LATVIJAS UNIVERSITĀTE
FIZIKAS UN MATEMĀTIKAS FAKULTĀTE
MATEMĀTIKAS NODAĻA

STATISTISKIE TOLERANCES INTERVĀLI

DIPLOMDARBS

Autors: **Margarita Fjodorova**

Stud. apl. mf05002

Darba vadītājs: doc. Dr.math. Jānis Valeinis

RĪGA 2010

Anotācija

Šajā diplomdarbā tiek aplūkoti vienpusējie un divpusējie tolerances intervāli parametriskajā un neparametriskajā gadījumā. Šo intervālu pielietojums praksē atšķiras no klasiskā ticamības intervālu pielietojuma. Ja ir nepieciešams noteikt, kāda procentuāla izlases daļa iekļausies kādās specifiskās robežās, tad tolerances intervāls spēj sniegt atbildi uz šo jautājumu. Ar piemēru palīdzību tiek attēloti dažādie tolerances intervālu pielietojumi. Šie intervāli var veiksmīgi tikt pielietoti dažādiem sadalījumiem. Aprēķinu veikšanai tiek izmantota brīvpieejas programma R.

Atslēgvārdi: Tolerances robežas, tolerances intervāli.

Abstract

This thesis deals with one-sided and two-sided tolerance intervals in the parametric and in the nonparametric case. The problem of tolerance intervals differs from that of classical confidence intervals. If there is a need to estimate what percental part of the sample will be included in some specific limits, then tolerance intervals can solve this problem. It is shown how to use tolerance intervals in different practical examples. Tolerance intervals can be successfully applied to various distributions. Calculations are performed using the open source program R.

Keywords: Tolerance limits, tolerance intervals.

Saturs

Ievads

Statistikā viens no pamatjautājumiem ir ticamības intervālu konstruēšana, kas parasti sniedz informāciju par kādu nezināmu populācijas parametru, piemēram, vidējo vērtību vai standartnovirzi. Tomēr ir gadījumi, kad ticamības intervāls nespēj sniegt pietiekoši daudz informācijas par visu populāciju kopumā. Kā izrādās, šādiem uzdevumiem ir paredzēti tolerances intervāli, kas parasti lekciju kursos netiek apskatīti.

Tolerances intervāliem ir daudzveidīgs pielietojums. Tos izmanto klīniskajos un rūpnieciskajos pētījumos, tādos kā kvalitātes kontrole, apkārtējās vides monitorings jeb vides kontrole un citur (sīkākas detaļas, uzskatāmi piemēri un pielietojumi pieejami autora Gibbons 1994.gada grāmatā [1], Gibbons un Coleman 2001.gada grāmatā [2], kā arī autoru Millard un Neerchal 2000.gada grāmatā [3]).

Tolerances robežas un tolerances intervāli pirmo reizi tika pieminēti jau 1941.gadā un 1942.gadā autora Wilks publikācijās [4] un [5]. Autors uzsvēra tolerances intervālu lietderīgumu industriālās ražošanas kvalitātes kontroles pētniecībā, parādīja atšķirību starp vienpusējo un divpusējo tolerances intervālu. Piemēram, ja, ražojot tērauda stieples, jānosaka spēks, kuru piemērojot, šīs stieples lūzīs, tad nozīmīgākā būs tieši apakšējā tolerances robeža. Šajā gadījumā apskata vienpusējo tolerances intervālu (sīkāk skatīt [4]). Pēc 1941.gada vairāk kā pusgadsimta garumā tolerances intervāli tika pielietoti dažādu problemātiku risināšanai. Sava diplomdarba rakstīšanai par teorētisko pamatu izmantoju 2009.gadā izdotu grāmatu [6], kas apliecina šī jautājuma aktualitāti.

Tiek gaidīts, ka tolerances intervāls aptvers vismaz noteiktu populācijas daļu, pie izvēlēta ticamības līmeņa. Piemēram, augšējā tolerances robeža ir tāda, ka ar dotu ticamības līmeni vismaz kāda specifiski izvēlēta populācijas procentuālā daļa atradīsies zem šīs robežas. Apakšējā tolerances intervāla robeža vai tolerances intervāls ar abām robežām tiek meklēti pēc līdzīgiem nosacījumiem.

Tolerances intervāli var tikt piemēroti dažādiem sadalījumiem, bet praksē visbiežāk pielietotais sadalījums ir normālais sadalījums, tādēļ sīkāk tiek apskatīti vienpusējie un divpusējie tolerances intervāli normālajam sadalījumam. Papildus tiek analizēti lognormālais un gamma sadalījumi. Tolerances intervālu konstruēšanai var tikt pielietotas parametriskās un neparametriskās metodes. Diplomdarbā tiks apskatītas abu veidu metodes, kā arī salīdzināts to praktiskais pielietojums.

Diplomdarba mērķi ir:

- Ticamības intervālu un tolerances intervālu salīdzināšana;
- Teorētiskā materiāla izpēte, analīze un tā pielietošana praksē;
- Tolerances intervālu atšķirīgu praktisko pielietojumu izpēte;
- Parametrisko un neparametrisko metožu salīdzināšana.

Diplomdarbā ir 5 galvenās nodaļas un pielikumi. 1.nodaļā īsā ievadā tiek parādītas galvenās atšķirības starp ticamības un tolerances intervāliem. Tālākās apakšnodaļās tiek iepazīstināts ar to, kas ir augšējā un apakšējā tolerances robeža, parādīta sakarība starp šīm robežām un kvantilēm. Sākumā aplūko vienpusējo tolerances intervālu. Līdzīgi arī tiek apskatīta vispārīga teorija par divpusējo tolerances intervālu. Nākošajā apakšnodaļā tiek novērtēta izdzīvošanas un pārsniegšanas varbūtība. 2.nodaļā tiek aplūkota situācija, kad dati atbilst lognormālajam vai gamma sadalījumam, un, kā ar datu transformācijas palīdzību var panākt pilnīgu vai tuvinātu datu atbilstību normālajam sadalījumam. 3.nodaļā tiek aplūkoti tolerances intervāli lineārās regresijas gadījumā. 4.nodaļā apkopots teorētiskais materiāls par neparametriskajiem tolerances intervāliem. Nodaļas sākumā tiek īsi pastāstīts, kādos gadījumos ir nepieciešams izmantot tieši neparametrisko tolerances intervālu. 5.nodaļā tiek analizēti 5 atšķirīgi piemēri, lai uzskatāmi parādītu tolerances intervālu lietderīgumu praksē. Tiek attēloti svarīgākie rezultāti, kas iegūti ar programmas R palīdzību, pielietojot teorētiskās formulas, turklāt, lai pārliecinātos, ka šie aprēķini sniedz precīzus rezultātus, tika atrasta speciāla pakete tolerances intervālu aprēķiniem ar attiecīgu nosaukumu *tolerance*. Tā tika izveidota 2009.gadā (visa nepieciešamā informācija pieejama [7]). Pielikumā atrodas pamatformulas tolerances intervālu konstruēšanai normālajam sadalījumam, tabulas izlases apjoma noteikšanai, kā arī izveidoto programmu kods.

1. Tolerances intervāli un izdzīvošanas varbūtība

Situācijā, kad uzdevums ir noteikt kāda novērtētā izlases parametra precizitāti, tiek pielietoti ticamības intervāli. Šie intervāli ar noteiktu ticamību (piemēram, 95%, 99%) ļauj noteikt, kādu vērtību robežās atradīsies īstā parametra vērtība. Bet situācijā, kad uzdevums ir noteikt, kādu vērtību robežās atradīsies ne mazāk kā noteikta procentuālā izlases daļa ar noteiktu ticamības līmeni, tad tiek pielietoti statistiskie tolerances intervāli.

1.1. Vienpusējie tolerances intervāli

Šajā nodaļā pārsvarā tiks izmantots teorētiskais materiāls no grāmatas [6].

Pieņem, ka X ir nepārtraukts gadījuma lielums ar kumulatīvo sadalījuma funkciju $F_X(x) = P(X \leq x)$. Pie dota skaitļa p, kur $(0 < p < 1)$, inversā kumulatīvā sadalījumam funkcija dota:

$$F_X^{-1}(p) = inf\{x : F_X(x) \geq p\}. \tag{1.1.1}$$

Lielums $F_X^{-1}(p)$ ir p - tā kvantile jeb $100p$ procentīle sadalījumam F_X. Kvantilei p lieto apzīmējumu q_p. Jāievēro, ka populācijas proporcija p (attiecībā pret sadalījuma funkciju F_X) ir mazāka vai vienāda ar q_p. Ja $F_X(x)$ ir stingri augoša funkcija no x, tad $F_X^{-1}(p)$ ir tāda x vērtība, kurai $F_X(x) = P(X \leq x) = p$.

Pieņemsim, ka $X_1, X_2, \ldots, X_n$ ir gadījuma izlase no F_X, lietosim pierakstu $\boldsymbol{X} = (X_1, X_2,$
$\ldots, X_n)$. Lai definētu tolerances intervālu, ir jāprecizē tā procentuālo apjomu un ticamību, ko attiecīgi apzīmē ar p un $1 - \alpha$, un tolerances intervāls tiek uzskatīts par procentuālā apjoma p un $(1 - \alpha)$ pārklājuma (vai apjoma p un $(1 - \alpha)$ ticamību) tolerances intervāls vai vienkārši kā $(p, 1 - \alpha)$ tolerances intervāls $(0 < p < 1,\ 0 < \alpha < 1)$. Praktiskos

pielietojumos p un $1-\alpha$ parasti pieņem vērtības no kopas $\{0.90, 0.95, 0.99\}$. Intervāls tiek konstruēts, izmantojot gadījuma izlasi $\boldsymbol{X}$, un tiek prasīts, lai tas satur ne mazāk kā proporciju p no izvēlētas populācijas ar ticamības līmeni $1-\alpha$. Formāli, vienpusējam tolerances intervālam $(p, 1-\alpha)$ formā $(-\infty, U(\boldsymbol{X})]$ jāatbilst nosacījumam

$$P_{\boldsymbol{X}}\left\{P_X\left(X \leq U(\boldsymbol{X})\middle|\boldsymbol{X}\right) \geq p\right\} = 1-\alpha, \tag{1.1.2}$$

kur ar $P_{\boldsymbol{X}}$ apzīmējām "ārējo" varbūtību, bet ar P_X apzīmējām "iekšējo" varbūtību. X un sekojoši F_X ir neatkarīgi no $\boldsymbol{X}$. Tas ir, $U(\boldsymbol{X})$ tiek noteikts tā, ka vismaz proporcija p ir mazāka vai vienāda ar $U(\boldsymbol{X})$ ar ticamību $1-\alpha$. Intervāls $(-\infty, U(\boldsymbol{X})]$ tiek saukts par vienpusējo augšējo tolerances robežu. Var atzīmēt, ka, pamatojoties uz p kvantiles q_p definīciju (1.1.1), (1.1.2) var pierakstīt sekojoši

$$P_{\boldsymbol{X}}\{q_p \leq U(\boldsymbol{X})\} = 1-\alpha. \tag{1.1.3}$$

No (1.1.3) redzam, ka $U(\boldsymbol{X})$ ir $1-\alpha$ augšējā ticamības robeža p kvantilei q_p.
$(p, 1-\alpha)$ vienpusējā apakšējā tolerances robeža $L(\boldsymbol{X})$ tiek definēta līdzīgi. Precīzāk, $L(\boldsymbol{X})$ nosaka sekojošā veidā

$$P_{\boldsymbol{X}}\left\{P_X\left(X \geq L(\boldsymbol{X})\middle|\boldsymbol{X}\right) \geq p\right\} = 1-\alpha,$$

vai ekvivalenti,

$$P_{\boldsymbol{X}}\{L(\boldsymbol{X}) \leq q_{1-p}\} = 1-\alpha.$$

$L(\boldsymbol{X})$ ir $1-\alpha$ apakšējā ticamības robeža kvantilei q_{1-p}.

1.2. Divpusējie tolerances intervāli

Eksistē divu veidu divpusējie tolerances intervāli. Pirmā veida intervālu konstruē tā, lai tas saturētu vismaz populācijas proporciju p ar ticamību $1-\alpha$, un tiek vienkārši uzskatīts par tolerances intervālu. Otrā veida tolerances intervāls tiek konstruēts tā, ka tam būtu jāsatur vismaz populācijas proporcija p no populācijas centra ar ticamību $1-\alpha$, un tas parasti tiek minēts kā vienādastu tolerances intervāls.
$(p, 1-\alpha)$ divpusējam tolerances intervālam $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ izpildās nosacījums

$$P_{\boldsymbol{X}}\left\{P_X\left(L(\boldsymbol{X}) \leq X \leq U(\boldsymbol{X})\middle|\boldsymbol{X}\right) \geq p\right\} = 1-\alpha, \tag{1.2.1}$$

vai ekvivalenti,

$$P_{\boldsymbol{X}}\{F_X(U(\boldsymbol{X})) - F_X(L(\boldsymbol{X})) \geq p\} = 1 - \alpha. \tag{1.2.2}$$

Citiem vārdiem, intervāls $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ tiek konstruēts tā, ka tas satur vismaz proporciju p no populācijas ar ticamību $1-\alpha$. Lielumi $L(\boldsymbol{X})$ un $U(\boldsymbol{X})$ tiek uzskatīti par tolerances robežām. $L(\boldsymbol{X})$ un $U(\boldsymbol{X})$ rēķināšana nereducējas uz ticamības robežu aprēķināšanu noteiktām procentīlēm.

Lai noteiktu vienādo astu tolerances intervālu, pieņem, ka $p > 0.5$. $(p, 1-\alpha)$ vienādo astu tolerances intervāls $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ ir tāds, ka ar ticamību $1-\alpha$ ne vairāk kā proporcija $\frac{1-p}{2}$ no populācijas ir mazāka par $L(\boldsymbol{X})$ un ne vairāk kā proporcija $\frac{1-p}{2}$ no populācijas ir lielāka par $U(\boldsymbol{X})$. Šo nosacījumu var uzrakstīt arī ar procentīļu palīdzību. Ievēro, ka nosacījums $L(\boldsymbol{X}) \leq q_{\frac{1-p}{2}}$ ir ekvivalents ar nosacījumu, ka ne vairāk kā proporcija $\frac{1-p}{2}$ no populācijas mazāka par $L(\boldsymbol{X})$, un nosacījums $q_{\frac{1+p}{2}} \leq U(\boldsymbol{X})$ ir ekvivalents ar nosacījumu, ka proporcija $1 - \frac{1+p}{2} = \frac{1-p}{2}$ no populācijas ir lielāka par $U(\boldsymbol{X})$. Konsekventi, intervālam $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ ir jābūt par $(p, 1-\alpha)$ vienādo astu tolerances intervālu un jāizpildās nosacījumam

$$P_{\boldsymbol{X}}\Big(L(\boldsymbol{X}) \leq q_{\frac{1-p}{2}} \;\text{ un }\; q_{\frac{1+p}{2}} \leq U(\boldsymbol{X})\Big) = 1 - \alpha. \tag{1.2.3}$$

1.3. Izdzīvošanas varbūtības novērtējums

Dažos praktiskos pielietojumos tiek prasīts novērtēt varbūtību, ka gadījuma lielums pārsniegs kādu konkrētu vērtību. Piemēram, ilgdzīvošanas datu analīzē, ir svarīgi novērtēt varbūtību, ka objekta dzīves ilgums pārsniegs noteiktu vērtību. Tāda varbūtība parasti tiek uzskatīta par izdzīvošanas varbūtību. Gadījumā, kad tiek veikta darba drošības kontrole kādā rūpnīcā, ir svarīgi novērtēt, vai kaitīgo vielu iedarbība uz strādnieku nepārsniedz noteiktu robežu (parasti to nosaka darba drošības un veselības administrācija). Šo uzskata par *pārsniegšanas varbūtību*. Šādā gadījumā mēs esam ieinteresēti noteikt tieši apakšējo tolerances robežu.

Pieņem, ka X ir nepārtraukts gadījuma lielums ar sadalījuma funkciju $F_X(x)$. Dotam t definē izdzīvošanas varbūtību $S_t = P(X > t) = 1 - F_X(t)$. Pieņem, ka $\boldsymbol{X}$ ir izlase no F_X, un $L(\boldsymbol{X}) = L(\boldsymbol{X}; p)$ ir $(p, 1-\alpha)$ apakšējā tolerances robeža X sadalījumam. Ja ir dota apakšējā tolerances robeža, tad

$$P_{\boldsymbol{X}}\{P_X(X \geq L(\boldsymbol{X}; p) | \boldsymbol{X}) \geq p\} = 1 - \alpha.$$

Tas ir,

$$P_{\boldsymbol{X}}\Big\{S_{L(\boldsymbol{X};p)} \geq p\Big\} = 1 - \alpha.$$

Ja $L(\boldsymbol{X};p) \geq t$, tad acīmredzami iegūst $S_t \geq S_{L(\boldsymbol{X};p)}$. Turklāt, ja $S_{L(\boldsymbol{X};p)} \geq p$, varam secināt, ka $S_t \geq p$, ja $L(\boldsymbol{X};p) \geq t$. Tātad tā ir maksimālā p vērtība, kurai $L(\boldsymbol{X};p) \geq t$ dod S_t apakšējo robežu $1-\alpha$, un apzīmējam ar p_l. Tas ir,

$$p_l = \max\{p : L(\boldsymbol{L};p) \geq t\}. \tag{1.3.1}$$

Vispārīgi runājot, $L(\boldsymbol{X};p)$ ir dilstoša funkcija no p, un tātad maksimums (1.3.1) tiek sasniegts, kad $L(\boldsymbol{X};p) = t$. Tas ir, p_l ir atrisinājums vienādībai $L(\boldsymbol{X};p) = t$. Apakšējā tolerances robeža var tikt izmantota arī vienpusējos hipotēžu testos attiecībā uz S_t. Tas ir, ja veicam testu

$$H_0 : S_t \leq p_0 \quad \text{un} \quad H_a : S_t > p_0$$

pie līmeņa α, tad H_0 tiks noraidīta, ja $(p_0, 1-\alpha)$ apakšējā tolerances robeža ir lielāka par t.

Augšējā ticamības robeža pārsniegšanas varbūtībai tiek bieži pielietota, lai novērtētu kaitīgo vielu ietekmi darba vietā. Piemēram, ja t apzīmē kaitīgo vielu iedarbības kritisko līmeni un X apzīmē vielu ietekmes mērvienību uz strādnieku, tad pārsniegšanas varbūtība tiktu definēta kā $P(X > t)$. Ja $U(\boldsymbol{X};p)$ ir $(p, 1-\alpha)$ augšējā tolerances robeža, un tā ir mazāka vai vienāda ar t, tad mēs varam secināt, ka $P(X > t)$ ir mazāka par $1-p$. Līdzīgi kā (1.3.1), secinām, ka, ja $p_u = \max\{p : U(\boldsymbol{X};p) \leq t\}$, tad $1-p_u$ ir $1-\alpha$ augšējā ticamības robeža pārsniegšanas varbūtībai. Vispārīgi runājot, $U(\boldsymbol{X};p)$ ir nedilstoša funkcija no p, un p_u ir atrisinājums vienādībai $U(\boldsymbol{X};p) = t$.

