Wykład 5: Podstawowe testy parametryczne dla dwóch populacji

UWAGA: Dwie tabele z testami znajdują się na końcu tego wykładu.

Przykład 5.1. 20 spośród 100 losowo wybranych studentów studiów zaocznych i 40 spośród 120 losowo wybranych studentów studiów dziennych zdało egzamin ze Statystyki w pierwszym terminie.
(a) Czy na podstawie powyższych danych możemy stwierdzić, że studenci studiów zaocznych gorzej przygotowują się do egzaminu ze Statystyki niż studenci dzienni? Przyjąć poziom istotności 0,01.
(b) Przypuszczamy, że zdawalność egzaminu ze Statystyki w pierwszym terminie wynosi dla studentów studiów zaocznych 0,2 a dziennych - 0,3. Ilu studentów studiów zaocznych i ilu studentów studiów dziennych trzeba by wylosować do próby by jednostronny test porównujący proporcje z poziomem istotności 0,01 miał moc 0,75.

Rozwiązanie przykładu 5.1:
Student, zarówno studiów zaocznych jak i dziennych, może zdać egzamin ze Statystyki w pierwszym terminie lub nie. Mamy zatem do czynienia z dwoma populacjami o rozkładach dwupunktowych. Oznaczmy:

$$X = \begin{cases} 1 & \textit{jeśli wybrany student studiów zaocznych zda egzamin w pierwszym terminie,} \\ 0 & \textit{jeśli wybrany student studiów zaocznych nie zda egzaminu w pierwszym terminie,} \end{cases}$$

$$Y = \begin{cases} 1 & \textit{jeśli wybrany student studiów dziennych zda egzamin w pierwszym terminie,} \\ 0 & \textit{jeśli wybrany student studiów dziennych nie zda egzaminu w pierwszym terminie,} \end{cases}$$

zaś p_z niech będzie prawdopodobieństwem trafienia na studenta studiów zaocznych, który zda egzamin w pierwszym terminie, tzn. $p_z = P(X = 1)$, a p_d niech będzie prawdopodobieństwem trafienia na studenta studiów dziennych, który zda egzamin w pierwszym terminie, tzn. $p_d = P(Y = 1)$.

Do rozważanego problemu pasuje model oznaczony jako model V w tabeli Weryfikacje hipotez dotyczących dwóch średnich.

(a) Stawiamy następujące hipotezy:
$H_0 : p_z = p_d$
$H_1 : p_z < p_d$ *(co by oznaczało, że studenci studiów zaocznych gorzej przygotowują się do egzaminu niż dzienni)*
Zauważmy, że

- $n_z\hat{p}_z = 20 \geq 5$ *($n_z\hat{p}_z$ to liczba sukcesów wśród studentów studiów zaocznych, czyli liczba studentów studiów zaocznych, którzy zdali egzamin w pierwszym terminie),*
- $n_z(1 - \hat{p}_z) = 80 \geq 5$ *($n_z(1 - \hat{p}_z)$ to liczba porażek wśród studentów studiów zaocznych, czyli liczba studentów studiów zaocznych, którzy nie zdali egzaminu w pierwszym terminie),*
- $n_d\hat{p}_d = 40 \geq 5$ *($n_d\hat{p}_d$ to liczba sukcesów wśród studentów studiów dziennych, czyli liczba studentów studiów dziennych, którzy zdali egzamin w pierwszym terminie),*
- $n_d(1 - \hat{p}_d) = 80 \geq 5$ *($n_d(1 - \hat{p}_d)$ to liczba porażek wśród studentów studiów dziennych, czyli liczba studentów studiów dziennych, którzy nie zdali egzaminu w pierwszym terminie).*

Zatem możemy użyć `prop.test()`*, czyli testu, w którym rozkład statystyki testowej jest przybliżany rozkładem normalnym:*

```
> prop.test(x=c(20,40),n=c(100,120),alternative="less")
```

