8. Hypotesetest.

8.1. Et innledende eksempel.

En dag du kjører i en 60-sone, blir du tatt i fartskontroll. Du prøver deg selvsagt med at "..dette må være målefeil. Jeg kjørte i nøyaktig 60 km/t..". Men politibetjenten forteller at farten var målt til 65 km/t, og skriver ubønnhørlig ut en bot.

La oss analysere denne situasjonen fra en statistikers synspunkt. Det er rimelig å anta at dersom det passerer mange biler som alle har *virkelig* fart på nøyaktig 60 km/t, så vil politiets fartsmålinger av disse *i gjennomsnitt* også være 60 km/t, men med noe spredning rundt denne verdien. La oss være mer presise, og anta at disse målingene er normalfordelt med forventningsverdi $\mu = 60\,\text{km/t}$, og at standardavviket for målingene er $\sigma = 2.0\,\text{km/t}$ (jeg aner ikke om denne verdien er realistisk).

Er det (under disse antakelsene) rimelig å tro at din *virkelige fart* var 60 km/t når *målt fart* var 65 km/t?

Vi starter med å beregne *z*-scoren for *målt* fart på 65 km/t, under forutsetning av at *virkelig* fart var 60 km/t:

$$z = \frac{x-\mu}{\sigma} = \frac{65-60}{2} = 2.5\,.$$

Dette tallet forteller at den målte verdien på 65 km/t ligger 2.5 standardavvik til side for forventningsverdien på 60 km/t.

Fra en standard normalfordelingstabell finner vi at sannsynligheten for å komme 2.5 standard-avvik eller mer til side for forventningsverdien er $1-0.9938 = 0.0062$, eller 0.62%. Så det var nok temmelig sannsynlig at du brøt fartsgrensen, og at politibetjenten var i sin fulle rett.

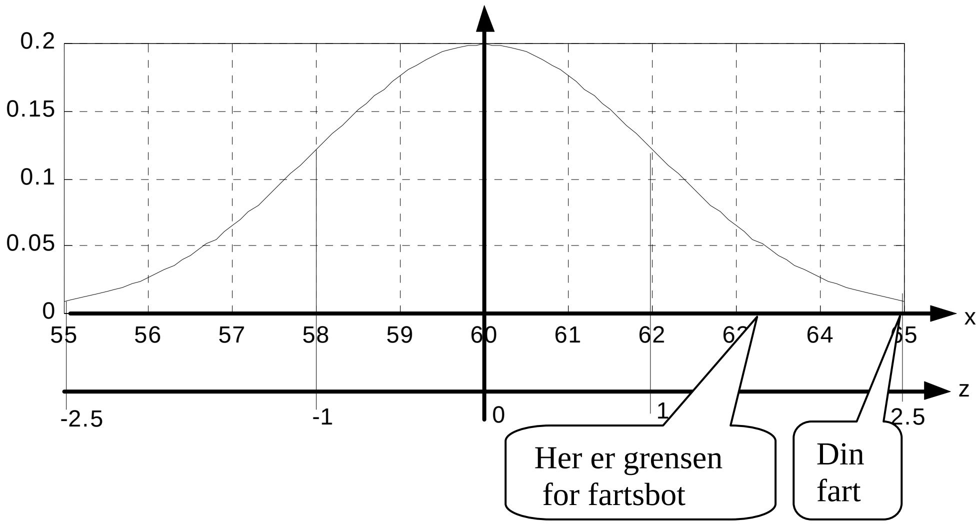
I slike situasjoner må vi *på forhånd* bestemme oss for hvor stor sannsynligheten skal være for at vi *feilaktig* gir fartsbot. La oss sette som grense at sannsynligheten skal være høyst 5 % for at vi *feilaktig* gir fartsbot, d.v.s. at en person som kjører med nøyaktig 60 km/t, skal får fartsbot. Dette gjøres enklest ved at vi fastsetter en grense for hvilken *målt fart* som skal gi fartsbot. Ved hjelp av en standard normalfordelingstabell fastsettes denne grensen slik:

$$z = \frac{x-\mu}{\sigma} = \frac{x-60}{2.0} = 1.645 \quad \Leftrightarrow \quad x-60 = 2.0\cdot 1.645 \quad \Leftrightarrow \quad x = \underline{\underline{63.3}}\,.$$

Dersom målt fart er større enn 63.3 km/t, blir det fartsbot.

