TESTOWANIE POJEDYNCZEGO PARAMETRU W KMNRL

(CZYLI O TEŚCIE t-STUDENTA)

Łukasz Kwiatkowski

Katedra Ekonometrii i Badań Operacyjnych

Plan wykładu

1) Wprowadzenie
2) Procedura testu statystycznego – ogólnie
3) Test *t*-Studenta
    a. Układ hipotez statystycznych
    b. Statystyka testowa
    c. Poziom istotności, wartość krytyczna i zbiór krytyczny
    d. Logika i konkluzja testu
4) Poziom istotności – *revisited*
5) Przykłady
6) *P-value* (pl. prawdopodobieństwo testowe)
7) Przykłady – c.d.

Wprowadzenie

- Przypomnijmy punkt wyjścia naszych rozważań, czyli zapis równania regresji (w domyśle spełniającej wszystkie założenia KMNRL):

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_t$$

UWAGA: W tym materiale, przyjęto zapisywać równanie regresji liniowej w tej właśnie, powyższej postaci (tj. z wyrazem wolnym na początku równania, oznaczonym symbolem β_0, zaś liczba WSZYSTYKICH – łącznie z wyrazem wolnym – parametrów strukturalnych wynosi $k = K + 1$). Możliwe są także alternatywne zapisy, takie jak:

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t \qquad \text{czy} \qquad y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k + \varepsilon_t$$

Przyjęcie takiej czy innej konwencji nie ma jednak większego znaczenia dla poniższych rozważań.

- **Pytanie**: do czego w praktyce można wykorzystać taki model (oczywiście, po jego oszacowaniu)?

  → Dwa główne zastosowania:

  1) Do ***prognozowania*** y_t (przy zadanych wartościach regresorów, czyli "iXsów") – o tym w niezbyt odległej przyszłości :)
  2) Do ***wnioskowania statystycznego*** o prawdziwych (zawsze nieznanych i niepoznawalnych) wartościach parametrów, które odzwierciedlają wpływ poszczególnych zmiennych objaśniających na zmienną objaśnianą

- Jak rozumieć pojęcie "***wnioskowanie statystyczne***"? Mówiąc nieformalnie, ***wnioskowanie statystyczne*** to formułowanie pewnych sądów (wniosków, zdań) o (z gruntu nieznanych) wartościach pewnych parametrów na podstawie ich oszacowań (i pewnych dodatkowych wielkości/obliczeń/procedur). Wnioskowanie to ma character ***statystyczny***, gdyż z uwagi na **losowość** (ang. *randomness*) obecną w modelowanych zjawiskach jest ono zawsze obarczone możliwością popełnienia błędu (z pewnym prawdopodobieństwem).

Wprowadzenie

- Do wnioskowania statystycznego o parametrach **nie wystarczą** same ich oszacowania. Przykładowo, sama ocena punktowa parametru β_1, wynosząca, dajmy na to, $\hat{\beta}_1 = 8$ jeszcze "wiosny nie czyni". Ważna jest przede wszystkim ***niepewność*** związana z tą oceną (czyli tzw. **niepewność estymacji**), mierzona **błędem średnim szacunku,** $d(\hat{\beta}_1)$, który – przypomnijmy – stanowi przeciętną (in.: spodziwaną, oczekiwaną) różnicę pomiędzy prawdziwą (nieznaną, oczywiście) wartością parametru β_1 a jego oceną punktową, $\hat{\beta}_1$. Jeśli dla $\hat{\beta}_1 = 8$ otrzymujemy $d(\hat{\beta}_1) = 12$, to relatywnie (w stosunku do owej oceny) jest to (bardzo) dużo i oznacza, że parametr β_1 *nie* jest szacowany precyzyjnie, a zatem i wnioskowanie o nim będzie takowe. Przy tak wysokiej niepewności estymacji "rozszerza" się spektrum wiarygodnych – jakoś możliwych do zaakceptowania/wyobrażenia sobie – wartości parametru. Przyjmując w dużym uproszczeniu, że zakres wiarygodnych wartości parametru to przedział wyznaczany wg reguły: *ocena parametru ± jego błąd średni szacunku*, wydaje się, że "równie wiarygodną" prawdziwą wartością parametru β_1 może tu być zarówno 8 (czy sama ocena), 4, 0, czy −4, jak i (idąc "w drugą stronę" od oceny z krokiem "co 4"): 12, 16 czy 20.[1] Widzimy zatem, że nawet wartości ujemne mieszczą się w tym zakresie, a po drodze także (dość kluczowa) wartość zera. Oczywiście, zupełnie inny obraz otrzymalibyśmy, gdyby $d(\hat{\beta}_1) = 1{,}2$ (czyli dokładnie o 1 rząd wielkości mniej)...

[1] Reguła: *ocena parametru ± jego błąd średni szacunku* jako "przepis" na wyznaczanie zakresu najbardziej wiarygodnych (w świetle danych i modelu) wartości parametru nie jest do końca poprawna. W praktyce zwykle stosuje się regułę: *ocena parametru ± 2 × błąd średni szacunku*, co, po pierwsze, jeszcze bardziej rozszerza(!) owy zakres, a po drugie – nawiązuje do kontrukcji tzw. przedziałów ufności (czyli takich właśnie zakresów "najbardziej wiarygodnych wartości danego parametru", tyle że skonstruowanych w sposób już formalny, a nie tylko "na chłopski rozum"). O przedziałach ufności – w osobnym odcinku :)

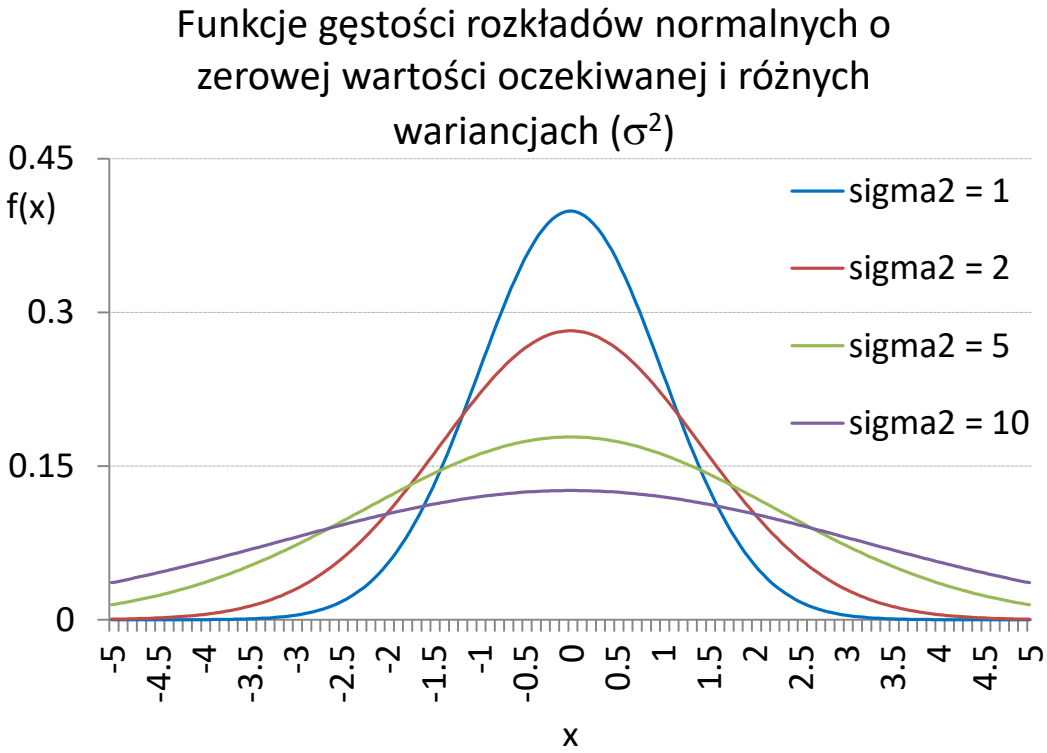
Wprowadzenie

- Dwie główne **techniki wnioskowania statystycznego**:
  - Testowanie (inaczej weryfkacja) hipotez statystycznych – o tym poniżej (w konkretnym kontekście testowania pojedynczego parametru KMNRL)
  - Budowanie przedziałów ufności – o tym przy innej okazji

- **UWAGA**: Do stosowania obydwu wymienionych wyżej technik wnioskowania statystycznego w ramach modelu regresji liniowej KLUCZOWE jest **dodatkowe, szóste założenie** (oprócz układu założeń 1-5 definiujących KMRL), zgodnie z którym przyjmujemy, że składniki losowe ($\varepsilon_t; t = 1,2, \dots, T$) mają **rozkład normalny**, co zapisujemy jako $\varepsilon_t \sim N$. Dołączenie tego założenia do układu założeń KMRL definiuje nam tzw. **Klasyczny Model NORMALNEJ Regresji Liniowej**, w skrócie: KM**N**RL. Zauważmy, że owo szóste założenie doprecyzowuje *typ* rozkładu składników losowych – o którym nic orzeka rozważany dotychczas układ założeń KMRL (w tym ostatnim przyjmujemy tylko pewne założenia odnośnie wartości oczekiwanej i wariancji ε_t, nic zaś nie zakładając o *typie* rozkładu zaburzeń). Zatem KM**N**RL to szczególny przypadek KMRL.