Powyżej argument `x` *to wektor z liczbami otrzymanych sukcesów, a* `n` *to wektor z licznościami pobranych prób. Można też za* `x` *podać macierz, która w pierszej kolumnie ma liczby sukcesów a w drugiej - liczby porażek, przy czym pierwszy wiersz dotyczy pierwszej populacji (tutaj studentów studiów zaocznych) a drugi - drugiej populacji (tutaj studentów stdiów dziennych). Jeśli jako* `x` *podamy macierz, to argument* `n` *jest ignorowany. Chcąc użyć tej drugiej konwencji potrzebujemy następującą macierz*

$$\begin{bmatrix} 20 & 80 \\ 40 & 80 \end{bmatrix}$$

Aby ją wpisać do R *użyjemy funkcji*
> `matrix((c(20,40,80,80),nrow=2)`
której dokładny opis znajduje się we Wstępie do środowiska R. *Reasumując, drugi sposób wywołania funkcji* `prop.test()` *jest następujący*
> `prop.test(x=matrix((c(20,40,80,80),nrow=2),alternative="less")`
Używając któregokolwiek z dwóch omówionych powyżej podejść, otrzymujemy

$$p - value = 0,01974 > \alpha = 0,01 \ \Rightarrow \ \textit{nie ma podstaw do odrzucenia } H_0,$$

gdzie $\alpha = 0,01$ *to poziom istotności testu. Wyciągamy więc wniosek, że studenci studiów zaocznych nie przygotowują się gorzej do egzaminu ze Statystyki niż studenci dzienni.*

UWAGA: Gdyby danych było za mało do użycia `prop.test()`, tzn., gdyby nie był spełniony którykolwiek z warunków: $n_z\hat{p}_z \geq 5$, $n_z(1-\hat{p}_z) \geq 5$, $n_d\hat{p}_d \geq 5$, $n_d(1-\hat{p}_d) \geq 5$, to należałoby użyć dokładnego testu Fishera:
> `fisher.test(x=matrix((c(20,40,80,80),nrow=2),alternative="less")`

(b) Zakładamy, że $p_z = 0,2$ *i* $p_d = 0,3$ *i szukamy* n_z *oraz* n_d *takich, aby moc testu wynosiła* $0,75$, *tzn. chcemy, aby* $moc.testu(0,2;0,3) = 0,75$.

Użyjemy funkcji `power.prop.test()`, *która działa analogicznie jak omówiona na poprzednim wykładzie funkcja* `power.t.test()`.
> `power.prop.test(power=0.75,p1=0.2,p2=0.3,sig.level=0.01,alternative="one.sided")`
Otrzymujemy $n = 336,6738$, *co oznacza, że potrzeba 337 studentów zaocznych i 337 studentów dziennych.*

UWAGA: Funkcja `power.prop.test()` obsługuje tylko `prop.test()` dla dwóch populacji, ponadto tylko sytuacje, gdy próbki z obu populacji są równoliczne (ich równe liczności to argument `n`).

Przykład 5.2. Losową grupę 5 osób poddano 6-tygodniowej diecie odchudzającej. Uzyskano następujące wyniki (waga przed i po kuracji [w kg]):

Przed kuracją	88	86	82	64	59
Po kuracji	75	76	83	65	58

Można założyć, że rozkład łączny wagi przed i po kuracji jest normalny.
(a) Czy powyższe wyniki potwierdzają skuteczność diety (przyjąć poziom istotności 0,05)?
(b) Przypuszczamy, że średnia różnica wagi sprzed i po kuracji wynosi 4 kg. Jakie jest prawdopodobieństwo, że test z punktu (a) potwierdzi skuteczność diety?
(c) Przypuszczamy, że średnia różnica wagi sprzed i po kuracji wynosi 4 kg. Ile osób trzeba by losowo wybrać do eksperymentu by test jednostronny o poziomie istotności 0,05 z prawdopodobieństwem 0,8 potwierdzał skuteczność diety?

