T-Test

Testy t-Studenta służą do porównania ze sobą dwóch grup. Korzystamy z nich wtedy, gdy mamy wyniki dla dwóch grup i chcemy porównać je ze sobą - tzn. stwierdzić, czy wyniki w jednej grupie są większe bądź mniejsze niż w drugiej grupie.

Ogólnie wykorzystujemy test t kiedy dostępne dane pochodzące z eksperymentu (badania) składają się z jednego parametru nominalnego (wyrażonego w skali nominalnej) oraz jednego parametru ilościowego (pomiaru), i kiedy chcemy porównać wartości średnie (parametru ilościowego). Parametr nominalny musi mieć jedynie dwie wartości, np. płeć (‘K’,’M’), leczenie (‘leczony’,’nieleczony’), lek (‘lek’,’placebo’), itd.

Rodzaje testów

Istnieją trzy rodzaje testu t-Studenta:

1. **test t dla prób niezależnych**
   Ideą testu t-Studenta dla prób niezależnych jest porównanie ze sobą dwóch różnych grup obserwacji. Jak sama nazwa testu wskazuje, grupy (próby) muszą być wobec siebie niezależne, czyli wyniki pomiaru jednej grupy nie są zależne wobec pomiaru drugiej grupy.
2. **test t dla prób zależnych**
   Ideą testu t-Studenta dla prób zależnych jest porównanie ze sobą dwóch różnych grup obserwacji, które są wobec siebie zależne. Przykładowo, obserwacje tej samej grupy osób, obserwacje dwukrotne, gdzie wynik w drugim badaniu jest zależny od wyniku w pierwszym badaniu, ponieważ dotyczy tej samej osoby, obserwacji. W takiej sytuacji celem testu t-Studenta jest określenie wielkości zmian (tendencji) w danym pomiarze wśród badanych osob, obserwacji.
3. **dla jednej próby**
   Ideą testu dla jednej próby jest porównanie ze sobą średniej i odchylenia standardowego wyznaczonej dla jednej grupy osób badanych z pewną założoną odgórnie wartością. W tym kontekście porównanie jednej grupy stanowi wyjątek w rodzinie testów t-Studenta. W głównej mierze testy dla jednej próby stosuje się w celu zweryfikowania poprawności wyników badań względem znanej, często uznanej w teorii, z innych badań i analiz wartością.

W zależności od rodzaju badania stosujemy w analizach jeden z tych testów.

Przypomnienie

Rozkład Studenta z ν stopniami swobody jest rozkładem zmiennej losowej t postaci:

$$t = \frac{U}{\sqrt{Z}}\sqrt{\nu}$$

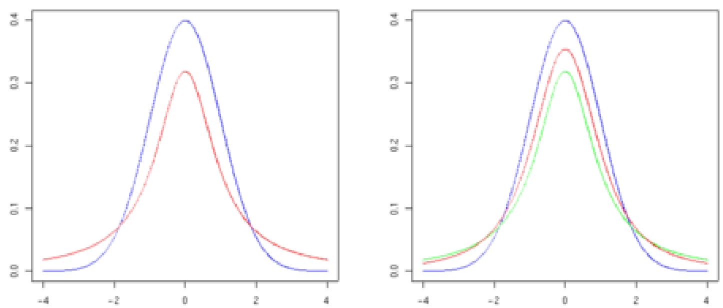
gdzie:

- U jest zmienną losową zestandaryzowaną, czyli mającą standardowy rozkład normalny $N(0,1)$,
- Z jest zmienną losową o rozkładzie chi kwadrat o ν stopniach swobody,
- U oraz Z są zmiennymi losowymi niezależnymi.

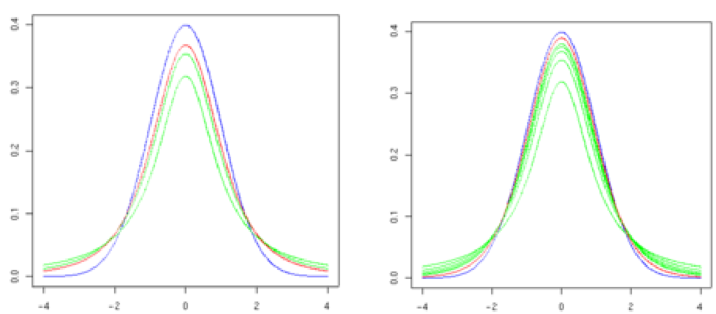
Zmienna losowa t określona powyżej ma gęstość prawdopodobieństwa opisaną wzorem:

$$f(t,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

gdzie $\Gamma(x)$ to funkcja gamma.



Rozkład T przy 2 (lewy) i 3 (prawy) stopniach swobody (niebieski - rozkład normalny)
Na zielono oznaczono rozkłady T o mniejszych stopniach swobody.



Rozkład T przy 5 (lewy) i 10 (prawy) stopniach swobody (niebieski - rozkład normalny)
Na zielono oznaczono rozkłady T o mniejszych stopniach swobody.