Situasjonen er illustrert på neste side, der grafen angir sannsynlighetstettheten for at en bil som kjører i nøyaktig 60 km/t skal får målt sin fart til *x*.

Situasjonen ovenfor er et enkelt eksempel på en **hypotesetest**. Politibetjentens hypotese var at du kjørte for fort. For å underbygge påstanden, antok han at du *ikke* kjørte for fort, og under-søkte om dette var sannsynlig. Alle skal som kjent betraktes som uskyldige inntil det motsatte er bevist. Konklusjonen var at det var *svært usannsynlig* at du holdt lovlig fart. Dermed ble det fartsbot.



7.2. Generelt om hypotesetest.

I fartsbot-eksemplet ovenfor framsatte vi en påstand om din *virkelige* fart, og testet den på grunnlag av en måling. Det er mye mer vanlig at vi framsetter en påstand om en eller annen parameter i en populasjon, og ønsker å teste den på grunnlag av observasjoner på et *tilfeldig utvalg* fra populasjonen. Men framgangsmåten ved testen er i prinsippet slik som med fartsbot-eksemplet.

Vi starter med å formulere en påstand om populasjons-parameteren, og vil undersøke hvor sannsynlig det er at påstanden er sann. Som alle gode forskere er vi forsiktige av oss, og vil ikke gå ut med påstander uten at vi er temmelig sikre på at vi har rett. Vi tester derfor vår hypotese slik: Vi *antar* at hypotesen vår *ikke* er rett. Dersom denne antakelsen er svært usannsynlig (d.v.s. at det er svært *usannsynlig* at hypotesen *ikke* er sann), aksepterer vi hypotesen. I motsatt fall våger vi ikke å framsette den.

Nå er tiden inne til litt terminologi:

Den hypotesen som vi ønsker å teste (”bevise”), kalles gjerne ***den alternative hypotese*** og gis symbolet H_1. Denne hypotesen får alltid form av at parameteren er ”større enn” eller ”mindre enn” eller ”forskjellig fra” en eller annen verdi (virkelig fart er *større enn* 60 km/t).

Den motsatte hypotesen (som vi under testen antar er sann), kalles gjerne ***nullhypotesen*** og gis symbolet H_0. Denne hypotesen vil alltid inneholde at parameteren ”er lik” en eller annen verdi.

I vårt fartsbot-eksempel blir:
H_0: Farten er *mindre enn eller lik* 60 km/t. I praksis tester vi med $v = 60\,\text{km/t}$.
H_1: Farten er *større enn* 60 km/t.

Noen lærebøker insisterer på at de to hypotesene skal være komplementære (d.v.s. at dersom den ene ikke er sann, må den andre være det). I praksis tester vi alltid mot ”er lik” i null-hypotesen. Det er derfor blitt vanlig å formulere nullhypotesen med bare likhetstegn.

Når vi gjennomfører en hypotesetest, kan vi gjøre to typer feil: Vi kan feilaktig forkaste nullhypotesen (bøtelegge en som kjører lovlig), eller vi kan feilaktig unnlate å forkaste nullhypotesen (unnlate å bøtelegge en som kjører for fort). Disse to feiltypene kalles henholdsvis ***Type I*** og ***Type II*** feil. Symbolene α og β brukes for å angi sannsynlighetene for å gjøre disse feilene. Vi kan systematisere situasjonen slik:

		Virkelig situasjon	
		H_0 er riktig (lovlig fart)	H_0 er ikke riktig (kjører for fort)
Beslutning	H_0 forkastes (gi bot)	Type I feil. Sannsynlighet α.	Korrekt.
	H_0 forkastes ikke (ingen bot)	Korrekt.	Type II feil. Sannsynlighet β.

Vanligvis er det mest alvorlig å begå en Type I feil (vi skal ikke bøtelegge folk som kjører lovlig). Vi bruker derfor å sette α forholdsvis lavt. Vanlige verdier er $\alpha = 0.05$, $\alpha = 0.02$ og $\alpha = 0.01$. Problemet er at jo mindre α settes, jo større blir β.

