F11, Hypotesprövning (en population)

Christian Tallberg

Statistiska institutionen

Stockholms universitet

Generellt ser en testvariabel ut på följande vis om punktskattningen (estimatorn) är normalfördelad (eller approximativt normalfördelad):

$$\text{testvariabel} = \frac{\text{punktskattning - nollhypotesens värde}}{\text{standardavvikelse för skattningen}}$$

(eller skattade standardavvikelsen för skattningen)

Normalfördelade variabler med okänd varians

Om $X_1, ..., X_n$ är ett stickprov av oberoende observationer från

$$X \sim N\left(\mu; \sigma^2\right)$$

så gäller att

$$\frac{\bar{X} - \mu}{\frac{\sigma}{sqrt(n)}} \sim N(0; 1)$$

Om σ^2 är okänd skattas den med s^2. Då gäller att

$$\frac{\bar{X} - \mu}{\frac{s}{sqrt(n)}} \sim t(n-1),$$

och testvariabeln ges då av

$$\frac{\bar{X} - \mu_0}{\frac{s}{sqrt(n)}} \sim t(n-1) \text{ då } H_0 \text{ är sann.}$$

Exempel 10 (Körner sid 203):

$X =$ Nikotininnehåll i cigaretter. En cigarettillverkare hävdar att hans cigaretter i genomsnitt innehåller mindre än 25 mg nikotin. Vi vill därför testa

$$\begin{aligned} H_0 &: \mu = 25 \\ H_1 &: \mu < 25 \end{aligned}$$

Låt nikotininnehåll vara normalfördelade variabler med okänt μ och okänt σ. Vi drar ett slumpmässigt stickprov av storlek $n = 15$ där $\bar{x} = 24.1$ mg och $s = 2$ mg. Ger data stöd åt hans påstående?

Testvariabeln är

$$\frac{\bar{X}-\mu_0}{\frac{s}{sqrt(n)}}, \text{ som är } t\,(n-1) \text{ då } H_0 \text{ är sann.}$$

Vi beräknar det observerade testvärdet i stickprovet

$$t_{obs} = \frac{\bar{x}-\mu_0}{\frac{s}{sqrt(n)}} = \frac{24.1-25}{\frac{2}{sqrt(15)}} = -1.74$$

Då p-värdet > 0.05 kan H_0 ej förkastas. Det kan ej anses statistiskt påvisat att genomsnittliga nikotininnehållet är mindre än 25 mg för $\alpha = 0.05$.

Stora stickprov

Om fördelningen för X är okänd, utnyttjar vi att fördelningen för $\bar{X}$ är approximativt normalfördelad om n är stort ($n \geq 30$). Dvs

$$\bar{X} \sim approx\ N\left(\mu; \frac{\sigma^2}{n}\right) \text{ enligt CGS.}$$

Om populationsvariansen σ^2 också är okänd skattas den med stickprovsvariansen s^2. Då gäller att

$$\frac{\bar{X}-\mu}{\frac{s}{sqrt(n)}} \sim approx\ N\,(0;1) \text{ enligt CGS,}$$

och testvariabeln ges då av

$$\frac{\bar{X}-\mu_0}{\frac{s}{sqrt(n)}} \sim approx\ N\,(0;1) \text{ enligt CGS då } H_0 \text{ är sann.}$$

Exempel:

Facket påstår att den genomsnittliga månadslönen för anställda i en viss sektor är 13500 kr. Du misstänker att den genomstittliga lönen är högre än så. Därför vill du testa

$$\begin{aligned} H_0 &: \mu = 13500 \\ H_1 &: \mu > 13500. \end{aligned}$$

Du drar ett slumpmässigt stickprov av storlek $n = 70$ och finner att medelvärdet och standardavvikelsen är $\bar{x} = 14100$ kr och $s = 1900$ kr. Ger data stöd åt din misstanke?

Testvariabeln

$$\frac{\bar{X}-\mu_0}{\frac{s}{sqrt(n)}} \sim approx\ N\,(0;1) \text{ enligt CGS då } H_0 \text{ är sann.}$$

Det observerade testvärdet i stickprovet är

$$z_{obs} = \frac{\bar{x}-\mu_0}{\frac{s}{sqrt(n)}} = \frac{14100-13500}{\frac{1900}{sqrt(70)}} = 2.64$$

Då p-värdet < 0.005 kan H_0 förkastas. Det kan anses statistiskt påvisat att genomsnittliga lönen är större än 13500 kr för alla $\alpha > 0.005$.

Sammanfattning: testvar för μ (jfr k.i. för μ)
1. Undersökningsvar är normfördelad
σ är känd: $\frac{\bar{x}-\mu_0}{\frac{\sigma}{sqrt(n)}} \sim N(0;1)$ då H_0 är sann.
σ är okänd: $\frac{\bar{x}-\mu_0}{\frac{s}{sqrt(n)}} \sim t(n-1)$ då H_0 är sann.
2. Förd undersökningsvar okänd, stort stickprov
σ är känd: $\frac{\bar{x}-\mu_0}{\frac{\sigma}{sqrt(n)}} \sim approx\ N(0;1)$ då H_0 är sann.
σ är okänd: $\frac{\bar{x}-\mu_0}{\frac{s}{sqrt(n)}} \sim approx\ N(0;1)$ då H_0 är sann.