Dwa podstawowe twierdzenia związane z rozkładem t-Studenta:

1 Niech zmienne losowe $X_1, X_2, ..., X_n$ mają jednakowy rozkład prawdopodobieństwa, rozkład normalny N(m, σ^2), oraz niech zmienna t będzie określona wzorem:

$$t = \frac{\overline{X} - m}{s} \cdot \sqrt{n}$$

, gdzie $\overline{X}$ jest wartością średnią z próby (empiryczną), zaś s - odchyleniem standardowym z próby (empirycznym). Wówczas zmienna t ma rozkład t-Studenta o $\nu = n - 1$ stopniach swobody (niezależny od wartości wariancji w populacji σ^2).

2 Jeżeli dwie próby o liczebnościach n_1 oraz n_2, wartościach średnich $\overline{X}_1$ oraz $\overline{X}_2$ i wariancjach wyznaczonych z próby s_1^2 oraz s_2^2 zostały wylosowane z populacji mających taki sam rozkład normalny, to zmienna t określona wzorem:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}(n_1 + n_2 - 2)}$$

ma rozkład t-Studenta o $\nu = n_1 + n_2 - 2$ stopniach swobody.

Test T dla prób niezależnych

Test t jest najbardziej powszechnie stosowaną metodą oceny różnic między średnimi w dwóch grupach. Test ten został opracowany, żeby wychwycić różnice w wartościach średnich pomiędzy badaniami pewnego atrybutu (jego wartości) względem dwóch różnych grup, lub dwóch atrybutów dla pojedynczej grupy.

Teoretycznie test t może być stosowany także w przypadku bardzo małych prób (np. o liczności 10, zaś niektórzy badacze twierdzą, że nawet w mniej licznych są możliwe); jedynym warunkiem jest normalność rozkładu zmiennych oraz brak istotnych różnic między wariancjami.

Jak zostało to wspomniane wcześniej założenie o normalności można sprawdzić przez analizę rozkładu danych (np. histogramów, q-q plot, etc.) lub przy pomocy testu normalności.

Założenie o równości wariancji sprawdzamy za pomocą testu F lub też przy pomocy mocniejszej opcji określonej jako test Levene'a.

Jeżeli warunki, o których mowa nie są spełnione, wówczas alternatywą pozostaje użycie jednego z testów nieparametrycznych alternatywnych w stosunku do testu t.

Założenia

Podajemy tutaj tylko najważniejsze założenia:

1 rozkład wyników zmiennej zależnej w każdej grupie jest zbliżony do rozkładu normalnego. W celu sprawdzenia, czy wyniki mają rozkład normalny można zastosować test K-S (Kołmogorowa-Smirnowa) bądź test Shapiro-Wilka

2 porównywane grupy mają podobną liczebność (ilość badanych osób). Aby dokładnie sprawdzić, czy grupy są równoliczne (podobna liczbeność) przeprowadza się test chi-kwadrat zgodności. Jednakże badacze, w szczególności nauk społecznych przyjmują, że gdy jedna z grup nie przekracza liczebnością drugiej dwukrotnie, to można przyjąć, że grupy są równoliczne.

3 wariancje w porównywanych grupach są do siebie podobne. Aby sprawdzić te założenie stosuje się dodatkowy test, przykładowo test Levene'a.

Hipoteza

H0: $\mu 1 = \mu 2$,
H1: $\mu 1 \neq \mu 2$ ('both') - dwustronny
H1': $\mu 1 > \mu 2$ ('right') lub $\mu 1 < \mu 2$ ('left') - jednostronny

Test (link)

Wykorzystywana statystyka T, w liczniku ma różnicę wartości średnich, zaś w mianowniku ma błąd standardowy (estymowane odchylenie standardowe) różnicy pomiędzy wartościami średnimi prób. W konsekwencji możemy zauważyć, iż T rośnie wraz z oddalaniem się wartości średnich od siebie i przy wariancjach będących blisko siebie.

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{x}_1 - \bar{x}_2}} \qquad S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Prawdopodobieństwo uzyskania zaobserwowanej wartości T pod warunkiem zerowej hipotezy wyznaczane jest na podstawie rozkładu T, który zależy od przyjętego stopnia swobody (zobacz poprzedni punkt).

Zazwyczaj przyjmuje się, iż wykorzystywana liczba stopni swobody zależy od ilości analizowanych obserwacji - w szczególności przyjmuje się liczbę wszystkich obserwacji z grup minus 2 : n1 + n2 - 2.

Podawany w wynikach testu t poziom p reprezentuje prawdopodobieństwo błędu związanego z przyjęciem hipotezy o istnieniu różnic między średnimi. Jest to prawdopodobieństwo popełnienia błędu polegającego na odrzuceniu hipotezy o braku różnicy między średnimi w dwóch badanych kategoriach obserwacji należących do populacji generalnej (reprezentowanych przez badane grupy) w sytuacji, gdy stan faktyczny w populacji jest taki, iż hipoteza ta jest prawdziwa.