Vi kaller α for testens ***signifikansnivå***. Det er prinsipielt viktig at α fastlegges *før* vi utfører testen.

Vi må også fastlegge en ***testobservator*** (”***Test Statistics***”). I fartsbot-eksemplet var det naturlig å bruke målt fart som testobservator. Det er imidlertid mer systematisk å bruke en testobservator som har en kjent sannsynlighetsfordeling, gjerne standard normalfordeling, *t*-fordeling eller χ^2-fordeling.

Det ***kritiske området*** eller ***forkastningsområdet*** består av de verdier av testobservatoren som fører til at vi forkaster nullhypotesen. Dersom vi i fartsbot-eksemplet bruker målt fart som testobservator, vil målt fart større enn 63.3 km/t være kritisk område. Hvis vi heller bruker *z*-scoren til målt fart som testobservator, vil *z*-verdier på mer enn 1.645 utgjøre det kritiske området. Den verdien av testobservatoren som avgrenser det kritiske området, kalles observatorens ***kritiske verdi***.

På grunnlag av testobservatorens verdi kan vi også beregne observasjonens ***P-verdi***. Dette er sannsynligheten for at testobservatoren skal få en verdi som er minst så ekstrem som den observerte. I fartsbot-eksemplet fant vi ut at det var en sannsynlighet på 0.0062 for at målt fart skal være minst 65 km/t når virkelig fart var 60 km/t. *P*-verdien er da 0.0062.

Så til poenget:

- Vi *forkaster* H_0 dersom testobservatoren befinner seg i det kritiske området (forkastningsområdet).

Eller:

- Vi *forkaster* H_0 dersom *P*-verdien er mindre enn signifikansnivået α.

Merknad: At vi forkaster H_0, betyr ikke at vi har *bevist* at vår alternative hypotese (H_1) er sann. Vi har kun slått fast at på vårt signifikansnivå er det ikke rimelig å holde fast på H_0.

La oss lage en oversikt over gangen i en slik hypotesetest. Jeg vil bruke eksemplet med farts-kontrollen til å illustrere trinnene.

Generell framgangsmåte:	Eksemplet med fartskontroll:
Formuler den hypotesen du vil teste, H_1, og den motsatte hypotesen, H_0.	H_1: Virkelig fart er større enn 60 km/t. H_0: Virkelig fart er mindre enn eller lik 60 km/t.
Fastsett testens **signifikansnivå** α.	Høyst 5% sannsynlighet for feilaktig å forkaste H_0 (feilaktig gi bot når virkelig fart er 60 km/t).
Samle inn og bearbeide data.	Måle fart til 65 km/t.
Angi en **testobservator**, og sett opp sannsynlighetsmodell for denne.	Bruker $z = \dfrac{v-60}{2.00}$ som testobservator, der v er målt fart. Antar at z er standard normalfordelt.
Utfør beregninger. Kan gjøres på to måter: 1) Beregn et **kritisk område**, som angir for hvilke verdier av testobservatoren vi skal forkaste H_0. 2) Beregn ***P*-verdien** for test-observatoren, d.v.s. den største sannsynligheten for å få de observerte verdiene dersom H_0 er sann.	1) Det kritiske området består av $z > z_{0.05} = 1.645$. 2) P-verdien for målt fart er 0.0062.
Trekk konklusjon: 1) Forkast H_0 dersom testobservatoren har verdi i det kritiske området. 2) Forkast H_0 dersom testens *P*-verdi er mindre enn testens signifikansnivå. Dersom H_0 forkastes, aksepteres H_1.	1) Forkaster H_0 fordi $z = 2.5 > z_{0.05} = 1.645$. 2) Forkaster H_0 fordi *P*-verdien på 0.0062 er mindre enn signifikansnivået $\alpha = 0.05$. H_0 forkastes, og det blir fartsbot.

Det er en nær sammenheng mellom hypotesetest og konfidensintervall. I fartsbot-eksemplet kunne vi konstruert et konfidensintervall med konfidensnivå $\alpha = 0.10$ (slik at $\alpha/2$ blir 0.05) rundt den observerte farten på 65 km/t. Med et standardavvik på 2.0 km/t for observert fart, blir dette konfidensintervallet

$$[65 - 1.645 \cdot 2.0,\ 65 + 1.645 \cdot 2.0] = [61.7,\ 68.3].$$

Høyeste lovlige fart på 60 km/t ligger utenfor dette intervallet. Da er det naturlig å konkludere med at du ikke holdt lovlig fart.