2. Tolerances intervāli normālajam, lognormālajam un gamma sadalījumam

Normālais sadalījums ir visbiežāk praktiski pielietotais sadalījums. Pirmās publikācijas par tolerances intervālu konstruēšanu normālajam sadalījumam tika publicētas 1941. un 1942. gadā (autors Wilks [4],[5]), 1943. gadā (autors Wald [8]) un 1946. gadā (autori Wald un Wolfowitz [9]). Praktiskajos piemēros bieži var sastapties ar situāciju, kad dati neatbilst normalitātes nosacījumiem. Šāda situācija var rasties, ja tiek izmantoti, piemēram, dzīves ilguma dati vai ienākumu dati. Šādos gadījumos tolerances intervāli normālajam sadalījumam var tikt piemēroti kādam citam sadalījumam, ja tam ir precīza vai vismaz tuvināta atbilstība normālajam sadalījumam. Piemēram, ja X atbilst lognormālajam sadalījumam, tad $ln(X)$ atbilst normālajam sadalījumam. Ja X atbilst gamma sadalījumam, tad $X^{\frac{1}{3}}$ tuvināti atbilst normālajam sadalījumam.

2.1. Vienpusējās tolerances robežas normāli sadalītai populācijai

Pieņem, ka $X_1, \ldots, X_n$ ir izlase no $N(\mu, \sigma^2)$ populācijas ar nezināmu vidējo vērtību μ un nezināmu dispersiju σ^2. Izlases vidējā vērtība $\bar{X}$ un izlases dispersija S^2 ir definētas sekojoši

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{un} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2. \tag{2.1.1}$$

Ar z_p apzīmē p kvantili standarta normālajam sadalījumam. Tad p-tā kvantile no $N(\mu, \sigma^2)$ ir formā

$$q_p = \mu + z_p\sigma.$$

$1-\alpha$ augšējā ticamības robeža kvantilei q_p ir $(p, 1-\alpha)$ vienpusējā augšējā tolerances robeža normāli sadalītai populācijai. Parasti praksē tiek prasīta augšējā robeža kvantilei q_p, ja $p > 0.5$, bet apakšējā robeža, ja $p < 0.5$.

Klasiskā pieeja[6]

Pieņemsim, ka $(p, 1-\alpha)$ augšējā tolerances robeža ir formā $\bar{X} + k_1 S$, kur S - standartnovirze. Faktors k_1 tiek saukts par *tolerances faktoru*, to nosaka tā, ka vismaz proporcija p no populācijas mērījumiem ir mazāka par $\bar{X} + k_1 S$ ar ticamību $1-\alpha$. Tas ir,

$$P_{\bar{X},S}\{P(X < \bar{X} + k_1 S | \bar{X}, S) > p\} = 1 - \alpha, \tag{2.1.2}$$

kur $X \sim N(\mu, \sigma^2)$. Tad $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$, $Z_n = \frac{\bar{X}-\mu}{\sigma} \sim N(0, \frac{1}{n})$, $U^2 = \frac{S^2}{\sigma^2} \sim \frac{\chi^2_{n-1}}{n-1}$, kur χ^2_m apzīmē hī-kvadrāta gadījuma lielumu ar m brīvības pakāpēm, (2.1.2) var pierakstīt sekojoši

$$P_{Z_n,U}\{P(Z < Z_n + k_1 U | Z_n, U) > p\} = P_{Z_n,U}\{\Phi(Z_n + k_1 U) > p\} = 1 - \alpha. \tag{2.1.3}$$

Tā kā $Z_n \sim N(0, \frac{1}{n})$ neatkarīgi no $U^2 \sim \frac{\chi^2_{n-1}}{n-1}$, var pielietot rezultātu (A0.2), kur $c = \frac{1}{n}$, $\gamma = 1 - \alpha$ un $m = n - 1$, lai iegūtu

$$k_1 = \frac{1}{\sqrt{n}} t_{n-1;1-\alpha}(z_p\sqrt{n}), \tag{2.1.4}$$

kur $t_{m,1-\alpha}(\delta)$ apzīmē $1-\alpha$ kvantili necentrālajam t sadalījumam ar brīvības pakāpēm m, un ar necentralitātes parametru δ. Visbeidzot, $(p, 1-\alpha)$ augšējā tolerances robeža ir dota ar

$$\bar{X} + k_1 S = \bar{X} + t_{n-1;1-\alpha}(z_p\sqrt{n})\frac{S}{\sqrt{n}}. \tag{2.1.5}$$

Tas pats faktors k_1 var tikt izmantots, lai iegūtu $(p, 1-\alpha)$ apakšējo tolerances robežu, kas dota ar $\bar{X} - k_1 S$.

Izdzīvošanas vai pārsniegšanas varbūtības novērtēšana normāli sadalītai populācijai[6]

Pieņemsim, ka $1-\alpha$ apakšējā ticamības robeža izdzīvošanas varbūtībai $S_t = P(X > t)$, kur X ir normāli sadalīts gadījuma lielums un t ir dots skaitlis (tā iegūšanu skatīt pielikumā A). Vērsīsim īpašu uzmanību uz to, ka $1-\alpha$ apakšējā ticamības robeža varbūtībai S_t ir atrisinājums (attiecībā uz p) vienādojumam

$$t_{n-1;1-\alpha}(z_p\sqrt{n}) = \frac{\bar{X}-t}{S/\sqrt{n}}. \tag{2.1.6}$$

Pie dota izlases lieluma, p, $1-\alpha$, $\bar{X}$, S un t, vērtība p, kurai atbilst nosacījums (2.1.6) var tikt iegūta, pirmkārt, risinot $z_p\sqrt{n}$ un tad, risinot rezultējošo p vienādojumu. Lielums S_t tiek uzskatīts par pārsniegšanas varbūtību, tā kā tā ir vienkārši varbūtība, ka X pārsniegs noteikto vērtību t.

2.2. Divpusējie tolerances intervāli normāli sadalītai populācijai

Tiek uzskatīts, ka kāds objekts ir piemērots paredzētajam mērķim, ja ar to saistītie mērījumi atrodas intervālā (L_l, L_u), kur L_l un L_u ir kāda specifiski noteikta apakšējā un augšējā robeža. Vispārīgi runājot, tehniskiem ražojumiem tiek prasīts, lai tie atbilstu kādiem specifiskiem nosacījumiem. Ja vairākums no objektiem (teiksim, proporcija p) lielā mērā atbilst noteiktajai prasībai, tad šie objekti tiek akceptēti. Šī vairākuma atbilstība var tikt noteikta, izmantojot atbilstošu divpusējo tolerances intervālu. Piemēram, ja $(0.95, 0.99)$ divpusējais tolerances intervāls ir ietverts (L_l, L_u), tad liela daļa tiks akceptēta. Tas ir tādēļ, ka vismaz 95% no šiem objektiem jeb produktiem iekrīt tolerances intervālā ar ticamību ne mazāku kā 99% , un tolerances intervāls ietilpst (L_l, L_u). Šajā sadaļā tiks aprakstītas metodes tolerances intervālu konstruēšanai, ka tie saturētu ne mazāk kā proporciju p no normāli sadalītas populācijas ar ticamību $1-\alpha$.

Divpusējais tolerances faktors k_2 tiek noteikts tā, lai intervāls $\bar{X} \pm k_2 S$ saturētu vismaz proporciju p no normāli sadalītas populācijas ar ticamību $1-\alpha$. Tas ir, k_2 nosaka sekojoši

$$P_{\bar{X},S}\{P_X(\bar{X}-k_2S \le X \le \bar{X}+k_2S|\bar{X},S) \ge p\} = 1-\alpha, \tag{2.2.1}$$

kur $X \sim N(\mu,\sigma^2)$. Iekšējā varbūtības nevienādība var tikt izteikta kā

$$P_X\Big(\frac{\bar{X}-\mu-k_2S}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{\bar{X}-\mu+k_2S}{\sigma}\Big) \ge p$$
$$\Leftrightarrow \Phi(Z_n+k_2U) - \Phi(Z_n-k_2U) \ge p, \tag{2.2.2}$$

kur Φ apzīmē standartu normālo kumulatīvo sadalījuma funkciju , $Z_n = \frac{\bar{X}-\mu}{\sigma} \sim N(0,\frac{1}{n})$ neatkarīgi no $U^2 = \frac{S^2}{\sigma^2} \sim \frac{\chi^2_m}{m}$ ar $m = n-1$. Izmantojot (2.2.2), mēs varam uzrakstīt (2.2.1) kā

$$P_{Z_n,U}(\Phi(Z_n + k_2U) - \Phi(Z_n - k_2U) > p) = 1 - \alpha. \tag{2.2.3}$$

tagad, pielietojot rezultātu (A0.4), kur $c = \frac{1}{n}$, varam redzēt, ka k_2 ir atrisinājums integrālajam vienādojumam

$$\sqrt{\frac{2n}{\pi}} \int_0^\infty P\left(\chi_m^2 > \frac{m\chi_{1;p}^2(z^2)}{k_2^2}\right) e^{-\frac{1}{2}nz^2} dz = 1 - \alpha. \tag{2.2.4}$$

Izmantojot rezultātu (A0.5) (ar $c = \frac{1}{n}$ un $\gamma = 1 - \alpha$), var atrast vienkāršu, bet nosacījumiem atbilstošu aproksimāciju

$$k_2 \simeq \left(\frac{m\chi_{1;p}^2(1/n)^2}{\chi_{m;\alpha}}\right)^{\frac{1}{2}}, \tag{2.2.5}$$

kur $\chi_{m;\alpha}^2$ apzīmē α kvantili Hī-kvadrāta sadalījumam ar m brīvības pakāpēm, un $\chi_{m;\alpha}^2(\delta)$ apzīmē α kvantili hī-kvadrāta sadalījumam ar brīvības pakāpēm m un necentralitātes parametru δ.

Iegūtā aproksimācija ir apmierinoša pat mazām izlasēm (apjomā 3), p un $1 - \alpha$ pieņem vērtības no kopas $\{0.9, 0.95, 0.99\}$. Lai parādītu šīs aproksimācijas precizitāti, var aprēķināt tolerances faktorus, kas apmierina (2.2.4) (skatīt pielikumā programmas kodu) un aproksimācijas, kas dotas (2.2.5), un n pieņem vērtības no 3 līdz 10, $p = 0.90, 0.95, 0.99$ un $1 - \alpha = 0.90, 0.95$. Šajā tabulā var redzēt, ka aproksimācija ir efektīva pat mazām

2.1. tabula Divpusējā tolerances faktora aproksimācija (a) un teorētiskais faktors (b).

	$1-\alpha=0.90$					
	p					
	0.90		0.95		0.99	
n	a	b	a	b	a	b
3	5.85	5.79	6.92	6.82	8.97	8.82
4	4.17	4.16	4.94	4.91	6.44	6.37
5	3.49	3.50	4.15	4.14	5.42	5.39
6	3.13	3.14	3.72	3.72	4.87	4.85
7	2.90	2.91	3.45	3.46	4.52	4.50
8	2.74	2.75	3.26	3.27	4.27	4.27
9	2.63	2.64	3.13	3.13	4.10	4.09
10	2.55	2.55	3.02	3.03	3.96	3.96

izlasēm. Ja $n \geq 10$, tad atšķirības starp precīzajiem un aproksimētajiem faktoriem retos gadījumos pārsniedz 0.01.

2.3. Vienādo astu tolerances intervāli normālajam sadalījumam

Šajā apakšnodaļā tiks aprakstīta metode tolerances intervāla (I_l, I_u) konstruēšanai, kas iekļautu vismaz $100p\%$ no normāli sadalītās populācijas "centra datiem" ar ticamību $1-\alpha$. Tas ir, intervāls (I_l, I_u) tiek konstruēts, izmantojot izlasi, tā, ka lielākā proporcija no normāli sadalītiem datiem, kas atrodas zem I_l ir $\frac{1-p}{2}$ un ka lielākā proporcija, kas atrodas virs I_u ir $\frac{1-p}{2}$, ar ticamību $1-\alpha$. Intervāls tiek meklēts tā, ka tas iekļautu intervālu $\left(\mu - z_{\frac{1+p}{2}}\sigma, \mu + z_{\frac{1+p}{2}}\sigma\right)$ ar ticamību $1-\alpha$. "Dabiski" par (I_l, I_u) izvēlēties $(\bar{X} - k_e S, \bar{X} + k_e S)$, kur k_e tiek noteikts sekojoši

$$P_{\bar{X},S}\Big(\bar{X} - k_e S < \mu - z_{\frac{1+p}{2}}\sigma \text{ un } \mu + z_{\frac{1+p}{2}}\sigma < \bar{X} + k_e S\Big) = 1 - \alpha. \tag{2.3.1}$$

Pēc $\bar{X}$ standartizēšanas un nosacījumu pārkārtošanas, varam redzēt, ka (2.3.1) ir ekvivalents ar

$$P_{Z,S}\left(\frac{Z/\sqrt{n} + z_{\frac{1+p}{2}}}{S/\sigma} < k_e \text{ un } \frac{Z/\sqrt{n} - z_{\frac{1+p}{2}}}{S/\sigma} \geq -k_e\right) = 1 - \alpha, \tag{2.3.2}$$

kur $Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0,1)$. Pieņemsim, ka $\delta = \sqrt{n} \times z_{\frac{1+p}{2}}$, un $U^2 = \frac{S^2}{\sigma^2}$. Izmantojot šos nosacījumus, varam (2.3.2) pierakstīt sekojoši

$$P_{Z,S}(Z < -\delta + k_e\sqrt{n}U \text{ un } Z > -\delta - k_e\sqrt{n}U) = 1 - \alpha. \tag{2.3.3}$$

Ievērosim, ka šīs nevienādības ir spēkā tad, ja $\delta - k_e\sqrt{n}U < -\delta + k_e\sqrt{n}U$ vai līdzvērtīgi $U^2 > \frac{\delta^2}{k_e^2 n}$. Tātad (2.3.3) var izteikt sekojoši

$$E_U\left[P_Z\left(\delta - k_e\sqrt{n}U < Z < -\delta + k_e\sqrt{n}U \middle| U^2 > \frac{\delta^2}{k_e^2 n}\right)\right] = 1 - \alpha, \tag{2.3.4}$$

kur E_U apzīmē matemātisko cerību attiecībā pret U. Zinot, ka $U^2 \sim \frac{\chi^2_{n-1}}{n-1}$, no (2.3.4) seko, ka k_e ir atrisinājums integrālvienādojumam

$$\frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}\int_{\frac{(n-1)\delta^2}{k_e^2 n}}^{\infty}\left(2\Phi\left(-\delta + \frac{k_e\sqrt{nx}}{\sqrt{n-1}}\right) - 1\right)e^{-x/2}x^{\frac{n-1}{2}-1}dx = 1 - \alpha, \tag{2.3.5}$$

kur $\Phi(x)$ apzīmē standarta normālo sadalījuma funkciju. Lai iegūtu (2.3.5) no (2.3.4), izmantojām sakarību $\Phi(x) = 1 - \Phi(-x)$.

2.4. Lognormālais sadalījums un gamma sadalījums

2.4.1. Lognormālais sadalījums

Dati no lognormālā sadalījuma var tikt veiksmīgi pielietoti metodēm, kas balstītas uz normālo sadalījumu, tiem tikai jāveic logaritmiskas transformācijas.

Gadījuma lielums Y ir lognormāli sadalīts (rakstīsim $Y \sim \log\mathrm{N}(\mu, \sigma^2)$), ja $X = \ln(Y) \sim N(\mu, \sigma^2)$. Varbūtību blīvuma funkcija no Y ir dota veidā

$$f(y|\,\mu, \sigma) = \frac{1}{\sqrt{2\pi}\, y\sigma} \exp\left[-\frac{(ln\, y-\mu)^2}{2\sigma^2}\right],\ y > 0,\ \sigma > 0,\ -\infty < \mu < \infty. \qquad (2.4.1)$$

$Y_1, \ldots, Y_n-$ izlase no $\log\mathrm{N}(\mu, \sigma^2)$ sadalījuma. $X_1 = \ln(Y_1), \ldots, X_n = \ln(Y_n)$ ir izlase no normālā sadalījuma ar vidējo vērtību μ un dispersiju σ^2. Tādejādi, pieejas, kas balstās uz normālo sadalījumu, var tik viegli pielietotas, konstruējot vienpusējās tolerances robežas, tolerances intervālus vai vienādo astu tolerances intervālus, kas balstīti uz izlasi $X_1, \ldots, X_n$. Lai to ilustrētu, atgādinam, ka

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{un} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

$\bar{X}+k_1\frac{S}{\sqrt{n}}$ ir vienpusējā augšējā tolerances robeža normālajam sadalījumam vai sadalījumiem, kas balstīti uz logtransformētām izlasēm. Tātad, $\exp\left(\bar{X} + k_1\frac{S}{\sqrt{n}}\right)$ ir vienpusējā augšejā tolerances robeža izvēlētajai lognormālai populācijai. Līdzīgi var iegūt tolerances vai venādastu tolerances intervālus. Turklāt, lai noteiktu izdzīvošanas laika brīdī t, jāatzīmē, ka $P(Y > t) = P(\ln(Y) > \ln(t)) = P(X > \ln(t))$, un rezultāti ar t aizvietotu ar $\ln(t)$ var tikt izmantoti. Īpaši, ja ir jāiegūst apakšējo ticamības robežu varbūtībai $P(Y > t)$.

2.4.2. Gamma sadalījums

Ja mainīgais atbilst gamma sadalījumam, tad šī mainīgā kubsakne atbilst tuvinātajam normālajam sadalījumam. Gamma sadalījums var tikt pielietots svarīgu praktisku problēmu un uzdevumu risināšanā, tādēļ šajā nodaļā tiks apskatīts atbilstošais teorētiskais materiāls.

Gamma saistītie sadalījumi tiek pielietoti reģiona diennakts nokrišņu daudzuma modelēšanā, un tiek piemēroti hidroloģiskajām datu kopām (skatīt [10], [11] un [12]). Divu parametru gamma tolerances robežas tiek pielietotas monitoringa un kontroles uzdevumos. Piemēram, vides monitoringā, augšejā tolerances robeža bieži tiek konstruēta, balstoties uz pamatdatiem (reģionālajiem virsējiem ūdeņiem, pazemes ūdeņiem vai gaisa monitoringa datiem) un tiek pielietota , lai noteiktu, vai potenciālais piesārņojuma avots (piemēram, izgāztuve, bīstamu materiālu pārvadāšana, rūpnīcas u.c.) ir postoši ietekmējis apkārtējo vidi (pielietojumu skatīt [13]).

