1. Vorlesung Statistik II
Zusammengesetzte Hypothesentests

Zusammenfassung Statistik I

Forschungsfrage

z.B. Wirksamkeit einer neuen Therapieform im Vergleich zur bisherigen Standardtherapie

Interessierende deskriptivstatistische Größen in der Population

$\bar{x}_{NT} - \bar{x}_{ST}$

Übersetzung in Parameter eines statistischen Modells

$\mu_{NT} - \mu_{ST}$

Schätzen

Konfidenzintervall für $\mu_{NT} - \mu_{ST}$

Testen

Hypothesentest für $H_1\colon \mu_{NT} - \mu_{ST} < 0$

Ziel von Statistik II:

Möglichkeit, noch mehr Arten von Forschungsfragen untersuchen zu können

- Schwierigkeiten bei komplexeren Fragestellungen (zusammengesetzte Hypothesen, Metaanalyse)
- Einführung weiterer statistischer Modellklassen (Varianzanalysen und Regressionsanalysen)

Plan für die ersten Vorlesungen

- Thema der ersten Vorlesungen: Interpretation von **Gruppen von Hypothesentests**:
  - Zusammengesetzte Hypothesentests
  - Mehrere nicht zusammengesetzte Hypothesentests aus einer oder mehreren Stichproben mit unterschiedlichen Hypothesen
  - Mehrere Hypothesentests aus verschiedenen Stichproben mit identischen Hypothesen

Plan heute

- Thema der ersten Vorlesungen: Interpretation von **Gruppen von Hypothesentests**:

    - **Zusammengesetzte Hypothesentests**

    - Mehrere nicht zusammengesetzte Hypothesentests aus einer oder mehreren Stichproben mit unterschiedlichen Hypothesen

    - Mehrere Hypothesentests aus verschiedenen Stichproben mit identischen Hypothesen

Zusammengesetzte Hypothesen I

- Ausgangslage: Wir haben eine Gruppe von N Hypothesentests:
  Test 1, Test 2, …, Test j, …, Test N.

- Mithilfe jedes einzelnen Hypothesentests j, können wir jeweils die Hypothesen H_{0j} und H_{1j} bei einem Signifikanzniveau von α überprüfen. Hierbei ist H_{0j} jeweils die logische Negation von H_{1j} (also das „Gegenteil“).

- Aber: Wir interessieren uns nicht für die einzelnen Hypothesen H_{0j} und H_{1j}, sondern für Hypothesen, die aus diesen **zusammengesetzt** sind.

Zusammengesetze Hypothesen II

- "Zusammengesetzt" bedeutet, dass wir uns für eine allgemeine Alternativhypothese (und durch Negation eine entsprechende Nullhypothese) der Form

$$H_0: H_{01}\ und\ H_{02}\ und\ \ldots\ und\ H_{0N}$$
$$H_1: H_{11}\ oder\ H_{12}\ oder\ \ldots\ oder\ H_{1N}$$

Mit dem Fall beschäftigen wir uns auf Folie 11 - 34

oder der Form

$$H_0: H_{01}\ oder\ H_{02}\ oder\ \ldots\ oder\ H_{0N}$$
$$H_1: H_{11}\ und\ H_{12}\ und\ \ldots\ und\ H_{1N}$$

interessieren.

Damit beschäftigen wir uns dann ab Folie 35

- Bemerkung: Das „$oder$" (in der zusammengesetzten Alternativhypothese im ersten Fall, bzw. der zusammengesetzten Nullhypothese im zweiten Fall) ist hier das logische oder, bedeutet also **nicht** „entweder oder".

Zusammengesetze Alternativhypothese mit „oder“ I

Beispiel 1:

- Wir interessieren uns für die erwartete Depressionsschwere nach einer Psychotherapie und gehen davon aus, dass diese zumindest in einer von zwei Therapieformen – Kognitive Verhaltenstherapie (KVT) und Tiefenpsychologische Therapie (TPT) - geringer als in der Kontrollgruppe (K) ist.

- Zusammengesetzte Nullhypothese: Keine der beiden Therapien wirkt.

- Zusammengesetzte Alternativhypothese: Mindestens eine der beiden Therapien wirkt.

Zusammengesetze Alternativhypothese mit „oder“ II

- Zusammengesetzte Hypothesen:

$$H_0: \mu_{KVT} \geq \mu_K \; und \; \mu_{TPT} \geq \mu_K$$
$$H_1: \mu_{KVT} < \mu_K \; oder \; \mu_{TPT} < \mu_K$$

- Einzelne Hypothesen:

$$H_{01}: \mu_{KVT} \geq \mu_K$$
$$H_{11}: \mu_{KVT} < \mu_K$$

$$H_{02}: \mu_{TPT} \geq \mu_K$$
$$H_{12}: \mu_{TPT} < \mu_K$$

- Hypothesentests für die einzelne Hypothesen: Zwei gerichtete t-Tests (für unabhängige Stichproben).