Niektórzy badacze uważają, że jeśli znak różnicy średnich jest zgodny z przewidywaniami, to można do testowania używać jedynie połowy (jednego "ogona") rozkładu prawdopodobieństwa i dzielić podawany poziom p wynikający z testu t (prawdopodobieństwo wyznaczone przez "obydwa ogony" rozkładu) przez dwa. Inni badacze uważają takie postępowanie za błędne i zalecają używać dwustronnego obszaru krytycznego.

Przykład w Matlabie (link)

Przeanalizujmy, czy wzrost studentów z dwóch grup zajęciowych, z godziny 13 i 17, jest znacząco rózny. Poniżej mamy dane, które wskazują, iż w pierwszej grupie średni wzrost to 169.051, zaś w drugiej 164.1475. Sprawdźmy, czy ta róznica jest faktycznie znacząca:

data13:
175.26,177.8,167.64000000000001,160.02,172.72,177.8,175.26,170.18,157.48,160.02,193.04,149.86,157.48,157.48,190.5,157.48,182.88,160.02
data17:
172.72,157.48,170.18,172.72,175.26,170.18,154.94,149.86,157.48,154.94,175.26,167.64000000000001,157.48,157.48,154.94,177.8

Mamy do dyspozycji zatem dwa atrybuty, jeden nominalny - grupa - oraz drugi ilościowy - wzrost. Zatem możemy przetestować hipotezę zerową, że wartości średnie wysokości studentów w obu grupach są takie same. W tym celu wykorzystamy funkcję matlaba 'ttest2':

```
[h,p,c,s] = ttest2(data13,data17)
```

otrzymujemy:

h = 0, p = 0.2067, c = -2.8464 12.6536, s = tstat: 1.2888 df: 32 sd: 11.0734

H=0 oznacza, że nie mamy podstaw do odrzucenia zerowej hipotezy (H=1 oznacza, że odrzucamy zerową hipotezę i akceptujemy H1), P=0.2067 to obliczona p-wartość, zaś C jest 95% przedziałem ufności dla różnicy wartości średnich.

Zatem pamiętając, iż zerowa hipoteza zakłada, że średni wzrost w obu grupach jest taki sam, rezultat testu nie pozwala na odrzucenie tejże hipotezy. Proszę zauważyć, że dane z 13 maja 18 obserwacji, zaś dane z 17 mają 16 obserwacji - co w sumie daje 34 obserwacje, a więc wykorzystujemy rozkład t o 32 stopniach swobody (stąd df=32).

Zadanie:

Przeanalizuj dodatkowe parametry dla funkcji 'ttest2', tzn. alpha, tail oraz vartype.

Załóżmy równą wariancję dla obserwacji tenperatury ciała dla kobiet i mężczyzn. Następnie sprawdźmy, korzystając z testu t przeanalizujmy, czy średnia temperatura ciała kobiety i mężczyzna jest taka sama:

```
[H,P,CI,STATS]=ttest2(men,women,0.05,'both','equal')
```

otrzymujemy:

H = 1, P = 0.0239, CI = [-0.5396, -0.0388], STATS = tstat: -2.2854, df: 128, sd: 0.7215.

Zatem możemy odrzucić hipotezę zerową, iż wartości średnie są sobie równe (przy poziomie istotności 0.05).

Następnie załóżmy, że wariancje w obu obserwacjach są różne i zastosować odpowiednią 'aproksymację' testu t (wersję zaimplementowaną w Matlabie):

```
[H,P,CI,STATS]=ttest2(men,women,0.05,'both','unequal')
```

otrzymujemy wówczas:

H = 1, P = 0.0239, CI =[-0.5396,-0.0388], STATS = tstat: -2.2854, df: 127.5103, sd: [2x1 double].

Zadanie:

Przeanalizuj wyniki dla ttestu2 przy założeniu nierówności wariancji ('unequal').

Zadanie:

Ludzie nerwowi mają więcej energii niż osoby spokojne i stonowane. Zastanówmy się zatem czy osoby nerwowe mają tendencję do widocznie silniejszej gestykulacji. W tym celu udało nam się zapytać 5 osób, które wskazują na wszelkie oznaki nerwowości, i zarejestrowaliśmy średnią ilość gestów jakie wykonują na minutę podczas swobodnej konwersacji. Następnie analogiczne obserwacje poczyniliśmy dla osób spokojnych. W ten sposób otrzymaliśmy następujące dane:

nerwowi = 3, 3, 4, 5, 5

spokojni = 4, 6, 7, 9, 9

Określ hipotezę jaką chcemy zweryfikować w naszym badaniu oraz określ hipotezę zerową dla testu ? Czy możemy zastosować test t, jeżeli tak to jaki jedno- czy dwustronny ? Jaki stopień swobody powinniśmy zastosować ? Przeprowadź test przy alpha = 0.05. Zinterpretuj pełny wynik analizy.