Normālā aproksimācija Gamma sadalījumam[6]

Blīvuma funkcija gamma sadalījumam ar formas parametru a un mēroga parametru b ir dota

$$f(y|a,b) = \frac{1}{\Gamma(a)b^a} e^{-y/b} y^{a-1}, \quad y > 0,\ a > 0,\ b > 0. \tag{2.4.2}$$

Alternatīvā metode, kā gamma sadalījums var tikt parametrizēts ar formas parametru a un inverso mēroga parametru $r = 1/b$, ko dēvē par proporcijas jeb lieluma (angļu val. *rate*) parametru, ir sekojoša

$$g(x;a,r) = \frac{r^a}{\Gamma(a)} x^{a-1} e^{-rx}, \qquad x > 0,$$

ja a ir pozitīvs vesels skaitlis, tad

$$\Gamma(a) = (a-1)!.$$

Nav pieejama tieša metode, kā konstruēt tolerances intervālus gamma sadalījumam, bet literatūrā tiek piedāvāti vairāki veidi, kā to veiksmīgāk izdarīt. Praktiskajos piemēros tiks izmantota metode, kas ir netikai vienkārša pielietošanā, bet arī sniedz labus rezultātus. Šī metode ir balstīta uz normālo aproksimāciju kubsaknei no gamma sadalījuma mainīgā, ko 1931.gadā piedāvāja autori Wilson un Hilferty (metodes pilnīgu aprakstu skatīt [14]). Normālā aproksimācija kubsaknei no Hī-kvadrāta sadalījuma gadījuma lielumam, izmantojot momentu atbilstības metodi, tiek iegūta sekojošā veidā. Pieņem, ka Y_a ir $gamma(a,1)$ gadījuma lielums. Tā kā Y_a sadalīts kā $\frac{1}{2}\chi^2_{2a}$, ir jāpaskaidro momentu atbilstības pieeju Hī-kvadrāta sadalījuma mainīgā pakāpē λ aproksimēšanai. Atzīmējam, ka vidējā vērtība un dispersija mainīgajam Y_a^λ ir dota

$$\mu_\lambda = \frac{\Gamma(a+\lambda)}{\Gamma(a)} \quad \text{un} \quad \sigma^2_\lambda = \frac{\Gamma(a+2\lambda)}{\Gamma(a)} - \mu^2_\lambda. \tag{2.4.3}$$

λ vērtība ir $\frac{1}{3}$, kas pēc autoru uzskatiem ir vispiemērotākā izvēle. Šajā gadījumā $Y_a^{\frac{1}{3}} \sim N\left(\mu_{\frac{1}{3}}, \sigma^2_{\frac{1}{3}}\right)$.

Lai konstruētu tolerances intervālu sadalījumam $gamma(a, b)$, sākumā jāatzīmē, ka $Y_{a,b}$ apzīmē gamma gadījuma lielumu, un, ka $Y_{a,b}$ ir sadalīts kā bY_a. *Wilson - Hilferty* aproksimācija nosaka, ka lielums $Y_{a,b}^{\frac{1}{3}}$ ir tuvināts normālajam ar vidējo vērtību un dispersiju

$$\mu = \frac{b^{\frac{1}{3}}\Gamma(a + 1/3)}{\Gamma(a)} \quad \text{un} \quad \sigma^2 = \frac{b^{\frac{2}{3}}\Gamma(a + 2/3)}{\Gamma(a)} - \mu^2. \tag{2.4.4}$$

Funkcionālās formas no μ un σ^2 (kā funkcijas no a un b), konstruējot tolerances robežas, var tikt ignorētas, pieļaujot nenozīmīgas neprecizitātes. Ja $Y_1, \ldots, Y_n$ ir izlase no $gamma(a, b)$ sadalījuma, tiek aplūkota transformēta izlase $X_1 = Y_1^{\frac{1}{3}}, \ldots, Y_n^{\frac{1}{3}}$ kā izlase no normālā sadalījuma ar patvaļīgu vidējo vērtību μ un dispersiju σ^2, tiek iegūts tolerances intervāls kā no normālā sadalījuma.

Tolerances intervāli un izdzīvošanas varbūtība[6]

Pieņem, ka $Y_1, \ldots, Y_n$ ir izlase no $gamma(a, b)$ sadalījuma. Lai pielietotu *Wilson-Hilferty* tuvinājumu, pieņem, ka $X_i = Y_i^{\frac{1}{3}}, \quad i = 1, \ldots, n$. Ja U ir $(p, 1 - \alpha)$ tolerances robeža, kas balstīta uz $\bar{X}$ un S^2, tad U^3 ir tuvināta $(p, 1 - \alpha)$ augšējā tolerances robeža $gamma(a, b)$ sadalījumam. Atsaucoties, ka normālajam sadalījumam $(p, 1 - \alpha)$ augšējā tolerances robeža U ir dota

$$U = \bar{X} + k_1 S, \quad \text{ar} \quad k_1 = \frac{1}{\sqrt{n}} t_{n-1;1-\alpha}(z_p\sqrt{n}), \tag{2.4.5}$$

kur z_p ir p kvantile no standarta normālā sadalījuma, un $t_{m;\alpha}(\delta)$ apzīmē α kvantili necentrālajam t sadalījumam ar $df = m$ un necentralitātes parametru δ. Jāatzīmē, ka U^3 ir tuvināta $1 - \alpha$ augšējā ticamības robeža kvantilei p no $gamma(a, b)$ sadalījuma. Līdzīgi, $(p, 1 - \alpha)$ apakšejā tolerances robeža ir arī $1 - \alpha$ apakšejā ticamības robeža kvantilei $(1 - p)$. Tātad, šajā gadījumā augšējā un apakšejā tolerances robeža atbilst tuvinātām ticamības robežām attiecīgām percentīlēm no gamma sadalījuma. Ja tuvinātā tolerances robeža ir negatīva, tad robeža tiek ņemta vienāda ar nulli.

Lai iegūtu tuvināto divpusējo tolerances intervālu $gamma(a, b)$ sadalījumam, pieņem, ka $L = \bar{X} - k_2 S$ un $U = \bar{X} + k_2 S$, kur k_2 ir iegūts no (2.2.4), kā arī vērtības k_2 var nolasīt no tabulām vai aprēķināt pēc aproksimācijas (2.2.5). Intervāls (L^3, U^3) ir $(p, 1 - \alpha)$ divpusējais tolerances intervāls $gamma(a, b)$ sadalījumam.

Izdzīvošanas varbūtības noteikšana

Pieņemsim, ka gribam noteikt izdzīvošanas varbūtību laikā t, balstītu uz izlasi no dzīves ilguma datiem $Y_1, \ldots, Y_n$ no gamma sadalījuma. Izdzīvošanas varbūtība $S_t =$

$P(Y > t) = P(Y^{\frac{1}{3}} > t^{\frac{1}{3}}) = P(X > t^{\frac{1}{3}})$, tuvināti, kur X ir gadījuma lielums no normālā sadalījuma, normālā aproksimācijas metode var tikt pielietota, lai izdarītu secinājumus par S_t. Patiešām, tuvinātā apakšejā ticamības robeža varbūtībai S_t var tikt noteikta kā risinājums (attiecībā uz p) sekojošai izteiksmei

$$t_{n-1;1-\alpha}(z_p\sqrt{n}) = \frac{\bar{X} - t^{\frac{1}{3}}}{S_x/\sqrt{n}}, \tag{2.4.6}$$

kur $\bar{X}$ un S ir definēti (2.1.1). Jāatzīmē, ka vienādojums (2.4.6) ir vienāds ar (2.1.6), ja t aizvieto ar $t^{\frac{1}{3}}$.

3. Tolerances intervāli lineārās regresijas modelim

Par pamatu teorētiskajam materiālam ņemtas grāmatas [6] un [15], kā arī pielietotas publikācijas [16] un [17].

Regresiju modeļus pielieto, lai attēlotu attiecības starp atbildes jeb atkarīgo mainīgo un vairākiem neatkarīgajiem mainīgajiem jeb regresoriem. Tiek aplūkota grupa, kas sastāv no n vienībām, kur ar Y_i apzīmē i-to atbildes jeb atkarīgo mainīgo i-tajai vienībai, ar x_i apzīmē atbilstošu $m \times 1$ regresoru vektoru. Vienkāršajā lineārās regresijas modelī pieņem, ka Y_i vidējā vērtība ir $x_i'\boldsymbol{\beta}$, kur $\boldsymbol{\beta}$ ir $m \times 1$ vektors, kas sastāv no nezināmiem parametriem. Šajā gadījumā tiek uzskatīts, ka Y_i atbilst normālajam sadalījumam, ar dispersiju σ^2. Ja $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)'$ apraksta $n \times 1$ vektoru, kas sastāv no apsekojumiem, un $\boldsymbol{X}$ apzīmē $n \times m$ matricu, kuras i-tā rinda ir vienāda ar x_i', tad vienkāršo lineārās regresijas modeli, pieņemot, ka izpildās normalitātes nosacījumi, pieraksta sekojoši

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n), \tag{3.0.1}$$

kur $\boldsymbol{\varepsilon}$ apzīmē kļūdu vektoru, un $\sigma^2 > 0$ arī ir nezināms parametrs, bet $\sigma^2 I_n$ ir matrica, kas sastāv no kļūdu vektoriem, kur katra kļūda atbilst normālajam sadalījumam. Tā kā pirmais saskaitāmais ir konstante, tad $\boldsymbol{X}$ pirmā kolonna sastāv no vieniniekiem, un matricas $\boldsymbol{X}$ rangs $R(\boldsymbol{X}) = m$.

Uzskata, ka $Y(\boldsymbol{x})$ apzīmē nākotnes novērojumu, kas atbilst regresoru vektoram $\boldsymbol{x}$. Pieņem, ka

$$Y(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(0, \sigma^2), \tag{3.0.2}$$

kur $Y(\boldsymbol{x})$ ir neatkarīgs no $\boldsymbol{Y}$ formulā (3.0.1). $Y(\boldsymbol{x})$ tolerances intervāls $(p, 1-\alpha)$, pie fiksēta $m \times 1$ vektora $\boldsymbol{x}$, ir intervāls, kas satur vismaz proporciju p no $Y(\boldsymbol{x})$ sadalījuma ar ticamību $1-\alpha$. Tolerances intervāls tiek konstruēts, izmantojot novērojumu vektoru

$\boldsymbol{Y}$ formulā (3.0.1). Jāpiezīmē, ka šajā gadījumā $\boldsymbol{x}$ ir dots vektors.

Pieņemsim, ka $\hat{\boldsymbol{\beta}}$ apzīmē $\boldsymbol{\beta}$ novērtējumu, bet S^2 apzīmē atlikumu vidējo kvadrātu modelī (3.0.1). Lai novērtētu parametrus $\boldsymbol{\beta}$ pielietojam mazāko kvadrātu metodi, kas minimizē Y_i kvadrātisko kļūdu summu. Tātad

$$\varepsilon_i = Y_i - E(Y_i)$$

apzīmē kļūdu. Šī summa matricu veidā

$$RSS = \varepsilon'\varepsilon = (\boldsymbol{Y} - \boldsymbol{X\beta})'(\boldsymbol{Y} - \boldsymbol{X\beta}). \tag{3.0.3}$$

Lai iegūtu $\boldsymbol{\beta}$ novērtējumu, apskatam RSS izvedumu

$$RSS = \boldsymbol{Y'Y} - \boldsymbol{\beta'X'Y} - \boldsymbol{Y'X\beta} + \boldsymbol{\beta'X'X\beta},$$

tā kā $\boldsymbol{\beta'X'Y}$ ir skalārs, tad $(\boldsymbol{\beta'X'Y}) = \boldsymbol{Y'X\beta}$, un RSS vienkāršojas uz

$$RSS = \boldsymbol{Y'Y} - 2\boldsymbol{Y'X\beta} + \boldsymbol{\beta'X'X\beta}.$$

Atvasina pēc $\boldsymbol{\beta}$:

$$\begin{aligned}\frac{\delta RSS}{\delta\beta} &= \frac{\delta}{\delta\boldsymbol{\beta}}(\boldsymbol{Y'Y}) - \frac{\delta}{\delta\boldsymbol{\beta}}(2(\boldsymbol{X'Y})'\boldsymbol{\beta}) - \frac{\delta}{\delta\boldsymbol{\beta}}(\boldsymbol{\beta'X'X\beta})\\ &= -2\boldsymbol{X'Y} - 2\boldsymbol{X'X\beta}\end{aligned}$$

Seko, ka

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y} \text{ un } S^2 = \frac{(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n-m}, \tag{3.0.4}$$

kur m ir atbilstoši $\boldsymbol{\beta}$ dimensija un arī rangs $n \times m$ vektoram matricā $\boldsymbol{X}$. Pilnīgu teorijas izklāstu par regresiju analīzi skatīt[15, 567–572 lpp.]

1951.gada publikācijā [16] autors Wallis prognožu rezultātu atainošanai ieteica alternatīvu metodi - tolerances intervālus. Ar šīs metodes palīdzību nosaka apgabalu, kurā iekļausies noteikta populācijas proporcija ar nepieciešamo ticamības līmeni. Tolerances intervāli veiksmīgi tika pielietoti ASV lauksaimniecības ekonomiskajos aprēķinos (pielietojumu skatīt [17]).

Divpusējais tolerances intervāls $Y(\boldsymbol{x})$ sadalījumam pieņem formu $\boldsymbol{x}'\hat{\boldsymbol{\beta}} \pm k(\boldsymbol{x})S$, kur $k(\boldsymbol{x})$ ir tolerances faktors, kas tiek noteikts atkarībā no satura un ticamības līmeņa nosacījumiem. Atsaucoties uz (3.0.2), $\boldsymbol{x}'\hat{\boldsymbol{\beta}}$ varam apzīmēt ar $\hat{Y}(\boldsymbol{x})$. Apzīmēsim ar $C(\boldsymbol{x}; \hat{\boldsymbol{\beta}}, S)$ tolerances intervāla "saturu", pie dotiem $\hat{\boldsymbol{\beta}}$ un S. Seko, ka

$$C(\boldsymbol{x}; \hat{\boldsymbol{\beta}}, S) = P_{Y(\boldsymbol{x})}\left(\boldsymbol{x}'\hat{\boldsymbol{\beta}} - k(\boldsymbol{x})S \leq \boldsymbol{x}'\hat{\boldsymbol{\beta}} + k(\boldsymbol{x})S \middle| \hat{\boldsymbol{\beta}}, S\right), \tag{3.0.5}$$

un tolerances faktors $k(\boldsymbol{x})$ atbilst nosacījumam

$$P_{\hat{\boldsymbol{\beta}},S}\Big(C(\boldsymbol{x};\hat{\boldsymbol{\beta}},S) \geq p\Big) = 1-\alpha. \tag{3.0.6}$$

3.1. Vienpusējie tolerances intervāli lineārajā regresijas modelī

Uzskata, ka $k(\boldsymbol{x})$ apzīmē tolerances faktoru, ko aprēķina, lai iegūtu vienpusējo tolerances intervālu. Tas ir, $\boldsymbol{x}'\hat{\boldsymbol{\beta}} + k(\boldsymbol{x})S$ ir augšējā tolerances robeža un $\boldsymbol{x}'\hat{\boldsymbol{\beta}} - k(\boldsymbol{x})S$ ir apakšējā tolerances robeža. Pirmkārt, jāizvērtē faktora $k(\boldsymbol{x})$ iegūšana, lai aprēķinātu vienpusējo tolerances intervālu $Y(\boldsymbol{x})$ sadalījumam, pie fiksēta $\boldsymbol{x}$.

Tolerances intervāla saturs, pie dotiem $\hat{\boldsymbol{\beta}}$ un S

$$C_1(\boldsymbol{x};\hat{\boldsymbol{\beta}},S) = P_{Y(\boldsymbol{x})}\Big(Y(\boldsymbol{x}) \leq \boldsymbol{x}'\hat{\boldsymbol{\beta}} + k(\boldsymbol{x})S \Big| \hat{\boldsymbol{\beta}}, S\Big),$$

kur faktors $k(\boldsymbol{x})$ tiek iegūts sekojoši

$$P_{\hat{\boldsymbol{\beta}},S}\Big(C_1(\boldsymbol{x};\hat{\boldsymbol{\beta}},S) \geq p\Big) = 1-\alpha. \tag{3.1.1}$$

Jāievēro, ka

$$Z = \frac{Y(\boldsymbol{x}) - \boldsymbol{x}'\boldsymbol{\beta}}{\sigma} \sim N(0,1), \tag{3.1.2}$$

$$Z_{\boldsymbol{X}} = \frac{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}}{\sigma} \sim N(0, (\boldsymbol{X'X})^{-1}), \tag{3.1.3}$$

$$U^2 = \frac{S^2}{\sigma^2} \sim \frac{\chi^2_{n-m}}{n-m}, \tag{3.1.4}$$

kur χ^2_r apzīmē centrālo Hī-kvadrātu gadījuma lielumu ar r brīvības pakāpēm, visi gadījuma lielumi ir neatkarīgi. Pie šo gadījuma lielumu nosacījuma, saturu var uzrakstīt sekojoši

$$C_1(\boldsymbol{x};\hat{\boldsymbol{\beta}},S) = P_Z(Z \leq \boldsymbol{x}'Z_{\boldsymbol{X}} + k(\boldsymbol{x})U | Z_{\boldsymbol{X}}, U). \tag{3.1.5}$$

Jāpiezīmē, ka $\boldsymbol{x}'Z_{\boldsymbol{X}} \sim N(0, \boldsymbol{x}'(\boldsymbol{X'X})^{-1}\boldsymbol{x})$ un uzskata, ka

$$d^2 = \boldsymbol{x}'(\boldsymbol{X'X})^{-1}\boldsymbol{x} \;\text{ un }\; V = \frac{\boldsymbol{x}'Z_{\boldsymbol{X}}}{d}, \tag{3.1.6}$$

kur $V \sim N(0,1)$. Izmantojot šos mainīgos, tolerances intervāla saturs var tikt izteikts sekojošā veidā

$$C_1(\boldsymbol{x};\hat{\boldsymbol{\beta}},S) = \Phi(dV + k_1(d)U) = C_1(d;V,U), \tag{3.1.7}$$

kur Φ apzīmē standarta normālo kumulatīvo sadalījuma funkciju, un esam lietojuši $k_1(d)$ pierakstu $k(\boldsymbol{x})$ vietā, kas $k(\boldsymbol{x})$ atkarīgs no $\boldsymbol{x}$ tikai caur d izteiksmē (3.1.6). Apzīmējumam $C_1(d;V,U)$, kas aizstāj $C_1(\boldsymbol{x};\hat{\boldsymbol{\beta}},S)$, nebūtu jārada apjukumu. Izmantojot (3.1.7) un (3.1.1), varam redzēt, ka faktors $k_1(d)$ atbilst nosacījumiem

$$P_{V,U}(\Phi(dV+k_1(d)U)\geq p)=1-\alpha,$$

pielietojot $X=dV\sim N(0,d^2)$ un $Q=U^2\sim\frac{\chi^2_{n-m}}{n-m}$, un tad, pielietojot formulas (A0.1) un (A0.2), iegūst

$$k_1(d)=d\times t_{n-m;1-\alpha}(z_p/d), \tag{3.1.8}$$

kur $t_{r;\gamma}(\eta)$ apzīmē γ kvantili necentrālajam t sadalījumam ar brīvības pakāpēm r un necentralitātes parametru η. Jāatzīmē, ka arī vienkāršajā lineārās regresijas modelī d^2 var tikt vienkāršots:

$$d^2=\left(\frac{1}{n}+c^2\right),\ \text{ kur }\ c^2=\frac{(x-\bar{x})^2}{\sum_{i=1}^n(x_i-\bar{x})^2}. \tag{3.1.9}$$

3.2. Divpusējie tolerances intervāli lineārajā regresijas modelī

Atsaucoties uz rezultātiem (3.0.1) un (3.0.2),tiek pieņemts, ka divpusējais tolerances intervāls pieņem formu $\boldsymbol{x}'\hat{\boldsymbol{\beta}}\pm k(\boldsymbol{x})S$, kur $\hat{\boldsymbol{\beta}}$ un S^2 tika definēti (3.0.4). Sākumā nosaka divpusējo tolerances intervālu $Y(\boldsymbol{x})$ sadalījumam pie fiksēta $\boldsymbol{x}$.