Hypothesentests für zusammengesetzte Hypothesen

- Wie können wir die zusammengesetzten Hypothesen bei einem Signifikanzniveau von α^* überprüfen?

- Zwei Möglichkeiten:

  - Möglichkeit 1: Es gibt einen sogenannten „Omnibustest“, der unsere zusammengesetzten Hypothesen durch einen einzigen Hypothesentest direkt überprüft.
    -> Vorlesungen zu den varianz- und regressionsanalytischen Modellen

  - Möglichkeit 2: Wir verwenden die Hypothesentests für die einzelnen Hypothesen, um die zusammengesetzten Hypothesen zu überprüfen.

- Falls es einen „Omnibustest“ zur Überprüfung der zusammengesetzten Hypothesen gibt, sollte dieser verwendet werden, da er über eine höhere Power verfügt.
- Wir beschäftigen uns in dieser Vorlesung mit dem Fall, in dem es **keinen** solchen Test gibt und wir auf Möglichkeit 2 zurückgreifen müssen (wie in unserem Beispiel 1).

Vorgehen bei einer mit ***oder*** zusammengesetzten Alternativhypothese

Naheliegende Idee:

- Wir entscheiden uns für die zusammengesetzte Nullhypothese

$$H_0: H_{01}\ und\ H_{02}\ und\ \ldots und\ H_{0N}$$

falls wir uns bei **allen** N einzelnen Hypothesentests für die H_{0j} entscheiden.

- Wir entscheiden uns für die zusammengesetzte Alternativhypothese

$$H_1: H_{11}\ oder\ H_{12}\ oder\ \ldots oder\ H_{1N}$$

falls wir uns bei **mindestens einem** der N einzelnen Hypothesentests für die H_{1j} entscheiden.

Vorgehen bei einer mit ***oder*** zusammengesetzten Alternativhypothese

In unserem Beispiel 1:

- Wir entscheiden uns für die zusammengesetzte Nullhypothese

$$H_0: \mu_{KVT} \geq \mu_K \ und \ \mu_{TPT} \geq \mu_K$$

falls wir uns bei **beiden** einzelnen t-Tests für die Nullhypothese entscheiden.

- Wir entscheiden uns für die zusammengesetzte Alternativhypothese

$$H_1: \mu_{KVT} < \mu_K \ oder \ \mu_{TPT} < \mu_K$$

falls wir uns bei **mindestens einem** der einzelnen t-Tests für die Alternativhypothese entscheiden.

Signifikanzniveau α^* des zusammengesetzten Hypothesentests

- Was ist die Wahrscheinlichkeit für einen **Fehler 1. Art** (d.h. das Signifikanzniveau α^* eines so konstruierten Hypothesentests), also die Wahrscheinlichkeit dafür, dass wir uns für die **zusammengesetzte** Alternativhypothese (mit *„oder“*) entscheiden, obwohl eigentlich die **zusammengesetzte** Nullhypothese (mit „*und*“) gilt?

- Anders ausgedrückt: Was ist bei einem so konstruierten Hypothesentest die Wahrscheinlichkeit dafür, dass wir uns für die zusammengesetzte Alternativhypothese

$$H_1\colon H_{11}\ oder\ H_{12}\ oder\ \ldots oder\ H_{1N}$$

entscheiden, obwohl eigentlich die zusammengesetzte Nullhypothese

$$H_0\colon H_{01}\ und\ H_{02}\ und\ \ldots und\ H_{0N}$$

gilt?

Signifikanzniveau α^* des zusammengesetzten Hypothesentests

Vorüberlegungen zur Bestimmung von α^*:

- Das Signifikanzniveau der **einzelnen** Hypothesentests sei gleich $\boldsymbol{\alpha}$.
- Wir betrachten den (unrealistischen, aber einfachen) Fall, in dem alle einzelnen Hypothesentests unabhängig voneinander sind. Dies bedeutet, dass ihre Teststatistiken unabhängig voneinander sind.
- Wir definieren für die folgenden Überlegungen für jeden Hypothesentest $j = 1, 2, \ldots, N$ eine Bernoulli-Variable T_j, die den Wert 1 annimmt, falls wir uns beim Hypothesentest j für die H_{1j} entscheiden und den Wert 0, falls wir uns beim Hypothesentest j für die H_{0j} entscheiden. Da alle Hypothesentests unabhängig sind, sind auch diese Zufallsvariablen unabhängig.