Wprowadzenie

- Przyjmując, że $\varepsilon_t \sim N$ zakładamy, że wartości składników losowych (choć same w sobie nieobserwowalne) "rozkładają" się według rozkładu normalnego – scentrowanego w zerze (zgodnie z założeniem 4. KMRL). Funkcja gęstości takiego rozkładu (obrazująca, które wartości ε_t są bardziej, a które mniej prawdopodobne) jest symetryczna (względem zera) i jednomodalna, z modalną (wierzchołkiem) właśnie w zerze (czyli zarazem w wartości oczekiwanej) – im większa wariancja rozkładu (σ^2), tym większe jego rozproszenie, a funkcja gęstości bardziej "rozpłaszczona":



Wprowadzenie

- Zatem mamy dwa układy założeń: KMRL i KM**N**RL
- Warto w tym miejscu sobie uporządkować kwestię tego, „co możemy, a czego nie możemy zrobić” w ramach „zwykłego” KMRL (bez zakładania normalności rozkładu zakłóceń losowych).
- W KMRL (czyli bez narzucania typu rozkładu zaburzeń losowych) możemy wykonywać wszystko to, czym zajmowaliśmy się dotychczas, a mianowicie:
  - przeprowadzać estymację parametrów
  - interpretować ich oceny
  - wyznaczać błędy średnie szacunku
  - oceniać dobroć dopasowania
- Natomiast w KM**N**RL możemy dodatkowo przeprowadzać **wnioskowanie statystyczne** (za pomocą wymienionych wyżej technik: testów statystycznych i przedziałów ufności)
- Tak więc w KM**N**RL zakładamy więcej o rzeczywistości (więc bardziej „narażamy się” na to, że nasze założenia – w szczególności normalność rozkładu składnika losowego – mogą być niespełnione, więc model może nie do końca odpowiadać realiom modelowanego zjawiska), ALE z drugiej strony, dzięki temu dodatkowemu założeniu „możemy więcej” i to w dodatku „znacznie więcej”, ponieważ bez możliwości wnioskowania statystycznego (niedającego się przeprowadzić w KMRL) nie możemy wnioskować o „prawdziwej naturze” modelowanego wycinka rzeczywistości.
- Ponadto, w KM**N**RL estymator MNK ma jeszcze lepsze własności niż w KMRL: jeśli bowiem założyć, że $\varepsilon_t \sim N$, wówczas estymator MNK jest najlepszy spośród wszystkich (i liniowych, i nieliniowych) nieobciążonych estymatorów wektora parametrów strukturalnych. W KMRL – przypomnijmy – jego „najlepszość” ograniczała się „tylko” do klasy liniowych nieobciążonych estymatorów.

Wprowadzenie

- Po tej dłuższej (ale ważnej) dygresji, wróćmy do wspomnianych dwóch głównych **technik wnioskowania statystycznego**:
  - Testowanie (inaczej weryfkacja) hipotez statystycznych – i tym zajmiemy się teraz
  - Budowanie przedziałów ufności – o tym przy innej okazji

- Testowanie hipotez statystycznych:
  - **Hipoteza statystyczna** (w skrócie: **hipoteza**) = zdanie/stwierdzenie, którego prawdziwość (albo raczej wiarygodność) podlega sprawdzeniu (testowi)
  - Testowaniu zawsze podlega **układ dwóch** wykluczających się wzajemnie **hipotez** dotyczących **parametrów** (a **nie** ich oszacowań); układ ten stanowią dwie hipotezy: tzw. **hipoteza zerowa** (oznaczana symbolem H_0) oraz **hipoteza alternatywna** (H_1). Szkielet zapisu układu ("vs." = łac. *versus*):

$$H_0\text{: ...} \quad \text{vs.} \quad H_1\text{: ...}$$

  - Testowaniu może podlegać **pojedyczy parameter** (np. β_1) lub pewna **ich grupa**; w tym materiale uwagę poświęcimy testowaniu pojedycznego współczynnika regresji (czyli ogólnie β_i, dla wybranego $i \in \{0, 1, \dots, K\}$); testowaniem grup parametrów zajmiemy się w kolejnym odcinku :)
  - Właściwym dla problemu testowania hipotez dot. pojedycznego parametru w KMNRL jest tzw. **test t-Studenta** (w skrócie: **test *t***)

Procedura testu statystycznego – ogólnie

- Na każdy **test statystyczny** (dowolny, nie tylko test *t*-Studenta) składa się **5 elementów**:
  1) Zapis testowanego **układu hipotez**
  2) Dobranie tzw. **statystyki testowej** (właściwej do problemu) – o konkretnym, znanym rozkładzie prawdopodobieństwa przy założeniu, że hipoteza zerowa (H_0) jest prawdziwa (i w domyśle: przy spełnionych założeniach KMNRL); obliczenie wartości tej statystyki
  3) Przyjęcie konkretnej wartości tzw. **poziomu istotności**, oznaczanego symbolem α
  4) Wyznaczenie tzw. **wartości krytycznej** i **zbioru krytycznego**
  5) Sformułowanie **konkluzji** testu

Test t-Studenta: układ hipotez

- Przypomnijmy:
  - zapis modelu: $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_t$
  - zajmujemy się wnioskowaniem o pojedynczym parametrze strukturalnym, ogólnie β_i, dla wybranego $i \in \{0, 1, \ldots, K\}$

- Interesuje nas rozstrzygnięcie, czy wiarygodna (w świetle posiadanych danych i w ramach oszacowanego modelu) jest hipoteza mówiąca, że parametr β_i jest równy pewnej zadanej przez nas, konkretnej, hipotetycznej jego wartości – tę określamy (bardzo intuicyjnym :) mianem **testowanej (in.: hipotetycznej) wartości** parametru β_i, i oznaczamy symbolem β_i^*. Tę równość podlegającą testowaniu ($\beta_i = \beta_i^*$) ujmujemy w postaci **hipotezy zerowej** (H_0), z którą możemy zestawić jedną z trzech postaci **hipotezy alternatywnej** (H_1). Stąd też trzy możliwe – alternatywne – układy hipotez, różniące się kierunkiem nierówności w H_1:

  - Test dwustronny: $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i \neq \beta_i^*$
  - Test lewostronny: $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i < \beta_i^*$
  - Test prawostronny: $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i > \beta_i^*$

Test t-Studenta: układ hipotez

- Test dwustronny: $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i \neq \beta_i^*$
- Test lewostronny: $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i < \beta_i^*$
- Test prawostronny: $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i > \beta_i^*$

➢ !!!Zauważmy!!! W **hipotezie zerowej** zawsze jest/musi być **znak równości** (=). Ponadto, w zapisie hipotez zawsze występują **parametry (bez daszka)**, a nie ich oceny (z daszkiem). Błędny byłby zapis np. $H_0: \hat{\beta}_i = \beta_i^*$ – przecież to, ile wynosi ocena parametru, to akurat wiadomo, bo ją możemy obliczyć, więc nie ma sensu o niej przeprowadzać wnioskowania statystycznego. To o parametrach chcemy wnioskować i to z dwóch powodów: 1) nie znamy (i nie poznamy) ich prawdziwych wartości (a chcemy jednak móc coś o nich powiedzieć = wnioskować), 2) interesują nas, bo mówią o tej prawdziwej naturze modelowanego zjawiska, o tych prawdziwych zależnościach, które tylko częściowo są odzwierciedlone w informacji zawartej w (zawsze przecież ograniczonej) próbie i tym samym w oszacowaniach parametrów.

➢ Co do postaci **hipotezy alternatywnej**, w praktyce **domyślnie** przyjmuje się dwustronny wariant testu. Zdecydowanie się na którychś z wariantów jednostronnych musi być podyktowane jakimiś dodatkowymi przesłankami. Zilustrujemy to za chwilę w **Przykładzie 1**.

Test t-Studenta: układ hipotez

- Co **w praktyce** oznacza zapis widniejący w H_0? Wartość parametru β_i (w H_0 równa/zadana na poziomie β_i^*) ma takie samo **znaczenie interpretacyjne** jak ocena tego parametru, tj. $\hat{\beta}_i$ (patrz zajęcia z interpretacji ocen parametrów w KMNRL). Zatem testując hipotezę $\beta_i = \beta_i^*$ chcemy w istocie rozsądzić, czy wiarygodne jest (z punktu widzenia danych i modelu) to, że "wraz ze wzrostem wartości zmiennej x_{ti} o 1 jednostkę (przy ustalonych wartościach pozostałych zmiennych objaśniających) wartość zmiennej objaśnianej y_t wzrasta (przy $\beta_i^* > 0$) / spada (przy $\beta_i^* < 0$) średnio o $|\beta_i^*|$ jednostek".[2] W grę wchodzą oczywiście wszystkie pozostałe schematy interpretacyjne, o których uczyliśmy się na wcześniejszych zajęciach.

- Wybór testowanej wartości parametru (β_i^*) i wariantu testu (dwustronny lub któryś jednostronny) zależy od **konkretnego problemu**. Zilustrujemy to na następującym przykładzie – patrz kolejna strona.