Tolerances faktors $k(\boldsymbol{x})$ atbilst nosacījumam

$$P_{\hat{\boldsymbol{\beta}},S}\Big(C_2(\boldsymbol{x};\hat{\boldsymbol{\beta}},S)\geq p\Big)=1-\alpha, \tag{3.2.1}$$

kur $C_2(\boldsymbol{x};\hat{\boldsymbol{\beta}},S)$ ir tolerances intervāla saturs, pie dotiem $\hat{\boldsymbol{\beta}}$ un S, kā definēts formulā (3.0.5). Uzskata, ka Z, $Z_{\boldsymbol{X}}$ un U definēti (3.1.2), (3.1.3) un (3.1.4). Balstoties uz šo mainīgo nosacījumiem, var rakstīt

$$\begin{aligned}C_2(\boldsymbol{x};\hat{\boldsymbol{\beta}},S)&=P_{Y(\boldsymbol{x})}\Big(\boldsymbol{x}'\hat{\boldsymbol{\beta}}-k(\boldsymbol{x})S\leq Y(\boldsymbol{x})\leq\boldsymbol{x}'\hat{\boldsymbol{\beta}}+k(\boldsymbol{x})S\mid\hat{\boldsymbol{\beta}},S\Big)\\&=P_Z\left(\boldsymbol{x}'Z_{\boldsymbol{X}}-k(\boldsymbol{x})\frac{S}{\sigma}\leq Z\leq\boldsymbol{x}'Z_{\boldsymbol{X}}+k(\boldsymbol{x})\frac{S}{\sigma}\,\middle|\,Z_{\boldsymbol{X}},S\right)\\&=P_Z\left(\boldsymbol{x}'Z_{\boldsymbol{X}}-k(\boldsymbol{x})U\leq Z\leq\boldsymbol{x}'Z_{\boldsymbol{X}}+k(\boldsymbol{x})U\mid Z_{\boldsymbol{X}},U\right).\end{aligned} \tag{3.2.2}$$

Atsaucoties uz to, ka $d^2 = \boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}$, $\quad \boldsymbol{x}'Z_{\boldsymbol{X}} \sim N(0, d^2)$ un $\quad V = \frac{\boldsymbol{x}'Z_{\boldsymbol{X}}}{d} \sim N(0,1)$, var vienkāršot $C_2(\boldsymbol{x}; \hat{\boldsymbol{\beta}}, S)$ vienādojumā (3.2.2) sekojošā veidā

$$
\begin{aligned}
C_2(\boldsymbol{x}; \hat{\boldsymbol{\beta}}, S) &= P_Z(dV - k_2(d)U \leq Z \leq dV + k_2(d)U \mid V, U) \\
&= \Phi(dV + k_2(d)U) - \Phi(dV - k_2(d)U),
\end{aligned} \tag{3.2.3}
$$

kur Φ apzīmē standarta normālo kumulatīvo funkciju, un $k(\boldsymbol{x})$ vietā lieto apzīmējumu $k_2(d)$. No vienādojumiem (3.2.1) un (3.2.3) var redzēt, ka $k_2(d)$ jāizvēlas tādu, lai tas atbilstu nosacījumam

$$
P_{V,U}(\Phi(dV + k_2(d)U) - \Phi(dV - k_2(d)U) \geq p) = 1 - \alpha. \tag{3.2.4}
$$

Pielietojot $X = dV \sim N(0, d^2)$ un $Q = U \sim \frac{\chi^2_{n-m}}{n-m}$, un ievērojot, ka X un Q ir neatkarīgi, varam pielietot formulas (A0.1) un (A0.2). Var redzēt, ka $k_2(d)$ ir integrālā vienādojuma atrisinājums

$$
\sqrt{\frac{2}{\pi d^2}} \int_0^\infty P\left(\chi^2_{n-m} > \frac{(n-m)\chi^2_{1;p}(x^2)}{(k_2(d))}\right) e^{-\frac{x^2}{2d^2}} dx = 1 - \alpha. \tag{3.2.5}
$$

3.2.1. Aproksimācija, balstīta uz vienpusējo tolerances faktoru

Vienpusējais tolerances faktors, kas pielāgots ticamības līmenim, var tikt pielietots kā aproksimācija faktoram, divpusējo tolerances robežu konstruēšanai. Aplūkojam vienpusējo tolerances faktoru (3.1.8) ar ticamības līmeni $(1 - \frac{\alpha}{2})$ kā $k_2(d)$ aproksimāciju. Rezultējošais tolerances faktors tiek apzīmēts ar $k_{2o}(d)$. Seko, ka

$$
k_{2o}(d) = d \times t_{n-m; 1-\frac{\alpha}{2}}(z_p/d). \tag{3.2.6}
$$

4. Neparametriskie tolerances intervāli

Ja mums ir dota izlase, kas pieder nepārtrauktai populācijai un neatbilst parametriskam modelim, vai atbilst parametriskam modelim, kuram ir grūti konstruēt tolerances intervālus, tad ir nepieciešams konstruēt neparametriskos tolerances intervālus. Mūsdienu literatūrā apskatītās neparametriskās metodes tolerances intervālu konstruēšanai galvenokārt balstās uz autora Wilks 1941.gada publikāciju [4].

Jebkurai fiksēta apjoma izlasei var arī neeksistēt sakārtotās statistikas, kas atbilst vajadzīgajiem vienpusējo tolerances robežu nosacījumiem. Wilks 1941.gadā risināja šo problēmu un pieņēma, ka, ja izlase ņemta no nepārtraukta sadalījuma, tad proporcijas sadalījums populācijai starp divām sakārtotām statistikām ir neatkarīgs no dotās populācijas, bet ir funkcija tikai no izvēlētajām statistikām (Wilks publikācijas tulkojumu skatīt pielikumā). Ir nepieciešams noteikt optimālu izlases apjomu.

4.1. Neparametriskie tolerances intervāli sakārtotām statistikām

Šajā nodaļā pārsvarā tiks pielietota teorija no grāmatas [6] un publikācijas [4]

Neparametriskie tolerances intervāli tiek balstīti uz sakārtotām statistikām, un tolerances intervāli tiek definēti sekojošā veidā. Uzskata, ka $\boldsymbol{X} = (X_1, \ldots, X_n)$ ir gadījuma izlase no nepārtraukta sadalījuma $F_X(x)$, un $X_{(1)} < \ldots < X_{(n)}$ ir izlases sakārtotās statistikas. Atsaucas, ka $(p, 1-\alpha)$ tolerances intervāls $(L(\boldsymbol{X}), U(\boldsymbol{X}))$ ir

$$P_{\boldsymbol{X}}\left\{P_X\left(L(\boldsymbol{X}) \leq X \leq U(\boldsymbol{X}) \;\middle|\; \boldsymbol{X}\right) \geq p\right\} = 1 - \alpha,$$

kur X arī atbilst nepārtrauktajam sadalījumam F_X, kas neatkarīgs no $\boldsymbol{X}$. Wilks rezultāts atļauj izvēlēties $L(\boldsymbol{X}) = X_{(r)}$ un $U(\boldsymbol{X}) = X_{(s)}$, $r < s$, un uzdevums ir noteikt r un s vērtības tā, lai

$$P_{X_{(r)},X_{(s)}} \left\{ P_X \left(X_{(r)} \leq X \leq X_{(s)} \;\middle|\; X_{(r)}, X_{(s)} \right) \geq p \right\} = 1 - \alpha. \tag{4.1.1}$$

Vienpusējā tolerances robeža ir līdzīgi definēta, balstoties uz vienu sakārtoto statistiku. Tālāk tiks apskatīti daži sagatavojoši rezultāti sakārtoto statistiku sadalījumam, kas būs noderīgi r un s (kas atbilst (4.1.1)) vērtību atrašanai, kā arī ir vajadzīgi vienpusējo tolerances robežu noteikšanai.

4.2. Sakārtotās statistikas un to sadalījumi

Pieņem, ka $X_1, \ldots, X_n$ ir izlase no populācijas ar nepārtrauktu sadalījuma funkciju $F_X(x)$. Uzskata, ka $X_{(i)}$ apzīmē i - to mazāko no $X_1, \ldots, X_n$. Sekojoši

$$X_{(1)} < X_{(2)} < \ldots < X_{(n)} \tag{4.2.1}$$

ir sakārtotās statistikas izlasei. Jāatzīmē, ka šis sakārtojums ir unikāls, jo F_X ir nepārtraukts sadalījums, un varbūtība, ka jebkuri divi gadījuma lielumi atbilst vienai un tai pašai vērtībai ir nulle. Statistika $X_{(r)}$ tiek saukta par r - to sakārtoto statistiku.

Tālāk aplūko sakārtoto statistiku sadalījumu rezultātus, kas tiek prasīti, lai konstruētu tolerances robežas.

Rezultāts 1. (Varbūtības integrālā transformācija)

Pieņem, ka X ir gadījuma mainīgais ar nepārtrauktu sadalījuma funkciju $F_X(x)$. Uzskata, ka $Y = F_X(X)$. Tad Y atbilst vienmērīgajam sadalījumam $Y \sim U(0,1)$.

Pierādījums. Inversā sadalījuma funkcija ir definēta kā

$$F_X^{-1}(y) = inf\{x : P(X \leq x) \geq y\},\ 0 < y < 1.$$

Sadalījuma $U(0,1)$ gadījuma mainīgais U, $F_U(u) = u$, katram y, ja izpildās $0 < y < 1$,

$$P_Y(Y \leq y) = P_X(F_X(X) \leq y) = P_X(X \leq F_X^{-1}(y)) = y,$$

tādā veidā $Y \sim U(0,1)$. $\square$

No šī rezultāta izriet sekojošais rezultāts.

Rezultāts 2.

Ja $X_1, \dots, X_n$ ir izlase no nepārtraukta sadalījuma F_X, tad

$$U_1 = F_X(X_1), \dots, U_n = F_X(X_n)$$

ir izlase no $U(0,1)$ sadalījuma. Turklāt, ja $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ ir sakārtotas statistikas izlasei $X_1, \dots, X_n$, tad

$$U_{(1)} = F_X(X_{(1)}), \dots, U_{(n)} = F_X(X_{(n)})$$

attiecīgi uzskata par sakārtotām statistikām izlasei $U_1, \dots, U_n$ no sadalījuma $U(0,1)$. Tas ir, $F_X(X_{(r)})$ ir sadalīts kā $U_{(r)}$.

Rezultāts 3. (Empīriskā sadalījuma funkcija)

Pieņem, ka $X_1, \dots, X_n$ ir izlase no nepārtrauktā sadalījuma F_X. Empīriskā sadalījuma funkcija ir soļu funkcija un definēta sekojoši

$$\hat{F}_n(x) = \frac{\text{skaits} X_i \leq x}{n}.$$

Tad $n\hat{F}_n(x) \sim bin(n, F_X(x))$.

Pierādījums. Uzskata, ka $Q_i = 1$, ja $X_i \leq x$, 0 pretējā gadījumā. Tā kā X_i - tie ir neatkarīgi un vienādi sadalīti, Q_i - tie ir neatkarīgi Bernulli gadījuma lielumi ar " veiksmes varbūtību" $F_X(x)$. Tādēļ $n\hat{F}_n(x) = \sum_{i=1}^{n} Q_i \sim bin(n, F_X(x))$, kur ar $bin(n, F_X(x))$ apzīmējam binomiālo sadalījumu. □

Rezultāts 4.

Uzskata, ka $X_{(r)}$ ir r - tā sakārtotā statistika n novērojumu izlasei no nepārtraukta sadalījuma F_X. $X_{(r)}$ blīvuma funkcija dota

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!}[F_X(x)]^{r-1}[1 - F_X(x)]^{n-r} f_X(x), \tag{4.2.2}$$

kur $f_X(x)$ ir blīvuma funkcija no X.

Pierādījums. $X_{(r)}$ blīvuma funkcija dota

$$\begin{aligned} F_{X_{(r)}}(x) &= P(X_{(r)} \le x) \\ &= P(\text{skaits} X_i \le x \text{ ir vismaz } r) \\ &= P(n\hat{F}_n(x) \ge r) \\ &= \sum_{k=r}^{n} \binom{n}{k} F_X(x)^k (1 - F_X(x))^{n-k} \\ &= n\binom{n-1}{r-1} \int_0^{F_X(x)} t^{r-1}(1-t)^{n-r} dt. \end{aligned}$$

□

Divu pēdējo izteiksmju vienādība var tikt pierādīta, integrējot pa daļām. Diferencējot pēdējo izteiksmi pēc x, iegūst blīvuma funkciju (4.2.2).

Pieņēmums 1. Ja $U_{(1)} < \ldots < U_{(n)}$ ir sakārtotās statistikas no standartnormālā sadalījuma, tad $U_{(r)}$ atbilst $beta(r, n-r+1)$ sadalījumam ar blīvuma funkciju

$$f_{U_{(r)}}(u) = \frac{1}{\beta(r, n-r+1)} u^{r-1}(1-u)^{n-r+1-1}, \qquad 0 < u < 1, \tag{4.2.3}$$

kur $\beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ ir beta funkcija.

Rezultāts 5.

$(X_{(r)}, X_{(s)})$ kopējā sadalījuma funkcija ir dota sekojošā veidā

$$f_{X_{(r)},X_{(s)}}(x,y) = \begin{cases} \frac{n!}{(r-1)!(s-r-1)!(n-s)!}[F_X(x)]^{r-1}[F_X(y) - F_X(x)]^{s-r-1} & \\ \times[1 - F_X(y)]^{n-s} f_X(x) f_X(y), & -\infty < x < y < \infty, \\ 0, & x \ge y. \end{cases} \tag{4.2.4}$$

Pierādījums. Pieņem, ka $r < s$ tā, ka $X_{(r)} < X_{(s)}$. Kā arī, pieņem, ka $x \ge y$. Jāatzīmē, ka $X_{(s)} \le y \Rightarrow X_{(r)} \le x$, tad

$$P(X_{(r)} \le x, X_{(s)} \le y) = P(X_{(s)} \le y) = F_{x_{(s)}}(y), \quad x \ge y. \tag{4.2.5}$$

Ja $x < y$ un $r < s$, tad

$$
\begin{aligned}
P(X_{(r)} \leq x, X_{(s)} \leq y) &= P(n\hat{F}_n(x) \geq r,\ n\hat{F}_n(y) \geq s) \\
&= \sum_{i=r}^{n}\sum_{j=s}^{n} P(n\hat{F}_n(x) = i,\ n\hat{F}_n(y) = j) \\
&= \sum_{i=r}^{n}\sum_{j=s}^{n} P(i - \text{tais no} X \leq x,\ j - \text{tais no} X \leq y) \\
&= \sum_{i=r}^{n}\sum_{j=s}^{n} P(i - \text{tais no} X \leq x,\ x < (j-i) - \text{tais no} X) \leq y, \\
&\quad n - j - \text{tais no} X \geq y) \\
&= \sum_{i=r}^{n}\sum_{j=s}^{n} \frac{n!}{i!(j-i)!(n-j)!}[F(x)]^i[F(y) - F(x)]^{j-i} \\
&\quad \times [1 - F(y)]^{n-j}, \quad x < y. \qquad (4.2.6)
\end{aligned}
$$

Jāatzīmē, ka varbūtību izteiksme (4.2.6) ir varbūtības masas funkcija trinomiālam sadalījumam. Šī izteiksme var tikt pārrakstīta sekojoši

$$
\begin{aligned}
F_{X_{(r)},X_{(s)}} &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \\
&\quad \times \int_0^{F(x)}\int_0^{F(y)} v^{r-1}(t-v)^{s-r-1}(1-t)^{n-s}dtdv. \qquad (4.2.7)
\end{aligned}
$$

Diferencējot pēc (x, y), iegūstam (4.2.4) □

Pielietojot šo rezultātu, varam redzēt, ka, ja $U_{(r)}$ un $U_{(s)}$ ir attiecīgi r - tā un s - tā sakārtotā statistika izlasei apjomā n no sadalījuma $U(0, 1)$, tad

$$
\begin{aligned}
f_{U_{(r)},U_{(s)}} &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} x^{r-1}[y-x]^{s-r-1} \\
&\quad \times [1-y]^{n-s}, \qquad 0 < x < y < 1. \qquad (4.2.8)
\end{aligned}
$$

Rezultāts 6.

Pieņem, ka X ir gadījuma lielums no binomiālā sadalījuma $bin(n, p)$, un U ir $beta(k, n - k + 1)$ gadījuma lielums. Tad pie dota k

$$P(X \geq k|n, p) = P(U \leq p), \quad k = 1, \ldots, n.$$

Kā arī

$$P(X \leq k - 1|n, p) = P(U \geq p), \quad k = 1, \ldots, n.$$

4.3. Vienpusējās tolerances robežas un pārsniegšanas varbūtības

Pieņem, ka $X_1, \ldots, X_n$ ir izlase no nepārtraukta sadalījuma F_X. Lai konstruētu neparametrisko $(p, 1-\alpha)$ apakšējo tolerances robežu, ir jāatrod pozitīvs vesels skaitlis k tāds, ka

$$P_{X_{(k)}}[1 - F(X_{(k)}) \geq p] = P_{X_{(k)}}[F(X_{(k)}) \geq 1 - p] = P(U_{(k)} \leq 1 - p).$$

Varbūtība labajā pusē var tikt izteikta kā

$$P_{X_{(k)}}[1 - F(X_{(k)}) \geq p] = P_{X(k)}[F(X_{(k)}) \leq 1 - p] = P(U_{(k)} \leq 1 - p),$$

kur $U_{(k)} = F(X_{(k)}) \sim beta(k, n-k+1)$. Izmantojot rezultātu 6., varam redzēt, ka

$$\begin{aligned}
P(U_{(k)} \leq 1 - p) &= 1 - P(U_{(k)} \geq 1 - p) \\
&= 1 - P(Y \leq k - 1 | n, 1 - p) \\
&= P(Y \geq k | n, 1 - p) \\
&= P(n - Y \leq n - k | n, 1 - p) \\
&= P(W \leq n - k | n, p), \qquad (4.3.1)
\end{aligned}$$

kur Y ir $bin(n, 1-p)$ gadījuma lielums, un $W = n - Y$ ir $bin(n, p)$ gadījuma lielums. Tad, ja k ir lielākais veselais skaitlis, kuram

$$P(Y \geq k | n, 1 - p) \geq 1 - \alpha, \qquad (4.3.2)$$

tad $X_{(k)}$ ir vajadzīgā $(p, 1-\alpha)$ apakšējā tolerances robeža.