Rozwiązanie przykładu 5.2:

Oznaczmy: X– *waga przed kuracją,* Y– *waga po kuracji. Dane mamy połączone w pary* (X, Y) *a pary są wzajemnie niezależne. Fakt, że dane są połączone w pary sugeruje, że przydatny będzie tzw. ,,paired t-test" opisany w modelu III w tabeli* Weryfikacje hipotez dotyczących dwóch średnich. *Sprawdźmy czy są spełnione założenia tego modelu.*

Z treści zadania wiemy, że łączny rozkład (X, Y) *to rozkład normalny (nie znamy jego parametrów), co pociąga za sobą, że rozkład* $X - Y$ *też jest rozkładem normalnym (o nieznanych parametrach). Zapisujemy to natępująco*

$$X - Y \sim \mathcal{N}(\mu_Z, \sigma_Z^2), \ \textit{gdzie } \mu_Z \textit{ i } \sigma_Z \textit{ są nieznane.}$$

Zatem rzeczywiście założenia są spełnione i możemy ,,paired t-test" użyć.

Wpisujemy dane do R:
> `waga.przed <- c(88,86,82,64,59)`
> `waga.po <- c(75,76,83,65,58)`

(a) Stawiamy hipotezy:
$H_0 : \mu_{przed} = \mu_{po}$ *(czyli dieta nie jest skuteczna)*
$H_1 : \mu_{przed} > \mu_{po}$ *(czyli średnio dieta powoduje spadek wagi)*
Do weryfikacji H_0 przeciwko H_1 użyjemy „paired t-test", czyli funkcji `t.test()` *z argumentem* `paired=TRUE`:

```
> t.test(waga.przed,waga.po,alternative="greater",paired=TRUE)
```

Powyżej mogliśmy nie podać wartości argumentu `mu`, *bo domyślnie przyjmuje on wartość 0 (taką wartość właśnie potrzebujemy, bo $H_0 : \mu_{przed} - \mu_{po} = 0$).*

Odczytujemy p-wartość i wyciągamy wnioski:

$$p - value = 0,1057 > \alpha = 0,05 \;\Rightarrow\; \textit{nie ma podstaw do odrzucenia } H_0.$$

Zatem uznajemy, że dieta nie jest skuteczna.

(b) Zakładamy, że $\mu_{przed} - \mu_{po} = 4$. Przy tym założeniu chcemy policzyć prawdopodobieństwo, że test potwierdzi skuteczność diety, czyli, że wtedy H_0 zostanie odrzucona. Zatem szukamy prawdopodobieństwa, że odrzucimy H_0 w sytuacji, gdy $\mu_{przed} - \mu_{po} = 4$:

$$P(\textit{odrzucimy } H_0 | \mu_{przed} - \mu_{po} = 4) = ?$$

Przypomniejmy, że

$$\textit{moc.testu}(\beta) = P(\textit{odrzucimy } H_0 | \textit{ badany parametr } = \beta).$$

Stąd

$$P(\textit{odrzucimy } H_0 | \; \mu_{przed} - \mu_{po} = 4) = \textit{moc.testu}(4).$$

Użyjemy funkcji `power.t.test()`, *która została dokładnie omówiona na poprzednim wykładzie. Teraz przypomnimy jedynie, że*

- *dla argumentu* `type` *mamy do wyboru trzy opcje* `type="one.sample"`, `"two.sample"` *lub* `"paired"`. *W tym zadaniu potrzebujemy oczywiście* `type="paired"`;
- *argument* `sd` *to odchylenie standardowe $X - Y$, którego nie znamy; pozostaje nam więc je oszacować na podstawie próby licząc*

  ```
  > sd(waga.przed-waga.po)
  ```

  i jednocześnie mając świadomość, że taki zabieg prowadzi do wyniku przybliżonego.

