

1. Postanowiono wyznaczyć faktyczny zasięg reklamy pralin w czasie jej pierwszej emisji. W tym celu określono grupę docelową: osoby w wieku od 10 do 30 lat. Z grupy tej wylosowano 1000 osób. Uzyskano wynik: reklama dotarła do 40% wylosowanych osób. Na poziomie ufności 95% oszacować faktyczny zasięg badanej reklamy w skali całej grupy docelowej. Zinterpretować wynik.

Rozwiązanie:

Populacja: grupa docelowa.

Próba: 1000 osób wylosowanych z populacji.

Cecha losowa Y : liczba osób w próbie, które miały styczność z reklamą.

Z treści zadania wynika, że $Y \sim B(1000, p)$, gdzie p oznacza prawdopodobieństwo wylosowania z populacji osoby, która miała styczność z reklamą.

Cel zadania: oszacować p na poziomie ufności 95%.

Do obliczeń wykorzystam wzór: $\hat{p} \pm u_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$.

Pod podany wzór podstawiam $\hat{p} = 0.40$, $n = 1000$ oraz $u_{1-\alpha/2} = u_{0.975} = 1.96$, otrzymując (z dokładnością do dwóch miejsc po przecinku) wynik: 0.40 ± 0.03 .

Wniosek: Faktyczny zasięg badanej reklamy (odsetek osób w grupie docelowej, które miały styczność z reklamą) wynosi co najmniej 37%, ale nie więcej niż 43%.

2. Każdy polityk wie, że nie należy zbyt długo odpowiadać na postawione przez dziennikarza pytanie, ponieważ można stracić wiarygodność, nie mówiąc już o tym, że długa wypowiedź może nie zostać zarejestrowana w umysłach wyborców. Zarejestrowano czas losowo wybranych wypowiedzi pewnego polityka. Uzyskano wyniki (w sekundach): 15, 10, 4, 7, 11, 12, 8. Na poziomie ufności 0.95 oszacować przeciętny czas wypowiedzi tego polityka.

Rozwiązanie:

Populacja: odpowiedzi polityka w sytuacjach, w których pytania zadawali dziennikarze.

Cecha losowa X : czas losowo wybranej wypowiedzi.

Założenie: zadanie zostanie rozwiązane przy założeniu, że czas wypowiedzi ma rozkład $N(\mu, \sigma^2)$.

Cel: oszacować μ (średni czas wypowiedzi) na poziomie ufności $1 - \alpha = 0.95$.

Zastosuję wzór: $\bar{x} \pm t(\alpha, n-1)s/\sqrt{n}$.

W wyniku obliczeń otrzymuję przedział (z dokładnością do sekundy): $< 6, 13 >$.

Wniosek: Przeciętny czas wypowiedzi wynosi co najmniej 6 sekund, ale nie więcej niż 13 sekund.

3. Na tysiąc wylosowanych studentów płci żeńskiej 348 miało styczność z dopalaczami, a na 915 studentów płci męskiej 315. Czy można uznać, że procent studentów, którzy mieli styczność z dopalaczami zależy od płci?

Rozwiązanie:

Populacje: pierwsza – studenci płci żeńskiej, druga - studenci płci męskiej.

Cecha losowa Y_1 : liczba osób, która miała styczność z dopalaczami (wśród osób wylosowanych z populacji pierwszej).

Cecha losowa Y_2 : liczba osób, która miała styczność z dopalaczami (wśród osób wylosowanych z populacji drugiej).

Zatem $Y_1 \sim B(1000, p_1)$ oraz $Y_2 \sim B(915, p_2)$.

Cel: Weryfikacja hipotezy $H_0 : p_1 = p_2$ (procent studentów, którzy mieli styczność z dopalaczami nie zależy od płci, tzn. jest identyczny w obu populacjach)

Zadaję poziom istotności $\alpha = 0.05$

Wyznaczam wartość

$$u_{emp} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \approx \frac{0.348 - 0.344}{\sqrt{0.346(1-0.346)(\frac{1}{1000} + \frac{1}{915})}} \approx 0.172$$

Weryfikacja hipotezy: Ponieważ $|u_{emp}| \approx 0.172 < u_{0.975} = 1.96$, brak podstaw do odrzucenia hipotezy.

