

Föreläsning 11: Mer om jämförelser och inferens

Matematisk statistik

David Bolin
Chalmers University of Technology
Oktober 8, 2018



Oberoende stickprov

Vi antar att vi har två oberoende stickprov

- n_1 observationer $X_{11}, X_{12}, \dots, X_{1n_1}$ från $N(\mu_1, \sigma_1^2)$.
- n_2 observationer $X_{21}, X_{22}, \dots, X_{2n_2}$ från $N(\mu_2, \sigma_2^2)$.

Vi vill nu testa om vi kan anta att $\mu_1 = \mu_2$.

Inför $\theta = \mu_1 - \mu_2$ som vi skattar med $\theta^* = \bar{X}_1 - \bar{X}_2$. Testa

$$H_0 : \theta = 0,$$

$$H_1 : \theta \neq 0$$

Alternativt en ensidig mothypotes, $\theta > 0$ eller $\theta < 0$.

Om det inte är känt från försökupställningen kan vi först behöva testa om vi kan anta $\sigma_1 = \sigma_2$. Vi utför hypotestestet

$$H_0 : \frac{\sigma_1}{\sigma_2} = 1$$

$$H_1 : \frac{\sigma_1}{\sigma_2} \neq 1$$

Ett exempel på stickprov i par

Man vill veta om en ny vetesort ger större skörd än den existerande sorten. Man väljer ut sex åkrar som skiljer sig i bördighet och klimat och delar in vardera åker i två delar där varsin sort odlas.

Åker nr	1	2	3	4	5	6
Skörd sort 1, kg/ha	7529	8913	6534	6503	6896	8023
Skörd sort 2, kg/ha	7239	8726	6129	6351	6644	7711
Skillnad D_i	290	187	405	152	252	312

Vi testar $H_0 : \Delta = 0$ mot $H_1 : \Delta \neq 0$.

Vi har $\bar{D} = 266.3$ och $s_D = 91$ och får

$$I_\Delta = (\bar{D} \pm t_{0.025}(5)s_D/\sqrt{6}) = (171, 362)$$

Eftersom $0 \notin I_\Delta$ kan vi förkasta H_0 .

Jämfördelse av variansen

Låt $F_\alpha(f_1, f_2)$ beteckna α -kvantilen i F-fördelningen. Ett konfidensintervall för σ_1^2/σ_2^2 ges av

$$I_{\sigma_1^2/\sigma_2^2} = \left[\frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1, n_2-1)}, \frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1, n_2-1)} \right]$$

För hypotestest av $\sigma_1^2/\sigma_2^2 = 1$ bildar vi teststorheten

$$T = s_1^2/s_2^2$$

som under H_0 är $F(n_1-1, n_2-1)$ -fördelad. För att underlätta beräkning av p-värdet, välj s_1 som den större av de två stickprovsvarianserna.

Fall 1: Kända σ_1 och σ_2

Om σ_1 och σ_2 är kända gäller att

$$D(\theta^*) = D(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

För hypotestest använder vi att under H_0 är

$$T = \frac{\bar{X}_1 - \bar{X}_2}{D(\theta^*)} \sim N(0, 1)$$

Så p-värdet ges av $p = 2(1 - \Phi(|T_{obs}|))$.

Ett konfidensintervall för $\theta = \mu_1 - \mu_2$ ges av

$$I_\theta = (\theta^* \pm z_{\alpha/2} D(\theta^*)) = \left(\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

Fall 2: $\sigma_1 = \sigma_2 = \sigma$ där σ är okänd

Sats: Poolad variansskattning

För k normalfördelade stickprov $N(\mu_j, \sigma^2)$, $j = 1, \dots, k$ fås en väntevärdesriktig skattning av σ^2 som

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_k - 1)s_k^2}{(n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1)}.$$

Dessutom gäller $(N - k)S_p^2/\sigma^2 \sim \chi^2(N - k)$, där $N = \sum_{i=1}^k n_i$.