Reasumując potrzebujemy:

```
> power.t.test(n=5, delta=4, sd=sd(waga.przed-waga.po), sig.level=0.05,
               type="paired",alternative="one.sided")
```

gdzie ewentualnie `sig.level=0.05` *można pominąć, bo tak jest ustawione domyślnie. Otrzymujemy power $\approx 0,30$. Jest to prawdopodobieństwo odrzucenia $H_0 : \mu_{przed} = \mu_{po}$ (czyli innymi słowy uznania, że dieta jest skuteczna) w sytuacji, gdy $\mu_{przed} - \mu_{po} = 4$. Zatem jeśli średnią stratę 4 kg w ciągu 6 tygodni uznajemy za zadawalający efekt diety, to chcielibyśmy, aby przy $\mu_{przed} - \mu_{po} = 4$ test z dużym prawdopodobieństwem odrzucał H_0. Tak niestety nie jest – otrzymane prawdopodobieństwo $\approx 0,30$ jest małe, czyli test nie radzi sobie z zauważeniem, że dieta jest skuteczna. Wynika to zapewne stąd, że jest on przeprowadzany na podstawie bardzo małej liczby obserwacji ($n = 5$). W kolejnym podpunkcie znajdziemy n gwarantujące, że wynik testu będzie bardziej wiarygodny.*

(c) Nadal zakładamy, że $\mu_{przed} - \mu_{po} = 4$. Przy tym założeniu szukamy n takiego by prawdopodobieństwo odrzucenia H_0 wynosiło 0,8:

$$\textit{szukamy } n \textit{ takiego by } P(\textit{odrzucimymy } H_0 | \mu_{przed} - \mu_{po} = 4) = 0,8$$

czyli

$$\textit{szukamy } n \textit{ takiego by } \textit{moc.testu}(4) = 0,8.$$

Użyjemy funkcji funkcji `power.t.test()`:

```
> power.t.test(power=0.8, delta=4, sd=sd(waga.przed-waga.po), sig.level=0.05,
               type="paired",alternative="one.sided")
```

Otrzymujemy $n \approx 18,36$ *co oznacza, że potrzebujemy próbkę o liczności* $n = 19$ *(wtedy moc testu trochę przekroczy* $0,8$*, ale dokładnej wartości* $0,8$ *nie da się uzyskać).*

Przykład 5.3. Dokonano pomiarów tego samego napięcia prądu: 10 pomiarów przy użyciu jednego woltomierza i niezależnie 10 pomiarów przy użyciu drugiego woltomierza. Dla pierwszego woltomierza otrzymano następujące wyniki

$$1.2, 1.0, 1.1, 1.4, 1.1, 1.2, 1.0, 0.9, 1.1, 1.2,$$

a dla drugiego:

$$1.3, 1.1, 1.4, 0.9, 1.4, 1.2, 1.3, 1.0, 1.2, 1.3.$$

Można założyć, że pomiary napięcia na badanych woltomierzach mają rozkłady normalne.
(a) Na poziomie istotności 0,01 zweryfikować hipotezę o jednakowych wynikach pomiaru napięcia przez oba woltomierze.
(b) Przypuszczamy, że średnia różnica pomiarów na obu woltomierzach to $0,1$. Ile pomiarów na każdym woltomierzu należy wykonać by moc dwustronnego testu o poziomie istotności 0,01 wynosiła nie mniej niż 0,8.

Rozwiązanie przykładu 5.3:

Oznaczmy: X *– napięcie prądu uzyskane przy użyciu pierwszego woltomierza,*
Y *– napięcie prądu uzyskane przy użyciu drugiego woltomierza.*
Pomiary na pierwszym i drugim woltomierzu były wykonywane niezależnie, co sugeruje, że przydatny będzie tzw. ,,unpaired t-test" opisany w modelu II w tabeli Weryfikacje hipotez dotyczących dwóch średnich. *Sprawdźmy czy są spełnione założenia tego modelu.*

Z treści zadania wiemy, że X *i* Y *mają rozkłady normalne. Parametrów tych rozkładów nie znamy. Zapisujemy to natępująco*

$$X \sim \mathcal{N}(\mu_1, \sigma_1^2),\ Y \sim \mathcal{N}(\mu_2, \sigma_2^2),\ \text{gdzie } \mu_1, \mu_2, \sigma_1 \text{ i } \sigma_2 \text{ są nieznane.}$$