[2] Moduł użyty tu w celu uniknięcia tej "niezręcznej" sytuacji, gdy $\beta_i^* < 0$ i bez zastosowania modułu musielibyśmy mówić o spadku o np. *minus* 5 jednostek (gdyby $\beta_i^* = -5$). Owszem, ujemny znak informuje o spadku, ale *mówiąc* o spadku zawsze wyrażamy się w kategoriach wartości dodatnich, czyli tu: spadek o $5 = |-5|$ jednostek.

Test t-Studenta: układ hipotez

Przykład 1. Rozważmy model, z którym zetknęliśmy się w materiale „Mierniki dopasowania modelu regresji liniowej do danych" – patrz tamtejszy Przykład 1. Zapiszmy model (nieoszacowany):

$$E_t = \beta_0 + \beta_1 I_t + \beta_2 H_t + \varepsilon_t,$$

gdzie:

- E_t – wielkość miesięcznych wydatków gospodarstwa domowego na transport (w zł/osobę w g.d.)
- I_t – wielkość miesięcznych dochodów gospodarstwa domowego (w setkach zł / osobę w g.d.)
- H_t – typ gospodarstwa domowego: zmienna typu 0-1, przy czym $H_t = 1$ dla gospodarstw pracowniczych, zaś $H_t = 0$ dla gospodarstw emerytów i rencistów

Proponuję małe **ćwiczonko** :) Dla testowania poniższych stwierdzeń (na kolejnej stronie) zapisz stosowny **układ hipotez**. KONIECZNIE wykonaj samodzielnie to ćwiczenie – dopiero po ostatnim punkcie zajrzyj do rozwiązania (na jeszcze następnym slajdzie).

Test t-Studenta: układ hipotez

Przykład 1. c.d.

UWAGA: punkty wyróżnione ~~przekreśleniem~~ NIE obowiązują Państwa (ani tu, ani też w dalszej części materiału, jeśli są do nich robione jakieś odniesienia)

1) Wielkość dochodu g.d. istotnie wpływa na wielkość wydatków g.d. na transport

2) Wielkość wydatków g.d. na transport w gospodarstwach emerytów i rencistów różni się od tej w gospodarstwach pracowników.

3) Wraz ze wzrostem wielkości dochodu g.d. wzrasta także wielkość wydatków g.d. na transport

~~4) W gospodarstwach pracowniczych wielkość wydatków na transport jest wyższa w stosunku do gospodarstw emerytów i rencistów średnio o 10 zł/osobę.~~

~~5) W gospodarstwach pracowniczych wielkość wydatków na transport jest wyższa w stosunku do gospodarstw emerytów i rencistów średnio o mniej niż 35 zł/osobę.~~

6) Wielkość wydatków gospodarstw domowych emerytów i rencistów na transport jest niższa od tej w gospodarstwach pracowniczych.

7) Typ gospodarstwa domowego nie determinuje wielkości ponoszonych przezeń wydatków na transport.

~~8) Wraz ze wzrostem wielkości dochodu g.d. o 100 zł/osobę, wielkość wydatków g.d. na transport wzrasta średnio o 7 zł/osobę.~~

Test t-Studenta: układ hipotez → odpowiedzi do Przykładu 1

1) Wielkość dochodu g.d. istotnie wpływa na wielkość wydatków g.d. na transport:

$H_0: \beta_1 = 0 \quad vs. \quad H_1: \beta_1 \neq 0$ (czyli: nie wpływa *vs.* wpływa; test istotności parametru β_1)

2) Wielkość wydatków g.d. na transport w gospodarstwach emerytów i rencistów różni się od tej w gospodarstwach pracowników:

$H_0: \beta_2 = 0 \quad vs. \quad H_1: \beta_2 \neq 0$ (czyli: nie różni się *vs*. różni się; test istotności parametru β_2)

3) Wraz ze wzrostem wielkości dochodu g.d. wzrasta także wielkość wydatków g.d. na transport:

$H_0: \beta_1 = 0 \quad vs. \quad H_1: \beta_1 > 0$ (czyli wydatki: nie zmieniają się *vs*. wzrastają; test dodatniości β_1)

4) W gospodarstwach pracowniczych wielkość wydatków na transport jest wyższa w stosunku do gospodarstw emerytów i rencistów średnio o 10 zł/osobę:

$H_0: \beta_2 = 10 \quad vs. \quad H_1: \beta_2 \neq 10$ (czyli wydatki: są wyższe o średnio 10 zł/os. *vs*. nieprawda, że H_0)

5) W gospodarstwach pracowniczych wielkość wydatków na transport jest wyższa w stosunku do gospodarstw emerytów i rencistów średnio o mniej niż 35 zł/osobę.

$H_0: \beta_2 = 35 \quad vs. \quad H_1: \beta_2 < 35$ (czyli wydatki są wyższe o średnio: 35 zł/os. *vs.* mniej niż 35 zł/os.)

Test t-Studenta: układ hipotez → odpowiedzi do Przykładu 1 – c.d.

6) Wielkość wydatków gospodarstw domowych emerytów i rencistów na transport jest niższa od tej w gospodarstwach pracowniczych:

$H_0: \beta_2 = 0 \quad vs. \quad H_1: \beta_2 > 0$ (czyli wydatki w obydwu typach g.d.: są jednakowe *vs.* g.d. emer. i rencistów wydają mniej, co jest równoważne – i to jest bezpośrednio zapisane w H_1 – temu, że g.d. pracowników wydatkują więcej; takie podejście jest tu konieczne z uwagi na przyjęte kodowanie wartości zmiennej H_t, która – przypomnijmy – wartość „1" przyjmuje dla gospodarstw pracowników

7) Typ gospodarstwa domowego nie determinuje wielkości ponoszonych przezeń wydatków na transport:

$H_0: \beta_2 = 0 \quad vs. \quad H_1: \beta_2 \neq 0$ (tylko inne brzmienie punktu 2, ale znaczenie identyczne)

8) Wraz ze wzrostem wielkości dochodu g.d. o 100 zł/osobę, wielkość wydatków g.d. na transport wzrasta średnio o 7 zł/osobę:

$H_0: \beta_1 = 7 \quad vs. \quad H_1: \beta_1 \neq 7$ (czyli wydatki: nie zmieniają się *vs*. wzrastają)

UWAGA: Jak pewnie Państwo zauważyli, w niektórych punktach aż się prosi, by „dofiksować" wartości pozostałych zmiennych objaśniających (w sumie jednej: albo wielkość dochodu g.d. albo typ g.d.). Przykładowo, w pkt 8 można by doprecyzować, że chodzi o „gospodarstwo domowe *danego typu*" (= ustalamy jego typ). W praktyce jednak, często ten zabieg – jakkolwiek ważny i nie do pominięcia na etapie formułowania samych *interpretacji* ocen parametrów – jest pomijany w słownym brzmieniu hipotez, które mają zostać poddane testowaniu. Niemniej, w istocie, nawet jeśli „fixing" wartości pozostałych zmiennych objaśniających jawnie nie wybrzmiewa w treści hipotezy, to i tak (domyślnie) obowiązuje.

Test t-Studenta: układ hipotez → test istotności

- Zanim przejdziemy dalej, omówmy pewien szczególny (i kluczowy z punktu widzenia praktyki) przypadek szczególny testu t – taki, w którym przedmiotem testowania jest zerowa wartość parametru:

$$H_0: \beta_i = 0 \quad \text{vs.} \quad H_1: \beta_i \neq 0$$

- Test, w którym sprawdzany jest powyższy układ hipotez nazywamy **testem istotności** (parametru β_i)
- O czym mówią tak postawione hipotezy:

$H_0: \beta_i = 0$	$H_1: \beta_i \neq 0$
Parametr β_i nie jest *statystycznie* różny od zera → Rzadziej używa się określenia: Parametr jest statystycznie równy zero	Parametr β_i jest statystycznie istotnie różny od zera
Parametr β_i jest *statystycznie* nieistotny → Równoważnie możemy powiedzieć: Parametr β_i nie jest *statystycznie* istotny	Parametr β_i jest statystycznie istotny
Pamiętamy, że współczynnik regresji β_i ujmuje wpływ odpowiedniej zmiennej objaśniającej, x_{ti}, na zmienną objaśnianą, y_t. Dlatego w powyższych sformułowaniach wyrażenie „Parametr β_i" można by zastąpić wyrażeniem „Wpływ zmiennej ... na ...", wstawiając w miejsce wielokropka nazwy stosownych zmiennych w modelu.	