Med $d(\theta^*) = s_p \sqrt{1/n_1 + 1/n_2}$ har vi att under H_0 är

$$T = \frac{\bar{X}_1 - \bar{X}_2}{d(\theta^*)} \sim t(n_1 + n_2 - 2)$$

Konfidensintervall ges nu av $I_\theta = (\theta^* \pm t_{\alpha/2}(n_1 + n_2 - 2)d(\theta^*))$.

Fall 3: $\sigma_1 \neq \sigma_2$ okända

Sats

Baserat på två normalfördelade stickprov gäller att

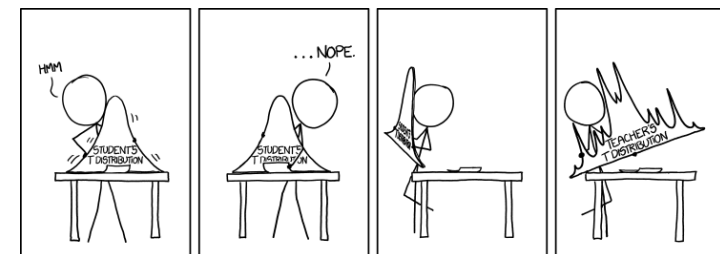
$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

är approximativt $t(f)$ -fördelad där

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

Vi kan nu bilda konfidensintervall och utföra hypotestest på samma sätt som tidigare: $I_\theta = (\theta^* \pm t_{\alpha/2}(f) \sqrt{s_1^2/n_1 + s_2^2/n_2})$.

Icke-parametriska metoder



- Antag att vi vill göra ett hypotestest men inte hittar en fördelning som passar till datan.
- Vi kan då använda en icke-parametrisk metod.
- Dessa är ofta baserade på medianen istället för väntevärdet.
- Medianen för en kontinuerlig fördelning är definierad som det tal M så att $P(X < M) = P(X > M) = 1/2$.

Teckentest för medianen

Givet ett stickprov av storlek n vill vi testa $H_0 : M = M_0$ mot en ensidig eller tvåsidig mothypotes. Vi kan då använda ett

- Låt Q_+ vara antalet observerade värden större än M .
- Låt Q_- vara antalet observerade värden mindre än M .
- Under H_0 är $Q_+ \sim \text{Bin}(n, 1/2)$ och $Q_- \sim \text{Bin}(n, 1/2)$.
 - Om $H_1 : M < M_0$, förkasta H_0 om Q_+ är för liten.
 - Om $H_1 : M > M_0$, förkasta H_0 om Q_- är för liten.
 - Om $H_1 : M \neq M_0$, förkasta H_0 om $\min(Q_+, Q_-)$ är för liten.
- Vi kan direkt beräkna p-värdet för testet!
- Om $X_i - M = 0$ räknar vi den observationen till den minsta av Q_+ eller Q_- eftersom det inte motsäger H_0 .
- Nackdelen med teckentest är att det endast använder tecknet på avvikelserna från medianen, och därför kan ha låg stryk.

Wilcoxon ranktest (I)

- Wilcoxon ranktest använder också storleken på avvikelserna, vilket ger högre styrka.
- Testet kräver att fördelningen är symmetrisk kring medianen.

Testet baseras på Wilcoxon teststorheter, som beräknas som:

- 1 Beräkna alla absolutdifferenser $|X_i - M_0|$ och ordna dessa i ökande storleksordning.
- 2 Låt R_i vara ranken för den i te absolutdifferensen gånger tecknet på motsvarande avvikelse.
- 3 Wilcoxon teststorheter ges nu av

$$W_+ = \sum_{i: R_i > 0} R_i, \quad |W_-| = \sum_{i: R_i < 0} |R_i|$$

Wilcoxon ranktest (II)

- Under H_0 borde det vara lika sannolikt att en viss rank får positivt som negativt tecken.
- Alltså, under H_0 borde W_+ och $|W_-|$ vara lika stora.
 - $H_1 : M < M_0$, förkasta H_0 om W_+ är för liten.
 - $H_1 : M > M_0$, förkasta H_0 om $|W_-|$ är för liten.
 - $H_1 : M \neq M_0$, förkasta H_0 om $W = \min(W_+, |W_-|)$ är för liten.
- Kritiska värden för teststorheterna finns tabulerade för olika n och α .
- Om två absolutdifferenser är lika stora så tilldelar vi dem medelvärde av motsvarande ranker.
 - Till exempel om vi observerar differenser 2 och -2 som skulle få rankerna 4 och 5 så tilldelar vi dem båda rank 4.5 gånger motsvarande tecken.
- För stora värden på n kan vi normalapproximera.