Wniosek: Nie można uznać, że procent studentów, którzy mieli styczność z dopalaczami zależy od płci.

4. W ankiecie rozesłanej wśród pracowników pewnego konsorcjum pytano, czy chcieliby zmienić obecne miejsce pracy. Uzyskano następujące wyniki. Czy chęć zmiany pracy zależy od aktualnych zarobków?

Zarobek aktualny	Odpowiedź	
	Tak	Nie
500–700	46	62
700–900	94	146
900–1100	249	501
1100–1300	126	326
1300–1500	43	135
1500–1700	26	70

Rozwiązanie:

Populacja: pracownicy pewnego konsorcjum.

Cechy: Zarobek (sześć kategorii), Chęć zmiany pracy (dwie kategorie).

Hipoteza H_0 : Chęć zmiany pracy nie zależy od wielkości zarobków.

Do weryfikacji tej hipotezy wybieram test chi-kwadrat na niezależność.

Zadaję poziom istotności $\alpha = 0.05$ i wyznaczam wartość statystyki $\chi_{emp}^2 = \sum_{i,j} \frac{(n_{ij} - n_{ij}^t)^2}{n_{ij}^t} = 20.553$

Weryfikacja: Ponieważ $\chi_{emp}^2 > \chi(0.05, 5)$, hipotezę odrzucam.

Wniosek: Chęć zmiany pracy zależy od wielkości zarobków. Oznacza to, że frakcja tych, którzy chcą zmienić pracę jest inna przy co najmniej dwóch poziomach zarobków. Jeżeli chcemy zbadać ten problem dokładniej, należy wyznaczyć reszty poprawione.

5. Badano wpływ koloru papieru na czas wypełniania pewnej ankiety. Do eksperymentu wylosowano 40 osób. Uzyskane wyniki znajdują się w pliku *kolor.wk1*. Jeżeli w wyniku zastosowania odpowiedniej analizy statystycznej okaże się, że kolor papieru ma wpływ na czas wypełniania ankiet, wykonać procedurę porównań szczegółowych i sformułować odpowiednie wnioski.

Rozwiązanie:

Populacja: zbiorowość osób.

Cecha: Czas wypełniania ankiety.

Czynnik: Kolor papieru.

Oznaczenia:

Y_N – czas wypełniania ankiety koloru niebieskiego,

Y_B – czas wypełniania ankiety koloru białego,

Y_Z – czas wypełniania ankiety koloru zielonego,

Y_S – czas wypełniania ankiety koloru szarego,

Założenia: $Y_N \sim N(\mu_N, \sigma^2)$, $Y_B \sim N(\mu_B, \sigma^2)$, $Y_Z \sim N(\mu_Z, \sigma^2)$, $Y_S \sim N(\mu_S, \sigma^2)$.

Stawiam hipotezę, że kolor papieru nie wpływa na czas wypełniania ankiet:

$$H_0: \mu_N = \mu_B = \mu_Z = \mu_S$$

Zadaję poziom istotności $\alpha = 0.05$ i weryfikuję tę hipotezę za pomocą testu jednoczynnikowej analizy wariancji. Wartość odpowiedniej statystyki wynosi:

$$F_{emp} = 6.12.$$

Weryfikacja hipotezy: Ponieważ $F_{emp} > F(0.05, 3, 36)$, hipotezę odrzucam.

Wniosek: czas wypełniania ankiety zależy od koloru papieru. Aby określić, który kolor sprzyja szybkiemu wypełnianiu ankiety, a który nie sprzyja, przeprowadzę procedurę porównań szczegółowych Tukey’a (na poziomie ufności $(1 - \alpha) = 0.95$). Najmniejsza istotna różnica dla tej procedury wynosi:

$$NIR = 2.85554.$$

Ponieważ $\bar{y}_N = 14.3$, $\bar{y}_B = 15.61$, $\bar{y}_Z = 17.64$, $\bar{y}_S = 18.34$, procedura Tukey’a daje nam dwie grupy jednorodne: $\{\mu_N, \mu_B\}$ oraz $\{\mu_B, \mu_Z, \mu_S\}$. Zatem kolor niebieski skraca czas wypełniania ankiet w porównaniu z kolorami zielonym i szarym. Problem, czy jest efekt koloru białego jest nierozstrzygnięty.