

DH19 Bedrijfsstatistiek MC, 2e Bach Hir, Juni 2009

Slechts één van de vier alternatieven is juist. Kruis het bolletje aan vóór het juiste antwoord. Indien je een meerkeuzevraag verkeerd beantwoordt, verlies je een derde van een punt (giscorrectie). Indien je een meerkeuzevraag juist beantwoordt, krijg je een punt. Indien je de vraag onbeantwoord laat, krijg je nul op die vraag.

Je kan de achterkant van de opgave gebruiken als kladpapier.

1. Gegeven, $H_0 : \mu = 8$ tov $H_a : \mu \neq 8$; normaal verdeelde populatie met $\sigma = 2$; steekproefgrootte $n = 16$ en $\alpha = 0,05$.

Het kritieke gebied wordt gegeven door

- ☐ (a) $\bar{x} \leq 7,02$ of $\bar{x} \geq 8,98$;
☐ (b) $7,02 \leq \bar{x} \leq 8,98$;
☐ (c) $7,18 \leq \bar{x} \leq 8,82$;
☐ (d) $\bar{x} \leq 7,76$ of $\bar{x} \geq 8,25$.

2. Een telefoonwinkel verkocht 150 GSM's van merk A, waarvan er 14 opnieuw werden binnengebracht wegens defecten. Van merk B verkocht de winkel 125 GSM's en hiervan keerden er 15 terug wegens niet correct functioneren. We wensen te testen of de fractie defecten bij merk A kleiner is dan bij merk B. Dus $H_0 : (p_A - p_B) = 0$ t.o.v $H_a : (p_A - p_B) < 0$ voor $\alpha = 0,05$.

Als waarde voor de toetsingsgrootte berekenden we $-1,72$.

Uit R weten we: $\text{pnorm}(0.95) = 0.83$ en $\text{qnorm}(0.95) = 1.645$

- ☐ (a) we verwerpen H_0 omdat $-1,72 < -1,645$;
☐ (b) we verwerpen H_0 niet omdat $-1,72 < 1,645$;
☐ (c) we verwerpen H_0 niet omdat $|-1,72| > 0,83$;
☐ (d) we verwerpen H_0 omdat $-1,72 < -0,83$.

3. Gegeven, volgende 9 waarnemingen, getrokken uit een normaal verdeelde populatie: 100, 93, 84, 120, 76, 69, 63, 108, 105.

Voor deze gegevens vinden we: $\bar{x} = 90,89$ en $s = 19,21$.

Om de hypothese $H_0 : \sigma = 14$ versus $H_a : \sigma > 14$ te testen gebruiken we een toetsingsgrootte die,

- ☐ (a) F -verdeeld is;
☐ (b) T -verdeeld is;
☐ (c) Chikwadraat verdeeld is;
☐ (d) standaardnormaal verdeeld is.

4. Onderstaande tabel geeft de opbrengst per aandeel van enkele distributiebedrijven voor twee opeenvolgende jaren. Onderstel de steekproeven lukraak getrokken uit normaal verdeelde populaties.

Distributiebedrijven	jaar 1	jaar 2	Samengevat:	jaar 1	jaar 2
Penney	4,12	4,52			
Federated Dprtmnt	4,01	4,31	Steekproefomvang	10	10
American Stores	3,14	4,36	Steekproefgemiddelde	3,479	3,836
Carter Hawley	2,49	2,67	Steekproefvariantie	2,925	2,622
Wiches	3,09	3,57			
Mc. Donald's	5,5	4,78			
Zayre	2,71	3,3			
Wal-Mart	0,93	2			
Mercantile	6,8	7,25			
Borman's	2	1,6			

We toetsen $H_0 : (\mu_1 - \mu_2) = 0$ t.o. $H_a : (\mu_1 - \mu_2) < 0$. Dan is $\sigma^2_{(\bar{x}_1 - \bar{x}_2)}$ gelijk aan

- 0 (a) 5,547;
 0 (b) 0,303;
 0 (c) 2,773;
 0 (d) geen van de vorige.
5. Beschouw een lukrake steekproef x_1, \dots, x_n uit een normale verdeling met ongekende parameters μ en σ^2 . We berekenden in de les de Maximum Likelihood schatter voor de variantie als $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$. Voor deze schatter geldt:
- 0 (a) $E(\hat{\sigma}^2) = \sigma^2$;
 0 (b) $E(\hat{\sigma}^2) = \frac{n-1}{n} \sigma^2$;
 0 (c) $E(\hat{\sigma}^2) = s^2$;
 0 (d) $E(\hat{\sigma}^2) = \frac{n}{n-1} \sigma^2$.
6. Gegeven, waarnemingen x_1, \dots, x_n , met n groot, en $E(x_i) = \mu$, $Var(x_i) = \sigma^2$, voor alle $1 \leq i \leq n$.
 We toetsen $H_0 : \mu = \mu_0$, versus $H_a : \mu < \mu_0$, links eenzijdig, en gebruiken de toetsingsgrootheid $z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$. Significantieniveau α is gegeven.
 We noteren $P_\mu(\dots)$ om aan te duiden dat de waarnemingen uit een populatie komen met verwachte waarde μ .
 Het onderscheidingsvermogen van de test wordt gegeven door:
- 0 (a) $1 - \beta(\mu) = P_\mu\left(\frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} > -z_\alpha\right)$;
 0 (b) $1 - \beta(\mu) = P_{\mu_0}\left(\frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} < -z_\alpha\right)$;
 0 (c) $1 - \beta(\mu) = P_\mu\left(\frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} < -z_\alpha\right)$;
 0 (d) $1 - \beta(\mu) = P_{\mu_0}\left(\frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} > z_\alpha\right)$.