Test av populationsandelar p

Om $X_1, ..., X_n$ är ett stickprov av oberoende variabler från

$$X \sim Bin(n;p)$$

så gäller att

$$\hat{P} = \frac{X}{n} \sim approx\ N\left(p; \frac{p(1-p)}{n}\right)$$

om $npq > 5$. Dessutom är då

$$\frac{\hat{P}-p}{sqrt\left(\frac{p(1-p)}{n}\right)} \sim approx\ N(0;1).$$

och testvariabeln ges då av

$$\frac{\hat{P}-p_0}{sqrt\left(\frac{p_0(1-p_0)}{n}\right)} \sim approx\ N(0;1) \text{ då } H_0 \text{ är sann.}$$

Exempel:

En stöddig partiledare (från Katrineholm?) påstår att mer än hälften av de som har rösträtt sympatiserar med "hans" parti. I ett slumpmässigt urval svarade 625 av 1200 personer att de sympatiserar med "hans" parti. Ger data stöd för hans påstående? Det vi vill testa är

$$\begin{aligned} H_0 &: p = 0.5 \\ H_1 &: p > 0.5 \end{aligned}$$

Andelen som sympatiserar med partiet i stickprovet är

$$\hat{p} = \frac{625}{1200} = 0.521.$$

Då $n\hat{p}\hat{q} = 299 > 5$ är

$$\hat{P} \sim approx\ N\left(p; \frac{p(1-p)}{n}\right).$$

Testvariabeln är då

$$\frac{\hat{P}-p_0}{sqrt\left(\frac{p_0(1-p_0)}{n}\right)} \sim approx\ N(0;1) \text{ då } H_0 \text{ är sann.}$$

Det observerade testvärdet i stickprovet är

$$z_{obs} = \frac{\hat{p}-p_0}{sqrt\left(\frac{p_0(1-p_0)}{n}\right)} = \frac{0.521-0.5}{sqrt\left(\frac{0.5\cdot 0.5}{1000}\right)} = 1.45$$

Då p-värdet $=$ Pr $(Z > 1.45\,|H_0$ är sann$)$ $= 0.073$ kan H_0 ej förkastas. Det kan ej anses statistiskt påvisat att genomsnittliga andelen partisympatisörer i populationen är större än 0.5 för alla $\alpha < 0.07$.

Test av populationsvariansen σ^2

Låt $X_1, X_2, ..., X_n$ vara ett stickprov av oberoende observationer från

$$X \sim N\left(\mu; \sigma^2\right).$$

Då är

$$\frac{(n-1)\,s^2}{\sigma^2} \sim \chi^2\,(n-1)\,,$$

och testvariabeln ges då av

$$\frac{(n-1)\,s^2}{\sigma_0^2} \sim \chi^2\,(n-1) \text{ då } H_0 \text{ är sann.}$$

Exempel:

En maskin fyller konservburkar och av erfarenhet vet man att vikten varierar från burk tilll burk. Väntevärdet är $\mu = 750$ gram och standardavvikelsen är $\sigma = 16$ gram. Man tänker köpa in en ny maskin om spridningen är mindre. Antag att vikten varierar som en normalfördelad variabel. Ett slumpmässigt stickprov av storlek $n = 9$ dras där $s = 12$. Vi vill testa

$$\begin{aligned} H_0 &: \sigma = 16 \\ H_1 &: \sigma < 16 \end{aligned}$$

Testvariabeln är

$$\frac{(n-1)\,s^2}{\sigma_0^2} \sim \chi^2\,(n-1) \text{ då } H_0 \text{ är sann,}$$

det observerade testvärdet i stickprovet är

$$\chi^2_{obs} = \frac{(n-1)\,s^2}{\sigma_0^2} = \frac{(9-1)\,12^2}{16^2} = 4.5$$

Då p-värdet är större än 0.05 kan H_0 ej förkastas. Det kan ej anses statistiskt påvisat att standardavvikelsen i populationen är mindre än 16.

Klassisk hypotesprövning - en sammanfattning

- Formulera det statistiska problemet
- Översätt till noll- och mothypotes
- Fastställ signifikansnivån (alternativt beräkna p-värde)
- Välj en lämplig testvariabel beroende bl a på förutsättningar
- Bestäm testvariabelns fördelning under nollhypotesen
- Bestäm det kritiska området (alternativt beräkna p-värde)
- Ta ett stickprov och utför beräkningar på detta
- Beräkna testfunktionens observerade värde
- Acceptera eller förkasta nollhypotesen
- Formulera en begriplig slutsats