- Bemerkung: Das Signifikanzniveau $\boldsymbol{\alpha}^*$ des **zusammengesetzten** Hypothesentests wird in der Literatur auch **Family-Wise-Error-Rate (FWER)** genannt. Das Signifikanzniveau $\boldsymbol{\alpha}$ der **einzelnen** Hypothesentests wird in diesem Kontext auch **Comparison-Wise-Error-Rate** genannt.

Wahrscheinlichkeit α für einen Fehler 1. Art bei **einzelnen** Hypothesentests

	Test 1		
Es gilt entweder	H_{01}	oder	H_{11}
		α	$1-\beta$
Testentscheidung für	H_{01}	oder	H_{11}
Hilfsvariable T_1	0	oder	1

Wahrscheinlichkeit α für Fehler 1. Art bei **mehreren einzelnen** Hypothesentests

	Test 1			Test 2			…	Test N		
Es gilt	H_{01}	oder	H_{11}	H_{02}	oder	H_{12}		H_{0N}	oder	H_{1N}
		α	$1-\beta$		α	$1-\beta$			α	$1-\beta$
Entscheidung	H_{01}	oder	H_{11}	H_{02}	oder	H_{12}		H_{0N}	oder	H_{1N}
Hilfsvariable T_j	0	oder	1	0	oder	1		0	oder	1

Wahrscheinlichkeit α für Fehler 1. Art bei **mehreren einzelnen** Hypothesentests

Wenn die H_0 gilt, gelten in diesem Beispiel **alle** H_{0j}

	Test 1			Test 2			…	Test N		
Es gilt	$\boxed{H_{01}}$	oder	H_{11}	$\boxed{H_{02}}$	oder	H_{12}		$\boxed{H_{0N}}$	oder	H_{1N}
		α			α				α	
Entscheidung	H_{01}	oder	H_{11}	H_{02}	oder	H_{12}		H_{0N}	oder	H_{1N}
Hilfsvariable T_j	0	oder	1	0	oder	1		0	oder	1

Wann würden wir uns trotzdem für die **zusammengesetzte** H_1 entscheiden, also dafür einen Fehler 1. Art begehen?

Wahrscheinlichkeit α^* für einen Fehler 1. Art bei **mehreren einzelnen** Hypothesentests

	Test 1			Test 2			…	Test N		
Es gilt	$\boxed{H_{01}}$	oder	H_{11}	$\boxed{H_{02}}$	oder	H_{12}		$\boxed{H_{0N}}$	oder	H_{1N}
		α			α				α	
Entscheidung	H_{01}	oder	H_{11}	H_{02}	oder	H_{12}		H_{0N}	oder	H_{1N}
Hilfsvariable T_j	0	oder	1	0	oder	1		0	oder	1

Entscheidung für H_1 falls

- Entscheidung für H_{11} oder H_{12} oder H_{13} , wenn die Summe der T_j also $\sum_{j=1}^{N} T_j = 1$ ist.
- Entscheidung für H_{11} **und** H_{12}, oder H_{11} **und** H_{13}, oder H_{12} **und** H_{13}, wenn die Summe der T_j also $\sum_{j=1}^{N} T_j = 2$ ist.
- Entscheidung für H_{11} **und** H_{12} **und** H_{13}, oder …, wenn die Summe der T_j also $\sum_{j=1}^{N} T_j = 3$ ist.
- …
- Entscheidung für **alle** H_{1j}, wenn die Summe der T_j also $\sum_{j=1}^{N} T_j = N$ ist.

Berechnung des Signifikanzniveaus α^* des zusammengesetzten Hypothesentests

- Unter der zusammengesetzten H_0 sind alle H_{0j} wahr und die Wahrscheinlichkeit dafür, dass wir uns im Hypothesentest j für die H_{1j} entscheiden ist für alle j gleich α, da α das Signifikanzniveau der einzelnen Tests ist.

- Das heißt: Die Wahrscheinlichkeit, dass die Bernoulli-Variable T_j den Wert 1 annimmt, ist unter der zusammengesetzten H_0 für alle j gleich α.

- Unter der zusammengesetzten H_0 gilt also:

$$T_j \overset{iid}{\sim} Be(\alpha)$$

- Bemerkung: Hier gilt $\pi = \alpha$, das heißt der Parameter π der Bernoulli-Verteilung entspricht dem Wert α.

Berechnung des Signifikanzniveaus α^* des zusammengesetzten Hypothesentests

- Wir entscheiden uns genau dann für die zusammengesetzte H_1, falls wir uns bei **mindestens einem** Hypothesentest j für die H_{1j} entscheiden, also wenn sich T_j für mindestens ein j in dem Wert 1 realisiert.