Zadanie:

Sprawdźmy hipotezę, iż osoby poniżej 30 są bardziej dowcipni niż osoby po 30. Sprawdźmy tę hipotozę na bazie wyników ankiety przeprowadzonej na 4 losowo wybranych osobach przed 30 i 4 po 30. Dla każdej osoby wyznaczamy współczynnik ich ‘rozbawienia’ na skali Likerta 1-10.

mniej30 = 6, 7, 10, 9

po30 = 5, 6, 2, 3

Określ hipotezę jaką chcemy zweryfikować w naszym badaniu oraz określ hipotezę zerową dla testu ? Czy możemy zastosować test t, jeżeli tak to jaki jedno- czy dwustronny ? Jaki stopień swobody powinniśmy zastosować ? Przeprowadź test przy alpha = 0.05. Zinterpretuj pełny wynik analizy.

Test T dla prób zależnych

W ogólności test ten można wykorzystać w celu weryfikacji czy wyniki danej grupy różnią się pomiędzy warunkami w jakich zostały one zebrane. Można go na przykład użyć do sprawdzenia różnicy w reakcji przeprowadzanej na grupie pacjentów poddanych działaniu jakiegoś leku w stosunku do tej samej grupy osób przed otrzymaniem leku. .

Zmienność wewnątrzgrupowa

Siła relacji pomiędzy dwiema zmiennymi, mierzona na przykład różnicą pomiędzy średnimi w dwóch grupach, zależy w dużej mierze od zmienności wartości wewnątrz grup. W zależności od tego, jak duża jest ta zmienność w obydwu grupach, taka sama co do "wartości różnica" między średnimi może wskazywać na silną lub słabą zależność pomiędzy zmienną zależną a niezależną (grupującą).

Na przykład jeśli średnia LBC (liczba białych ciałek) wynosi 102 u mężczyzn i 104 u kobiet, wówczas różnica "jedynie" 2 punkty musi zostać oceniona jako niezwykle istotna w przypadku, gdyby u wszystkich mężczyzn wyniki zawierały się w granicach 101 do 103, zaś u wszystkich kobiet w granicach 103 do 105. W takim wypadku moglibyśmy precyzyjnie przewidzieć wartość LBC przy pomocy zmiennej Płeć. Jeśliby jednakże taka sama różnica równa 2 została uzyskana w próbach o dużej zmienności (np. o zakresie zmienności 0 - 200), wówczas każdy skłonny byłby taką różnicę ocenić jako nieistotną. Możemy zakończyć następującą konkluzją: zmniejszenie zmienności wewnątrzgrupowej zwiększa czułość naszego testu.

Test t dla prób zależnych pozwala na wykorzystanie pewnego specyficznego typu układu eksperymentalnego, w którym ważne źródło zmienności wewnątrzgrupowej (lub tzw. błędu) może zostać łatwo zidentyfikowane i wykluczone z analizy. W szczególności, jeśli dwie grupy obserwacji (które mają zostać porównane) zostały oparte na tej samej grupie obiektów zmierzonych dwukrotnie (np. przed i po zabiegu), to wówczas znaczna część zmienności wewnątrzgrupowej w obydwu grupach wyników może zostać przypisana początkowej indywidualnej różnicy pomiędzy obiektami. Zauważmy, że w pewnym sensie fakt ten jest podobny do sytuacji, kiedy obydwie grupy są całkowicie niezależne i indywidualne różnice również wnoszą wkład do składnika błędu. W tym ostatnim przypadku nie możemy jednak zidentyfikować (lub, inaczej mówiąc," wyeliminować") wariancji pochodzącej od indywidualnych różnic poszczególnych obiektów.

Jeśli jednak ta sama próba została zmierzona dwukrotnie, to wariancję tę łatwo jest zidentyfikować ("wyeliminować"). Można w szczególności, zamiast analizować oddzielnie każdy z pomiarów, brać do analizy różnice pomiędzy wynikami,(przed i po) dla każdego z obiektów pomiarowych. Przez procedurę odejmowania wyniku przed zabiegiem od wyniku po zabiegu i analizowaniu "czystych różnic" dokonujemy wyeliminowania tej części wariancji w naszym zbiorze danych, która pochodzi od różnic w wartościach bezwzględnych poszczególnych obiektów pomiarowych. Dokładnie tak przebiega procedura w opcji test t dla prób zależnych i w porównaniu z testem t dla prób niezależnych daje ona zawsze "lepsze" wyniki (w tym sensie, że jest bardziej czuła).

Założenia teoretyczne występujące w przypadku testu testu t dla prób niezależnych mają również zastosowanie w teście dla prób zależnych; to znaczy różnice pomiędzy parami pomiarów powinny mieć rozkład normalny. Jeśli założenie to jest zdecydowanie niespełnione, wówczas należy zastosować jeden z alternatywnych testów nieparametrycznych.