Lai konstruētu $(p, 1-\alpha)$ augšējo tolerances robežu, ir jāatrod pozitīvs vesels skaitlis m tāds, ka

$$P_{X_{(m)}}[P_X(X \leq X_{(m)} | X_{(m)}) \geq p] = 1 - \alpha.$$

Turpinot līdzīgi kā, meklējot apakšējo tolerances robežu, var parādīt, ka $X_{(n-k+1)}$ ir $(p, 1-\alpha)$ augšējā tolerances robeža, kur k ir lielākais veselais skaitlis, kas atbilst nosacījumam (4.3.2).

Apakšejās robežas pārsniegšanas varbūtībai

Nepārtrauktam gadījuma lielumam X gribam novērtēt $P(X > t)$, kur t ir noteikts skaitlis. Ja $t \leq X_{(n)}$, tad $P(X > t)$ apakšējā robeža $1 - \alpha$ var tikt iegūta sekojošā veidā.

Pieņem, ka $X_{(r)}$ ir mazākā sakārtotā statistika , kas ir lielāka par t. Pieņem, ka p ir tāds, ka $X_{(r)}$ ir $(p, 1-\alpha)$ apakšejā tolerances robeža. Tā kā $t < X_{(r)}$,

$$P(X > t) \geq p$$

at varbūtību vismaz $1-\alpha$. Tādēļ mums ir jāatrod tādu vērtību p, ka $X_{(r)}$ ir $(p, 1-\alpha)$ apakšējā tolerances robeža sadalījumam X. Pēc (4.3.1) redzam, ka p tiek noteikts tā, ka

$$P(W \leq n-r|n,p) \geq 1-\alpha \Leftrightarrow P(U > p|n-r+1, r) \geq 1-\alpha, \tag{4.3.3}$$

kur W ir $bin(n,p)$ gadījuma lielums, un U ir $beta(n-r+1, r)$ gadījuma lielums. (4.3.3) tika iegūts izmantojot Wilks rezultātu. Tātad p ir dots kā $beta(\alpha; n-r+1, r)$, α kvantile no $beta(n-r+1, r)$ sadalījuma. Tas ir, $1-\alpha$ apakšejā $P(X > t)$ tolerances robeža $P(X > t)$ ir $beta(\alpha; n-r+1, r)$.

4.4. Tolerances intervāli

Lai konstruētu $(p, 1-\alpha)$ neparametrisko tolerances intervālu nepārtrauktam sadalījumam, ir jānosaka sakārtoto statistiku pāris $X_{(r)}$ un $X_{(s)}$, kur $r < s$, tāds, ka intervāls $(X_{(r)}, X_{(s)})$ saturētu vismaz populācijas proporciju p ar ticamību $1-\alpha$. Ir jānosaka vērtības $r < s$, ka

$$P_{X_{(r)},X_{(s)}}\left\{P_X[X_{(r)} \leq X \leq X_{(s)}|X_{(r)}, X_{(s)}] \geq p\right\} = 1-\alpha. \tag{4.4.1}$$

Tādēļ, ka r un s ir veseli skaitļi, tos ir jānosaka tā, ka $r-s$ ir minimums, un pārklājuma varbūtība ir vismaz $1-\alpha$, un ir pēc iespējas tuvāk $1-\alpha$. Jāatzīmē, ka iekšejā varbūtība ir

$$P_X[X_{(r)} \leq X \leq X_{(s)}] = F(X_{(s)}) - F(X_{(r)}).$$

Tātad varam pārrakstīt (4.4.1) kā

$$P_{X_{(r)},X_{(s)}}\left\{F(X_{(s)}) - F(X_{(r)}) \geq p\right\} = 1-\alpha, \tag{4.4.2}$$

vai ekvivalenti

$$P_{U_{(r)},U_{(s)}}\left\{U_{(s)} - U_{(r)} \geq p\right\} = 1-\alpha. \tag{4.4.3}$$

$U_{(s)} - U_{(r)}$ blīvuma funkcija var tikt noteikta no $U_{(r)}$ un $U_{(s)}$ kopējās blīvuma funkcijas izteiksmē (4.2.8). Pieņem, ka $u = y-x$ un $v = y$ tā, ka inversā transformācija ir $x = v-u$

un $y = v, \quad 0 < u < v < 1$. Transformācijas jakobiāns ir viens, ja sākumā atvasina pēc v. Pielietojot šos rezultātus (4.2.8), iegūstam

$$f_{U,V}(u,v) = c(v-u)^{r-1}u^{s-r-1}(1-v)^{n-s}, \quad 0 < u < v < 1,$$

kur $c = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}$, kas ir konstants nosacījums izteiksmē (4.2.8). Robežu blīvuma funkcija no $U = U_{(s)} - U_{(r)}$ ir dota

$$f_U(u) = cu^{s-r-1}\int_u^1 (v-u)^{r-1}(1-v)^{n-s}dv.$$

Šī ir beta funkcija. Lai to parādītu, pieņem, ka $v - u = t(1-u), \quad 0 < t < 1$. Tas sevī ietver arī $(1-v) = (1-u) - t(1-u) = (1-u)(1-t)$. Kā arī $dv = (1-u)dt$. Šo jauno mainīgo ietvaros, varam pārrakstīt blīvuma funkciju sekojoša veidā

$$\begin{aligned}
f_U(u) &= cu^{s-r-1}\int_0^1 (1-u)^{r-1}t^{r-1}(1-u)^{n-s}(1-t)^{n-s}(1-u)dt \\
&= cu^{s-r-1}(1-u)^{n-s+r}\int_0^1 (1-t)^{n-s}dt.
\end{aligned}$$

Jāatzīmē, ka $\int_0^1 t^{r-1}(1-t)^{n-s}dt = \beta(r, n-s+1) = \frac{\Gamma(r)\Gamma(n-s+1)}{\Gamma(n-s+r+1).}$ Aizvietojot ar šo izteiksmi daļu no iekpriekšējās izteiksmes, iegūstam

$$\begin{aligned}
f_U(u) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!}u^{s-r-1}(1-u)^{n-s+r}\frac{(r-1)!(n-s)!}{(n-s+r)!} \\
&= \frac{1}{\beta(s-r, n-s+r+1)}u^{s-r-1}(1-u)^{n-s+r}, \quad 0 < u < 1. \qquad (4.4.4)
\end{aligned}$$

Tātad $U = U_{(s)} - U_{(r)}$ ir sadalīts kā beta gadījuma lielums ar parametriem $s - r$ un $n - s + r + 1$. Pielietojot (4.3.1), var secināt, ka

$$P_{U_{(r)},U_{(s)}}\left\{U_{(s)} - U_{(r)} \geq p\right\} = P(X \leq s-r-1), \qquad (4.4.5)$$

kur $X \sim bin(n,p)$. Jāatzīmē, ka (4.4.5) atkarīgs tikai no starpības $s - r$, bet nav atkarīgs no s un r vērtības. Pieņem, ka $k = s - r$ ir mazākā vērtība, kurai

$$P(X \leq s-r-1) \geq 1-\alpha. \qquad (4.4.6)$$

Tad jebkurš intervāls $(X_{(r)}, X_{(s)})$ ir $(p, 1-\alpha)$ tolerances intervāls, balstīts uz $1 \leq r < s \leq n$ un $s - r = k$.

Ir ierasti pielietot $s = n - r + 1$ tā, ka $(X_{(r)}, X_{(n-r+1)})$ ir $(p, 1-\alpha)$ tolerances intervāls (skatīt Wilks publikācijas tulkojumu pielikumā). Autors Wilks atsaucas uz šo intervālu kā nošķeltās izlases apgabalu, jo tas ir veidots ar r - to mazāko un r - to lielāko novērojumu no izlases ar apjomu n. Ir iespējams arī īsāks intervāls, ja mēs neņemam vērā šo nosacījumu. Piemēram, ja $n = 38$ un $(p, 1-\alpha) = (0.80, 0.90)$, īsākais intervāls formā $(X_{(r)}, X_{(n-r+1)})$, kas atbilst varbūtības prasībām (4.4.6), ir $(X_{(2)}, X_{(37)})$ ar pārklājuma varbūtību 0.9613. Var pārbaudīt, ka $(X_{(2)}, X_{(36)})$ arī ir $(p, 1-\alpha) = (0.80, 0.90)$ tolerances intervāls ar pārklājuma varbūtību 0.9014. Līdzīgi, kad $n = 69$ un $(p, 1-\alpha) = (0.80, 0.99)$, vienīgais īsākais intervāls $(X_{(r)}, X_{(n-r+1)})$, kas atbilst nosacījumam (4.4.6) ir $(X_{(3)}, X_{(67)})$ ar pārklājuma varbūtību 0.9968, kamēr $(X_{(3)}, X_{(66)})$ ir ar pārklājuma varbūtību 0.9908.

4.5. Ticamības intervāli populācijas kvantilēm

Pieņem, ka k_p apzīmē p kvantili nepārtrauktam sadalījumam F_X, un $X_{(1)}, \ldots, X_{(n)}$ ir sakārtoto statistiku kopums no F_X. Gadījumā, kad $0 < p < 1$, pieņem, ka $r = np$, ja np ir vesels skaitlis, un $r = [np]$ pretējā gadījumā, kur $[x]$ ir lielākais veselais skaitlis, kas nepārsniedz x. Sakārtotā statistika $X_{(r)}$ ir izlases p - tā kvantile, kas ir punkta k_p novērtējums.

Lai konstruētu $1-\alpha$ ticamības intervālu kvantilei k_p, kas balstīts uz sakārtotām statistikām, ir jānosaka r un s, $r < s$, vērtības, lai izpildās $P(X_{(r)} \leq k_p \leq X_{(s)}) = 1 - \alpha$. Jāpiezīmē, ka notikums $X_{(r)} \leq k_p \leq X_{(s)}$ ir ekvivalents ar "skaits $X_i < k_p$ ir vismaz r un ne vairāk kā $s - 1$." Izmantojot šo attiecību, varam redzēt, ka

$$\begin{aligned}
P(X_{(r)} \leq k_p \leq X_{(s)}) &= P(r \leq n\hat{F}_n(k_p) \leq s-1) \\
&= \sum_{i=r}^{s-1} \binom{n}{i} [F(k_p)^i][1 - F(k_p)]^{n-i} \\
&= \sum_{i=r}^{s-1} \binom{n}{i} p^i [1-p]^{n-i}, \qquad \text{jo} F(k_p) = p \\
&= 1 - \alpha. \qquad (4.5.1)
\end{aligned}$$

Tātad $(X_{(r)}, X_{(s)})$ ir $1-\alpha$ ticamības intervāls kvantilei k_p, kur r un s atbilst nosacījumam (4.5.1) un starpība $s - r$ ir pēc iespējas mazāka. Dotiem n un $1 - \alpha$, var arī neeksistēt tādi r un s, ka $P(X_{(r)} \leq k_p \leq X_{(s)}) \geq 1 - \alpha$.

Atsaucamies uz to, ka $1 - \alpha$ vienpusējā ticamības robeža kvantilei k_p ir arī $(p, 1-\alpha)$ vienpusējā tolerances robeža sadalījumam F_X. Sakārtotā statistika $X_{(k)}$, kur k ir iegūts no (4.3.2), ir $1-\alpha$ apakšējā ticamības robeža kvantilei k_{1-p}, un $X_{(n-k+1)}$ ir $1-\alpha$ augšējā ticamības robeža kvantilei k_p.

4.6. Izlases apjoma noteikšana

Jebkurai fiksēta apjoma izlasei var neeksistēt sakārtotā statistika, kas dotu vajadzīgo $(p, 1-\alpha)$ tolerances intervālu. Pie dotiem p un $1-\alpha$ ir svarīgi noteikt tādu izlases apjomu, kuram eksistētu nepieciešamās sakārtotās statistikas, tolerances intervāla noteikšanai.

Izlases apjoms tolerances intervāliem formā $(X_{(1)}, X_{(n)})$

Sākumā nosakam izlases apjomu n tā, lai $(X_{(1)}, X_{(n)})$ saturētu vismaz populācijas proporciju p ar ticamības līmeni $1-\alpha$. Tātad

$$P_{X_{(1)},X_{(n)}}\left\{P_X[X_{(1)} \leq X \leq X_{(n)}|X_{(1)}, X_{(n)}] \geq p\right\} = 1-\alpha. \tag{4.6.1}$$

Ievietojot $s = n$ un $r = 1$ izteiksmē (4.4.6), varam redzēt ka izteiksmes (4.6.1) prasības vienkāršojas uz

$$P(X \leq n-2) \geq 1-\alpha \Leftrightarrow (n-1)p^n - np^{n-1} + 1 \geq 1-\alpha, \tag{4.6.2}$$

kur X ir $bin(n,p)$ gadījuma lielums. Pieņem, ka n_0 ir mazākā n vērtība, kas atbilst nosacījumam (4.6.2). Tad, mazākā sakārtotā statistika $X_{(1)}$ un lielākā sakārtotā statistika $X_{(n_0)}$ no izlases ar apjomu n_0, veido $(p, 1-\alpha)$ tolerances intervālu. Turklāt, jebkuram izlases apjomam $n \geq n_0$, eksistē vismaz viens sakārtoto statistiku pāris, kas veido $(p, 1-\alpha)$ tolerances intervālu.

Apskatīsim vienpusējās augšejās tolerances robežas gadījumu (kas ir arī vienpusējā augšejā ticamības robeža kvantilei k_p). Šo tolerances robežu var noteikt, atrodot mazāko n vērtību

$$P(k_p < X_{(n)}) \geq 1-\alpha.$$

Pielietojot (4.5.1), kad $r = 0$ un $s = n$, varam redzēt, ka

$$P(k_p < X_{(n)}) = \sum_{i=0}^{n-1} \binom{n}{i} p^i [1-p]^{n-i} = 1-p^n. \tag{4.6.3}$$

$X_{(n)}$ ir $(p, 1-\alpha)$ augšejā tolerances robeža, kad izlases apjoms n ir mazākā vērtība, kurai

$$1-p^n \geq 1-\alpha \qquad \text{vai} \qquad n \geq \frac{ln(\alpha)}{ln(p)}.$$

Lai nodrošinātu pārklājuma varbūtību, izvēlamies $n = \left[\frac{ln(\alpha)}{ln(p)}\right]_+$, kur $[x]_+$ ir mazākais veselais skaitlis, kas ir lielāks vai vienāds ar x. Arī, izvēloties šādu n, $X_{(1)}$ ir mazākā sakārtotā statistika izlasē ar apjomu n un ir $(p, 1-\alpha)$ apakšējā tolerances robeža, vai ekvivalenti, $X_{(1)}$ ir $1-\alpha$ apakšējā ticamības robeža kvantilei k_{1-p}.

Jāpiezīmē, ka izlases apjoma diskrētuma dēļ, īstā pārklājuma varbūtība var būt nedaudz lielākā par ticamības līmeni $1-\alpha$. Piemēram, kad $p = 0.90$ un $1-\alpha = 0.90$, vajadzīgais izlases apjoms divpusējam tolerances intervālam ir 38. Aizvietojot n ar 38 un p ar 0.90 izteiksmē (4.6.2), iegūstam īsto pārklājuma varbūtību 0.9047.

Vērtības n dotas tabulā 4.1. tādā veidā, ka ekstremālās sakārtoto statistiku vērtības veido $(p, 1-\alpha)$ vienpusējās tolerances robežas, kā arī tādas n vērtības, ka $(X_{(1)}, X_{(n)})$ veido $(p, 1-\alpha)$ divpusējo tolerances intervālu. Piemēram, ja izlase sastāv no 5 novērojumiem no F_X, tad intervāls $(X_{(1)}, X_{(5)})$ saturēs vismaz 50% no populācijas ar ticamību 0.80. Ja dota izlase no 459 vērtībām, tad vismaz 99% no populācijas vērtībām pārsniegs $X_{(1)}$ ar ticamību 0.99.

Izlases apjoms tolerances intervāliem formā $(X_{(r)}, X_{(s)})$

Kā pieminēts agrāk, sakārtoto statistiku pārim, lai veidotu $(p, 1-\alpha)$ tolerances intervālu, izlases apjomam n ir jāatbilst nosacījumam

$$(n-1)p^n - np^{n-1} + 1 \geq 1-\alpha. \tag{4.6.4}$$

Ja n_0 ir mazākā n vērtība, kas atbilst (4.6.4), tad $(X_{(1)}, X_{(n_0)})$ ir $(p, 1-\alpha)$ tolerances intervāls. Turklāt, dotiem $(p, 1-\alpha)$ un $n \geq n_0$, eksistē sakārtoto statistiku pāris, kas veido $(p, 1-\alpha)$ tolerances intervālu. Tātad, mērķis ir atrast tādu mazāko izlases apjomu, ka sakārtoto statistiku pāris veido tolerances intervālu, un jebkuram n, kas lielāks par šo mazāko vērtību, nosaka sakārtoto statistiku pāri, kas veido $(p, 1-\alpha)$ tolerances intervālu ar pārklājuma varbūtību, kas stipri nepārniedz $1-\alpha$.

Dotam m izlases apjoms n tiek aprēķināts kā mazākā vērtība, kas atbilst nosacījumam

$$P(X \leq n-m-1|n,p) \geq 1-\alpha. \tag{4.6.5}$$

Jebkurš sakārtoto statistiku pāris, teiksim, $X_{(l)}$ un $X_{(o)}$ veido $(p, 1-\alpha)$ tolerances intervālu ar nosacījumu $l - o = n - m$. Turklāt, $\left(X_{(\frac{m}{2})}, X_{(n-\frac{m}{2})}\right)$ - ja m ir pāra skaitlis, un $\left(X_{(\frac{m+1}{2})}, X_{(n-\frac{m+1}{2}+1)}\right)$ - ja m ir nepāra skaitlis, ir $(p, 1-\alpha)$ tolerances intervāli. Acīmredzami, ka izlases apjoms n, kas iegūts (4.6.5), ir atkarīgs no m. Jāatzīmē, ka dotam $(p, 1-\alpha)$, ja n_{m_1} ir izlases apjoms, kas iegūts, kad $m = m_1$ un n_{m_1+1} ir izlases apjoms

4.1. tabula: n vērtības, ka (a) $(X_{(1)}, X_{(n)})$ ir divpusējais tolerances intervāls, (b) $X_{(1)}$ ir $(p, 1-\alpha)$ vienpusējā apakšejā robeža, vai ekvivalenti $X_{(n)}$ ir $(p, 1-\alpha)$ vienpusējā augšejā robeža.

	Intervāla	$1-\alpha$			
p	veids	0.80	0.90	0.95	0.99
0.50	vienpusējs	3	4	5	7
	divpusējs	5	7	8	11
0.75	vienpusējs	6	9	11	17
	divpusējs	11	15	18	24
0.80	vienpusējs	8	11	14	21
	divpusējs	14	18	22	31
0.90	vienpusējs	16	22	29	44
	divpusējs	29	38	46	64
0.95	vienpusējs	32	45	59	90
	divpusējs	59	77	93	130
0.99	vienpusējs	161	230	299	459
	divpusējs	299	388	473	662

$m = m_1 + 1$, tad $\left(X_{(m_1)}, X_{(n^*)}\right)$ ir $(p, 1-\alpha)$ tolerances intervāls jebkuram n^*, kas atbilst $n_{m_1} \leq n^* < n_{m_1+1}$.