Test t-Studenta: statystyka testowa

- Podsumowując wcześniejsze rozważania, w pierwszym etapie testu zapisujemy układ hipotez. Przypomnijmy, że w kontekście rozważań prowadzonych w tym materiale, poświęconych testowaniu pojedynczego parametru regresji, układ ten możemy zapisać na 3 sposoby, w zależności od testowanego stwierdzenia:
  - Test dwustronny: $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i \neq \beta_i^*$
  - Test lewostronny: $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i < \beta_i^*$
  - Test prawostronny: $H_0: \beta_i = \beta_i^*$ vs. $H_1: \beta_i > \beta_i^*$

- W kolejnym etapie sięgamy po odpowiednią **statystykę testową** – czyli pewną *funkcję* wyników estymacji parametrów. W przypadku testu pojedynczego parametru (β_i) – **bez względu na wariant testu** (dwustronny czy jednostronny) – przyjmuje ona postać:

$$t(\beta_i) = \frac{\hat{\beta}_i - \beta_i^*}{d(\hat{\beta}_i)}$$

(czyli od oceny parametru odejmujemy jego testowaną wartość, a całą różnicę dzielimy przez błąd średni szacunku parametru). Statystykę tę określamy mianem **statystyki *t*-Studenta** (w skrócie: **statystyki *t***). W szczególnym przypadku, jakim jest **test istotności** (i tym samym: **nieistotności**) parametru β_i, statystyka *t* przyjmuje postać tzw. **ilorazu *t*** (in.: ilorazu *t*-Studenta, ilorazu Studenta; ang. ***t-ratio***), czyli „ocena przez błąd”:

$$t(\beta_i) = \frac{\hat{\beta}_i}{d(\hat{\beta}_i)}$$

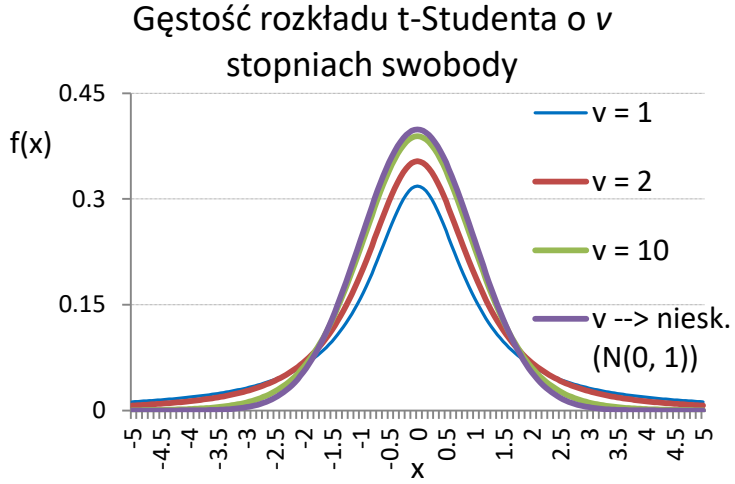
Test t-Studenta: statystyka testowa

- W ogólności (i przy założeniach KMNRL), można pokazać, że **statystyka *t*** – przy prawdziwości hipotezy zerowej – ma **rozkład *t*-Studenta o *T-k* stopniach swobody**, co zapisujemy:

$$t(\beta_i) = \frac{\widehat{\beta}_i - \beta_i^*}{d(\widehat{\beta}_i)} \,|H_0 \sim St(T-k),$$

gdzie – przypomnijmy – T oznacza liczbę obserwacji, zaś k – liczbę wszystkich parametrów strukturalnych regresji (łącznie z wyrazem wolnym, $k = K + 1$).

- Rozkład *t*-Studenta jest podobny do rozkładu normalnego – symetryczny i jednomodalny, tu w dodatku scentrowany w zerze. Ma jednak grubsze **ogony** (ang. *fat tails*), co oznacza, że w tym rozkładzie wartości dalekie na lewo i na prawo od modalnej (która znajduje się w zerze) są bardziej prawdopodobne aniżeli w rozkładzie normalnym. Warto jednak podkreślić, że wraz ze wzrostem **liczby stopni swobody** (wynoszącej $T-k$), która to jest jedynym parametrem tego rozkładu, rozkład *t*-Studenta dąży do standardowego rozkładu normalnego, $N(0,1)$:



Test t-Studenta: statystyka testowa

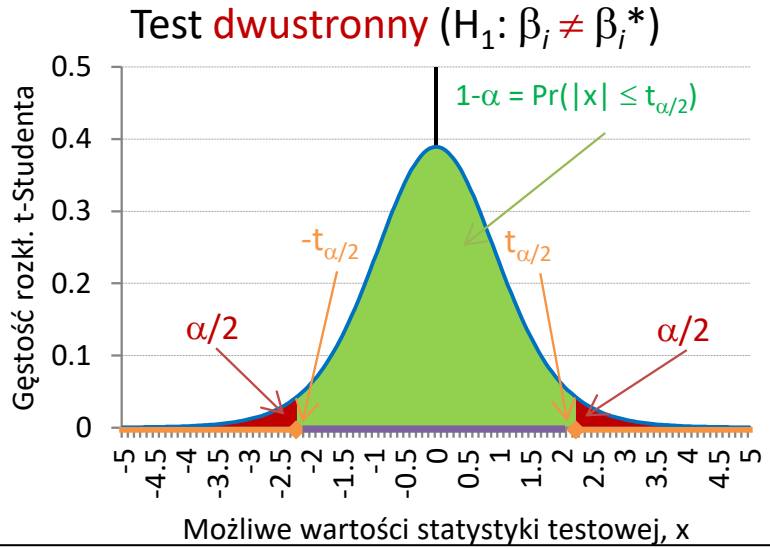
- Co to w ogóle oznacza, że **statystyka *t*** – przy prawdziwości hipotezy zerowej (i w domyśle: przy założeniach KMNRL) – ma **rozkład *t*-Studenta** (o *T-k* stopniach swobody)? Uświadomijmy sobie, że cała „klasyczna" statystyka opiera się na fundamentalnym założeniu o próbkowaniu, zgodnie z którym modelowany zbiór danych traktuje się jako pewną próbę losową z (teoretycznie) nieograniczonej populacji (czyli jak zbudowaliśmy model na przykład dla 10 gospodarstw domowych, to te 10 gospodarstw traktujemy jako próbę losową z teoretycznie nieskończonej całej populacji takowych gospodarstw). Zatem dla konkretnego zbioru danych otrzymujemy konkretne wyniki liczbowe estymacji (oceny parametrów i błędy średnie szacunku). Jeśli jednak „wylosowalibyśmy" inną próbę, to otrzymalibyśmy (mniej lub bardziej) inne wyniki estymacji – oceny i błędy średnie. I gdybyśmy tak losowali w nieskończoność, to moglibyśmy wyznaczyć cały **rozkład** otrzymywanych wartości – w postaci funkcji gęstości, czyli takiego „uciąglonego histogramu". Taki rozkład (jego funkcja gęstości, w skrócie: gęstość) – na przykład ocen parametru β_1 – informowałby nas o tym, których z tych wartości należałoby się spodziewać bardziej (wysokie wartości gęstości), a których rzadziej czy prawie w ogóle (niskie i zdążające do zera wartości gęstości). Oczywiście, „ogonowe" wartości ocen parametru β_1 (czyli te, dla których wartości funkcji gęstości wygasają do zera) to te, które bardzo rzadko występują, i są nietypowe (w danym rozkładzie).

  Podobnie ma się sprawa z rozkładem (gęstością) statystyki testowej – obrazuje on, które wartości tej statystyki są „typowe" (tam gdzie gęstość przyjmuje wysokie wartości) i w związku z tym należałoby się ich spodziewać bardziej, a które „nietypowe" (te w ogonach) – UWAGA – przy założeniu, że hipoteza zerowa jest prawdziwa (i założenia KMNRL są spełnione). Dlaczego przy tych założeniach? Bo tylko przy nich statystyk matematyczny jest w stanie wykazać formalnie (udowodnić) to, że faktycznie „tak, a nie inaczej" wygląda rozkład wartości statystyki testowej.

Test t-Studenta: poziom istotności, wartość krytyczna, zbiór krytyczny

- Zatem w rozkładzie statystyki testowej możemy wyodrębnić dwa **obszary (in.: zbiory)** jej wartości:
  - Zbiór wartości **typowych** (w świetle H_0) – tych wokół modalnej (wysokie wartości gęstości)
  - Zbiór wartości **nietypowych** (w świetle H_0) – znajdujących się w ogonach rozkładu

- Przyjrzyjmy się poniższemu wykresowi – ilustruje on rozkład statystyki testowej t-Studenta, z zaznaczeniem wspomnianych obszarów (odcinki o odpowiednich kolorach na osi OX) w przypadku **testu dwustronnego** (przypadki jednostronne – na kolejnym slajdzie).



Pole pod krzywą gęstości nad danym zakresem wartości stat. test. jest **prawdopodobieństwem** tego, że wartość stat. test. znajduje się w tym zakresie. **Zbiór kryt.** to zbiór $\boldsymbol{\alpha \cdot 100\%}$ **najbardziej nietypowych** w świetle H_0 wartości statystyki test.