- Dies ist genau dann der Fall, wenn sich die Summe S der T_j

$$S = \sum_{j=1}^{N} T_j$$

in einem Wert größer als 0 realisiert.

Berechnung des Signifikanzniveaus α^* des zusammengesetzten Hypothesentests

- Das Signifikanzniveau α^* des zusammengesetzten Hypothesentests ist also

$$\alpha^* = P(S > 0) = 1 - P(S = 0)$$

wobei P die Wahrscheinlichkeit unter der zusammengesetzten H_0 ist.

- Da die T_j unabhängig sind und unter der zusammengesetzten H_0 jeweils einer Bernoulli-Verteilung mit Parameter α folgen, ist die Summe S der T_j unter der zusammengesetzten H_0 binomialverteilt mit Parametern $n = N$ und $\pi = \alpha$.

Berechnung des Signifikanzniveaus α^* des zusammengesetzten Hypothesentests

- Wir können also mithilfe der Wahrscheinlichkeitsfunktion

$$f(s) = \binom{N}{s} \alpha^s (1-\alpha)^{N-s}$$

der Binomialverteilung mit Parametern N und α das Signifikanzniveau α^* des zusammengesetzten Hypothesentests berechnen:

$$\alpha^* = 1 - P(S=0) = 1 - f(s) = 1 - \binom{N}{0} \alpha^0 (1-\alpha)^N = 1 - (1-\alpha)^N$$

- Also:
  - α^* ist für $N > 1$ nicht gleich α
  - Je höher N, also je mehr einzelne Tests, desto höher α^*

Signifikanzniveau des zusammengesetzten Hypothesentests

- Beispiel für $\alpha = 0.005$ und verschiedene N:

N	α^*
2	0.010
10	0.049
30	0.140
50	0.222

- Intuition: Je mehr Hypothesentests man durchführt, desto wahrscheinlicher wird es, auch wenn alle einzelnen Nullhypothesen wahr sind, dass man irgendwann einen Fehler 1. Art macht.

$\boldsymbol{\alpha}$-Fehler-Kumulierung

- Das Signifikanzniveau α^* des zusammengesetzten Hypothesentests ist also höher als das Signifikanzniveau α der einzelnen Tests, falls diese unabhängig sind.

- Man kann zeigen, dass dies auch bei abhängigen Hypothesentests in der Regel so ist (außer in einigen Spezialfällen).

- Dieses Problem wird **$\boldsymbol{\alpha}$-Fehler-Kumulierung** genannt.

- Die Frage ist also: Wie können wir das Signifikanzniveau α^* des zusammengesetzten Hypothesentests kontrollieren, d.h. auf einen bestimmten Wert von z.B. $\alpha^* = 0.005$ festlegen?

Sidak Methode I

- Für unabhängige Hypothesentests ist dies sehr einfach: Wir wählen das Signifikanzniveau α der einzelnen Hypothesentests einfach so, dass für α^* der gewünschte Wert herauskommt.

- Diese Methode wird **Sidak-Korrektur** genannt.

- Auflösen der Formel von Folie 22 nach α ergibt:

$$\alpha^* = 1 - (1 - \alpha)^N$$

$$(1 - \alpha)^N = 1 - \alpha^*$$

$$(1 - \alpha) = \sqrt[N]{1 - \alpha^*}$$

$$\alpha = 1 - \sqrt[N]{1 - \alpha^*}$$

Sidak Methode II

- Beispiel: Wir haben $N = 5$ unabhängige einzelne Hypothesentests und wollen das Signifikanzniveau des zusammengesetzten Hypothesentests auf $\alpha^* = 0.005$ festlegen.

- Wir müssen also als Signifikanzniveau für die einzelnen Hypothesentests

$$\alpha = 1 - \sqrt[N]{1 - \alpha^*} = 1 - \sqrt[5]{1 - 0.005} \approx 0.001$$

wählen.

Sidak Methode III

- Problem: Die einzelnen Hypothesentests sind in den wenigsten Fällen unabhängig. Wir können nur dann sicherstellen, dass sie unabhängig sind, falls wir sie alle in unterschiedlichen Stichproben durchführen. Das wäre jedoch sehr ineffizient.

- Falls wir von einer Abhängigkeit der Hypothesentests ausgehen müssen, gibt es zwei Möglichkeiten:

  - Wir wählen ein allgemeines Verfahren, das α^* auch bei Abhängigkeit der einzelnen Hypothesentests kontrolliert.