7. Het verschil tussen kwantitatieve en kwalitatieve factoren is:
- 0 (a) kwalitatieve factoren zijn onafhankelijk, kwantitatieve factoren zijn afhankelijk;
 - 0 (b) kwalitatieve factoren worden gemeten op een numerieke schaal, kwantitatieve factoren niet;
 - 0 (c) kwantitatieve factoren worden gemeten op een numerieke schaal, kwalitatieve factoren niet;
 - 0 (d) kwantitatieve factoren zijn onafhankelijk, kwalitatieve factoren zijn afhankelijk .
8. De omzet van een bedrijf wordt beïnvloed door de uitgaven aan reclame (in euro's). We willen ook rekening houden met het gebruikte medium waarmee men reclame voert. Er zijn 3 verschillende media die gebruikt worden: televisie, radio of de krant. Om dit probleem te bestuderen stellen we een regressiemodel op, waarin we ook toelaten dat er interactie is tussen het gebruikte medium en de uitgaven aan reclame. (We onderstellen dat een bedrijf in de steekproef slechts gebruikt maakt van één enkel medium waarmee men reclame voert). In ons model nemen we ook een intercept (of constante term) op. Hoeveel regressieparameters bevat ons model dan (constante term meerekenen!)?
- 0 (a) 5
 - 0 (b) 6
 - 0 (c) 7
 - 0 (d) 8
9. Beschouw een t -verdeelde kansvariabele en $\alpha = 5\%$. Indien het aantal vrijheidsgraden toeneemt, zullen de waarden $-t(\alpha/2)$ en $t(\alpha/2)$
- 0 (a) zich van elkaar verwijderen;
 - 0 (b) dichterbij elkaar komen, maar niet dichterbij 2;
 - 0 (c) naar 0 convergeren;
 - 0 (d) geen van de vorige.
10. Een onderzoeker gebruikt het model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$. Hoe zal het effect van een toename van x met één eenheid op de verwachte waarde van y veranderen als de waarde van x toeneemt en $\beta_2 < 0$?
- 0 (a) het effect neemt af;
 - 0 (b) het effect neemt toe;
 - 0 (c) het effect verandert niet;
 - 0 (d) dit hangt ook af van de waarde van ε .
11. Bereken de determinatiecoëfficiënt voor een enkelvoudige regressie (met constante term) met respons y en verklarende variabele x . De varianties van y en x zijn beide 1, en hun gemiddeldes nul. Verder is de schatting $\hat{\beta}_1 = 0,5$.
- 0 (a) 0,25;
 - 0 (b) 0,75;
 - 0 (c) 0,5;
 - 0 (d) 1.

12. Noteer P_t de grootte van de bevolking in Mexico in jaar t , gemeten in miljoenen inwoners. We schatten het regressiemodel

$$\log P_t = \alpha + \beta t + \varepsilon_t,$$

en dit op basis van jaarlijkse data van 1980 tot 2008. We bekomen $\hat{\beta} = 0,03$. De interpretatie is dat

- 0 (a) de verwachte jaarlijkse bevolkingsaangroei is 30000 extra inwoners per jaar;
 0 (b) de verwachte jaarlijkse bevolkingsaangroei is 3 miljoen extra inwoners per jaar;
 0 (c) de verwachte jaarlijkse bevolkingsgroei is 3%;
 0 (d) de verwachte jaarlijkse bevolkingsgroei is 0,03%.
13. Vertrekkend van de definitie $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$ kunnen we de stelling bewijzen dat $MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2$. Het bewijs vertrekt van:

$$MSE(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}))^2] + E[2(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)] + E[(E(\hat{\theta}) - \theta)^2]$$