Exempel för stickprov i par

I fallet med parade observationer $(X_1, Y_1), \dots, (X_m, Y_m)$ bildar vi först differenserna $X_i - Y_i$ och använder sedan ett ranktest på differenserna för att testa om differensernas median är noll.

Vi vill testa hur mycket minne två statistiska paket använder vid analys av ett dataset. Vi gör mätningar (i Kb).

Program	Paket X	Paket Y	$D_i = X_i - Y_i$
1	512	500	12
2	650	600	50
3	890	890	0
4	410	400	10
5	1050	1025	25
6	1500	1400	100
7	600	625	-25
8	750	710	40

Exempel (forts)

- Vi vill testa

$$H_0 : M_X = M_Y$$

$$H_1 : M_X \neq M_Y$$

- Vi ordnar differenserna och beräknar rankerna

$ D_i $:	0	10	12	25	25	40	50	100
Rank:	1	2	3	4.5	4.5	6	7	8
R_i :	-1	2	3	-4.5	4.5	6	7	8

- Vi har nu

$$W_+ = 2 + 3 + 4.5 + 6 + 7 + 8 = 30.5$$

$$\text{och } |W_-| = |-1| + |-4.5| = 5.5.$$

- Eftersom vi har ett tvåsidigt test använder vi $W = \min(|W_-|, W_+) = 5.5$.
- Enligt tabell är 6 det kritiska värdet med $\alpha = 0.1$.
- $W = 5.5 < 6$, vi kan förkasta H_0 .

Exempel (I)

Vi har två olika typer av värmepannor och vill testa om märke A värmer upp ett rum snabbare än märke B. Vill testa

$$H_0 : M_A = M_B$$

$$H_1 : M_A < M_B$$

Mätningarna är

Märke B:	69.3	56.0	22.1	47.6	53.2	48.1	23.2	13.8
	52.6	34.4	60.2	43.8				
Märke A:	28.6	25.1	26.4	34.9	29.8	28.4	38.5	30.2
	30.6	31.8	41.6	21.1	36.0	37.9	13.9	

Wilcoxon's ranksummetest

- Vi har två oberoende stickprov $X_1, \dots, X_m$ och $Y_1, \dots, Y_n$ från några fördelningar X respektive Y .
- Vill vi testa $H_0 : M_X = M_Y$ mot ensidig eller tvåsidig mothypotes.

Testa om fördelningarna för X och Y kan antas vara lika:

- 1 Ordna de $m + n$ värdena i ökande storleksordning och ge varje observation en rank R_i från 1 till $m + n$.
- 2 Om vi har värden som är lika stora tilldelar vi dem medelranken precis som i rank-testet.
- 3 Vi antar att $m \leq n$ och bildar teststorheten W_m som summan av rankerna associerade med variablerna $X_1, \dots, X_m$.
- 4 Förkasta H_0 om W_m ligger i det kritiska området för testet.

Det kritiska värdet för W_m finns tabulerat för olika värden på m , n , och α .

Exempel (II)

- Vi poolar mätningarna och ordnar dem i stigande storleksordning:

Obs:	13.8	13.9	21.1	22.1	23.2	25.1	26.4	28.4	...
Märke:	B	A	A	B	B	A	A	A	...
Rank:	1	2	3	4	5	6	7	8	...

- Stickprov B är det mindre stickprovet och teststorheten blir därför

$$W_m = 1 + 4 + 5 + \dots = 212$$

- Vi slår upp det kritiska värdet för teststorheten i tabell. Med $m = 12$, $n = m + 3 = 15$ och $\alpha = 0.05$ är det kritiska värdet 202.
- Eftersom $W_m = 212 > 202$ kan vi förkasta H_0 .