Hipoteza

H0: μ1 - μ2 = 0,
H1: μ1 - μ2 ≠ 0 (dwustronny)
H1’: μ1 - μ2 > 0 lub μ1 - μ2 < 0 (jednostronny)

Test

Wykorzystywana statystyka T, w liczniku ma średnią różnicę wartości próby, zaś w mianowniku ma błąd standardowy (estymowane odchylenie standardowe) obliczone jako nieobciążona średnia różnica pomiędzy różnicą wartości a średnią różnicą wartości próby. W konsekwencji możemy zauważyć, iż T rośnie wraz z oddalaniem się wartości średnich od siebie i przy ‘estymowanych wariancjach’ będących blisko siebie.

$$T = \frac{\overline{D}}{S_D/\sqrt{n}} \qquad \overline{D} = \frac{1}{n}\sum_{i=1}^{n}\left(x_{1i} - x_{2i}\right) \qquad S_D = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{1i} - x_{2i} - \overline{D}\right)^2}$$

Prawdopodobieństwo uzyskania zaobserwowanej wartości T pod warunkiem zerowej hipotezy wyznaczane jest na podstawie rozkładu T, który zależy od przyjętego stopnia swobody (zobacz poprzedni punkt).

Podawany w wynikach testu t poziom p reprezentuje prawdopodobieństwo błędu związanego z przyjęciem hipotezy o istnieniu różnic między średnimi. Jest to prawdopodobieństwo popełnienia błędu polegającego na odrzuceniu hipotezy o braku różnicy między średnimi w dwóch badanych kategoriach obserwacji należących do populacji generalnej (reprezentowanych przez badane grupy) w sytuacji, gdy stan faktyczny w populacji jest taki, iż hipoteza ta jest prawdziwa.

Przykład w Matlabie (link)

Przeanalizujmy, czy przeprowadzony przez nas kurs szybkiego pisania na klawiaturze. Przeprowadziliśmy badanie szybkości pisania na klawiaturze przed jak i po kursie. Poniżej mamy dane 16 uczestników naszego kursu. Dane te wskazują, iż po zakończeniu kursu średnia liczba słów na minutę wzrosła do 68.3512, zaś przed kursem wynosiła jedynie 63.7475. Sprawdźmy, czy ta róznica jest faktycznie znacząca:

dataAfter:
74.86,77.4,67.24,59.62,72.32,77.4,74.86,69.78,57.08,59.62,92.64,49.46,57.08,57.08,90.1,57.08
dataBefore:
72.32,57.08,69.78,72.32,74.86,69.78,54.54,49.46,57.08,54.54,74.86,67.24,57.08,57.08,54.54,77,4

Mamy do dyspozycji zatem dwa zbiory danych. Zatem możemy przetestować hipotezę zerową, że wartości średnie w obu grupach są takie same. W tym celu wykorzystamy funkcję Matlaba 'ttest':

```
[h,p,c,s] = ttest(dataAfter,dataBefore)
```

otrzymujemy:

h = 0, p = 0.2450, c = -3.5053   12.7128, s = tstat: 1.2101 df: 15 sd: 15.2179

H=0 oznacza, że nie mamy podstaw do odrzucenia zerowej hipotezy (H=1 oznacza, że odrzucamy zerową hipotezę i akceptujemy H1), P=0.2450 to obliczona p-wartość, zaś C jest 95% przedziałem ufności dla różnicy wartości średnich.

Zweryfikujmy dodatkowo jako alternatywną hipotezę, że średnia (wśród uczestników) liczba słów na minutę zwiększyła się po ukończonym kursie:

```
[h,p,c,s] = ttest(dataAfter,dataBefore,0.05,'right')
```

otrzymujemy:

h = 0, p = 0.1225, c = -2.0657   Inf, s = tstat: 1.2101 df: 15 sd: 15.2179

Zatem pamiętając, iż zerowa hipoteza zakłada, że średnia liczba słów na minutę w obu grupach (przed i po kursie) jest taka sama, rezultat testu nie pozwala na odrzucenie tej hipotezy.

Po pierwsze wykorzystamy dane odnośnie temperatury ciała zmierzonej wśród kobiet i mężczyzn. Możemy wykorzystać test t dla prób zależnych ponieważ liczba obserwacji w obu zbiorach jest taka sama i ponieważ możemy obserwacje pogrupować, tak że Zi = Xi - Yi (gdzie równice te mają rozkład normalny). Zatem zakładamy hipotezę zerową, iż wartości średnie są sobie równe, a więc μx = μy (μz = 0), i możemy zastosować test t:

```
[H,P,CI,STATS]=ttest(men,women,0.05,'both')
```

otrzymujemy:

H = 1, P = 3.9773e-019, CI = [-0.3348,-0.2437] STATS = tstat: -12.6858, df: 64, sd: 0.1838.

Zatem możemy odrzucić hipotezę zerową - w szczególności niska wartość p (rzędu 10-19) jest silnym dowodem przeciw hipotezie zerowej.