Izlases apjoms vienpusējām tolerances robežām

Lai konstruētu $(p, 1-\alpha)$ vienpusējo apakšējo tolerances robežu, ir jāatrod pozitīvs vesels skaitlis, ka

$$P_{X_{(k)}}\left(P_X(X \geq X_{(k)}|X_{(k)}) \geq p\right) = 1-\alpha$$

Kā pieminēts iepriekš, varbūtības prasības vienkāršojas uz

$$P(X \leq n-(k-1)-1|n, p) \geq 1-\alpha, \tag{4.6.6}$$

kur $X \sim binom(n, p)$. Ja m atbilst nosacījumam (4.6.5), tad $k-1 = m$ jeb $k = m+1$ atbilst (4.6.6). Interesants fakts ir tas, ja $(X_{(m)}, X_{(n)})$ ir $(p, 1-\alpha)$ tolerances intervāls, tad $X_{(m+1)}$ ir $(p, 1-\alpha)$ vienpusējā apakšējā robeža un $X_{(n-m)}$ ir $(p, 1-\alpha)$ vienpusējā augšējā tolerances robeža.

5. Praktiskais pielietojums

Piemērs 1. *Piesārņojuma līmeņa novērtēšana.*

Vienpusējās tolerances robežas parasti tiek lietotas, lai novērtētu piesārņojuma līmeni darba vietā vai kādā reģionā. Šajā piemērā novērtēsim svina līmeni gaisā. Tabulā 5.1. attēloti svina līmeņi gaisā, kas apkopoti Nacionālajā darba drošības un veselības institūta laboratorijā, veselības riska novērtēšanai [6]. Šie līmeņi tika noteikti 15 pēc iespējas dažādākās kādas laboratorijas vietās.

5.1. tabula :Svina līmenis gaisā(μ/m^3)

Oriģinālie dati	200	120	15	7	8	6	48	61
	380	80	29	1000	350	1400	110	
Logtransformētie dati	5.298	4.788	2.708	1.946	2.079	1.792	3.871	4.111
	5.940	4.382	3.367	6.908	5.858	7.244	4.700	

1. *Datu normalitātes pārbaude:*

   Normālais sadalījums labi atbilst logtransformētiem svina līmeņiem, bet pirms tam jāveic datu normalitātes pārbaudi. Šajā piemērā izmantosim *Lilliefors(Kolmogorova - Smirnova)* normalitātes testu. Šis tests ir Kolmogorova - Smirnova tests saliktai hipotēzei par normalitāti. Testa aprakstu skatīt [18]. Izvirzam hipotēzi:

   $$H_0 : X \sim N(\mu, \sigma^2) \quad \text{un} \quad H_1 : X \nsim N(\mu, \sigma^2).$$

   Tiek pielietota programmā R iebūvētā komanda *lillie.test*. Rezultātā iegūstam, ka p - vērtība ir vienāda ar 0.9416, tātad nav pamata noraidīt nulles hipotēzi, jo šī vērtība ir lielāka par jebkuru no α vērtībām $0.01, 0.05$ vai 0.1.

2. *Augšējā tolerances robeža:*

   Sākumā aprēķinājām augšējo tolerances robežu balstītu uz logtransformētiem datiem, lai novērtētu maksimālo svina līmeni laboratorijā. Aprēķini tika veikti, izmantojot programmu R (programmas kods pielikumā). Izlases vidējā vērtība un standartnovirze logtransformētiem datiem ir sekojoša: $\bar{X} = 4.333$ un $S = 1.739$. $(0.95, 0.90)$ augšējā tolerances robeža svina līmenim gaisā (skatīt (2.1.5)) ir $\bar{X} + k_1 S = 4.333 + 2.329(1.739) = 8.383$. Tolerances faktoru $k_1 = 2.329$ aprēķinām pēc formulas (2.1.4). Tātad, $\exp(8.383) = 4376.39$ ir $(0.95, 0.90)$ augšējā tolerances robeža svina līmeņiem gaisā. Darba drošības un veselības administrācijas noteiktā robeža svina līmenim gaisā ir $50\mu g/m^3$. Darba vieta tiek uzskatīta par drošu, ja augšējā tolerances robeža nepārsniedz šo noteikto robežu. Acīmredzami, ka augšējā robeža 4376.39 to stipri pārsniedz, tātad nevar uzskatīt, ka darba vieta ir droša.

3. *Rezultāta pārbaude:*

   Lai pārliecinātos, ka augšējā tolerances robeža ir aprēķināta korekti, pielietosim programmā R iebūvēto paketi *tolerance*. Šī pakete sniedz iespēju aprēķināt tolerances intervālus un tolerances robežas datiem, kas atbilst normālajam sadalījumam, kā arī lognormālajam sadalījumam. Pielietojot abus aprēķinu paņēmienus, iegūstam rezultātus, ko apkopojam tabulā 5.2.. Varam secināt, ka rezultāti sakrīt.

5.2. tabula Tolerances robežas normālajam un lognormālajam sadalījumam

Sadalījums	α	p	$\bar{X}$	Apakšējā robeža	Augšējā robeža
Normālais	0.1	0.95	4.332862	0.2817462	8.383979
Lognormālais	0.1	0.95	76.16198	1.325442	4376.386

4. Neparametriskā metode:

   Teorētiskajā materiālā tika aprakstīta metode neparametrisko tolerances intervālu konstruēšanai, kas balstīta uz autora Wilks publikāciju. Programmas R paketes *tolerance* iebūvētā komanda *nptol.reg* sniedz iespēju aprēķināt tolerances robežas un intervālus neparametriski, turklāt automātiski atsaucoties uz Wilks metodi. Aprēķināsim augšējo neparametrisko tolerances robežu svina līmeņa datiem. Aprēķinot šo robežu sākotnējiem datiem, iegūst robežas vērtību 1400, kas ir lielākā vērtība šajā izlasē. Augšējā neparametriskā tolerances robeža logtransformētajiem datiem ir 7.24423, kas arī šajā gadījumā ir lielākā šo datu vērtība. Pārbaudīsim, vai izlases apjoms atbilst nosacījumam (4.6.4). Tātad jāizpildās

$$(n-1)p^n - np^{n-1} + 1 \geq 1 - \alpha,$$

kad $n = 15$ un $1 - \alpha = 0.9$. Iegūstam rezultātu $0.17095 \ngeq 0.9$, tātad šoreiz izlases apjoms neatbilst nosacījumiem, un nevar pieņemt, ka neparametriskās tolerances robežas aprēķinātas korekti. Paeksperimentējot ar p, ieguvām, ka pie vērtības $p = 0.8$ vēlamais rezultāts netika sasniegts $0.83287 \ngeq 0.9$, bet pie $p = 0.75$ tika uzrādīts apmierinošs rezultāts $0.91982 \geq 0.9$. Ievietojam jauno p iebūvētajās komandās. Rezultātā secinām, ka ar ticamību 90% var apgalvot, ka vismaz $p = 0.75$ nepārniegs izlases vērtību 1000.

5. *Pārsniegšanas varbūtība:*

   Lai novērtētu varbūtību, ka svina līmenis nejauši izvēlētā vietā pārsniedz noteikto robežu, aprēķināsim 95% apakšējo ticamības robežu pārsniegšanas varbūtībai $P(X > R) = P(X > 50)$. Izmantojot (1.1.3), meklējam $(p, 0.95)$ apakšējo tolerances robežu vienādu ar $\ln(50)$, un p atrisinājumu. Tas ir, 95% apakšējā tolerances robeža varbūtībai $P(X > 50)$ ir atrisinājums vienādojumam

$$\bar{X} - \frac{1}{\sqrt{n}}\, t_{n-1;0.95}(z_p\sqrt{n})S = 4.333 - \frac{1}{\sqrt{15}}\, t_{14;0.95}(z_p\sqrt{15}) \times 1.739 = \ln(50).$$

   Lai atrisinātu šo vienādojumu, pirmkārt, nosakam $t_{14;0.95}(z_p\sqrt{15}) = 0.9376$. Tālāk aprēķinām $z_p\sqrt{15} = -0.7486$ jeb $p = 0.423$. Tātad, $P(X > 50)$ vismaz izlases daļa apjomā 0.423 ar ticamību 0.95.

Piemērs 2. *Uzpildes mašīnu monitorings.*

Mašīnai ir uzlikts iepildīt plastmasas pudelē litru piena. Maiņas beigās tika atlasīta izlase no 20 pudelēm, un tika smalki izmērīts katras pudeles piena daudzums. Šie mērījumi doti tabulā 5.3.

5.3. tabula :Piena daudzums pudelēs (litros)

0.968	0.982	1.030	1.003	1.046	1.020	0.997	1.010	1.027	1.010
0.973	1.000	1.044	0.995	1.020	0.993	0.984	0.981	0.997	0.992

1. *Datu normalitātes pārbaude:*

   Lai pārliecinātos, ka varam pielietot formulas, kas atbilst normālajam sadalījumam, ir jāpārbauda datu normalitāte. Šādiem uzdevumiem ir paredzēti normalitētes testi.

Šajā piemērā izmantosim *Lilliefors(Kolmogorova - Smirnova)* normalitātes testu. Izvirzam hipotēzi:

$$H_0 : X \sim N(\mu, \sigma^2) \quad \text{un} \quad H_1 : X \nsim N(\mu, \sigma^2).$$

Rezultātā iegūstam, ka p - vērtība ir vienāda ar 0.6668, tātad nav pamata noraidīt nulles hipotēzi, jo šī vērtība ir lielāka par jebkuru no α vērtībām $0.01, 0.05$ vai 0.1.

2. *Tolerances intervāls:*

Lai novērtētu uzpildes mašīnas precizitāti, aprēķināsim divpusējo tolerances intervālu, izmantojot dotās tabulas datus. Izlases vidējā vērtība $\bar{X} = 1.0036$ un standartnovirze $S = 0.0221$. Izmantojot šos lielumus, aprēķināsim $(0.99, 0.95)$ divpusējo tolerances intervālu. Šim nolūkam ar programmas R palīdzību tika aprēķināta tolerances faktora k_2 aproksimācija, kas atbilst nosacījumam (2.2.5).
Iegūstam, ka $k_2 = 3.615$, pie nosacījumiem $(n = 20, p = 0.99, 1 - \alpha = 0.95)$. Tolerances intervāls $\bar{X} \pm k_2 S = 1.0036 \pm 3.615(0.0221) = 1.0036 \pm 0.07989$. Tātad, vismaz 99% pudeļu ir piepildītas ar piena apjomu, kas ir robežās no 0.9237 līdz 1.0835 litriem ar ticamību 0.95.

3. *Rezultātu pārbaude:*

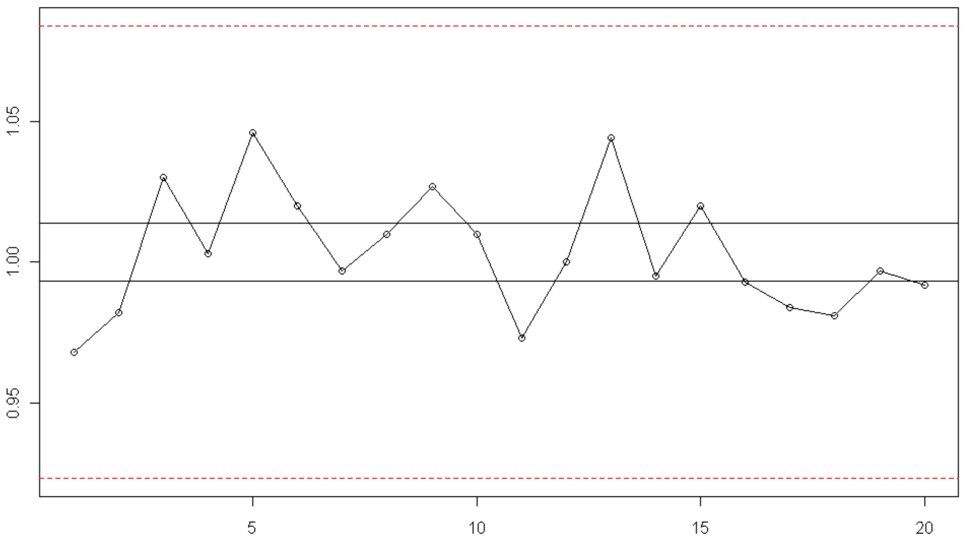
Iegūtos rezultātus pārbaudam ar iebūvētās komandas *normtol.int* palīdzību. Iegūtie rezultāti redzami tabulā 5.4., var secināt, ka ir vērojamas minimālas atšķirības.

5.4. tabula Tolerances intervāls normālajam sadalījumam piena datiem

Sadalījums	α	p	$\bar{X}$	Tolerances intervāls
Normālais	0.05	0.99	1.0036	$(0.923346, 1.083854)$

4. *Ticamības intervāls:*

Intereses pēc varam salīdzināt iegūtos rezultātus ar ticamības intervālu. Pie ticamības līmeņa 95% to varam izdarīt ar R programmā iebūvētās komandas *t.test* palīdzību. Rezultātā iegūstam, ka izlases vidējā vērtība 1.0036 atrodas intervālā $(0.9933, 1.0139)$. Ar programā R iebūvēto komandu *plottol* attēlosim ticamības un tolerances intervālus kopējā grafikā 5.1.. Varam redzēt, ka tolerances intervāls (pārtrauktās līnijas) attēlo, ka vismaz 99% no datiem atradīsies robežās $(0.923346, 1.083854)$ ar ticamību



5.1. att. Ticamības intervāls un tolerances intervāls piena datiem

95%, bet ticamības intervāls (nepārtrauktās līnijas) parāda, ka vidējā vērtība ar ticamību 95% atradīsies starp robežām $(0.9933, 1.0139)$. Varam secināt, ka ticamības intervāls nesniedz pietiekoši daudz informācijas par visiem datiem kopumā.

5. *Neparametriskā metode:*

Izmēģinām neparametrisko tolerances intervālu konstruēšanas metodi uz piena datiem. Rezultātā, pie $\alpha = 0.05$ un $p = 0.99$ divpusējais neparametriskais tolerances intervāls ir $(0.968, 1.046)$. Varam secināt, ka šajā piemērā neparametriskais tolerances intervāls ir mazākā izlases datu vērtība un lielākā izlases datu vērtība. Dotie aprēķini tika veikti ar programmā R iebūvētās komandas palīdzību. Pārbaudīsim, vai izlases apjoms atbilst nosacījumam (4.6.4). Rezultātā iegūst, ka $0.01686 \ngeq 0.95$. Acīmredzami, ka šajā piemērā neparametriskā metode nevar tikt veiksmīgi pielietota, ja $p = 0.99$ un $1 - \alpha = 0.95$. Pamainot p, ieguvām, ka tad, kad $p = 0.8$ rezultāts atbilst nosacījumiem. Tātad var secināt, ka ar ticamību 0.95% vismaz izlases daļa $p = 0.8$ iekļausies robežās $0.968, 1.046$.

Piemērs 3. *Alkohola koncentrācija asinīs*

Šis piemērs bastīts uz apsekojumu, kas iegūts Kriminālistikas laboratorijā, Luiziānā, ASV. Tā mērķis bija salīdzināt elpas novērtējumu alkohola koncentrācijai asinīs (iegūti, izmantojot elpas analizatoru) ar asins analīzēm laboratorijā. Dati ņemti no [19]. Tika veikti 15 mērījumi. Tabulā 5.5. tika atainoti elpas novērtējumi apzīmēti ar Y un laboratorijas asins testa rezultāti apzīmēti ar x. Šie skaitļi ataino procentuālo alkohola koncetrāciju asinīs.

5.5. tabula : Alkohola koncentrācijas asinīs mērījumu dati

Apsekojums	Alkohola koncentrācija asinīs (x)	Elpas novērtējums (Y)
1	0.160	0.145
2	0.170	0.156
3	0.180	0.181
4	0.100	0.108
5	0.170	0.180
6	0.100	0.112
7	0.060	0.081
8	0.100	0.104
9	0.170	0.176
10	0.056	0.048
11	0.111	0.092
12	0.162	0.144
13	0.143	0.121
14	0.079	0.065
15	0.006	0.000

Vienkāršais lineārās regresijas modelis labi atbilst šiem datiem, un novērtētā regresijas taisne izskatās sekojoši

$$\hat{Y} = 0.00135 + 0.958x. \tag{5.0.1}$$

Daudzos ASV štatos pieļaujamā alkohola koncentrācija asinīs, braucot ar automašīnu, nepārsniedz 0.1. Tā vietā, lai veiktu asins analīzes, alkohola līmeņa noteikšanai, daudz vienkāršāk ir iegūt elpas novērtējumu. Uzdevums ir novērtēt alkohola koncentrāciju asinīs x, kad ir iegūts elpas novērtējums Y.

1. *Datu normalitātes pārbaude:*

   Lai aprēķinos izmantotu formulas normālajam sadalījumam, Y datiem jāveic normalitātes pārbaudi (pamatojumu skatīt [20]). Ar programmas R palīdzību tika veikts *Shapiro-Wilk* tests (testa aprakstu skatīt [15, 461 lpp.]). Tika izvirzīta hipotēze

$$H_0 : Y \sim N(\mu, \sigma^2) \quad \text{un} \quad H_1 : Y \nsim N(\mu, \sigma^2).$$

Rezultātā ieguvām, ka testa statistika $W = 0.9532$ un p - vērtība ir 0.576. Rezultātu salīdzinam ar *Lielliefors (Kolmogorova - Smirnova)* testu, iegūstam, ka p - vērtība ir vienāda ar 0.8414. Sekojot hipotēžu testu nosacījumiem, varam secināt, ka šajā gadījumā nav pamata noraidīt nulles hipotēzi, jo p - vērtība ir lielāka par jebkuru nozīmības līmeni $\alpha = 0.01, 0.05, 0.1$.

2. *Augšējā tolerances robeža:*

Ir jāaprēķina $(0.90, 0.95)$ apakšējā tolerances robeža sadalījumam $Y(\boldsymbol{x})$, kad $\boldsymbol{x}' = (1, 0.10)$. Atbilstošais modelis ir $\hat{Y} = 0.00135 + 0.958x$, un nepieciešamie lielumi tolerances robežas aprēķināšanai ir sekojoši

$$(\boldsymbol{X'X})^{-1} = \begin{pmatrix} 0.4264 & -3.0538 \\ -3.0538 & 25.9240 \end{pmatrix},$$

$$d = (\boldsymbol{x}'\boldsymbol{X'X})^{-1}\boldsymbol{x})^{\frac{1}{2}} = 0.273643.$$

Tika iegūts, ka $\boldsymbol{x}'\hat{\boldsymbol{\beta}} = 0.00135 + 0.958(0.10) = 0.0971$ un $S = 0.01366$. Kritiskā vērtība $t_{n-m;1-\alpha}(z_p/d) = 2.1172$, kur $z_p = z_{0.9} = 1.2816$. $(0.90, 0.95)$ apakšējā tolerances robeža ir

$$\boldsymbol{x}'\hat{\boldsymbol{\beta}} - k_1(d)S = 0.0971 - 2.1172 \times 0.01366 = 0.068.$$

Tātad ar ticamību 95% vismaz 90% elpas alkohola novērtējumiem pārsniedz 0.068, kad alkohola līmenis asinīs ir 0.10.