Nie wiemy jednak czy $\sigma_1 = \sigma_2$ *i to musimy sprawdzić na samym początku. Zatem stawiamy hipotezy:*
$H_0 : \sigma_1 = \sigma_2$ *(lub równoważnie* $\sigma_1^2 = \sigma_2^2$*),*
$H_1 : \sigma_1 \neq \sigma_2$ *(lub równoważnie* $\sigma_1^2 \neq \sigma_2^2$*).*
Powyższe hipotezy będziemy weryfikować używając testu F*,opisanego w tabeli* Weryfikacja hipotezy dotyczącej równości dwóch wariancji. *Zalecane jest przeprowadzanie tego testu na poziomie istotności* $0,1$.

Wpisujemy dane do R:

```
> wolt1 <- c(1.2, 1.0, 1.1, 1.4, 1.1, 1.2, 1.0, 0.9, 1.1,1.2)
> wolt2 <- c(1.3, 1.1, 1.4, 0.9, 1.4, 1.2, 1.3, 1.0, 1.2, 1.3)
```

Test F przeprowadzamy używąc funkcji `var.test()`:

```
> var.test(wolt1,wolt2,alternative="two.sided")
```

gdzie `alternative="two.sided"` *można pominąć, bo tak jest ustawione domyślnie. Pamiętając, że zalecany poziom istotności testu* F *to* $\alpha = 0,1$*, wyciągamy wnioski*

$$p - value = 0,6136 > \alpha = 0,1 \Rightarrow \text{ nie ma podstaw do odrzucenia } H_0.$$

Zatem uznajemy, że wariancje są równe $\sigma_1^2 = \sigma_2^2$ *(równoważnie, że odchylenia standardowe są równe:* $\sigma_1 = \sigma_2$*). Reasumując, wnioskujemy, że do problemu rozważanego w tym zadaniu możemy użyć ,,unpaired t-test" z równymi wariancjami. Wiedząc to przechodzimi do podpunktu (a).*

(a) Stawiamy hipotezy:
$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$
i do ich weryfikacji używamy ,,unpaired t-test" z równymi wariancjami, czyli funkcji `t.test()` *z argumentem* `paired=FALSE` *i argumentem* `var.equal=TRUE`:

```
> t.test(wolt1,wolt2,alternative="two.sided",paired=FALSE,var.equal=TRUE)
```

przy czym `alternative="two.sided",paired=FALSE` *można pominąć, bo tak jest ustawione domyślnie.*

Odczytujemy p-wartość i wyciągamy wnioski:

$$p - value = 0,2068 > \alpha = 0,01 \;\Rightarrow\; \textit{nie ma podstaw do odrzucenia } H_0.$$

Zatem uznajemy, że oba woltomierze dają średnio takie same wyniki.

(b) Zakładamy, że $\mu_1 - \mu_2 = 0,1$. Przy tym założeniu chcemy wyznaczyć minimalną liczbę pomiarów do wykonania na każdym woltomierzu, gwarantującą moc testu nie mniejszą niż 0,8:

$$\textit{szukamy } n_1 = n_2 \textit{ takiego by } moc.testu(0,1) \geq 0,8.$$

Użyjemy funkcji `power.t.test()`, *pamiętając, że*

- *dla argumentu* `type` *mamy do wyboru trzy opcje* `type="one.sample"`, `"two.sample"` *lub* `"paired"`; *w przypadku ,,unpaired t-test" z równymi wariancjami wybieramy* `type="two.sample"`;
- *argument* `n` *oznacza równe liczności prób pobranych z pierwszej i drugiej populacji (obsługiwany jest tylko przypadek, gdy te liczności są równe);*
- *argument* `sd` *to odchylenie standardowe X i Y, czyli nieznana wielkość $\sigma_1 = \sigma_2$; szacujemy ją wykorzystując obie próby i używając następującego estymatora*

  $$\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

  i jednocześnie mając świadomość, że taki zabieg prowadzi do wyniku przybliżonego.