Som regel vil konfidensintervall-metoden gi samme resultat som vanlig hypotesetest. Men merk at mens konfidensintervall tar utgangspunkt i *estimatet* av parameterverdien, (målt fart), så tar hypotesetest utgangspunkt i den parameterverdien som nullhypotesen angir (lovlig fart). Denne forskjellen *kan* gi litt forskjellige verdier for standardavviket til parameteren, noe som i grensetilfeller *kan* gi ulike konklusjoner med de to metodene. I praksis forekommer imidlertid dette svært sjelden.

Til slutt noen ord om Type II feil. Vi definerer begrepet ***styrkefunksjon*** som sannsynligheten for å *forkaste* en feil nullhypotese. Denne sannsynligheten blir $1-\beta$. I fartsbot-eksemplet vil styrkefunksjonen gi sannsynligheten for at en bilfører som kjører for fort får bot. Denne sannsynligheten avhenger av bilens *virkelige* fart. Generelt vil styrkefunksjonen til en test være en funksjon av den virkelige verdien til den populasjonsparameteren som testes.

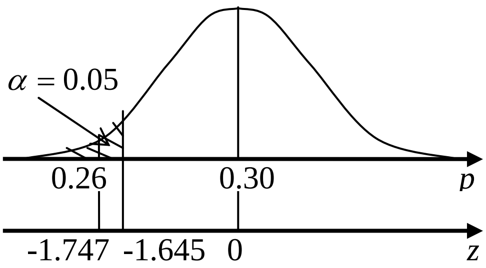
7.3. Test av prosentandel.

Når vi nå skal gå i gang med å benytte disse prinsippene, skal vi starte med en test av andel *p*. Vi går rett løs på et eksempel. Anta at X-partiet fikk 30 % av stemmene i en by ved siste valg. Foran neste valg utfører lokalavisa det en meningsmåling der tilfeldige velgere spørres om hvilket parti de vil stemme på. Av de 400 om svarte, var det 104 som sa at de ville stemme på X-partiet. Et matematikk-kyndig medlem av redaksjonen fant etter hvert ut at dette svarte til en andel på $\frac{104}{400} = 0.26$, eller 26 %. Dagen etter utbasunerte avisa at ”X -partiet går klart tilbake”. Kunne avisa framsette en slik påstand på bakgrunn av spørreundersøkelsen?

Vi starter med å formulere hypotesene. Den påstanden som avisa framsetter, formuleres som alternativ hypotese. Nullhypotesen blir da at X-partiet har (minst) samme oppslutning som før. Vi får:

H_0: X-partiet har oppslutning fra minst 30 % av velgerne ($p = 0.30$).
H_1: X-partiet har oppslutning fra mindre enn 30 % av velgerne ($p < 0.30$).

Videre bestemmer vi oss for et 5 % signifikansnivå slik at $\alpha = 0.05$. Under testen forutsetter vi at H_0 er sann slik at $p = 0.30$. Vi skal undersøke sannsynligheten for at et estimat av *p* kan bli så lavt som 0.26. Vi opererer da på negativ side av standard normalfordelingen. Det kritiske området blir da $z < z_{0.05} = -1.645$. Merk fortegnet.



Vi vet fra før at standardavviket for estimatet av *p* er

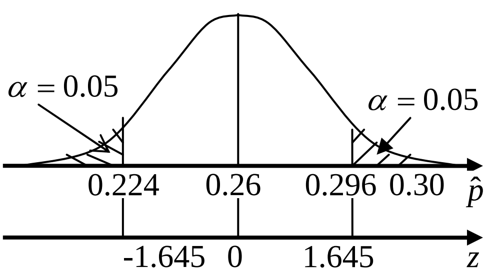
$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.30 \cdot 0.70}{400}} = \underline{0.0229}.$$

Vår testobservator blir

$$z = \frac{\hat{p} - p}{\sigma_p} = \frac{0.26 - 0.30}{0.0229} = \underline{-1.747}.$$

Testobservatoren ligger i det kritiske området, slik at vi forkaster H_0. Konklusjonen er altså at på 5 % signifikansnivå kan vi ikke holde fast på at oppslutningen om X-partiet fremdeles er (minst) 30 %.