Zbiór będący sumą pomarańczowych odcinków nazywamy **zbiorem krytycznym** (in.: obszarem krytycznym) – stanowią go „nietypowe" (w świetle H_0) wartości statystyki testowej, czyli tzw. ogony rozkładu. Przyjmuje on tu postać:

$$Z_k = (-\infty; -t_{\alpha/2}) \cup (t_{\alpha/2}; +\infty)$$

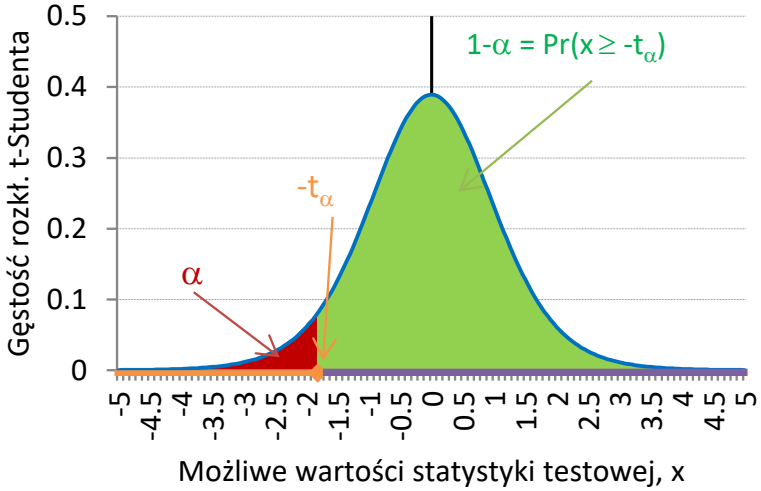
$t_{\alpha/2}$ – **wartość krytyczna** (odczytywana z tablic rozkładu Studenta)

α – tzw. **poziom istotności**, czyli jakieś niskie prawdopodobieństwo owych „nietypowych" wartości statystyki testowej; określa ono „gdzie są ogony", od której wartości (wartości krytycznej) się te ogony zaczynają. UWAGA: poziom istotności zadajemy samodzielnie – to nasza decyzja. Standardowo przyjmuje się którąś z trzech wartości: 0,01, albo 0,05, albo 0,1 (**domyślnie** przyjmuje się $\boldsymbol{\alpha = 0,05}$). Dopiero przy zadanym α znajdujemy $t_{\alpha/2}$.

Test t-Studenta: poziom istotności, wartość krytyczna, zbiór krytyczny

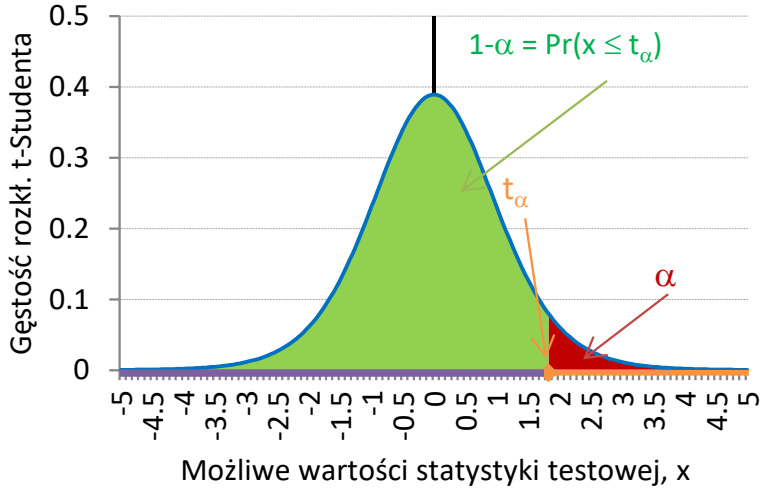
➢ Sytuacja w **jednostronnych** wariantach testu *t*:

Test lewostronny (H_1: $\beta_i < \beta_i^*$)



$Z_k = (-\infty; -t_\alpha)$

Test prawostronny (H_1: $\beta_i > \beta_i^*$)



$Z_k = (t_\alpha; +\infty)$

➢ Uwagi:

- $t_{\alpha/2}$ – dwustronna wartość krytyczna

  $\Rightarrow Z_k = (-\infty; -t_{\alpha/2}) \cup (t_{\alpha/2}; +\infty)$ – dwustronny zbiór krytyczny
- t_α – jednostronna wartość krytyczna:

  $\Rightarrow Z_k = (-\infty; -t_\alpha)$ – lewostronny zbiór krytyczny

  $\Rightarrow Z_k = (t_\alpha; +\infty)$ – prawostronny zbiór krytyczny
- Lewo- i prawostronny zbiór krytyczny są symetryczne względem 0, co wynika z symetrii rozkładu

Test t-Studenta: poziom istotności, wartość krytyczna, zbiór krytyczny

- Jak wspomniano wcześniej, poziom istotności określa sam badacz, przyjmując jedną z typowych jego wartości (0,01; 0,05; 0,1), zaś stosowną wartość krytyczną (jedno- lub dwustronną, przy danym α i liczbie stopni swobody, $T - k$) odczytujemy z tablic rozkładu t-Studenta lub korzystamy z odpowiedniej funkcji w arkuszu Excel:

  =ROZKŁAD.T.ODW(poziom_istotności; stopnie_swobody)

  Powyższa funkcja zwraca zawsze **dwustronną** wartość krytyczną („rozkładając" zadaną wartość poziomu istotności symetrycznie, po $\alpha/2$ w lewym i prawym ogonie rozkładu).

  → **Pytanie „na pomyślunek":** Co trzeba by zrobić z argumentem „poziom_istotności" w tej funkcji, aby – dla danego poziomu istotności α – zwróciła ona jednak wartość jednostronną? Każdy niech spróbuje pomyśleć :)

- Mając wartość krytyczną (w postaci konkretnej liczby), możemy zapisać odpowiedni zbiór krytyczny, w takiej postaci, jaka jest stosowna dla przeprowadzanego wariantu testu (dwu- lub jednostronny).

  → **Pytanie „na pomyślunek":** Co się stanie z wartością krytyczną i zbiorem krytycznym jeśli zwiększyć poziom istotności (np. z 0,05 do 0,1)?

- No dobra... ALE PO CO TO WSZYSTKO?? Do czego to nas prowadzi?? Te pytania, nurtujące się w sercu każdego Czytelnika domagają się wyjaśnienia LOGIKI testu statystycznego – patrz kolejny slajd.

Test t-Studenta: logika testu statystycznego

- W praktyce zazwyczaj (i tak też załóżmy dalej) mamy do dyspozycji **jedną próbę** (jeden zbiór danych), więc możemy obliczyć tylko **jedną, konkretną wartość statystyki testowej** (do jej wzoru podstawiając otrzymaną ocenę danego parametru i błąd średni szacunku). Zatem w owej praktyce nie otrzymujemy „całego rozkładu" wartości statystyki testowej. Rozkład ten – wyznaczony na „poziomie teorii" – pozwala nam tylko „rozeznać", gdzie należy się spodziewać, że ta jedna konkretna wartość statystyki testowej (którą uzyskujemy dla konkretnego zbioru danych i w konkretnym modelu) może się ulokować z mniejszym lub większym prawdopodobieństwem: czy w obszarze typowych jej wartości, czy też w zbiorze krytycznym. Rozkład ten – pamiętajmy – że jest wyznaczony przy założeniu prawdziwości H_0 i „unaocznia" on, których wartości statystyki testowej się należy najprawdopodobniej spodziewać (z wysokim prawdopodobieństwem $1-\alpha$), a które są raczej „nietypowe" (= zbiór krytyczny) i raczej mało prawdopodobne (jeśli H_0 miałaby być prawdziwa).

Test t-Studenta: logika i konkluzja testu statystycznego

- W związku z powyższym **kluczowe** w testach statystycznych (wszelakiej maści, nie tylko w teście *t*) jest sprawdzenie, gdzie znajduje się otrzymana (na podstawie konkretnych wyników estymacji) wartość statystyki testowej: czy w zbiorze jej typowych („typowo spodziewanych") wartości, czy też w zbiorze krytycznym, a jeśli w tym ostatnim – UWAGA! – to wtedy stwierdzamy, że w świetle hipotezy zerowej zdarzyło się coś nietypowego, coś zbyt mało prawdopodobnego, aby hipoteza zerowa mogła być uznana za prawdziwą (czy mniej ostro: wiarygodną). Dlatego też, jeśli otrzymana wartość statystyki testowej **„wpada"** do zbioru krytycznego, wówczas **odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną**. Co jednak, gdy wartość statystyki testowej „nie wpada" do zbioru krytycznego? Czy możemy powiedzieć, że uznajemy H_0 za prawdziwą i ją przyjmujemy? Niestety NIE do końca... Na gruncie głębokiej epistemologii (wiem, powiało i pompą, i grozą jednocześnie... ;) wskazuje się, że w takiej sytuacji, tj. gdy wartość statystyki testowej **„nie wpada"** do zbioru krytycznego (czyli w świetle H_0 zdarzyło się coś w dużej mierze prawdopodobnego, typowego, „tego się należało spodziewać"), wówczas po prostu nie zdarzyło się coś, co by przeczyło H_0, więc konkludujemy wtedy tylko, że **nie mamy podstaw do odrzucenia hipotezy zerowej.** Zatem... – patrz następny slajd.

