

Tentamen (Wiskundige) Statistiek Module 5 Statistiek en Analyse

Bachelor 2 Applied Mathematics

Modulecode: 201400218
Datum: dit is **voorbeeldtentamen A**
Tijd: tijdsduur is 3 uur
Blokcoördinator: Dr. P.K. Mandal
Docent: Dr. K. Poortema

Toegestane middelen: alleen gewone rekenmachine (geen GR)

Bijlagen: standaardnormale tabel,
t-tabel,
chi-kwadraat-tabel,
Tabel F verdeling (1 pagina, $\alpha = 0.05$)
formuleblad Wiskundige Statistiek (2 pagina's)

Normering:

1	2a	2b	2c	3a	3b	3c	4a	4b	5a	5b	5c	6a	6b	7
2	2	2	3	2	3	3	1	3	2	2	2	4	1	4

Totaal: 36 punten

Opmerking 1:

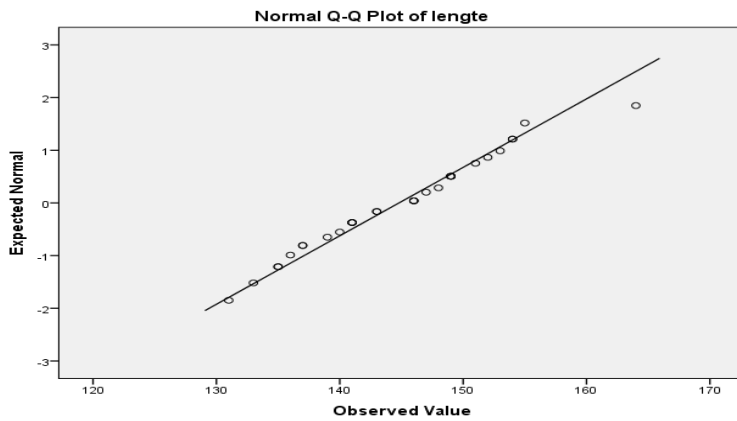
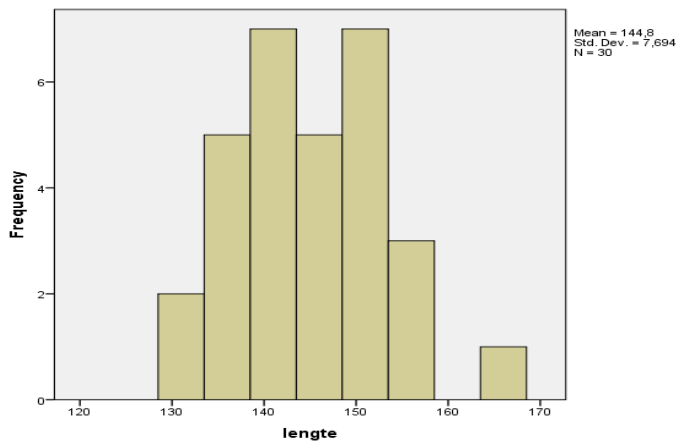
Voor dit voorbeeldtentamen hebben we zo veel mogelijk gebruik gemaakt van werkcollegeopgaven. Engelstalige opgaven hebben we gemakshalve niet vertaald. De opgaven van de tentamens van studiejaar 2016/2017 zijn in het Nederlands gesteld, met als enige mogelijke uitzondering beschrijvingen van data sets (die onvertaald overgenomen worden).

Opmerking 2:

De te gebruiken tabellen komen uit het dictaat van Statistiek(TBK, INF,BIT), maar zijn hier NIET daadwerkelijk toegevoegd. De vier tabellen worden beschouwd als de standaard set van tabellen.

Opgave 1

We beschouwen de lengtemetingen van een steekproef van 30 elfjarige meisjes van een school uit Engeland (Heaton Middle School, Bradford). We willen de verdeling bestuderen aan de hand van de volgende plots.



Beoordeel de drie plots. Kunnen we een normale verdeling toepassen? Motiveer kort je antwoord. Zijn er uitschieters?

Opgave 2

X_1, \dots, X_n is a random sample of X , which has a normal distribution with unknown parameters μ and σ^2 . The “standard” unbiased estimator S^2 is not the same as the maximum likelihood estimator $\widehat{\sigma}^2$, but what is the best estimator?