Reasumując, najpierw policzymy estymator wielkości $\sigma_1 = \sigma_2$:

$$\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{\frac{(10-1)s_1^2 + (10-1)s_2^2}{10+10-2}} = \sqrt{\frac{9s_1^2 + 9s_2^2}{18}} = \sqrt{\frac{s_1^2 + s_2^2}{2}},$$

kończąc rachunki w R

```
> estymator.odchylenia <- sqrt((var(wolt1)+var(wolt2))/2)
```

Następnie wykorzystamy ten estymator jako argument funkcji `power.t.test()`

```
> power.t.test(power=0.8, delta=0.1, sd= estymator.odchylenia, sig.level=0.01,
               type="two.sample",alternative="two.sided")
```

Otrzymujemy n ≈ 56,83, z czego wynika, że potrzebujemy 57 pomiarów wykonanych przy użyciu pierwszego woltomierza i 57 pomiarów wykonanych przy użyciu drugiego woltomierza.

UWAGA: Gdybyśmy, rozwiązując przykład 5.3, odrzucili hipotezę $H_0 : \sigma_1 = \sigma_2$ i przyjęli $H_1 : \sigma_1 \neq \sigma_2$, to zamiast użyć ,,unpaired t-test" z równymi wariancjami trzeba by zastosować ,,unpaired t-test" z różnymi wariancjami, czyli zastosować `t.test()` z argumentemi `paired=FALSE`, `var.equal=FALSE`.

Funkcja `power.t.test()` niestety ,,unpaired t-test" z różnymi wariancjami nie obsługuje.

Przykład 5.4. Pełnomocnik rządu Alfalandii d/s równego statusu kobiet i mężczyzn podejrzewa, że mężczyźni pracujący jako modele średnio zarabiają mniej niż kobiety modelki. Czy na poziomie istotności 0,01 można uznać to podejrzenie za słusznene, jeśli średni miesięczny dochód 100 losowo wybranych modeli wyniósł 480 dukatów z odchyleniem standardowym 60 dukatów, a średni miesięczny dochód 100 losowo wybranych modelek to 600 dukatów z odchyleniem standardowym 200 dukatów?

Rozwiązanie przykładu 5.4:

Niech X to zarobki mężczyzny modela, zaś Y to zarobki modelki. Nie mamy informacji na temat rozkładu X ani Y, ale dysponujemy dużymi próbami: $n_M = 100 \geq 100$ i $n_K = 100 \geq 100$. Pasuje nam zatem model oznaczony jako model IV w tabeli Weryfikacje hipotez dotyczących dwóch średnich.

Będziemy testować $H_0 : \mu_M = \mu_K$ przeciwko $H_1 : \mu_M < \mu_K$.

Liczymy wartość statystyki testowej $U = \frac{\overline{x}-\overline{y}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$:

```
> sredniaM <- 480
> sredniaK <- 600
> odchylenieM <- 60
> odchylenieK <- 200
> nM <- 100
> nK <- 100
> (sredniaM-sredniaK)/sqrt((odchylenieM^2/nM) + (odchylenieK^2/nK))
```

Otrzymujemy $U \approx -5,747$ i przechodzimy do wyznaczenia zbioru krytycznego $W = (-\infty; -u_{1-\alpha}\rangle$. Skoro poziom istotności testu wynosi $\alpha = 0,01$, to potrzebujemy $u_{1-\alpha} = u_{0,99}$:

```
> qnorm(0.99)
```

co daje $\approx 2,326$. Zatem $W \approx (-\infty; -2,326\rangle$ i widzimy, że $U \approx -5,747 \in W \approx (-\infty; -2,326\rangle$, więc odrzucamy H_0 i stwierdzamy, że rzeczywiście mężczyźni pracujący jako modele średnio zarabiają mniej niż kobiety modelki.