3. *Rezultātu pārbaude:*

Lai pārbaudītu iegūto rezultātu, salīdzinam to ar programmā R iebūvēto komandu *regtol.int*. Programma izdod tabulu ar Y vērtībām, šo vērtību novērtējumiem $\hat{Y}$ un tolerances robežu vērtībām. Uzskatāmākā tabulas daļa redzama attēlā 5.6.. Varam redzēt, ka pie $\hat{Y}$ vērtības 0.0971 apakšejā tolerances robeža sakrīt ar mūsu aprēķināto 0.068, pie alkohola koncentrācijas, noapaļojot, 0.1.

5.6. tabula Tolerances robežas balstītas uz Wilson-Hilferty aproksimāciju

p	α	Y	$\hat{Y}$	Apakšējā robeža	Augšējā robeža
0.9	0.05	0.108	0.097147450	0.06822952	0.12606538
0.9	0.05	0.112	0.097147450	0.06822952	0.12606538
0.9	0.05	0.104	0.097147450	0.06822952	0.12606538

Piemērs 4. *Viskozitātes dati*

Tiek apskatītas saistības starp viskozitāti un diviem procesa mainīgajiem: reakcijas temperatūru un katalizatora caurplūdi. Dati ņemti no [21, 394–395]. Doti 16 rādījumi, kas atainoti tabulā 6.3. Y attēlo viskozitātes datus ($cSt/100°C$), x_1 - temperatūru pēc Celsija, un x_2 - katalizatora caurplūdi (lb/h).

5.7. tabula : Viskozitātes dati

Apsekojums Nr.	Temperatūra (x_1)	Katalizatora caurplūde (x_2)	Viskozitāte (Y)
1	80	8	2256
2	93	9	2340
3	100	10	2426
4	82	12	2293
5	90	11	2330
6	99	8	2368
7	81	8	2250
8	96	10	2409
9	94	12	2364
10	93	11	2379
11	97	13	2440
12	95	11	2364
13	100	8	2404
14	85	12	2317
15	86	9	2309
16	87	12	2328

Regresijas vienādojums attiecībā uz viskozitāti Y, temperatūru x_1 un katalizatora caurplūdi x_2 ir sekojošs

$$\hat{Y} = 1566.078 + 7.621x_1 + 8.585x_2, \tag{5.0.2}$$

ar kvadrātu korelāciju 0.927. Atlikumu vidējā vērtība kvadrātā S^2 pieņem vērtību 267.604.

Pēc datiem var redzēt, ka viskozitāte palielinās līdz ar temperatūru. Šo faktu vajadzētu ignorēt, lai varētu veiksmīgi veikt analīzi un uzskatīt, ka dispersija ir konstanta. Šajā gadījumā tolerances intervāls viskozitātes datiem ir intervāls, kas satur vismaz specifisku proporciju no viskozitātes vērtībām ar noteiktu ticamības līmeni pie fiksēta temperatūras un katalizatora caurplūdes vērtību pāra.

1. *Datu normalitātes pārbaude:*

   Ir nepieciešams novērtēt normalitāti. Šim nolūkam tiek pielietots *Shapiro-Wilk* normalitātes tests. Tiek izvirzīta hipotēze

   $$H_0 : Y \sim N(\mu, \sigma^2) \quad \text{un} \quad H_1 : Y \nsim N(\mu, \sigma^2).$$

   Ar programmā R iegūvēto komandu *shapiro.test* iegūstam, ka testa statistikas vērtība W ir vienāda ar 0.9718 un p - vērtība ir 0.8663. Salīdzinam ar *Lielliefors (Kolmogorova - Smirnova)* testu, iegūstam, ka p - vērtība ir vienāda ar 0.8872. Tātad varam secināt, ka nav pamata noraidīt nulles hipotēzi, jo p - vērtība ir lielākā par jebkuru nozīmības līmeni $\alpha = 0.01, 0.05, 0.1$.

2. *Augšējā tolerances robeža:*

   Piemēram, ir jāaprēķina $(0.90, 0.95)$ vienpusējo augšējo tolerances robežu sadalījumam $Y(\boldsymbol{x})$ pie $\boldsymbol{x}' = (1, 88, 9)$. Jāpiezīmē, ka $n = 16$ un $m = 3$, un atbilstošais modelis ir atbilstoši (5.0.2). Nepieciešamie lielumi, lai aprēķinātu tolerances robežas

   $$(\boldsymbol{X'X})^{-1} = \begin{pmatrix} 14.176 & -0.1297 & -0.2235 \\ -0.1297461 & 0.00143 & -0.00005 \\ -0.2234529 & -0.00005 & 0.02222 \end{pmatrix}, \tag{5.0.3}$$

   $$d = (\boldsymbol{x}'\boldsymbol{X'X})^{-1}\boldsymbol{x})^{\frac{1}{2}} = 0.33289 \text{ un } S = 16.3585. \tag{5.0.4}$$

   Turklāt $\boldsymbol{x}'\hat{\boldsymbol{\beta}} = 2314.02$, $z_{0.9} = 1.2816$ un $t_{n-m;1-\alpha}(z_p/d) = t_{13;0.95}(3.850) = 6.6023$. Tādā veidā $(0.90, 0.95)$ tolerances faktors $k_1(d) = d \times t_{13;0.95}(3.850) = 2.1977$ un atbilstošā augšējā tolerances robeža ir dota

   $$\boldsymbol{x}'\hat{\boldsymbol{\beta}} + k_1(d)S = 2314.02 + 2.1977 \times 16.3585 = 2349.97$$

   Tātad ar ticamību 95% vismaz 90% no sadalījuma $Y(\boldsymbol{x})$ mērījumiem nepārsniedz vērtību 2349.97, kad $\boldsymbol{x}' = (1, 88, 9)$.

3. Šajā piemērā iebūvēto programmas R komandu *regtol.int* nepielietosim, jo ir vairāk kā viena neatkarīgo mainīgo kolonna, tādēļ rezultāti var būt neprecīzi.

Piemērs 5. *Sārmainības koncentrācija gruntsūdeņos*

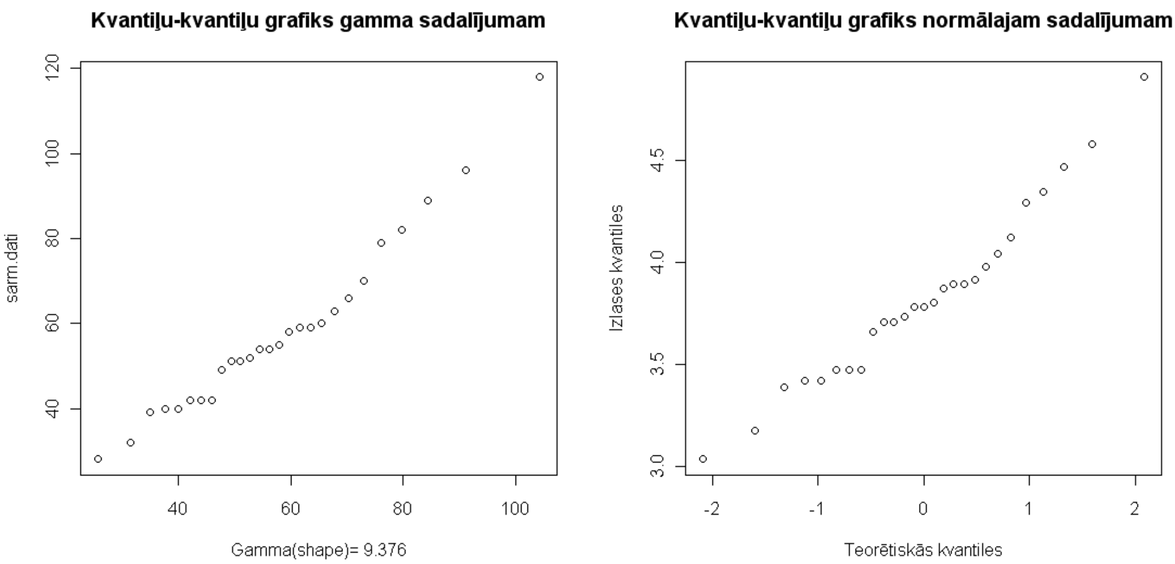
Mērījumi tabulā (5.8.), ataino sārmainības koncentrācijas (mg/L) gruntsūdeņos. Tie

5.8. tabula : Sārmainības koncentrācija gruntsūdeņos (mg/L)

Y :	28	32	39	40	40	42	42	42	49	51	51	52	54	54
	55	58	59	59	60	63	66	70	79	82	89	96	118	

iegūti no "*greenfield*" mājas lapas, kas ataino datus par atkritumu izmantošanu, izvietošanu izgāztuvēs u.c. Konkrētie dati ņemti no [1, 216 lpp.]

Šajā piemērā vēlamies pielietot *Wilson - Hilferty* aproksimāciju, kas atbilst gamma sadalījumam, lai parādītu, ka gamma transformētie dati dod līdzvērtīgus rezultātus normālajam sadalījumam (skatīt apakšnodaļu 2.4.2). Varbūtību grafiki oriģinālajiem gamma datiem un pārveidotajiem datiem ir doti attēlā 5.2. Jāatzīmē, ka šie divi varbūtību grafiki



5.2. att. Kvantiļu-kvantiļu grafiki gamma sadalījumam un normālajam sadalījumam

ir ļoti līdzīgi, un tie parāda, ka gamma kubsakņu transformētie dati labi atbilst normāli sadalītiem datiem.

Datu normalitātes pārbaude:

Lai pārliecinātos, ka dati patiešām labi atbilst normālajam sadalījumam, iegūtajām datu transformācijām veiksim datu normalitātes pārbaudi. Līdzīgi kā iepriekšējos piemēros pielietosim *Lielliefors (Kolmogorova - Smirnova)* normalitātes testu. Izvirza hipotēzi

$$H_0 : Y \sim N(\mu, \sigma^2) \quad \text{un} \quad H_1 : Y \nsim N(\mu, \sigma^2).$$

Rezultātā ieguvām, ka p- vērtība ir 0.3675. Sekojot hipotēžu testu nosacījumiem, varam secināt, ka šajā gadījumā nav pamata noraidīt nulles hipotēzi, jo p - vērtība ir lielāka par jebkuru nozīmības līmeni $\alpha = 0.01, 0.05, 0.1$

Lai pielietotu *Wilson - Hilferty* aproksimāciju, tika aprēķināta vidējā vērtība un standartnovirze transformētajiem kubsaknes datiem $\bar{Y} = 3.8274$ un $S_y = 0.4298$.

1. *Tolerances robežas*:

   Tabulā 5.9. attēloti 95% vienpusējās tolerances robežas un divpusējie tolerances intervāli ar atbilstošām tolerances faktoru vērtībām.

5.9. tabula Tolerances robežas balstītas uz *Wilson - Hilferty* aproksimāciju

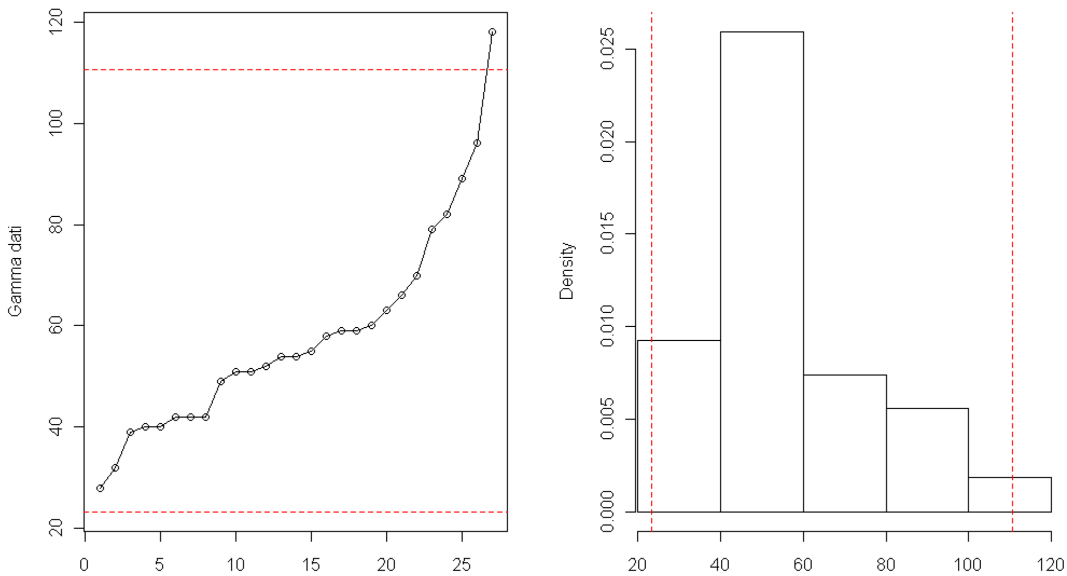
$(p, 1-\alpha)$	Vienpusējais faktors	Apakšējā robeža	Augšējā robeža	Divpusējais faktors	Divpusējais tolerances intervāls
$(0.9, 0.95)$	1.8114	28.341	97.714	2.178	$(24.17, 108.089)$
$(0.95, 0.95)$	2.26	23.297	110.505	2.5949	$(19.949, 120.751)$
$(0.99, 0.95)$	3.1165	15.4	137.938	3.4093	$(13.179, 148.264)$

Lai pārliecinātos, ka teorētiskās formulas sniedz pietiekoši precīzus rezultātus, salīdzināsim tos ar programmā R iebūvētajām komandām tolerances faktoru un intervālu aprēķināšanai. Iebūvēto komandu izdotos rezultātus apvienojam tabulā 5.10.

5.10. tabula Tolerances robežas ar R programmā iebūvētajām komandām

$(p, 1-\alpha)$	Vienpusējais faktors	Apakšējā robeža	Augšējā robeža	Divpusējais faktors	Divpusējais tolerances intervāls
$(0.9, 0.95)$	1.8114	28.780	96.8298	2.184	$(23.3268, 110.5421)$
$(0.95, 0.95)$	2.26	23.773	109.2969	2.602	$(19.061, 123.9004)$
$(0.99, 0.95)$	3.1165	15.894	135.9791	3.42	$(12.2895, 153.0323)$

Iebūvētajās komandās tolerances faktora aprēķināšanai tiek izmantota cita metode, kas tiek uzskatīta par visprecīzāko. Šo metodi sauc par *Howe* metodi (metodes aprakstu skatīt [22]). Varam redzēt, ka iegūvētās komandas izdod nedaudz atšķirīgāku rezultātu. Šī ir tā novirze, kas tika aprakstīta teorētiskajā materiālā par gamma sadalījumu. Iebūvētā komanda vidējo vērtību un dispersiju rēķina pēc formulām (2.4.4), turpretim mēs pielietojām $\bar{Y} = 3.8274$ un $S_y = 0.4298$. Uzskatīsim iebūvēto komandu gadījumu par precīzāku, jo kubsaknes no datiem, kas atbilst gamma sadalījumam pilnīgi precīzi nesakrīt ar datiem, kas atbilst normālajam sadalījumam. Lai rezultāts būtu uzskatāms, divpusējais $(0.9, 0.95)$ tolerances intervāls gamma sadalītajiem sārmainības datiem attēlots grafiski 5.3..



5.3. att. Tolerances intervāls sārmainības datiem ar $p = 0.90$ un $\alpha = 0.05$

2. *Varbūtība sliekšņa vērtības pārsniegšanai*: Atradīsim 95% apakšejo robežu varbūtībai, ka izlases sārmainības koncentrācija pārsniegs 41(mg/L), tas ir, $P(Y > t) = P(X > 41^{\frac{1}{3}})$. Pielietojot (2.4.6), iegūstam

$$t_{26;0.95}(z_p\sqrt{27}) = \frac{3.8274 - 41^{\frac{1}{3}}}{0.4298/\sqrt{27}} = 4.584.$$

Atrodam necentralitātes parametru $z_p\sqrt{27} = 2.601$. Tātad $z_p = 0.5006$ un pēc tabulām atrodam $p = \Phi(0.5006) = 0.692$. Rezultātā secinam, ka varbūtība sārmainības koncentrācijai pārsniegt 41(mg/L) izlasē ir vismaz 0.692 ar ticamību 95%.

3. *Neparametriskā metode*:

Šiem pašiem tabulas 5.8. datiem vēlamies konstruēt neparametriskos tolerances intervālus. Sākumā tiek aprēķināta 95% augšējā ticamības robeža sārmainības koncentrācijas proporcijai, kas iekļauta intervālā $(X_{(1)}, X_{(n)})$. Tā kā kreisā izteiksmes daļa (4.6.4) ir dilstoša funkcija no p, uzdevums ir noteikt maksimālo p vērtību, kas atbilst nevienādībai (4.6.4). Tātad, ja mūsu konkrētajā gadījumā $n = 27$, tad

$$26p^{27} - 27p^{27-1} + 1 \geq 0.95. \tag{5.0.5}$$

Paeksperimentējot ar dažādām varbūtībām p, var secināt, ka lielākā varbūtība p, kas atbilst nevienādībai (5.0.5) ir 0.836. Jāpiezīmē, ka sārmainības datiem $X_{(1)} = 28$ un $X_{(n)} = 118$. Tātad var teikt, ka ne vairāk kā 83.6% no sārmainības koncentrāciju datiem atrodas starp 28 un 118, ar ticamības līmeni 0.95.

Šim piemēram neeksistē sakārtoto statistiku pāris, kas veidotu $(p, 0.95)$ tolerances intervālu, kad $p > 0.836$. Tādēļ konstruēsim $(0.75, 0.95)$ tolerances intervālu. Aplūkojot kolonnu, kas atbilst $(p, 1-\alpha) = (0.75, 0.95)$ tabulā C1., varam redzēt, ka izlases apjoms, kas ir mazāks par 27 ir 23, un tas atbilst vērtībai $m = 2$. Tad intervāls $(X_{(2)}, X_{(n)}) = (32, 118)$ vai $(X_{(1)}, X_{(n-1)}) = (28, 96)$ ir tolerances intervāls. Vienpusējā apakšējā tolerances robeža ir $X_{(m+1)} = X_{(3)} = 39$ un vienpusējā augšejā tolerances robeža ir $X_{(n-m)} = X_{(25)} = 89$.

Konstruēsim $(0.75, 0.95)$ tolerances robežas, kas pamatojas uz normālo sadalījumu. Divpusējo tolerances faktoru aprēķina pēc (2.2.5), tātad $k_2 \simeq 1.523$. Ņemot vērā, ka $\bar{Y} = 3.8274$ un $S_y = 0.4298$, varam rēķināt tolerances intervālu $(3.8274 \pm 1.523 \times 0.4298)^3 = (31.94, 90.03)$. Parametriskais tolerances intervāls ir šaurāks par abiem iegūtajiem neparametriskajiem tolerances intervāliem. Tolerances faktors $(0.75, 0.95)$ vienpusējām tolerances robežām var tikt aprēķināta pēc (2.1.4), tātad $k_1 = 1.083$. Apakšejā tolerances robeža sārmainības koncentrācijām ir $(3.8274 - 1.083 \times 0.4298)^3 = 38$ un augšejā tolerances robeža ir vienāda ar $(3.8274 + 1.083 \times 0.4298)^3 = 79.11$. Varam secināt, ka tolerances robeža, balstīta uz normālo sadalījumu ir vienāda ar 38, bet neparametriskā ir lielāka - 39 (apakšējai robežai ir labāk, ja tā ir lielāka), bet augšejā tolerances robeža 79.11 ir mazāka par 89 (augšējai tolerances robežai ir labāk, ja tā ir mazāka).