Dersom vi benytter *P*-verdi, går vi fram slik: For den observerte andelen X-parti-velgere får vi en *z*-verdi på -1.747. Av tabell finner vi at sannsynligheten for å få en slik (eller mer ekstrem) *z*-verdi er 0.040. Dette er mindre enn $\alpha = 0.05$. Derfor forkaster vi H_0.



Dersom vi benytter konfidensintervall, tar vi utgangspunkt i den *observerte* verdien $\hat{p} = 0.26$. Estimatet av standardavviket blir nå

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.26 \cdot 0.74}{400}} = \underline{0.0219}$$

som er litt forskjellig fra den verdien vi fikk i sted.

For å få et konfidensintervall som kan sammenliknes med hypotesetesten, må vi benytte $\alpha = 0.10$ slik at $\alpha/2 = 0.05$ og $z_{\alpha/2} = 1.645$. Konfidensintervallet for X-partiets oppslutning den dagen spørreundersøkelsen ble utført, er

$$\left[\hat{p} - z_{\alpha/2}\hat{\sigma}_{\hat{p}},\ \hat{p} + z_{\alpha/2}\hat{\sigma}_{\hat{p}}\right] = \left[0.26 - 1.645 \cdot 0.0219,\ 0.26 + 1.645 \cdot 0.0219\right] = \underline{\left[0.224,\ 0.296\right]}.$$

Vi ser at $p = 0.30$ ligger utenfor dette konfidensintervallet. Da konkluderer vi med at på dette signifikansnivået er oppslutningen om X-partiet lavere enn $p = 0.30$.

Et par sluttmerknader: Strengt tatt burde vi benyttet binomisk fordeling, og undersøkt sannsynligheten for å få 104 *eller færre* X-parti-velgere når $n = 400$ og $p = 0.30$. Men betingelsen for at binomisk fordeling er tilnærmet lik normalfordelingen er klart oppfylt, slik at det er enklere å benytte normalfordeling enn binomisk fordeling (i alle fall dersom man ikke har dataverktøy til disposisjon). Prinsipielt burde vi også benyttet kontinuitetskorreksjon når vi går over fra binomisk til normalfordeling, men det er ikke vanlig å gjøre det fordi denne korreksjonen som regel har liten betydning.

Vi har testet om $p < p_0$ der $p_0 = 0.30$ i vårt eksempel. I andre situasjoner kan vi teste om $p > p_0$. Begge disse situasjonene fører til en *ensidig test* fordi vi bare bruker den ene ”halen” av normalfordelingen. Men noen ganger tester vi om $p \neq p_0$. Da må vi bruke *tosidig test* der vi plasserer $\alpha/2$ i hver ”hale” på samme måte som når vi konstruerer konfidensintervall.

Vi oppsummerer teknikken: Når vi ønsker å teste andel *p* i en populasjon, trekker vi *n* objekter tilfeldig fra populasjonen. Anta at *x* av disse har ønsket egenskap. Da er

$$\hat{p} = \frac{x}{n}.$$

Dersom betingelsene for normaltilnærmelse er oppfylt, bruker vi testobservatoren

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}.$$

Forkast nullhypotesen dersom *z* er mer ekstrem enn z_α ved ensidig test, mer ekstrem enn $z_{\alpha/2}$ ved tosidig test.

7.4. Test av populasjonens middelverdi.

Også nå går vi rett løs på et praktisk problem: En storforbruker av en spesiell type snacks har mistanke om at det er for lite i de posene hun kjøper. De skal inneholde 500 gram hver. Hun veier innholdet i 12 tilfeldig valgte poser med en nøyaktig vekt, og finner disse verdiene (gitt i gram):

493, 482, 502, 493, 467, 496, 510, 508, 493, 476, 489, 501.

Ved hjelp av Excel finner hun at gjennomsnittsvekten av disse 12 posene er $\bar{x} = 492.5$ gram, og at standardavviket for vektene er $s = 12.67$ gram. Kan hun på dette grunnlaget påstå at det er for lite i posene? Testen skal gjennomføres med signifikansnivå $\alpha = 0.02$.