Test t-Studenta: logika i konkluzja testu statystycznego

➢ **Reguła decyzyjna** testu statystycznego (dowolnego, nie tylko t-Studenta):

- Jeżeli wartość statystyki testowej należy („wpada”) do zbioru krytycznego, wówczas – przy danym poziomie istotności – odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej

  W teście t-Studenta statystykę testową oznaczamy symbolem $t(\beta_i)$, więc powyższe stwierdzenie możemy zapisać bardziej symbolicznie/skrótowo:

  $$t(\beta_i) \in Z_k \Longrightarrow \text{odrzucamy } H_0 \text{ na rzecz } H_1 \text{ (przy zadanym } \alpha)$$

  (czyli test „opowiada się” za hipotezą alternatywną, którą przyjmujemy)

- Jeżeli wartość statystyki testowej NIE należy („NIE wpada”) do zbioru krytycznego, wówczas – przy danym poziomie istotności – nie mamy podstaw do odrzucenia hipotezy zerowej:

  $$t(\beta_i) \notin Z_k \Longrightarrow \text{nie mamy podstaw do odrzucenia } H_0 \text{ (przy zadanym } \alpha)$$

  (czyli test „opowiada się” za hipotezą zerową, choć nie uprawnia nas to jeszcze do jej przyjęcia)

➢ **Zauważmy**, że ostateczna konkluzja testu może ZALEŻEĆ od wyboru poziomu istotności i wraz z jego zmianą może się zmienić.

→ **Pytanie „na pomyślunek”:** Jeśli zwiększyć poziom istotności (np. z 0,05 do 0,1), to odrzucenie H_0 staje się bardziej czy mniej prawdopodobne???

Test t-Studenta: poziom istotności – *revisited*

- Na chwilę powróćmy do poziomu istotności (α) i roli, którą odgrywa w testowaniu. Wiemy, że:
  1) Wyboru jego konkretnej wartości dokonuje sam badacz (czyli my :), wybierając jedną z trzech typowych: 0,01; 0,05; 0,1. Jeśli nic nie sugeruje inaczej (w szczególności treść zadania :), wówczas domyślnie przyjmujemy $\alpha = 0{,}05$.
  2) Jest prawdopodobieństwem nietypowych, mało prawdopodobnych / wiarygodnych z punktu widzenia H_0 wartości stastyki testowej – wartości, które w związku z tym traktujemy jako przeczące samej H_0 (na rzecz hipotezy alternatywnej).
  3) Wpływa na to, gdzie w rozkładzie statystyki testowej jest położona wartość krytyczna i jak wpływa to na „długość" zbioru krytycznego („długość" w cudzysłowie, bo zasadniczo jego długość – z uwagi na samą jego konstrukcję – jest nieskończona, nawet jeśli przesuniemy wartość krytyczną w jedną czy drugą stronę)
  4) W związku z powyższym, jego modyfikacja może również zmienić konkluzję (dlatego w praktyce można wykonać test przy różnych wartościach α, sprawdzając „wrażliwość" konkluzji testu)

Test t-Studenta: poziom istotności – *revisited*

- Na podstawie powyższej uwagi 2) możemy mieć małe „ALE” do procedury/logiki testu statystycznego. Z tą uwagą się, owszem, zgadzamy, ale pamiętajmy, że wartości w zbiorze krytycznym – choć mało prawdopodobne w świetle H_0 – tak jednak przecież też mogą „wynikać” z H_0, statystyka testowa może je przyjmować także gdy H_0 jest w istocie prawdziwa – tyle tylko, że z bardzo małym prawdopodobieństwem. Tak więc jeśli otrzymamy wartość statystyki testowej „wpadającą” do zbioru krytycznego, wówczas H_0 też może być prawdziwa – ALE ponieważ może to mieć miejsce tylko z bardzo niskim prawdopodobieństwem (równym przyjętemu poziomowi istotności, α), to UZNAJEMY, że to zbyt mało prawdopodobne, aby H_0 była jednak prawdziwa, i dlatego ją odrzucamy (na rzecz alternatywy). Możemy się jednak... mylić!
- Zatem pomyłka polega tu na tym, że „niechcący” odrzucamy hipotezę (zerową), która może być prawdziwa.
- W statystyce taką pomyłkę określamy mianem **błędu I rodzaju** = odrzucenie prawdziwej H_0 („niechcący”)
- W takim razie **poziom istotności** to prawdopodobieństwo popełnienia przez nas błędu I rodzaju. Sami je zadajemy, sami się na nie godzimy – to „cena”, którą (trochu „w ciemno”) godzimy się zapłacić. Zysk jednak polega na tym, że mamy w ogóle możliwość skorzystania z testu statystycznego – postępowania/procedury, która: 1) ma porządne postawy teoretyczne i dzięki temu jest sformalizowana, 2) służy naszemu poznaniu (z pewnymi ograniczeniami, oczywiście, ale jednak)
- W teście statystycznym możemy popełnić także tzw. **błąd II rodzaju** = nieodrzucenie *fałszywej* H_0, co wydaje się mieć poważniejsze skutki, lecz nie będziemy tu rozwijać tego wątku... (Uszami wyobraźni słyszę donośne: „Uff...!!” ;)

(Czas na...) Przykłady

Powróćmy do **Przykładu 1** i przeprowadźmy wybrane trzy (spośród sformułowanych tam) układy hipotez: nr 1, 3 i 5 (przytoczone poniżej). Przypomnijmy model (z oszacowaniami parametrów – patrz Przykład 1 w materiale o miernikach dopasowania):

$$E_t = -34{,}2 + 7{,}1I_t + 27{,}3H_t + \hat{\varepsilon}_t \qquad \text{(oszacowany na podstawie } T = 10 \text{ obserwacji)}$$

Zatem:

$$\hat{\beta}_0 = -34{,}2;\ \hat{\beta}_1 = 7{,}1;\ \hat{\beta}_2 = 27{,}3.$$

Do testów będziemy potrzebowali jeszcze błędów średnich szacunku. Przyjmijmy, że te wyniosły kolejno:

$$d(\hat{\beta}_0) = 1{,}2;\ d(\hat{\beta}_1) = 3{,}5;\ d(\hat{\beta}_2) = 4{,}3.$$

Przykłady

1) Wielkość dochodu g.d. istotnie wpływa na wielkość wydatków g.d. na transport:

- Zapis **układu hipotez**: $H_0: \beta_1 = 0 \quad vs. \quad H_1: \beta_1 \neq 0$ (test **dwustronny**)
- Zapis **statystyki testowej**:

$$t(\beta_1) = \frac{\hat{\beta}_1}{d(\hat{\beta}_1)} | H_0 \sim St(T-k), \qquad \text{przy czym } T-k = 10-3 = 7$$

- Obliczenie **wartości statystyki testowej**:

$$t(\beta_1) = \frac{\hat{\beta}_1}{d(\hat{\beta}_1)} = \frac{7{,}1}{3{,}5} \approx 2{,}029$$

- Ustalenie **poziomu istotności**: $\alpha = 0{,}05$ (domyślny, gdyż w zadaniu nie zasugerowanego innego)
- Odczytanie **wartości krytycznej** (tu: dwustronnej) – za pomocą funkcji w Excelu: =ROZKŁAD.T.ODW(0,05; 7) $\Rightarrow \quad t_{\alpha/2} = 2{,}364$
- Zapisanie zbioru krytycznego (dwustronnego): $Z_k = (-\infty; -2{,}364) \cup (2{,}364; +\infty)$
- Sprawdzenie, czy statystyka testowa „wpada" do Z_k – nie wpada, więc zapisujemy:

$$t(\beta_1) = 2{,}029 \notin Z_k \Rightarrow \text{brak podstaw do odrzucenia } H_0$$

- Konkluzja testu – pełna obejmuje **dwie części**. Na początek komunikujemy tylko, co – na zadanym poziomie istotności – reguła decyzyjna testu nakazuje zrobić z H_0 (odrzucamy na rzecz H_1 albo nie mamy podstaw po temu). Następnie ustosunkowujemy się do treści testowanego stwierdzenia (tu: „Wielkość dochodu g.d. istotnie wpływa ..."). Zatem:

  Na poziomie istotności $\alpha = 0{,}05$ nie mamy podstaw do odrzucenia H_0 (na rzecz H_1)*. Stwierdzamy zatem, że wielkość dochodu g.d.* ***NIE wpływa istotnie*** *na wielkość wydatków g.d. na transport.*

Przykłady

3) Wraz ze wzrostem wielkości dochodu g.d. wzrasta także wielkość wydatków g.d. na transport:

- Zapis **układu hipotez**: $H_0: \beta_1 = 0 \quad vs. \quad H_1: \beta_1 > 0$ (test **prawostronny**)
- Zapis **statystyki testowej** (nic się tu nie zmienia w stosunku do poprzedniego przykładu):

$$t(\beta_1) = \frac{\hat{\beta}_1}{d(\hat{\beta}_1)} | H_0 \sim St(T-k), \quad \text{przy czym } T-k = 10-3 = 7$$

- Obliczenie **wartości statystyki testowej** (komentarz j.w.):

$$t(\beta_1) = \frac{\hat{\beta}_1}{d(\hat{\beta}_1)} = \frac{7{,}1}{3{,}5} \approx 2{,}029$$

- Ustalenie **poziomu istotności**: $\alpha = 0{,}05$ (domyślny, gdyż w zadaniu na zasugerowanego innego)
- Odczytanie **wartości krytycznej** (tu: jednostronnej) – za pomocą funkcji w Excelu:
  =ROZKŁAD.T.ODW(**2***0,05; 7) $\Rightarrow \quad t_\alpha = 1{,}895$ (funkcja rozdziela **2***0,05 po połowie...)
- Zapisanie zbioru krytycznego (prawostronnego): $Z_k = (1{,}895; +\infty)$
- Sprawdzenie, czy statystyka testowa „wpada" do Z_k – owszem, więc zapisujemy:

$$t(\beta_1) = 2{,}029 \in Z_k \Rightarrow \text{odrzucamy } H_0$$

- Konkluzja testu (przypomnijmy: pełna obejmuje **dwie części**):

  Na poziomie istotności $\alpha = 0{,}05$ odrzucamy H_0 na rzecz H_1. Stwierdzamy zatem, że – owszem – wraz ze wzrostem wielkości dochodu g.d. wzrasta także wielkość wydatków g.d. na transport.