  - Wir wählen ein modellspezifisches Verfahren, dass die genaue Form der Abhängigkeiten der Hypothesentests berücksichtigt.

Bonferroni Methode I

- Ein allgemeines Verfahren zur Kontrolle von α^* bei abhängigen einzelnen Hypothesentests ist die sogenannte **Bonferroni-Korrektur**.
- Hierbei gibt man ein α' vor und wählt dann ein daraus berechnetes Signifikanzniveau α für die einzelnen Tests:

$$\alpha = \frac{\alpha'}{N}$$

- Diese Korrektur ist nicht in allen Fällen exakt, aber es lässt sich mathematisch zeigen, dass das Signifikanzniveau α^* des zusammengesetzten Hypothesentests, das wir eigentlich kontrollieren wollen, immerhin kleiner oder gleich α' ist:

$$\alpha^* \leq \alpha'$$

- Falls wir also z.B. $\alpha' = 0.005$ wählen und das Signifikanzniveau α der einzelnen Tests nach der Formel oben berechnen, ist das Signifikanzniveau α^* des zusammengesetzten Hypothesentests zwar nicht exakt gleich 0.005, aber es ist auf keinen Fall höher.

Bonferroni Methode II

- Falls man für jeden einzelnen Hypothesentest j einen p-Wert p_j berechnet hat, können wir wegen

$$p_j \leq \frac{\alpha'}{N} \Leftrightarrow N \cdot p_j \leq \alpha'$$

auch alternativ alle einzelnen p-Werte mit N multiplizieren und dann die so korrigierten p-Werte direkt mit α' vergleichen.

- Die multiplizierten p-Werte werden auch **Bonferroni-korrigierte p-Werte** genannt.

Bonferroni Methode III

- In unserem Beispiel 1 haben wir $N = 2$ gerichtete t-Tests. Wir wollen das Signifikanzniveau des zusammengesetzten Hypothesentests auf $\alpha^* = 0.005$ begrenzen.

- Die Bonferroni-Korrektur ergibt als Signifikanzniveau für die einzelnen t-Tests

$$\alpha = \frac{\alpha'}{N} = \frac{0.005}{2} = 0.0025$$

- Falls wir beispielsweise beim ersten t-Test einen p-Wert von $p_1 = 0.003$ und beim zweiten t-Test einen p-Wert von $p_2 = 0.004$ erhalten hätten, würden wir uns
  - beim ersten t-Test für die Nullhypothese entscheiden, da $p_1 > 0.0025$ ist und
  - beim zweiten t-Test für die Nullhypothese entscheiden, da $p_2 > 0.0025$ ist

- Wir würden uns also für in diesem Fall für die zusammengesetzte Nullhypothese entscheiden, da alle einzelnen Nullhypothesen angenommen wurden.

Bonferroni Methode IV

- Alternativ können wir auch die p-Werte korrigieren und mit $\alpha' = 0.005$ vergleichen:

$$p_{1\_Bonferroni} = N \cdot p_1 = 2 \cdot 0.003 = 0.006$$

$$p_{2\_Bonferroni} = N \cdot p_2 = 2 \cdot 0.004 = 0.008$$

- Wir kommen zum gleichen Ergebnis, da wir uns
  - beim ersten t-Test wegen $p_{1\_Bonferroni} > 0.005$ für die Nullhypothese entscheiden und
  - beim zweiten t-Test wegen $p_{2\_Bonferroni} > 0.005$ ebenfalls für die Nullhypothese entscheiden.

Bonferroni Methode V

- Berechnung der korrigierten p-Werte in R:

```
> p_werte <- c(0.003, 0.004)
> p.adjust(p_werte, method = 'bonferroni‘)
[1] 0.006 0.008
```

Ausblick: Tukey Methode

- Falls sich die einzelnen Hypothesen, aus denen die zusammengesetzten Hypothesen zusammengesetzt sind, alle auf Parameter des gleichen varianz- oder regressionsanalytischen Modells beziehen, kann die genaue Form der Abhängigkeiten der einzelnen Hypothesentests hergeleitet und bei der Korrektur der p-Werte berücksichtigt werden.

- Die hierfür verwendete Methode wird **Tukey-Korrektur** genannt.

- Sie führt zu einer höheren Power des zusammengesetzten Hypothesentests im Vergleich zur Bonferroni-Methode.

- Berechnung der Tukey-korrigierten p-Werte in R mithilfe des *multcomp* Pakets:
  -> Siehe Vorlesung und Übungsblatt 5 zur Varianzanalyse.

