei ist, dass unsere Nullhypothese ein ganzes Intervall abdeckst. Wir ersetzen die Nullhypothese für den Moment durch $\tilde{H}_0 : \mu_X = \mu_Y$ bzw. analog $\tilde{H}_0 : \mu_X - \mu_Y = 0$
- Im Klartext bedeutet die modifizierte Nullhypothese, dass wir von völlig identischen Verteilungen der Fleischmengen pro Portion Gulasch ausgehen.

# Die modifizierte Nullhypothese grafisch



# Wir gehen davon aus, dass die modifizierte Nullhypothese zutrifft. . .

. . . also davon, dass der Schöpfmechanismus beim Gulasch Schöpfen sich in den beiden Personengruppen gar nicht unterscheidet.

- Die Teststatistik $Z := \bar{X} - \bar{Y}$ erfüllt die Eigenschaft, dass sie sensibel ist für unser Testproblem. Genau aus dem Testproblem haben wir sie ja abgeleitet.
- Sie erfüllt noch nicht die Eigenschaft, dass ihre Verteilung unter $H_0$ bekannt ist.
- Diese Eigenschaft muss erfüllt sein. Nur so können wir die Wahrscheinlichkeit für den Fehler 1. Art berechnen.

# Die Verteilung der Teststatistik $Z := \bar{X} - \bar{Y}$

- Die Verteilung der Teststatistik $Z := \bar{X} - \bar{Y}$ hängt von der Standardabweichung $\sigma := \sigma_X := \sigma_Y$ ab.
- Diese Standartabweichung kennen wir nicht.
- Daher wird eine neue Teststatistik $T$ definiert,

$$T := \frac{\bar{X} - \bar{Y}}{S \cdot \left(\frac{1}{n} + \frac{1}{m}\right)},$$

  wobei $S$ ein Schätzer für den Modellparameter $\sigma$ ist.
- Diese Teststatistik erfüllt die beiden gewünschten Eigenschaften.

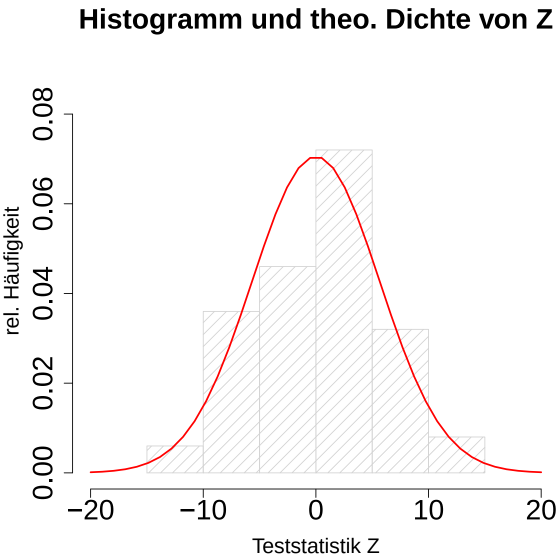
# Zur Sensibilität der Teststatistik $Z$

- Unter $\tilde{H}_0$ kann man, für $Z = \bar{X} - \bar{Y}$
- Stellen Sie sich vor sie erheben Daten für je 10 Personen pro Gruppe. Wo würden Sie persönlich die Grenze ziehen und die Nullhypothese verwerfen?
- In einer fiktiven Erhebung ergibt sich die mittlere Fleischmenge $\bar{x} = 237.6$ g und $\bar{x} = 252.0$ g. Also $z = \bar{x} - \bar{y} = -14.4$
- Würden Sie die Nullhypothese weiter für möglich halten?