- Verify that $\text{var}(S^2) = \frac{2\sigma^4}{n-1}$, using the χ^2 -distribution.
- Is S^2 a better estimator for σ^2 than $\widehat{\sigma}^2$?
- For what real value $a \in \mathbb{R}$ is $T = a \cdot S^2$ the best estimator of σ^2 ?

Opgave 3

De verdeling van de s.v. X heeft de volgende dichtheid met onbekende parameter $\theta \geq 0$:

$$f(x; \theta) = \begin{cases} e^{-(x-\theta)}, & \text{als } x \geq \theta \\ 0, & \text{elders} \end{cases}$$

X_1, X_2, \dots, X_n is een aselechte steekproef van X .

- Toon aan dat $\hat{\theta} = \min(X_1, \dots, X_n)$ de meest aannemelijke schatter van θ is.
- Ga na of $\hat{\theta}$ een consistente schatter van θ is.
- Bepaal aannemelijkheidsquotiënttoets voor $H_0: \theta = 0$ tegen $H_1: \theta > 0$ bij gegeven α_0

Opgave 4

Volgens de theorie van (multiple) regressie geldt een gelijkheid die vaak als volgt opgeschreven wordt:

$$SS(\text{total}) = SS(\text{regression}) + SS(\text{error})$$

- Geef voor elke term, elke Sum of Squares, een formule en geef aan hoe je de term kunt interpreteren.
- Bewijs de gelijkheid, uitgaande van formules van “kleinste kwadraten”.

Opgave 5

Het aantal uren zonneshijn in de maand juli in De Bilt van 1964 t/m 1983 is als volgt:

Jaar	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973
uren zon	188.4	146.2	154.9	250.6	205.4	186.5	158.8	249.1	171.4	181.1
Jaar	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983
uren zon	158.0	214.7	251.8	183.8	167.1	144.4	131.0	167.8	231.3	252.5

Steekproefgemiddelde en steekproefstandaardafwijking zijn 189.74 resp. 39.50

- Bepaal het 99%-betrouwbaarheidsinterval voor het verwachte aantal uren zonneshijn in de maand juli. Op welke veronderstellingen is dit interval gebaseerd?
- Leg in gewone woorden uit wat het in a gevonden resultaat betekent.
- Voor 1984 blijkt het aantal uren zonneshijn in juli in De Bilt gelijk te zijn aan 164.1. Is dit een uitzonderlijk lage waarde of niet? Motiveer je antwoord.

Opgave 6

De effectiviteit van reclame wordt gewoonlijk beoordeeld op grond van de mate waarin de verkoop toeneemt. Voor een bepaald product houdt een grote winkelketen een reclamecampagne. De volgende resultaten van de wekelijkse verkoop (aantal verkochte exemplaren) zijn verkregen voor 7 (vergelijkbare) filialen van de winkelketen.

	1	2	3	4	5	6	7
voor reclamecampagne (x)	3419	4135	4979	3752	6222	4047	3720
na reclamecampagne (y)	4340	5269	6061	4011	5749	4814	3642

	steekproefgemiddelde	steekproefstandaardafwijking
x	4324.9	970.8
y	4840.9	901.3
$z = y - x$	516.0	622.7

- Onderzoek m.b.v. een geschikte toets of er het verwachte wekelijkse verkoopaantal is toegenomen na de reclamecampagne. Ga uit van redelijke statistische veronderstellingen om dit toetsingsprobleem aan te pakken. Neem als onbetrouwbaarheidsdrempel $\alpha_0 = 0.05$ en volg het schema van acht stappen (zie formuleblad).
- Als de vraagstelling in a. was geweest: “test of er verschil is in (verwachte) verkoopaantallen vóór en na de reclamecampagne”, welke wijzigingen moet je dan aanbrengen in je oplossing?

Opgave 7

Aan 200 aselekt gekozen bedrijven van ieder van drie verschillende groottes waren enquêtes toegezonden. Voor de representativiteit is het van belang te weten of de mate van beantwoording gelijk is voor de drie verschillende bedrijfsgroottes. Van de 200 aselekt gekozen kleine bedrijven reageerden er 98, van de 200 middelgrote bedrijven waren dat er 79 en van de 200 grote bedrijven kwamen 71 reacties binnen.

Ga met behulp van de geschikte statistische toets na of er verschil is in respons tussen de drie verschillende bedrijfsgroottes.

Volg het schema van 8 stappen (zie formuleblad) en neem $\alpha = 1\%$.