Fazit: Zusammengesetze Alternativhypothese mit „oder“

- Zusammenfassung der verschiedenen Korrekturmethoden:
  - Sidak-Methode:
    - Nicht zu empfehlen, da sie Unabhängigkeit der Hypothesentests voraussetzt.
  - Bonferroni-Methode:
    - Setzt keine Unabhängigkeit der einzelnen Hypothesen voraus.
    - Modellunabhängig.
  - Tukey-Methode:
    - Setzt keine Unabhängigkeit der einzelnen Hypothesen voraus.
    - Nur in varianzanalytischen Modellen und Regressionsmodellen möglich. Hier aber der Bonferroni-Methode vorzuziehen, da sie eine höhere Power aufweist.

Zusammengesetze Alternativhypothese mit „und“ I

- Bisher haben wir zusammengesetzte Hypothesentests mit Hypothesen der Form

$$H_0: H_{01}\ und\ H_{02}\ und\ \ldots und\ H_{0N}$$
$$H_1: H_{11}\ oder\ H_{12}\ oder\ \ldots oder\ H_{1N}$$

betrachtet.

- Jetzt betrachten wir den umgekehrten Fall:

$$H_0: H_{01}\ oder\ H_{02}\ oder\ \ldots oder\ H_{0N}$$
$$H_1: H_{11}\ und\ H_{12}\ und\ \ldots und\ H_{1N}$$

- Unterschied: Jetzt sind statt den einzelnen Nullhypothesen die einzelnen Alternativhypothesen mit „und“ verknüpft.

Zusammengesetze Alternativhypothese mit „und“ II

Beispiel 2:

- Wir interessieren uns für die erwartete Depressionsschwere nach einer Psychotherapie und gehen davon aus, dass diese nach einer kognitiven Verhaltenstherapie (KVT) geringer ist als nach einer Tiefenpsychologischen Therapie (TPT) **und** nach einer solchen wiederum geringer als in der Kontrollgruppe (K) ist.

- Zusammengesetzte Alternativhypothese: KVT wirkt besser als TPT **und** TPT wirkt besser als gar keine Therapie

- Zusammengesetzte Nullhypothese: KVT wirkt nicht besser als TPT **oder** TPT wirkt nicht besser als gar keine Therapie.

Zusammengesetze Alternativhypothese mit „und“ III

- Zusammengesetzte Hypothesen:

$$H_0: \mu_{KVT} \geq \mu_{TPT} \; oder \; \mu_{TPT} \geq \mu_K$$
$$H_1: \mu_{KVT} < \mu_{TPT} \; und \; \mu_{TPT} < \mu_K$$

- Einzelne Hypothesen:

$$H_{01}: \mu_{KVT} \geq \mu_{TPT}$$
$$H_{11}: \mu_{KVT} < \mu_{TPT}$$

$$H_{02}: \mu_{TPT} \geq \mu_K$$
$$H_{12}: \mu_{TPT} < \mu_K$$

- Hypothesentests für die einzelnen Hypothesen: Zwei gerichtete t-Tests (für unabhängige Stichproben).

Hypothesentests für zusammengesetzte Hypothesen

- Wie können wir diese zusammengesetzten Hypothesen bei einem Signifikanzniveau von α^* überprüfen?

- Wie im ersten Fall bei zusammengesetzten Alternativhypothesen mit „*oder*“:
  Wir verwenden die Hypothesentests für die einzelnen Hypothesen, um die zusammengesetzten Hypothesen zu überprüfen.

Mögliches Vorgehen

Idee wie bei Alternativhypothesen mit „*oder*“, nur umgekehrt:

- Wir entscheiden uns für die zusammengesetzte Nullhypothese

$$H_0 \colon H_{01}\ oder\ H_{02}\ oder\ \ldots oder\ H_{0N}$$

falls wir uns bei **mindestens einem** der N einzelnen Hypothesentests für die H_{0j} entscheiden.

- Wir entscheiden uns für die zusammengesetzte Alternativhypothese

$$H_1 \colon H_{11}\ und\ H_{12}\ und\ \ldots und\ H_{1N}$$

falls wir uns bei **allen** N einzelnen Hypothesentests für die H_{1j} entscheiden.

Mögliches Vorgehen

In unserem Beispiel 2:

- Wir entscheiden uns für die zusammengesetzte Nullhypothese

$$H_0: \mu_{KVT} \geq \mu_{TPT} \; oder \; \mu_{TPT} \geq \mu_K$$

falls wir uns bei **mindestens einem** der einzelnen t-Tests für die Nullhypothese entscheiden.

- Wir entscheiden uns für die zusammengesetzte Alternativhypothese

$$H_1: \mu_{KVT} < \mu_{TPT} \; und \; \mu_{TPT} < \mu_K$$

falls wir uns bei **beiden** t-Tests für die Alternativhypothese entscheiden.