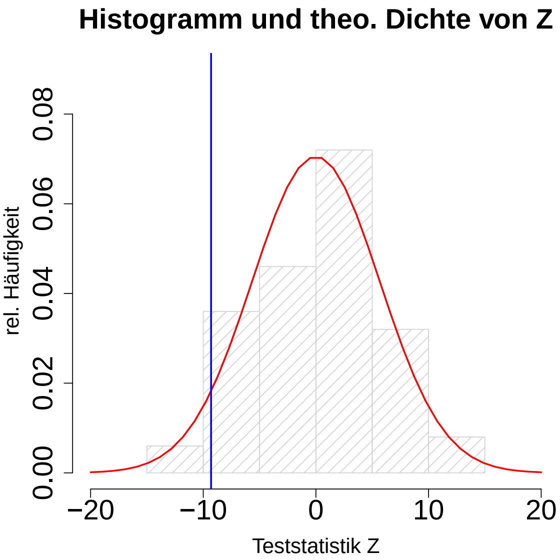
Die folgenden Grafiken zeigen 100-fache Wiederholung der Experiments, (für $\sigma = 40$).
In der zweiten Grafik ist ein Wert $k$ eingezeichnet, der die Eigenschaft besitzt, dass

$$\mathbb{P}(Z < k) = 5\%.$$

# Verteilung der Teststatistik $Z$



# Verteilung der Teststatistik $Z$ und kritischer Wert $k$



# Von der modifizieren Nullhypothese zu Nullhypothese

Die modifizieren Nullhypothese lautet:

$$\tilde{H}_0 : \mu_X = \mu_Y.$$

Die eigentliche Nullhypothese lautet

$$H_0 : \mu_X \geq \mu_Y$$

Es genügt, wenn man sich mit der Verteilung der Teststatistik $T$ (bzw. $Z$) unter der modifizierten Nullhypothese zu beschäftigen, denn wenn die empirische Information genügt, die modifizierte Nullhypothese zu verwerfen, wird jede Hypothese aus der Menge der eigentlichen Nullhypothese erst Recht verworfen.

# Die Verteilung der Teststatistik $T$

- die Verteilung dieser Dichte unter $\tilde{H}_0$ ist berechenbar und damit bekannt.
- Sie hängt nicht von den Modellparametern $\mu_X$, $\mu_Y$ oder $\sigma$ ab.
- Sie hängt ab von $n_1$ und $n_2$.
- Sie heißt t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden.
- Auch hier lässt sich ein Wert $k$ berechnen, so dass

$$\mathbb{P}(T < k) = 5\%.$$

- Dieser Wert hat die Eigenschaft, dass unter $H_0$ die Wahrscheinlichkeit für einen Fehler 1. Art höchstens 5% beträgt.

# Ungerichtet Hypothesen (vgl. LMLG S. 142,143)

Diese Teststatistik $T$ ist die Teststatistik des doppelten t-Test. Diesen Test kann man für gerichtete und ungerichtet Hypothesen formulieren.

**Definition: Ungerichtete Hypothese über zwei Mittelwerte $\mu_X$ und $\mu_Y$**

Bei einer ungerichteten oder zweiseitigen Hypothese lautet die Forschungs- oder Alternativhypothese:

$$H_1 : \mu_X < \mu_Y.$$

Die Nullhypothese lautet dementsprechend:

$$H_0 : \mu_X \geq \mu_Y.$$

# Gerichtete Hypothesen (vgl. LMLG S. 142,143)

Diese Teststatistik $T$ ist die Teststatistik des doppelten t-Test. Diesen Test kann man für gerichtete und ungerichtet Hypothesen formulieren.

**Definition: Gerichtete Hypothese über zwei Mittelwerte $\mu_X$ und $\mu_Y$**

Bei einer gerichteten oder einseitigen Forschungshypothese behautet entweder:

$$H_1 : \mu_X > \mu_Y.$$

mit der Nullhypothese:

$$H_0 : \mu_X \leq \mu_Y.$$

oder

$$H_1 : \mu_X < \mu_Y.$$

mit der Nullhypothese:

$$H_0 : \mu_X \geq \mu_Y.$$

# Der Zwei-Stichproben t-Test I (vgl. LMLG S. 144)

**Definition: Teststatistik des Zwei-Stichproben t-Test**

Weist das untersuchte Merkmal in beiden Gruppen die gleiche Varianz auf (und zwar die der Grundgesamtheit), so folgt die Statistik

$$T := \frac{\bar{X} - \bar{Y}}{S \cdot \left(\frac{1}{n} + \frac{1}{m}\right)},$$

einer t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden.