En liten presisering: Vi antar at de 12 posene er et tilfeldig utvalg fra en populasjon som består av alle de posene som omsettes. Vi skal prøve å si noe om hele populasjonen, ikke bare om de 12 posene som hun har veid opp.

Vi starter med å sette opp hypotesene. Siden vi skal undersøke om posene er undervektige, blir nullhypotesen at de posene som omsettes i gjennomsnitt holder korrekt vekt. Den alternative hypotesen blir da at posene er undervektige.

H_0: De posene som omsettes inneholder i gjennomsnitt minst 500 gram ($\mu = 500$).
H_1: De posene som omsettes inneholder i gjennomsnitt mindre enn 500 gram ($\mu < 500$).

Den naturlige testobservatoren er

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}},$$

der

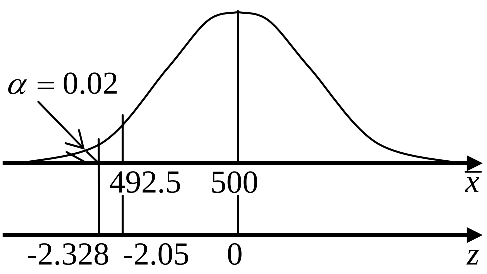
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

Her er σ standardavviket for vekten av alle posene i populasjonen, og *n* er antall poser som inngår i beregningen av $\bar{x}$. Men her dukker det opp to problemer:

1. For å kunne beregne $\sigma_{\bar{x}}$, må du kjenne σ. Det gjør du ikke.
2. Er det sikkert at $\bar{x}$ er normalfordelt?

La oss ta det siste problemet først. Vi vet at dersom vektene i populasjonen er normalfordelt, så er også middelverdien det. Selv om disse vektene ikke er normalfordelt, vil *middelverdien* nærme seg en normalfordeling når *n* øker. Det er rimelig å anta at vektene i populasjonen er såpass nær normalfordelt at $\bar{x}$ vil være normalfordelt selv om den magiske grensen på $n \geq 30$ ikke er passert.

Problem 1 er verre. Vi er nødt til å bruke det *beregnede* standardavviket som estimat for σ. Men da innfører du en ekstra usikkerhet. Denne usikkerheten kompenserer du for med å bruke *t*-fordeling istedenfor standard normalfordeling på samme måte som da du konstruerte konfidensintervall. Med $n = 12$, får du 11 frihetsgrader.



Vi summerer opp: Som testobservator bruker vi

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

der

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{12.67}{\sqrt{12}} = \underline{3.66}.$$

Under H_0 er $\mu = 500$, slik at

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}} = \frac{492.5 - 500}{3.66} = \underline{-2.05}.$$

Vårt signifikansnivå $\alpha = 0.02$ står ikke i tabellen i boka. For å finne kritisk område, bruker vi Excel-funksjonen

=TINV(0,04;11)

der vi må bruke $2\alpha = 2 \cdot 0.02 = 0.04$ som sannsynlighet fordi Excel bruker tosidig *t*-fordeling. Resultatet blir kritisk område når $t \leq -2.328$. Vår observator ligger ikke i det kritiske området, slik at vi kan ikke forkaste nullhypotesen. På vårt signifikansnivå har vi derfor ikke grunnlag til å hevde at posene som omsettes er undervektige.

Dersom vi bruker *P*-verdi, får vi *P*-verdien til $t = -2.05$ med Excel-funksjonen
=TFORDELING(2,05;11;1)
der tallene står for:
2.05 er absoluttverdien av *t*-verdien.
11 er antall frihetsgrader.
1 angir at vi bruker ensidig verdi.
Vi får en *P*-verdi på 0.0325. Dette er større enn vårt signifikansnivå på $\alpha = 0.02$. Vi kan ikke forkaste nullhypotesen.

Vi summerer opp: Dersom populasjonens standardavvik σ er kjent, og $\bar{x}$ er normalfordelt, brukes testobservatoren

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$ og standard normalfordelingstabell.