Jednakże, analizując dane tak silny wynik wydaje się być bardzo specyficznym rezultatem (Patrz zadanie poniżej).

Zadanie:

Przeprowadź test t dla prób niezależnych i przeanalizuj wynik w kontekście wyniku dla testu t dla prób zależnych.

Przeanalizuj dokładnie dane i zastanów się, czy spełnione są wszystkie założenia dla przeprowadzenia testu t dla prób zależnych ? Jeżeli tak to zinterpretuj wyniki, jeżeli nie to zastanów się jak poprawić dane, żeby móc zastosować test t dla prób zależnych.

Test T dla jednej próby

Założenia

Podajemy tutaj tylko najważniejsze założenia:

1 rozkład wyników zmiennej zależnej jest zbliżony do rozkładu normalnego.

Hipoteza

H0: μ1 = μ,
H1: μ1 ≠ μ (dwustronny)
H1': μ1 > μ lub μ1 < μ (jednostronny)

Test

Wykorzystywana statystyka T, w liczniku ma różnicę wartości średniej próby (empiryczną wartość średnią) od zadanej średnij μ, zaś w mianowniku ma estymowane odchylenie standardowe obliczone (nieobciążone) obliczone z próby. W konsekwencji możemy zauważyć, iż T rośnie wraz z oddalaniem się wartości średnich od siebie i przy 'estymowanych wariancjach' będących blisko siebie.

$$T = \frac{\overline{X_1} - \mu}{Sx}$$

Prawdopodobieństwo uzyskania zaobserwowanej wartości T pod warunkiem zerowej hipotezy wyznaczane jest na podstawie rozkładu T, który zależy od przyjętego stopnia swobody (zobacz poprzedni punkt).

Podawany w wynikach testu t poziom p reprezentuje prawdopodobieństwo błędu związanego z przyjęciem hipotezy o istnieniu różnic między średnimi. Jest to prawdopodobieństwo popełnienia błędu polegającego na odrzuceniu hipotezy o braku różnicy między średnimi w dwóch badanych kategoriach obserwacji należących do populacji generalnej (reprezentowanych przez badane grupy) w sytuacji, gdy stan faktyczny w populacji jest taki, iż hipoteza ta jest prawdziwa.

Przykład w Matlabie (link)

Wykorzystajmy przykład temperatury ciała człowieka, zakładając że wczytane dane są zapisane w zmiennej 'normtemp' (dane zarejestronwanej temperatury ciała 130 osób). Zbadajmy teraz czy średnia temperatura ciała jest równa do 98.6:

[H,P,CI,STATS] = ttest(normtemp,98.6,0.05,'both')

Otrzymujemy:

H = 1, P = 2.4106e-007, CI = [98.1220, 98.3765], STATS = tstat: -5.4548, df: 129, sd: 0.7332.

Zatem μ = 98,6 przy poziomie istotności 0.05 pozwala na odrzucenie hipotezy zerowej (o równości wartości średnich) na rzecz hipotezy alternatywnej (nierówności średnich - 'both'). H=1 oznacza, że odrzucamy zerową hipotezę i akceptujemy H1 (H=0 oznacza, że nie mamy podstaw do odrzucenia zerowej hipotezy), zaś C jest 95% przedziałem ufności dla różnicy wartości średniej od μ=98.6.

Jeżeli chcemy zweryfikować hipotezę, że wartość średnia jest blisko 98.6 wówczas wykorzytujemy:

[H,P,CI,STATS] = ttest(normtemp(1:130,1),98.6,0.05,'left')

co daje nam:

H = 1, P = 1.2053e-007, CI = [-Inf, 98.3558], STATS = tstat: -5.4548, df: 129, sd: 0.7332.

Fakt, że ponownie odrzucamy hipotezę zerową oznacza, iż istnieją silne przesłanki przeciw niej.

Ćwiczenie

Przeanalizujmy, czy średni wzrost studentów dla grupy z godziny 13 jest 169.051:

data13:
175.26,177.8,167.64000000000001,160.02,172.72,177.8,175.26,170.18,157.48,160.02,193.04,149.86,157.48,157.48,190.5,157.48,182.88,160.02

Ćwiczenie

Przeanalizujmy, czy średni wzrost studentów dla grupy z godziny 17 jest 164.1475:

data17:
172.72,157.48,170.18,172.72,175.26,170.18,154.94,149.86,157.48,154.94,175.26,167.64000000000001,157.48,157.48,154.94,177.8

Mann-Whitney U test (link)

Test U Manna-Whitneya jest jedną z najpopularniejszych alternatyw dla testu t-Studenta dla prób niezależnych. Jeżeli dane nie spełniają założeń dla zastosowania testu t-Studenta, a wobec czego, nie możemy go zastosować, warto skorzystać z testu U Manna-Whitneya, gdy chcemy porównać ze sobą dwie niezależne wobec siebie grupy. Podstawową zaletą tego testu są niewielkie wymogi, założenia do zastosowania testu U Manna-Whitneya.