Signifikanzniveau des zusammengesetzten Hypothesentests I

- Die Frage ist auch hier wieder: Was ist das Signifikanzniveau α^* eines so konstruierten Hypothesentests, d.h. die Wahrscheinlichkeit dafür, dass wir uns für die zusammengesetzte Alternativhypothese entscheiden, obwohl eigentlich die zusammengesetzte Nullhypothese gilt?

- Anders ausgedrückt: Was ist bei einem so konstruierten Hypothesentest die Wahrscheinlichkeit dafür, dass wir uns für die zusammengesetzte Alternativhypothese

$$H_1: H_{11}\ und\ H_{12}\ und\ \ldots und\ H_{1N}$$

entscheiden, obwohl eigentlich die zusammengesetzte Nullhypothese

$$H_0: H_{01}\ oder\ H_{02}\ oder\ \ldots oder\ H_{0N}$$

gilt?

Signifikanzniveau des zusammengesetzten Hypothesentests II

- Wir betrachten für die Überlegung wieder den einfachsten Fall, in dem die einzelnen Hypothesentests voneinander unabhängig sind.

- Die zusammengesetzte Nullhypothese

$$H_0: H_{01}\ oder\ H_{02}\ oder\ \ldots oder\ H_{0N}$$

bedeutet, dass mindestens eine der einzelnen H_{0j} wahr ist.

- Hierfür gibt es sehr viele Möglichkeiten. Wir müssen uns also verschiedene Fälle anschauen.

Signifikanzniveau des zusammengesetzten Hypothesentests III

- Fall 1: Alle einzelnen H_{0j} sind wahr.

- Wir entscheiden uns genau dann fälschlicherweise für die zusammengesetzte H_1, falls wir uns in allen einzelnen Hypothesentests für die H_{1j} entscheiden.

- Falls die einzelnen Hypothesentests unabhängig voneinander sind und alle ein Signifikanzniveau gleich α aufweisen, ist die Wahrscheinlichkeit α^* hierfür unter der Bedingung von Fall 1

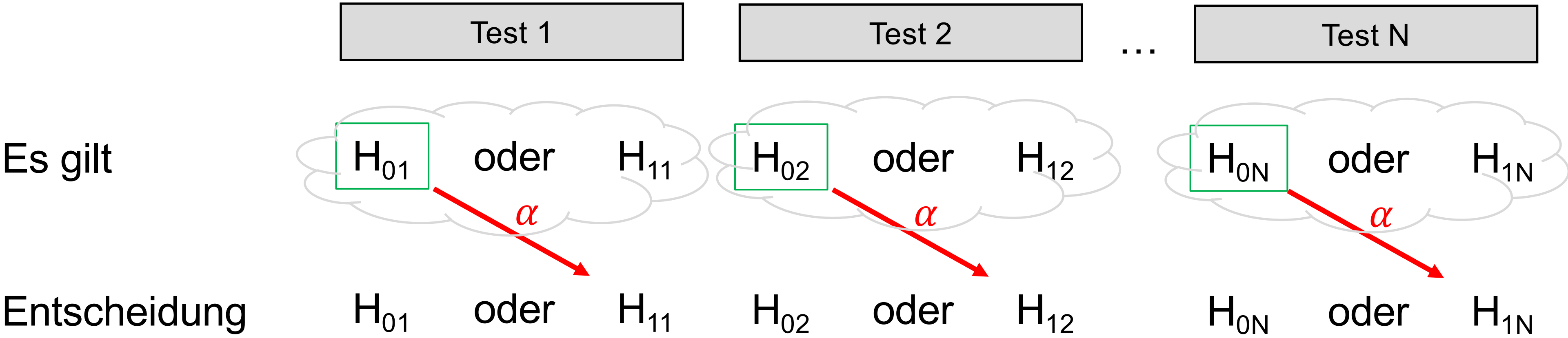
$$\alpha^* = \alpha^N$$

da die Wahrscheinlichkeiten für unabhängige Ereignisse einfach multipliziert werden können.

- Wegen $\alpha < 1$ ist $\alpha^N < \alpha$ und deswegen $\alpha^* < \alpha$.