Dersom populasjonens standardavvik σ *ikke* er kjent, og $\bar{x}$ er normalfordelt, brukes testobservatoren

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$ og *t*-fordelingstabell med $n-1$ frihetsgrader,

der *s* er utvalgets standardavvik. Forkast nullhypotesen dersom testobservatoren er mer ekstrem enn z_α (eller t_α) ved ensidig test, mer ekstrem enn $z_{\alpha/2}$ (eller $t_{\alpha/2}$) ved tosidig test.

7.5. Test av varians.

Vår snacks-elskende venninne var lite fornøyd med resultatet av testen ovenfor. Hun mente at det var alt for stor variasjon i vektene av snacks-posene, og at denne variasjonen førte til at nullhypotesen ikke kunne forkastes. Hun kontaktet derfor produsenten, som svarte at det ”av pakke-tekniske grunner” ikke var mulig å få alle posene like tunge. Men etter produsentens egne målinger var standardavviket for posevektene høyst 8.0 gram.

Vår venninne vil nå teste denne påstanden. Hun fant at vekten til de 12 posene hadde et standardavvik på $s = 12.67$ gram. Vi setter da opp disse hypotesene:

H_0: Vekten til de posene som omsettes har standardavvik $\sigma \leq 8.0$ gram.
H_1: Vekten til de posene som omsettes har standardavvik $\sigma > 8.0$ gram.

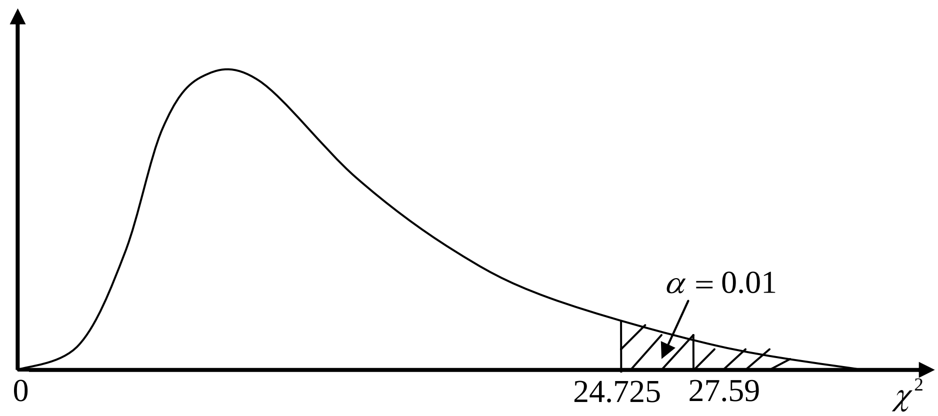
For å teste varians, må vi bruke kji-kvadrat-fordelingen. Den krever at dataene er normal-fordelt, og er ganske følsom dersom denne betingelsen ikke er oppfylt. Mens vi under testen av middelverdi tok lett på kravet om at vektene til posene som omsettes er normalfordelt, må vi her forutsette at dette kravet er oppfylt. Under denne forutsetningen er

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

kji-kvadrat-fordelt med $n-1$ frihetsgrader. Det er derfor naturlig å bruke denne størrelsen som testobservator. For sikkerhets skyld (i tilfelle dataene ikke er helt normalfordelt) velger vi et signifikansnivå $\alpha = 0.01$. Av tabell finner vi at med 11 frihetsgrader er kritisk område $\chi^2 \geq 24.725$. Vi får

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{11 \cdot 12.67^2}{8.0^2} = \underline{27.59}\,.$$

Denne verdien befinner seg i det kritiske området, slik at vi må forkaste nullhypotesen. Dataene tyder på at standardavviket for vekten av posene er større enn 8.0 gram slik at variasjonen i posevekt er for stor.



Ved hjelp av Excel kan vi også finne *P*-verdi. Vi benytter da funksjonen

=KJI.FORDELING(27,59;11)

og får *P*-verdien 0.00374. Dette er klart lavere enn vår verdi $\alpha = 0.01$.

Oppsummering:

Dersom variansen s^2 er beregnet for et utvalg på *n* objekter fra en *normalfordelt* populasjon med populasjonsvarians σ^2, brukes testobservatoren

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$ som er kji-kvadrat-fordelt med $n-1$ frihetsgrader.

Forkast nullhypotesen dersom denne testobservatoren er større enn den kritiske verdien.