Przykłady

5) W gospodarstwach pracowniczych wielkość wydatków na transport jest wyższa w stosunku do gospodarstw emerytów i rencistów średnio o mniej niż 35 zł/osobę.

- Zapis **układu hipotez** : $H_0: \beta_2 = 35 \quad vs. \quad H_1: \beta_2 < 35$ (test **lewostronny**)
- Zapis **statystyki testowej**:

$$t(\beta_2) = \frac{\hat{\beta}_2 - \beta_2^*}{d(\hat{\beta}_2)} | H_0 \sim St(T-k), \quad \text{przy czym } \beta_2^* = 35 \text{ oraz } T-k = 10-3 = 7$$

- Obliczenie **wartości statystyki testowej** (komentarz j.w.):

$$t(\beta_2) = \frac{\hat{\beta}_2 - \beta_2^*}{d(\hat{\beta}_2)} = \frac{27{,}3 - 35}{4{,}3} \approx -1{,}791$$

- Ustalenie **poziomu istotności**: $\alpha = 0{,}05$ (domyślny, gdyż w zadaniu na zasugerowanego innego)
- Odczytanie **wartości krytycznej** (tu: jednostronnej) – za pomocą funkcji w Excelu: =ROZKŁAD.T.ODW(**2***0,05; 7) $\Rightarrow \quad t_\alpha = 1{,}895$ (funkcja rozdziela **2***0,05 po połowie...)
- Zapisanie zbioru krytycznego (lewostronnego): $Z_k = (-\infty; -1{,}895)$
- Sprawdzenie, czy statystyka testowa „wpada" do Z_k – nie wpada, więc zapisujemy:

$$t(\beta_2) = -1{,}791 \notin Z_k \Rightarrow \text{brak podstaw do odrzucenia } H_0$$

- Konkluzja testu (przypomnijmy: pełna obejmuje **dwie części**):

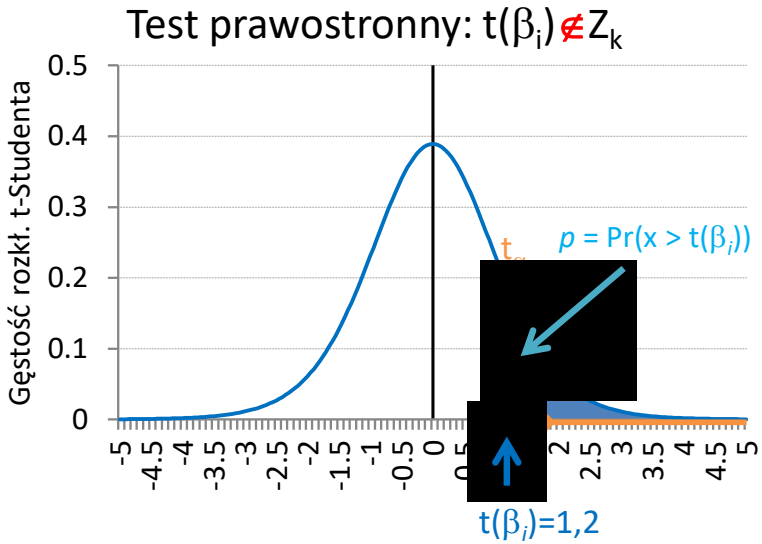
Na poziomie istotności $\alpha = 0{,}05$ nie mamy podstaw do odrzucenia H_0 na rzecz H_1. Zatem nie mamy podstaw twierdzić, że w gospodarstwach pracowniczych wielkość wydatków na transport jest wyższa w stosunku do gospodarstw emerytów i rencistów średnio o mniej niż 35 zł/os.

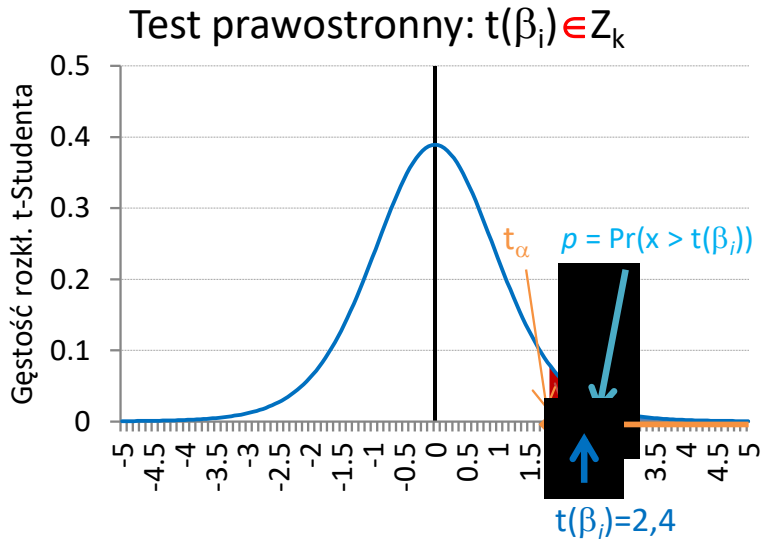
P-value (pl. prawdopodobieństwo testowe)

- W praktyce przeprowadzania testów statystycznych (dowolnych, nie tylko t-Studenta) zwykle raportowana jest dodatkowa wielkość, tzw. ***p-value***. Ten anglojęzyczny termin jest powszechnie stosowany także w języku polskim (rzadziej natomiast jego polskojęzyczne tłumaczenia: **prawdopodobieństwo testowe**, **wartość *p*** czy ***p*-wartość**).

- Czym jest ***p-value***? To prawdopodobieństwo (czyli pole pod krzywą funkcji gęstości) tych wszystkich możliwych wartości statystyki testowej w jej rozkładzie, które są bardziej skrajne (czyli w kierunku wytyczanym przez zbiór krytyczny) od obliczonej wartości statystyki testowej (tej jednej konkretnej, uzyskanej dla analizowanego zbioru danych). Najprościej zilustrować to graficznie – tu tylko dla przypadku testu prawostronnego (test lewostronny stanowi lustrzane odbicie prawostronnego, zaś test dwustronny „dzieli wszystko” po połowie, symetrycznie względem zera) – patrz kolejny slajd.

P-value (pl. prawdopodobieństwo testowe)





- Powyżej rozważono dwie sytuacje: nieodrzucania H_0 (ponieważ $t(\beta_i) \notin Z_k$) oraz odrzucania H_0 (ponieważ $t(\beta_i) \in Z_k$). Niebieskim, przeźroczystym kolorem zaznaczono właśnie *p-value* (w skrócie *p*), czerwonym natomiast – poziom istotności. Pomarańczowym – wartość krytyczną i zbiór krytyczny.

- Na podstawie powyższych wykresów daje się sformułować zauważyć pewna **reguła** (**bardzo ważna!**):
  - Gdy $t(\beta_i) \notin Z_k$ (a zatem gdy **nie mamy podstaw do odrzucenia H_0**), wówczas ***p-value* jest WIĘKSZE od poziomu istotności** (niebieskie pole **>** czerwone pole; wykres po lewej stronie)
    → w skrócie: $p > \alpha \iff$ NIE odrzucamy H_0

  - Gdy $t(\beta_i) \in Z_k$ (a zatem gdy **odrzucamy H_0**), wówczas ***p-value* jest MNIEJSZE (lub równe) od poziomu istotności** (niebieskie pole **<** czerwone pole; wykres po prawej stronie)
    → w skrócie: $p \leq \alpha \iff$ Odrzucamy H_0

P-value (pl. prawdopodobieństwo testowe)

- Te reguły należy sobie **dobrze przyswoić** :) (jak cytrynian magnezu ;) Dlaczego? Ponieważ w pakietach statystycznych (i pracy zawodowej analityków danych) komunikowanie wyników testów statystycznych zwykle ogranicza się właśnie do podania wartości *p-value*. Zapytacie z niedowierzaniem i nadzieją zarazem: „To to wystarczy?" Odpowiedź brzmi: „Zaiste!" (lub prawie zaiste). Podajmy sobie przykład. Załóżmy, że testowany jest układ hipotez:

$$H_0\text{: Wiesiek nie lubi czekolady} \quad \textit{vs.} \quad H_1\text{: Wiesiek lubi czekoladę}$$