# Der Zwei-Stichproben t-Test II

1) Man überlegt ob die Annahmen des Zwei-Stichproben t-Tests zu den Daten passen und formuliert die Forschungshypothese und Nullhypothese.
2) Man legt das Signifikanzniveau (=Irrtumswahrscheinlichkeit) fest.
3) Durch die Festlegung der Irrtumswahrscheinlichkeit (=Signifikanzniveau) und der des Auswahl der Teststatistik, ergibt sich ein kritischer Wert.
4) Die Statistik-Software berechnet diesen Wert und informiert uns darüber ob wir die Nullhypothese beibehalten, oder ob sich die Forschungshypothese nachweisen lässt.

# Die vier Schritte im Signifikanztest

## Vorgehen im Signifikanztest

Die Daten liegen entweder schon vor oder werden nach dem 2. Schritt erhoben. Wichtig ist, dass er erste Schritt erfolgt, bevor die Daten erhoben wurden.

1) Man stellt seine Hypothesen (Null- und Alternativhypothese) auf und legt das Signifikanzniveau $\alpha$ fest
2) Man wählt einen geeigneten Test aus.
3) Man berechnet anhand der Daten die die Teststatistik
4) Man prüft anhand des Wertes dieser Teststatistik, ob man die Nullhypothese ablehnt oder beibehält.

# Der Ablehnbereich (vgl. LMLG S. 144)

- Am Beispiel der Gulaschausgabe hatten wir gesehen: Das Ergebnis eines statistischen Tests gilt dann als unwahrscheinlich, wenn der aus der Stichprobe berechnete Wert der Teststatistik einen kritischen Wert $k$ überschreitet.
- Wenn wir in jeder Personengruppe 10 Personen haben, beträgt dieser Wert $k = -1.73$. Wir rechnen also den Wert der Teststatistik aus und lehnen die Nullhypothese ab, wenn $t < 1.73$.
- Man sagt dann der Wert der Teststatistik liegt im Ablehnbereich. Diese Aussage ist völlig äquivalent damit zu sagen, dass Ergebnis ist signifikant.

# Der p-Wert (vgl. LMLG S. 150)

In der Praxis wird ein Test häufig durch die Bestimmung seiner Überschreitungswahrscheinlichkeit (p-Wert) durchgeführt. Bei Statistik-Software ist es üblich, die Testentscheidung beruhend auf p-Werten zu treffen.

## Der p-Wert

Der p-wert gibt die Wahrscheinlichkeit an, dass die Teststatistik einen (im Lichte der Nullhypothese) so extremen oder noch extremeren Wert annimmt.

# Der p-Wert als Grundlage für die Testentscheidung

- Man kann die Testentscheidung auch beruhend auf dem p-Wert treffen.
- Wichtig ist allerdings, dass der Test nur dann gültig ist, wenn man das Signifikanzniveau $\alpha$ festlegt, bevor man den Test durchführt.
- Wenn man sich am p-Wert orientiert, gilt ein Test als signifikant, wenn der p-Wert kleiner ist als+das Signifikanzniveau.
- Ist das Signifikanzniveau $5\% = 0.05$ und der p-Wert 0.03, so ist das Ergebnis signifikant. Ist der p-Wert hingegen 0.23, so ist das Ergebnis nicht signifikant.

# Ausblick Übung

Wir behandeln:

1) Wir werden für Kreuztabellen am Beispiel des Thüringen-Monitor den $\chi^2$-Test besprechen in `STATA` umsetzen und interpretieren.
2) Ebenfalls am Beispiel des Thüringen-Monitor besprechen wir den (approximative) Binominaltest und rechnen mit `STATA`.
3) Wir werden bei der Frage, ob Mädchen besser lesen als Jungen, den doppelten t-Test anwenden und in `STATA` durchführen.
4) Wir werden auch den einfachen t-Test am Beispiel der PISA-Daten behandeln.