<table>
<tr><th colspan="2">Weryfikacja hipotezy dotyczącej równości dwóch wariancji na poziomie istotności α</th></tr>
<tr><td colspan="2">Model I. (test F: var.test)
$X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, μ_X, μ_Y, σ_X, σ_Y - nieznane,
dysponujemy niezależnymi próbami losowymi z tych populacji.
Hipoteza zerowa $H_0 : \sigma_X^2 = \sigma_Y^2$.</td></tr>
<tr><td>Hipoteza alternatywna $H_1 : \sigma_X^2 \neq \sigma_Y^2$
Statystyka testowa $F = s_1^2/s_2^2$ (w liczniku jest większa z wariancji).
Zbiór krytyczny
$W = \langle F(1-\alpha/2, n_1-1, n_2-1); +\infty)$,
gdzie n_1 oznacza liczność próby o większej wariancji próbkowej.</td><td>Hipoteza alternatywna $H_1 : \sigma_X^2 > \sigma_Y^2$
Statystyka testowa $F = s_X^2/s_Y^2$
Zbiór krytyczny
$W = \langle F(1-\alpha, n_X-1, n_Y-1); +\infty)$</td></tr>
<tr><td colspan="2">Powyżej $F(\alpha, n, m)$ oznacza kwantyl rozkładu F-Snedecora: < qf(α, n, m)</td></tr>
<tr><td colspan="2">UWAGA: jeżeli wyznaczona wartość statystyki testowej F należy do W, to H_0 odrzucamy.</td></tr>
</table>