Wahrscheinlichkeit α für Fehler 1. Art bei **mehreren einzelnen** Hypothesentests

Wenn die H_0 gilt, gelten im ersten Fall dieses Beispiels **alle** H_{0j}



- Wann würden wir uns trotzdem für die **zusammengesetzte** H_1 entscheiden, also dafür einen Fehler 1. Art begehen?
- Nur dann, wenn wir uns in **allen** der **N** einzelnen Hypothesentests fälschlicherweise für die H_{1j} entscheiden.
- Die Wahrscheinlichkeit dafür ist **jeweils** α.
- Die Gesamtwahrscheinlichkeit α^* für diesen Fall (bei unabhängigen Hypothesen) ist

$$\alpha^* = \alpha \cdot \alpha \cdot \ldots \cdot \alpha = \alpha^N$$

Signifikanzniveau des zusammengesetzten Hypothesentests IV

- Fall 2: $k \geq 1$ der einzelnen H_{0j} sind wahr. Damit ist in $N - k$ Fällen die H_{1j} wahr.

- Auch hier entscheiden wir uns genau dann fälschlicherweise für die zusammengesetzte H_1, falls wir uns in allen einzelnen Hypothesentests für die H_{1j} entscheiden.

- In den $\boldsymbol{k}$ Fällen, in denen H_{0j} wahr ist, entscheiden wir uns jeweils mit einer Wahrscheinlichkeit von α für die H_{1j}, da α das Signifikanzniveau der einzelnen Tests ist.

- In den $\boldsymbol{N - k}$ Fällen, in denen die H_{1j} wahr ist, entspricht die Wahrscheinlichkeit dafür, dass wir uns für die H_{1j} entscheiden, der Power $1 - \beta_j$ des einzelnen Tests j.

Signifikanzniveau des zusammen-gesetzten Hypothesentests V

- Falls die einzelnen Hypothesentests unabhängig voneinander sind, ist die Wahrscheinlichkeit α^* dafür, dass wir uns fälschlicherweise für die zusammengesetzte H_1 entscheiden unter der Bedingung von Fall 2

$$\alpha^* = \underbrace{\alpha \cdot \ldots \cdot \alpha}_{k \text{ mal}} \cdot \underbrace{\left(1-\beta_j\right) \cdot \ \ldots \cdot \left(1-\beta_j\right)}_{(N-k) \text{ mal}} = \alpha^k \cdot (1-\beta_j)^{N-k}$$

da die Wahrscheinlichkeiten für unabhängige Ereignisse einfach multipliziert werden können.

- Bemerkung I: Wegen $\alpha < 1$ und $\left(1-\beta_j\right) < 1$ für alle j gilt daher auch in Fall 2: $\alpha^* < \alpha$.
- Bemerkung II: In der Regel ist es völlig unbekannt, wie viele einzelne H_{0j} und H_{1j} tatsächlich richtig sind. Deshalb ist die Berechnung und damit Kontrolle von α^* für eine daraus zusammengesetzte Hypothese in der Praxis nicht möglich.

Fazit: Zusammengesetze Alternativhypothese mit „und“

- Falls die einzelnen Tests unabhängig sind, ist das Signifikanzniveau α^* des zusammengesetzten Hypothesentests in jedem Fall **kleiner** als das Signifikanzniveau α der einzelnen Tests.

- Anders gesagt:
  Die Wahrscheinlichkeit für einen Fehler 1. Art bei der zusammengesetzten Hypothese ist in diesem Fall immer kleiner als die Wahrscheinlichkeit für einen Fehler 1. Art bei einer einzelnen Hypothese.

- Dies gilt auch dann, wenn die einzelnen Hypothesen abhängig sind. Der Beweis hierfür ist allerdings etwas aufwendiger.

- Da in der Regel ein **kleines Signifikanzniveau** gewünscht ist, müssen wir das Signifikanzniveau der einzelnen Hypothesentests bei einer zusammengesetzten Alternativhypothese mit *„und“* also **nicht korrigieren**.

Zusammenfassung: Zusammengesetzte Hypothesentests

- Im Fall von

$$H_0: H_{01} \ und \ H_{02} \ und \ \dots und \ H_{0N}$$
$$H_1: H_{11} \ oder \ H_{12} \ oder \ \dots oder \ H_{1N}$$

ist eine Korrektur notwendig.

- Im Fall von

$$H_0: H_{01} \ oder \ H_{02} \ oder \ \dots oder \ H_{0N}$$
$$H_1: H_{11} \ und \ H_{12} \ und \ \dots und \ H_{1N}$$

ist keine Korrektur notwendig.

Plan für nächste Woche

- Thema der ersten Vorlesungen: Interpretation von **Gruppen von Hypothesentests**:

  ✓ Zusammengesetzte Hypothesentests

  - **Mehrere nicht zusammengesetzte Hypothesentests aus einer oder mehreren Stichproben mit unterschiedlichen Hypothesen**

  - **Mehrere Hypothesentests aus verschiedenen Stichproben mit identischen Hypothesen**