Zmienna zależna musi być mierzona na skali co najmniej porządkowej (może być również mierzona na skali ilościowej). Jest to podstawowy warunek dla zastosowania tego testu. Co ciekawe, z tego testu możemy skorzystać również, gdy zmienna jest mierzona na skali dychotomicznej (czyli 0-1), dlatego, że jest to przypadek zmiennej nominalnej, która jest zarazem zmienną porządkową. Zastosowanie testu U Manna-whitneya nie wymaga równoliczności grup, rozkładu normalnego czy też homogenicznych wariancji. To sprawia, że może być on szeroko stosowany.

Test U Manna-Whitneya polega na rangowaniu wyników zmiennej zależnej (od najmniejszej do największej) w badanych grupach, a następnie grupy są ze sobą porównywane.

W statystyce, test U Manna–Whitneya (zwany również testem Manna-Whitneya-Wilcoxona lub czasem testem sumy rang Wilcoxona dla dwu próbek, albo po prostu testem Wilcoxona dla dwóch próbek) jest nieparametrycznym testem do sprawdzenia czy wartości próbek pobranych z dwóch niezależnych populacji są jednakowo duże. Jest jednym z najbardziej popularnych nieparametrycznych testów znamienności. Zaproponowany pierwotnie jako test przesunięcia dla dwóch równolicznych próbek przez Franka Wilcoxona w 1945, uogólniony następnie przez H. B. Manna i D. R. Whitneya (1947) dla przypadku różnolicznych próbek oraz do testowania równości stochastycznej.

Założenia

W ogólnym przypadku zakłada się, że:

1. wszystkie obserwacje, dla obydwu grup, są niezależne statystycznie,
2. zmienne X i Y mierzone są na skali porządkowej, a więc dla dowolnej pary obserwacji, można określić ich uporządkowanie: stwierdzić ich równość lub wskazać na większą spośród nich.
3. testowaną hipotezą zerową jest symetria względem prawdopodobieństwa większej wartości jednej ze zmiennych, a więc hipoteza zerowa zakłada jednakowe prawdopodobieństwo X > Y i Y > X: $P(X > Y)=P(Y > X)$.
4. hipotezą alternatywną jest asymetria względem prawdopodobieństwa większej wartości jednej ze zmiennych, a więc w wersji dwustronnej testu, że prawdopodobieństwo X > Yjest różne od prawdopodobieństwa Y > X (w wersji jednostronnej, hipotezą alternatywną jest $P(X > Y) > P(Y > X)$ lub $P(X > Y) < P(Y > X)$).

Test Manna-Whitneya w ogólnej wersji, sformułowanej powyżej, ma słabe założenia i może być stosowany w bardzo szerokim zakresie. W literaturze jednak często zdarza się, że hipotezy zerowa i alternatywna formułowane są w sposób mniej ogólny, stosowalność testu Manna-Whitneya do testowania tych hipotez jest wówczas ograniczona. Tak więc czasem zakłada się, że wg hipotezy zerowej obydwie zmienne X i Y mają identyczny rozkład, a wg hipotezy alternatywnej ich rozkłady są jednakowe z wyjątkiem przesunięcia o stałą δ (tzn. $F1(x) = F2(x + \delta)$). Taka interpretacja testu Manna-Whitneya nazywa się czasem testowaniem przesunięcia. W praktyce zwykle zmienne różnić się mogą nie tylko pod względem przesunięcia ale i innych parametrów np. wariancji, wówczas stosowalność tak interpretowanego testu Manna-Whitneya jest wątpliwa, a test może mieć małą moc dla odrzucenia hipotezy zerowej. Np. test nie jest w stanie odrzucić hipotezy zerowej o identyczności rozkładów gdy X i Y są zmiennymi normalnymi o tej samej średniej ale różnych wariancjach.

Innym dodatkowym założeniem dodawanym czasem dla wygody obliczeniowej jest brak obserwacji związanych: tzn. zerowe prawdopodobieństwo otrzymania dwu obserwacji o jednakowej wartości. Jednak założenie braku obserwacji związanych nie jest konieczne, test Manna-Whitneya stosować można w obecności obserwacji związanych używając poprawek obliczeniowych na obserwacje związane lub przeprowadzając obliczenia przy pomocy testu permutacji.

Hipotezy analitycznie w ogólności najłatwiej wyrazić za pomocą tzw. stochastycznej równości (bądź nierówności) dystrybuant. Nie będziemy jej tutaj dodatkowo definiować i pozostaniemy przy oryginalnej reprezentacji.

Hipoteza

H0: $P(X > Y)=P(Y > X)$,
H1: $P(X > Y)\neq P(Y > X)$ (dwustronny)
H1’: $P(X > Y)>P(Y > X)$ lub $P(X > Y)<P(Y > X)$ (jednostronny)

Test

Test wymaga obliczenia statystyki, zwykle oznaczonej U, której rozkład jest znany w przypadku hipotezy zerowej. W przypadku małych próbek, rozkład statystyki U jest tabularyzowany, ale dla próbek powyżej ~20 obserwacji znane są dobre przybliżenia oparte na rozkładzie normalnym. Czasem w miejsce statystyki U używa się równoważnej statystyki, np. sumy rang jednej z próbek.