De middelste term in deze uitdrukking kunnen we herschrijven als

- 0 (a) $2(\hat{\theta} - E(\hat{\theta}))E[E(\hat{\theta}) - \theta]$;
 0 (b) $-2E[(E(\hat{\theta}) - \theta)^2]$;
 0 (c) $-2E[(E(\hat{\theta}) - \hat{\theta})^2]$;
 0 (d) $2(E(\hat{\theta}) - \theta)E[\hat{\theta} - E(\hat{\theta})]$.
14. Voor het toetsen van $H_0 : \mu_1 = \mu_2 = \dots = \mu_p$, gebruiken we als toetsingsgrootheid $F = \frac{MST}{MSE}$. Voor $p = 2$, en dus voor het toetsen van $H_0 : \mu_1 = \mu_2$, gebruiken we de toetsingsgrootheid $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$ (onafhankelijke steekproeven). We kunnen bewijzen dat $t^2 = F$. Hiervoor hebben we het verband nodig tussen het totale gemiddelde \bar{x} en de resp. gemiddelden \bar{x}_i . Er geldt:
- 0 (a) $\bar{x}_1 = \bar{x} + \frac{n_2}{n_1}(\bar{x} - \bar{x}_2)$;
 0 (b) $\bar{x}_1 = \bar{x} + \frac{n_1}{n_2}(\bar{x} - \bar{x}_2)$;
 0 (c) $\bar{x}_1 = \bar{x} - \frac{n_2}{n_1}(\bar{x} - \bar{x}_2)$;
 0 (d) $\bar{x}_1 = \bar{x} - \frac{n_1}{n_2}(\bar{x} - \bar{x}_2)$.
15. Een onderzoeker wil toetsen of de consument een bepaalde koffiesoort kiest onafhankelijk van het merk. Er zijn 3 verschillende soorten koffie en 4 verschillende merken. Op basis van een steekproef wordt een kruistabel opgesteld, en de χ^2 toetsingsgrootheid levert een P-waarde van 0,009. De onderzoeker besluit dat
- 0 (a) er geen significante resultaten gevonden werden;
 0 (b) de consumenten een significant grotere voorkeur hebben voor een bepaald type merk **of** een bepaald type koffie;
 0 (c) de consumenten een significant grotere voorkeur hebben voor een bepaald type merk **en** een bepaald type koffie;
 0 (d) er een significant verband is tussen de keuze voor de koffiesoort en de keuze voor het merk.

Oplossingen

1. a

2. a

3. c

4. d

De steekproeven zijn gepaarde waarnemingen.

$$\sigma^2_{(\bar{x}_1 - \bar{x}_2)} = \sigma^2_{\bar{x}_d} = \sigma^2_d / 10 = 0.344$$

5. b

$E(\hat{\sigma}^2 * \frac{n}{n-1}) = E(s^2) = \sigma^2$ (want $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ en cursusslides p. 108 (opgelet: op deze pagina in de cursus is $\hat{\sigma}$ anders gedefinieerd dan in deze oefening))

$$\text{Dus: } E(\hat{\sigma}^2) = \sigma^2 * \frac{n-1}{n}$$

6. c

$$(\text{Verwerp } H_0 | H_a \text{ geldt}) = P\left(\frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} < -z_{\alpha} | H_a \text{ geldt}\right) = P_{\mu}\left(\frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} < -z_{\alpha}\right)$$

7. c

8. b

Een mogelijke regressievergelijking is:

$$Omzet = \beta_0 + \beta_1 * Dummy_{television} + \beta_2 * Dummy_{radio} + \beta_3 * Reclame * Dummy_{television} + \beta_4 * Reclame * Dummy_{radio} + \epsilon$$

Merk op dat we Krant als basisniveau wordt gebruikt (Opmerking: Het basisniveau is vrij te kiezen).

9. b

Naarmate een t-verdeelde kansvariabele meer vrijheidsgraden heeft, zal ze meer de normale verdeling benaderen en dus minder 'dikke staarten' hebben. Het gevolg is dat de percentielen $-t(\alpha/2)$ en $t(\alpha/2)$ dichter bij elkaar komen, maar niet dichter dan de percentielen $-z(\alpha/2)$ en $z(\alpha/2)$. Dit is ongeveer gelijk aan $1.96 * 2 = 3.92$ en dus groter dan 2.

10. a

$$\text{Het effect} = \frac{E(y)}{dx} = \beta_1 + 2\beta_2 x.$$

11. a

$$\text{Voor enkelvoudige regressie geldt: } R^2 = \text{Corr}(x, y)^2 = \left[\frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} \right]^2 = \frac{SS_{xy} SS_{xy}}{SS_{xx} SS_{yy}}$$

$$\text{Nu is } SS_{xx} = \sum_i (x_i - \hat{x})^2 = \sum_i (y_i - \hat{y})^2 = SS_{yy}$$

$$\text{Dus } R^2 = \frac{SS_{xy}^2}{SS_{yy}^2} = [\beta_1]^2 = 0.5^2 = 0.25$$

12. c

13. d

14. c

$$\text{Vertrek van de volgende vergelijking: } \bar{x} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

15. d