

Opracowanie danych

1. Opis zagadnienia:

Np. badamy ceny mleka „łaciego” w małych sklepach i supermarketach.

2. Zestawienie danych: obliczenie średniej i wariancji / odchylenia standardowego w próbach:

Zbieramy dane X_1 w $n_1 = 25$ małych sklepach i X_2 w $n_2 = 10$ supermarketach. Obliczamy $\bar{X}_1$, $\bar{X}_2$, s_1 i s_2 w próbach.

3. Przedziały ufności liczone dla średniej ceny:

95%-przedział ufności $\bar{X}_1 \pm 1.96 \frac{s_1}{\sqrt{n_1}}$ gdy próba jest duża lub liczony za pomocą wartości krytycznych rozkładu t -Studenta ($df = n_1 - 1$) gdy próba jest mała tak jak w naszym przykładzie. To samo dla $\bar{X}_2$.

99%-przedział ufności – tjw.

Testowanie hipotez

4. Hipoteza zerowa

Ceny w małych sklepach i supermarketach są takie same.

5. Hipoteza alternatywna

Ceny w supermarketach są niższe.

6. Zmienna losowa $Y = \bar{X}_1 - \bar{X}_2$

Różnica średnich z zebranych danych. Zmienna ma rozkład t -Studenta z $df = n_1 + n_2 - 2 = 33$ stopniami swobody, lub jeśli danych jest dużo ma rozkład normalny.

7. Estymacja odchylenia standardowego σ_Y :

1) zakładamy, że nie ma różnicy między wariancjami w populacjach

$$\sigma_Y = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

2) nie zakładamy, że nie ma różnicy między wariancjami w populacjach

$$\sigma_Y = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

8. Ustalamy obszar krytyczny $t = \frac{Y}{\sigma_Y}$ dla różnych poziomów istotności (odczytujemy z tablicy wartości krytyczne):

W naszym przykładzie stosujemy test jednostronny np dla poziomu istotności 5% i 1%.

9. Podejmujemy decyzję o odrzuceniu lub nie hipotezy zerowej.

10. W przypadku odrzucenia hipotezy zerowej szacujemy błąd I rodzaju:

Szukamy najmniejszego poziomu istotności, przy którym odrzucimy hipotezę zerową

11. Dla porównania testujemy hipotezę zakładając, że Y ma rozkład normalny i obliczamy błąd I rodzaju w przypadku odrzucenia hipotezy zerowej.

Przy dużej ilości danych powinniśmy otrzymać zbliżony obszar krytyczny, a błąd I rodzaju będzie można obliczyć dużo dokładniej.