<table>
<tr><td colspan="3" align="center">Weryfikacje hipotez dotyczących dwóch średnich na poziomie istotności α</td></tr>
<tr><td colspan="3">UWAGA: jeżeli wyznaczone wartości statystyk (U lub T) należą do odpowiednich zbiorów krytycznych, to H_0 odrzucamy.</td></tr>
<tr><td colspan="3">Model I. $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, μ_1, μ_2 - nieznane, σ_1, σ_2 - znane; dysponujemy niezależnymi próbami losowymi z tych populacji.
Hipoteza zerowa $H_0 : \mu_1 = \mu_2$. Statystyka testowa $U = \frac{\overline{x}-\overline{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$.</td></tr>
<tr><td>Hipoteza alternatywna
$H_1 : \mu_1 \neq \mu_2$
Zbiór krytyczny
$W = (-\infty; -u_{1-\alpha/2}\rangle \cup \langle u_{1-\alpha/2}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu_1 > \mu_2$
Zbiór krytyczny
$W = \langle u_{1-\alpha}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu_1 < \mu_2$
Zbiór krytyczny
$W = (-\infty; -u_{1-\alpha}\rangle$</td></tr>
<tr><td colspan="3">Model II.(unpaired t-test: t.test(...,paired=FALSE,var.equal=TRUE))
$X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, μ_1, μ_2 - nieznane, σ_1, σ_2 - nieznane, ale takie, że $\sigma_1 = \sigma_2$; dysponujemy niezależnymi próbami losowymi z tych populacji.
Hipoteza zerowa $H_0 : \mu_1 = \mu_2$. Statystyka testowa $T = \frac{\overline{x}-\overline{y}}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$.</td></tr>
<tr><td>Hipoteza alternatywna
$H_1 : \mu_1 \neq \mu_2$
Zbiór krytyczny
$W = (-\infty, -t_{1-\alpha/2,n_1+n_2-2}\rangle \cup \langle t_{1-\alpha/2,n_1+n_2-2}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu_1 > \mu_2$
Zbiór krytyczny
$W = \langle t_{1-\alpha,n_1+n_2-2}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu_1 < \mu_2$
Zbiór krytyczny
$W = (-\infty, -t_{1-\alpha,n_1+n_2-2}\rangle$</td></tr>
<tr><td colspan="3">Jeśli w modelu II nie jest spełnione założenie, że $\sigma_1 = \sigma_2$, to zamiast t.test(...,paired=FALSE,var.equal=TRUE) należy użyć t.test(...,paired=FALSE,var.equal=FALSE)</td></tr>
<tr><td colspan="3">Model III.(paired t-test: t.test(..., paired=TRUE))
$X - Y \sim N(\mu_Z, \sigma_Z^2)$, μ_Z, σ_Z - nieznane; dysponujemy parami obserwacji, gdzie pary są wzajemnie niezależne.
Hipoteza zerowa $H_0 : \mu_1 = \mu_2$. Statystyka testowa $T = \frac{\overline{z}}{s_z}\sqrt{n}$, gdzie $z_i = x_i - y_i, i = 1, 2, \ldots, n$.</td></tr>
<tr><td>Hipoteza alternatywna
$H_1 : \mu_1 \neq \mu_2$
Zbiór krytyczny
$W = (-\infty; -t_{1-\alpha/2,n-1}\rangle \cup \langle t_{1-\alpha/2,n-1}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu_1 > \mu_2$
Zbiór krytyczny
$W = \langle t_{1-\alpha,n-1}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu_1 < \mu_2$
Zbiór krytyczny
$W = (-\infty; -t_{1-\alpha,n-1}\rangle$</td></tr>
<tr><td colspan="3">Model IV. Cechy X, Y mają rozkłady dowolne ($n_1 \geq 100$, $n_2 \geq 100$), μ_1, μ_2, σ_1, σ_2 - nieznane; dysponujemy niezależnymi próbami losowymi z tych populacji.
Hipoteza zerowa $H_0 : \mu_1 = \mu_2$. Statystyka testowa $U = \frac{\overline{x}-\overline{y}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$.</td></tr>
<tr><td>Hipoteza alternatywna
$H_1 : \mu_1 \neq \mu_2$
Zbiór krytyczny
$W = (-\infty; -u_{1-\alpha/2}\rangle \cup \langle u_{1-\alpha/2}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu_1 > \mu_2$
Zbiór krytyczny
$W = \langle u_{1-\alpha}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : \mu_1 < \mu_2$
Zbiór krytyczny
$W = (-\infty; -u_{1-\alpha}\rangle$</td></tr>
<tr><td colspan="3">Model V (prop.test). Cechy X, Y mają rozkłady dwupunktowe, $P(X = 1) = p_1 = 1 - P(X = 0)$, $P(Y = 1) = p_2 = 1 - P(Y = 0)$, p_1, p_2 - nieznane, $n_1\hat{p}_1 \geq 5$ i $n_1(1 - \hat{p}_1) \geq 5$ i $n_2\hat{p}_2 \geq 5$ i $n_2(1 - \hat{p}_2) \geq 5$.
Hipoteza zerowa $H_0 : p_1 = p_2$. Statystyka testowa $U = \frac{\hat{p}_1-\hat{p}_2}{\sqrt{\frac{\overline{p}\overline{q}}{n}}}$, gdzie $\hat{p}_1 = \frac{k_1}{n_1}$, $\hat{p}_2 = \frac{k_2}{n_2}$, $\overline{p} = \frac{k_1+k_2}{n_1+n_2}$, $\overline{q} = 1- \overline{p}$, $n = \frac{n_1 n_2}{n_1+n_2}$.</td></tr>
<tr><td>Hipoteza alternatywna
$H_1 : p_1 \neq p_2$
Zbiór krytyczny
$W = (-\infty; -u_{1-\alpha/2}\rangle \cup \langle u_{1-\alpha/2}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : p_1 > p_2$
Zbiór krytyczny
$W = \langle u_{1-\alpha}; +\infty)$</td><td>Hipoteza alternatywna
$H_1 : p_1 < p_2$
Zbiór krytyczny
$W = (-\infty; -u_{1-\alpha}\rangle$</td></tr>
<tr><td colspan="3">Jeśli w modelu V nie jest spełnione założenie, że n_1, n_2 są wystarczająco duże, to zamiast prop.test należy zastosować dokładny test Fishera fisher.test oparty na rozkładzie hipergeometrycznym.</td></tr>
</table>