Statystyka U określa liczbę wszystkich par obserwacji (x,y) (x jest obserwacją z pierwszej próbki, y jest obserwacją z drugiej próbki) takich, że $x > y$. W wypadku obecności obserwacji związanych do U dodaje się również połowę liczby par (x,y) takich że $x=y$.

Wzór testu U Manna-Whitneya ma postać:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R$$
,

gdzie R oznacza sumę rang n1, n2 oznacza liczebność w badanych grupach

Należy podkreślić, że należy obliczyć statystykę U zarówno dla R1 (suma rang w I grupie) jak i dla R2 (suma rang w II grupie). Mniejsza z dwóch wartości U stanowi statystykę U., a istotność statystyczna odczytywana jest z tabel.

Dalej, dla próby większej 20, stosuje się inny wzór, z założeniem, że dla próby większej niż 20, rozkład U jest w przybliżeniu normalny. Wzór ten ma postać:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 [n_1 + (n_2+1)]}{12}}}$$

Przykład w Matlabie (link)

Wykorzystamy dane dotyczące procentowej aktywności dobowej koali zapisane w pliku koala.mat. Dane przedstawiają dwie niezależne próby pochodzące z Grampianów i z Wilsons Promontory. Należy sprawdzić, czy wartości próbek pobranych z obu populacji są jednakowo duże. Wczytujemy dane

```
load('koala.mat')
```

Rozkłady danych nie są zbliżone do rozkładu normalnego i decydujemy się na test U Manna-Whitneya. Sprawdziliśmy, że zarówno różnica średnich arytmetycznych, jak i median wynosi ok. 3. Sprawdzimy, czy jest to istotna różnica i czy potwierdzi się to w teście.

Stawiamy hipotezę zerową, że próbki z obu populacji są jednowko duże, przeciwko hipotezie alternatywnej, że nie są jednakowo duże. Uruchamiamy test dla obu populacji

```
mwwtest(koala_grampians,koala_wilsonsprom);
```

i otrzymujemy następujące wyniki:

```
-----------------------------------------------------------------------------------------
Sample size is good enough to use the normal distribution approximation
 T              U              mT              sT              zT        p-value (1-tailed)
-----------------------------------------------------------------------------------------
14576.0000     6701.0000      15687.5000      571.6841        1.9434          0.0260
-----------------------------------------------------------------------------------------
```

Pierwszy wynik to statystyka Mann-Whitney-Wilcon równa sumie rang elementów policzona dla krótszej z serii danych (zob. tiedrank). Drugi wynik to wartość statystyki U, wyznaczanej przez Matlaba jako U=T-n*(n+1)/2, gdzie n to mniejsza z liczności poszczególnych próbek. L próbek jest wystarczająca do tego, żeby wykorzystać rozkład normalny do aproksymacji rozkładu

Mann-Whitney-Wilcoxon. Kolejne trzy wartości, tj. mT, sT, zT, to wyliczane na te potrzeby parametry. Ostatecznie wyświetlana jest wartość p wyliczona jako

```
p=1-normcdf(zT);
```

Prawdopodobieństwo wystąpienia takiej zależności między próbami przy prawdziwości hipotezy zerowej wynosi 0.0260 i jest mniejsza od standardowo przyjmowanego przez nas poziomu istotności 0.05, co powoduje, że mamy podstawy do odrzucenia hipotezy zerowej i przyjęcia hipotezy alternatywnej.

Zadanie:
Zastanówmy się ponownie czy osoby nerwowe mają tendencję do widocznie silniejszej gestykulacji. Wykorzystajmy wcześniejsze dane o ludziach spokojnych i nerwowych:

nerwowi = 3, 3, 4, 5, 5
spokojni = 4, 6, 7, 9, 9

Określ hipotezę jaką chcemy zweryfikować w naszym badaniu oraz określ hipotezę zerową dla testu ? Czy możemy zastosować test U? Przeprowadź test przy alpha = 0.05. Zinterpretuj pełny wynik analizy.

Zadanie
W celu analizy wzrostu studentów różnych grup zajęciowych pozyskaliśmy następujące próbki:

data13:
175.26,177.8,167.64,160.02,172.72,177.8,175.26,170.18,157.48,160.02,193.04,149.86,157.48,157.48,190.5,157.48,182.88,160.02
data17:
172.72,157.48,170.18,172.72,175.26,170.18,154.94,149.86,157.48,154.94,175.26,167.64,157.48,157.48,154.94,177.8

Zbadaj z pomocą testu U, czy wzrost studentów z jednej grupy jest równie duży jak wzrost studentów drugiej grupy. Sformułuj hipotezy i przeprowadź test przy alpha = 0.05. Zinterpretuj pełny wynik analizy.