Jakiś badacz przeprowadził stosowny test (na Wieśku) i otrzymał wartość $p = 0{,}012$. Co teraz? Teraz od razu (bez jakiejkolwiek znajomości obliczonej „po drodze" wartości statystyki testowej, wartości krytycznej czy zbioru krytycznego) można sformułować konkluzję testu – choć (UWAGA!) ta zależy od przyjętego przez nas poziomu istotności, α. Jeśli bowiem przyjąć domyślnie, że (zgadzamy się na możliwość odrzucenia „przez przypadek" prawdziwej H_0 z prawdopodobieństwem) $\alpha = 0{,}05$, to wtedy zachodzi $p \leq \alpha$, co oznacza, że wartość statystyki testowej musi przy tym poziomie istotności wpadać do zbioru krytycznego, a zatem odrzucamy H_0 i przyjmujemy, że Wiesiek istotnie lubi czekoladę (co, swoją drogą, nikogo nie zaskakuje, bo od dawna wiadomo, że co jak co, ale czekolada to nie ma wrogów). Jeśli natomiast przyjmiemy $\alpha = 0{,}01$, to zachodzi $p > \alpha$ i konkluzja się zmienia (i co gorsza, trzeba jeszcze oznajmić samemu Wieśkowi, że coś z nim jest nie tak ;).

P-value (pl. prawdopodobieństwo testowe)

- W powyższym przykładzie (wnioskowanie o Wieśku) konkluzja testu faktycznie zależy od wyboru poziomu istotności (jednego z trzech konwencjonalnych: 0,01; 0,05; 0,1). W praktyce nierzadko otrzymujemy wynik, w którym:

  a) *p-value* jest WIĘKSZE niż każdy konwencjonalny poziom istotności, np. $p = 0{,}48$

  (czyli równoważnie: $p > 0{,}1$ = najwyższy z konwencjonalnych poziomów istotności)

  Wtedy powiemy, że na każdym konwencjonalnym poziomie istotności nie mamy podstaw do odrzucenia H_0.

  b) *p-value* jest MNIEJSZE niż każdy konwencjonalny poziom istotności, np. $p = 0{,}0003$

  (czyli równoważnie: $p < 0{,}01$ = najniższy z konwencjonalnych poziomów istotności)

  Wtedy powiemy, że na każdym konwencjonalnym poziomie istotności należy odrzucić H_0 na rzecz H_1.

- I jeszcze jedna, prawie ostatnia już myśl. Zauważmy, że niskie wartości *p-value* świadczą przeciwko H_0, wysokie zaś – opowiadają się za H_0. Można powiedzieć, że im niższa (bliższa 0) wartość *p-value* (i mniejsza od 0,01), tym silniej H_0 jest odrzucana. I w drugą stronę: im wyższa (bliższa 1) wartość *p-value* (i wyższa od 0,1), tym H_0 jest silniej „wspierana” przez dane.

P-value (pl. prawdopodobieństwo testowe)

- Już na absolutnie sam koniec powiedzmy sobie, jak obliczyć *p-value* w teście t-Studenta za pomocą arkusza Excel. W tym celu korzystamy z funkcji: =ROZKŁAD.T(*x*; *stopnie_swobody*; *ślady*), przy czym:
  - w miejsce argumentu *x* podajemy wartość bezwzględną wartości statystyki testowej; formuła na wartość bezwzględną: MODUŁ.LICZBY(...)
  - jako *stopnie_swobody* podajemy oczywiście $T - k$
  - dziwnie brzmiący argument *ślady* to „stronność" testu:
    - jeśli przeprowadzamy test jednostronny (czy to lewo-, czy prawostronny), podajemy wartość 1
    - jeśli przeprowadzamy test dwustronny, podajemy wartość 2
  - bardzo ważne jest zachowanie dokładnej pisowni nazwy powyższej funkcji. W pakiecie Excel jest bowiem zaimplementowanych wiele funkcji o podobnym brzmieniu (np. =ROZKŁ.T(...)), które choć łudząco podobne w pisowni, służą (nieco) innym celom
  - dla porównania z powyższą, przypomnijmy jeszcze tylko zapis funkcji – podanej wcześniej w niniejszym materiale – służącej wyznaczaniu wartości krytycznej:

    =ROZKŁAD.T.ODW(*prawdopodobieństwo*; *stopnie_swobody*)

    Funkcja ta domyślnie zwraca dwustronną wartość krytyczną dla argumentu *prawdopodobieństwo* ustalonego na poziomie α. Jeśli chcemy wyznaczyć wartość jednostronną, wówczas należy podać

    $$prawdopodobieństwo = 2 \cdot \alpha.$$

    Oczywiście to tylko pomocniczy zabieg, mający na celu „obejście" ograniczeń związanych z działaniem tej funkcji (nie ma niestety w niej możliwości podania argumentu *ślady*, jak w przypadku funkcji do obliczania *p-value*). Otrzymana wartość jest w dalszym ciągu wartością jednostronną na poziomie istotności α (a nie 2α).

Przykłady – *c.d.*

➢ Kontynuując wcześniejszy przykład, przypomnijmy podstawowe rezultaty w każdym z trzech testów:

1) Wielkość dochodu g.d. istotnie wpływa na wielkość wydatków g.d. na transport

- Zapis układu hipotez: $H_0: \beta_1 = 0 \quad vs. \quad H_1: \beta_1 \neq 0$
- Poziom istotności: $\alpha = 0{,}05$
- Zbiór krytyczny: $Z_k = (-\infty; -2{,}364) \cup (2{,}364; +\infty)$
- $t(\beta_1) = 2{,}029 \notin Z_k \Rightarrow$ brak podstaw do odrzucenia H_0

3) Wraz ze wzrostem wielkości dochodu g.d. wzrasta także wielkość wydatków g.d. na transport

- Zapis układu hipotez: $H_0: \beta_1 = 0 \quad vs. \quad H_1: \beta_1 > 0$
- Poziom istotności: $\alpha = 0{,}05$
- Zbiór krytyczny: $Z_k = (1{,}895; +\infty)$
- $t(\beta_1) = 2{,}029 \in Z_k \Rightarrow$ odrzucamy H_0

5) W gospodarstwach pracowniczych wielkość wydatków na transport jest wyższa w stosunku do gospodarstw emerytów i rencistów średnio o mniej niż 35 zł/osobę

- Zapis układu hipotez: $H_0: \beta_2 = 35 \quad vs. \quad H_1: \beta_2 < 35$
- Poziom istotności: $\alpha = 0{,}05$
- Zbiór krytyczny: $Z_k = (-\infty; -1{,}895)$
- $t(\beta_2) = -1{,}791 \notin Z_k \Rightarrow$ brak podstaw do odrzucenia H_0

→ **Czas na pomyślunek**! Zanim podamy i skomentujemy wartości *p-value* **ZASTANÓW SIĘ**(!): jakich wartości *p-value* należy się spodziewać w powyższych testach – większej czy mniejszej niż przyjęty poziom istotności? Samo spojrzenie na konkluzję testu powinno wystarczyć :)

Przykłady – *c.d.*

- Wartości p-value uzyskujemy za pomocą stosownej funkcji w Excelu (omówionej wyżej):

1) Wielkość dochodu g.d. istotnie wpływa na wielkość wydatków g.d. na transport: $p - value = 0{,}082$

P-value jest tu większe niż $\alpha = 0{,}05$, zatem *na tym poziomie istotności* nie mamy podstaw do odrzucania H_0 (co faktycznie stanowi już wcześniej sformułowaną konkluzję testu)

3) Wraz ze wzrostem wielkości dochodu g.d. wzrasta także wielkość wydatków g.d. na transport: $p - value = 0{,}041$

P-value jest tu mniejsze niż $\alpha = 0{,}05$, zatem *na tym poziomie istotności* odrzucamy H_0 (co faktycznie stanowi już wcześniej sformułowaną konkluzję testu)

$\rightarrow$ **Pytanie „na pomyślunek”**: zauważmy, że tu *p-value* jest dwukrotnie mniejsze niż w pkt 1 – czy to przypadek? :)

5) W gospodarstwach pracowniczych wielkość wydatków na transport jest wyższa w stosunku do gospodarstw emerytów i rencistów średnio o mniej niż 35 zł/osobę: $p - value = 0{,}058$

P-value jest tu większe niż $\alpha = 0{,}05$, zatem *na tym poziomie istotności* nie mamy podstaw do odrzucania H_0 (co faktycznie stanowi już wcześniej sformułowaną konkluzję testu)

- **Ważniejszym** jest ju jednak następujący **komentarz**: w każdym przypadku *p-value* znajduje się gdzieś pomiędzy którymiś z dwóch konwencjonalnych poziomów istotności: albo między 0,05 i 0,1 (przypadki nr 1 i 5), albo między 0,01 i 0,05 (przypadek nr 3). Oznacza to, że w każdym przypadku ostateczna konkluzja testu *zależy* od przyjętego poziomu istotności. Tym samym, zmiana zakładanego poziomu istotności prowadzić będzie do zmiany konkluzji.