4. *Varbūtība sliekšņa vērtības pārsniegšanai (neparametriskajā gadījumā)*

Gribam aprēķināt 95% apakšejo robežu varbūtībai, ka izlases sārmainības koncentrācija pārsniegs 41(mg/L) neparametriskajā gadījumā pie $p = 0.75$, tas ir, $P(X > 41)$. Pēc tabulas (C1.) nosakām, ka mazākā sakārtotā statistika, kas ir lielāka par 41 ir $X_{(6)}$. Tātad no apakšnodaļas 4.4 seko

$$P(X > 41) \geq beta(\alpha; n - k + 1, k) = beta(0.05; 22, 6) = 0.649.$$

Jāuzsver, ka apakšējā ticamības robeža varbūtībai $P(X > 41)$, kas balstīta uz gamma sadalījumu ir vienāda ar 0.692, kas ir lielāks skaitlis nekā neparametriskā apakšejā ticamības robeža 0.649.

Secinājumi

Izpētot teorētisko materiālu par tolerances robežām un intervāliem, var secināt, ka labi pazīstamie ticamības intervāli problemātikas ziņā ir atšķirīgi. Tolerances intervāli ir lieterīgi, ja ir jānosaka, kādu vērtību robežās atradīsies vismaz noteikta populācijas procentuālā daļa ar noteiktu ticamību, turpretim ticamības intervāli sniedz informāciju par kādu nezināmu parametru.

Diplomdarbā ir sniegts plašs teorētiskais materiāls par tolerances robežām un intervāliem normālajam sadalījumam. Ir analizēts gadījums, kad transformētie dati no lognormālā vai gamma sadalījuma atbilst normāli sadalītiem datiem. Ir apskatīti tolerances intervāli lineārās regresijas gadījumam, kā arī neparametriskajam gadījumam. Teorija par neparametriskajām tolerances robežām un tolerances intervāliem galvenokārt balstās uz 1941.gada publikāciju. Ir jāsecina, ka neparametrisko tolerances robežu konstruēšana, balstoties uz sakārtotām statistikām ir lielā mērā atkarīga no dotās izlases apjoma. Tika sniegta teorija optimālai izlases apjoma atrašanai.

Ar piemēru palīdzību ir izdevies parādīt, ka tolerances robežām un intervāliem ir daudzveidīgs un lietderīgs praktisks pielietojums. Tos pielieto vides piesārņojuma novērtēšanā (šobrīd ļoti aktuāla tēma visā pasaulē), ražošanas kvalitātes kontrolē, prognožu veikšanai lineārās regresijas gadījumā. Parametriskās metodes tika salīdzinātas ar neparametriskām. Var secināt, ka atsevišķā gadījumā neparametriskā metode uzrāda labāku rezultātu, jo ļoti liela nozīmē ir pareizai izlases apjoma atrašanai, pretējā gadījumā neparametriskā metode nedod vēlamo rezultātu.

Nepieciešamie aprēķini tika veikti ar brīvpieejas programmas R palīdzību. Lai pārliecinātos, ka iegūtie rezultāti ir korekti, tika atrasta speciāla statistiskā pakete tolerances intervālu konstruēšanai ar nosaukumu *tolerance*.

Nākotnē būtu lietderīgi izpētīt tolerances intervālu konstruēšanu ar citām neparametriskajām metodēm, piemēram, ar *bootstrap* metodi. Šī problemātika ir pietiekoši sarežģīta, to pielieto bioinženierijā, medicīniskajos pētījumos, kā arī riska novērtēšanā (uzskatāmu publikāciju skatīt [23]).

Izmantotā literatūra un avoti

[1] R.D. Gibbons. *Statistical methods of Groundwater Monitoring.* Wiley, New York, 1994.

[2] Coleman D.E. Gibbons, R.D. *Statistical methods for Detection and Quantification of Environmental Contamination.* Wiley, New York, 2001.

[3] Neerchal N.K. Millard, S.P. *Environmental Statistics with S - plus.* CRC Press, New York, 2000.

[4] S.S. Wilks. *Determination of sample sizes for setting tolerance limits.* 1941.

[5] S.S. Wilks. *Statistical prediction with special reference to the problem of tolerance limits.* 1942.

[6] Thomas Mathew Krishnamoorthy, K. *Statistical Tolerance Regions.* John Wiley & Sons,INC., New Jersey, 2009.

[7] `http://cran.r-project.org/web/packages/tolerance/tolerance.pdf`.

[8] A. Wald. *An extension of Wilks' method for setting tolerance limits.* 1943.

[9] Wolfowitz J. Wald, A. *Tolerance limits for a normal distribution.* 1946.

[10] Ouarda T.B.M.J. Ashkar, F. *Aproximate confidence intervals for quantiles of gamma and generalized gamma distributions*, pages 43–51. 1998.

[11] Bobee B. Ashkar, F. *Confidence intervals for flood events under pearson distribution*, pages 639–650. 1988.

[12] H. Aksoy. *Use of gamma distribution in hydrological analysis*, pages 419–428. 2000.

[13] Gibbons R.D. Bhaumik, D.K. *One sided approximate prediction intervals for at least p of m observations from a gamma population at each of r locations.* 2006.

[14] Hilferty M.M. Wilson, E.B. *The distribution of chi-squares.* 1931.

[15] Ana F.Militino Alan T.Arnholt Ugante, M.D. *Probability and statistics with R.* Taylor & Francis group, Florida, 2008.

[16] W.A. Wallis. *Tolerance Interval for Linear Regression.* 1951.

[17] John Y. Lu. *Tolerance Interval for Multiple Regression.* 1960.

[18] H. Lilliefors. *On the Kolmogorov Smirnov test for normality with mean and variance unknown.* 1967.

[19] Kulkarni P.M. Mathew T. Krishnamoorthy, K. *Multiple use one-sided hypotheses testing in univariate linear calibration*, pages 219–220. 2001.

[20] `http://www.variation.com/da/help/hs144.htm`.

[21] Douglas C. Montgomery. *Design and Analysis of Experiments.* 7th edition.

[22] W.G. Howe. *Two-Sided Tolerance Limits for Normal Populations-Some Improvements.* 1969.

[23] Jose R.Fernandez Diana E.Alaya Ramon, C.Hermida. *Circadian Time - Qualified Tolerance Intervals for Ambulatory Blood Pressure Monitoring in the Diagnosis of Hypertension.* 2004.

[24] W.A. Shewhart. *Economic Control of Quality of Manufactured Product.* 1931.

A Pamatformulas

Šajā nodaļā tiks apskatīts un pierādīts rezultāts, kuram ir svarīga loma tolerances intervālu iegūšanā normālajam sadalījumam un citiem normāli bāzētiem modeļiem.

Pieņemsim, ka $X \sim N(0, c)$ neatkarīgi no $Q \sim \frac{\chi^2_m}{m}$, kur χ^2_ν apzīmē hī-kvadrāta gadījuma mainīgo ar brīvības pakāpēm m. $0 < p < 1$, $0 < \gamma < 1$ un ar Φ apzīmēsim standarta normālo sadalījuma funkciju.

(i) Faktoram k_1 atbilst nosacījums

$$P_{X,\,Q}\Big(\Phi\Big(X + k_1\sqrt{Q}\Big) \geq p\Big) = \gamma \tag{A0.1}$$

ir dots sekojošā veidā

$$k_1 = \sqrt{c} \times t_{m;\,\gamma}\left(\frac{z_p}{\sqrt{c}}\right), \tag{A0.2}$$

kur z_p apzīmē p kvantili standarta nomālajam sadalījumam, un $t_{\nu;\,\eta}(\delta)$ apzīmē η kvantili necentrētam t sadalījumam ar brīvības pakāpēm ν un necentralitātes parametru δ.

(ii) Faktors k_2, kas apmierina

$$P_{X,\,Q}\Big(\Phi(X + k_2\sqrt{Q} - \Phi(X - k_2\sqrt{Q}) \geq p\Big) = \gamma \tag{A0.3}$$

ir atrisinājums integrālajam vienādojumam

$$\sqrt{\frac{2}{\pi c}} \int_0^\infty P_Q\left(Q \geq \frac{\chi^2_{1;\,p}(x^2)}{k_2^2}\right) e^{-\frac{x^2}{2c}} dx = \gamma, \tag{A0.4}$$

kur $\chi^2_{\nu;\,\eta}(\delta)$ apzīmē η kvantili necentrālajam hī-kvadrāta sadalījumam ar brīvības pakāpēm ν un necentralitātes parametru δ.

(iii) k_2 aproksimācija, kurai atbilst nosacījums (A0.4) ir dota sekojošā veidā

$$k_2 \simeq \left(\frac{m\chi^2_{1;\,p}(c)}{\chi^2_{m;\,1-\gamma}} \right)^{\frac{1}{2}}, \tag{A0.5}$$

kur $\chi^2_{\nu;\,\eta}$ apzīmē η kvantili hī- kvadrāta sadalījumam ar brīvības pakāpēm ν (un $\simeq$ nozīmē ekvivalents, atbilstošs).

Pierādījums. (i) Jāievēro, ka (A0.1) satur iekšējo varbūtības nevienādību tad un tikai tad, ja $X + k_1\sqrt{Q} \geq z_p$. Tātad var (A0.1) pierakstīt kā

$$\begin{aligned} P_{X,\,Q}\big(X + k_1\sqrt{Q} \geq z_p\big) &= P_{X,\,Q}\left(\frac{X - z_p}{\sqrt{Q}} \geq -k_1 \right) \\ &= P_{X,\,Q}\left(\sqrt{c}\,\frac{X/\sqrt{c} + z_p/\sqrt{c}}{\sqrt{Q}} \leq k_1 \right) \\ &= \gamma. \end{aligned} \tag{A0.6}$$

Lai iegūtu otro soli formulā (A0.6), jāizmanto fakts, ka X un $-X$ ir identiski sadalīti. Tādēļ, ka $X/\sqrt{c} \sim N(0, c)$ neatkarīgi no $Q \sim \frac{\chi^2_m}{m}$, iegūstam

$$\frac{X/\sqrt{c} + z_p/\sqrt{c}}{\sqrt{Q}} \sim t_m\left(\frac{z_p}{\sqrt{c}} \right),$$

kur $t_\nu(\delta)$ apzīmē necentrālu t gadījuma mainīgo ar brīvības pakāpēm ν un necentralitātes parametru δ. Tādēļ, k_1, kas apmierina (A0.6) ir dots ar (A0.2).

(ii) Ievērosim, ka fiksētam X, $\Phi(X + r) - \Phi(X - r)$ ir augoša funkcija no r. Tādēļ, fiksētam X, $\Phi(X + k_2\sqrt{Q}) - \Phi(X - k_2\sqrt{Q}) \leq p$ tad un tikai tad, ja $k_2\sqrt{Q} > r$ vai $Q > \frac{r^2}{k_2^2}$. kur r ir atrisinājums vienādojumam

$$\Phi(X + r) - \Phi(X - r) = p,$$

vai ekvivalenti,

$$P_Z\big((Z - X)^2 \leq r^2|X\big) = p, \tag{A0.7}$$

Z ir standarta normālais gadījuma lielums. Fiksētam X, $(Z - X)^2 \sim \chi^2_1(X^2)$, kur $\chi^2_m(\delta)$ apzīmē necentrālu hī-kvadrāta gadījuma mainīgo ar necentralitātes parametru δ. Tādēļ, nosacīti dotais X^2, r^2, kuram atbilst nosacījums (A0.7) ir p kvantile no

$\chi^2_1(X^2)$, kuru apzīmēsim ar $\chi^2_{1;\,p}(X^2)$. Lietojot šos rezultātus, un, ievērojot, ka r funkcija no X^2 un p, iegūsim

$$\begin{aligned} P_Q\left(\Phi(X + k_2\sqrt{Q}) - \Phi(X - k_2\sqrt{Q}) > p \;\middle|\; X\right) &= P_Q\left(Q > \frac{r^2}{k_2^2}\middle| X\right) \\ &= P_Q\left(Q > \frac{\chi^2_{1;\,p}(X^2)}{k_2^2}\middle| X\right). \end{aligned}$$

Ņemot matemātisko cerību attiecībā uz sadalījumu no X, varam redzēt, ka faktoram k_2 izpildās

$$E_X\left[P\left(Q > \frac{\chi^2_{1;\,p}(X^2)}{k_2^2}\right)\right] = \gamma. \tag{A0.8}$$

Zinot, ka $X \sim N(0, c)$, var pārrakstīt (A0.8) kā

$$\begin{aligned} &\frac{1}{\sqrt{2\pi c}}\int_{-\infty}^{\infty} P\left(Q > \frac{\chi^2_{1;\,p}(x^2)}{k_2^2}\right) e^{-\frac{x^2}{2c}}dx \\ &= \sqrt{\frac{2}{\pi c}}\int_{0}^{\infty} P\left(Q > \frac{\chi^2_{1;\,p}(x^2)}{k_2^2}\right) e^{-\frac{x^2}{2c}}dx \\ &= \gamma. \end{aligned} \tag{A0.9}$$

(iii) k_2 aproksimācija var tikt iegūta no (A0.8) sekojoši. Pieņemsim, ka $V = X^2$ un $g(V) = P_Q\left(Q > \frac{\chi^2_{1;\,p}(V)}{k_2^2}\right)$. Lietojot izvedumu Teilora rindā pie $V = E(V) = c$, iegūst

$$g(V) = g(c) + (V - c)\frac{\partial g(V)}{\partial V}\bigg|_{V=c} + \frac{(V-c)^2}{2!}\frac{\partial^2 g(V)}{\partial V^2}\bigg|_{V=c} + \dots$$

Ievērojot, ka $\frac{V}{c} \sim \chi^2_1$, un, ņemot matemātisko cerību abām pusēm, iegūstam

$$\begin{aligned} E(g(V)) &= g(c) + c^2\frac{\partial^2 g(V)}{\partial V^2}\bigg|_{V=c} + \dots \\ &= g(c) + O(c^2). \end{aligned}$$

Visbeidzot, $E(g(V)) \simeq g(c)$, un lietojot aproksimāciju (A0.8), iegūstam

$$P_Q\left(Q > \frac{\chi^2_{1;\,p}(c)}{k_2^2}\right) \simeq \gamma. \tag{A0.10}$$

Tā kā $Q \sim \frac{\chi^2_m}{m}$, tad no izteiksmes (A0.10) seko, ka $\frac{\chi^2_{1;\,p}(c)}{k_2^2} \simeq \frac{\chi^2_{m;\,1-\gamma}}{m}$, un kā atrisinājumu k_2 mēs iegūstam (iii).

$\square$

B Izlases apjoma noteikšana tolerancs robežu konstruēšanai

Šajā pielikumā aplūkota teorija no Wilks publikācijas [4]

Vēsturisks ieskats

1931.gadā autors Shewhard [24] izvirzīja jautājumu par to, kā noteikt konkrēta produkta vai ierīces kvalitātes svārstību cēloni, pēc kādiem noteiktiem kritērijiem masveida ražošanas kvalitātes noteikšanai. Šīs svārstības atspoguļo mainīgais x. Piemēram, x varētu būt stiepes izturība alumīnija stieplei, kas izveidota pēc kādiem specifiskiem nosacījumiem. x mainās no stieples uz stiepli. Šīs kvalitātes svārstības tiek uzskatītas kā "gadījuma rakstura". Tas ļauj mums uzskatīt, ka mums ir zināma telpa, kurā x ir gadījuma lielums ar kādu sadalījuma likumu $f(x)$. Šis sadalījums $f(x)$ ir nezināms. Praksē visbiežāk sastopamas divu veidu situācijas, kad x pieņem diskrētas vērtības, vai arī x ir nepārtraukts lielums kādā noteiktā apgabalā ar atbilstošu blīvuma funkciju $f(x)$. Apskatīsim otro gadījumu.

Problēma rodas tad, kad mums ir jākonstruē tolerances intervāls (L_1, L_2) mainīgajam x un jānosaka, cik lielai jābūt izlasei, lai tolerances intervāls atbilstu kādam noteiktam stabilitātes rādītājam. Sīkāk, pie dotas metodes tolerances robežu aprēķināšanai, cik lielai jābūt izlasei, pie nosacījuma, ka proporcija P no telpas atradīsies starp L_1 un L_2, vidējā vērtība būs a, un ar varbūtību vismaz p proporcija P iekļausies starp divām vērtībām b un c? Piemēram, ja tolerances apgabals tiek noteikts, izmantojot nošķeltās izlases apgabalu, tas ir, uzskatot L_1 par lielāko no r vismazākajām vērtībām izlasē un L_2 par mazāko no r vislielākajām vērtībām. Piemēram, izvēloties r tā, ka $E(P) = 0.99$, cik lielai jābūt izlasei n, lai ar varbūtību 0.90 proporcija P atrastos starp vērtībām 0985 un 0.995? Līdzīgs jautājums var tikt uzdots, konstruējot tikai vienu tolerances robežu.

B0.1. Tolerances apgabals nošķeltai izlasei

Uzskata, ka nekas nav zināms par sadalījuma blīvuma funkciju $f(x)$, izņemot to, ka tā ir nepārtraukta. Pieņem, ka a ir vidējā vērtība, ko satur P, un izlase apjomā n iekļaujas U tā, ka $[(1-a)(n+1)]/2 = r$ ir vesels pozitīvs skaitlis. Pieņem, ka $x_1, x_2, \ldots, x_n$ ir izlases x vērtības augošā secībā. Uzskata, ka $L_1 = x_r$ un $L_2 = x_{n-r+1}$. Sadalījuma likums $g(P)$ no proporcijas P, kas iekļaujas starp divām robežām L_1 un L_2 ir dots sekojošā veidā

$$g(P)dP = \frac{\Gamma(n+1)}{\Gamma[a(n+1)]\Gamma[(1-a)(n+1)]} P^{a(n+1)-1}(1-P)^{(1-a)(n+1)-1}dP. \qquad \text{(B0.1)}$$

Tas seko no kopējā x_n un x_{n-r+1} sadalījuma likuma, kas var tikt pierakstīts sekojoši

Pieņēmums 2. Uzskatam, ka x ass tiek sadalīta k neatkarīgos intervālos $I_1, I_2, \ldots, I_k$ ar attiecīgām varbūtībām $p_1, p_2, \ldots, p_k$, kā arī $\left(\sum_1^k p_i = 1\right)$. Izlasei apjomā n varbūtības, ka $n_1, n_2, \ldots, n_k$, kur $\left(\sum_1^k n_i = n\right)$ vērtības no izlases x iekļausies intervālos $I_1, I_2, \ldots, I_k$, ir dotas ar multinomiālo sadalījuma likumu

$$\frac{n!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}. \qquad \text{(B0.2)}$$

Lai iegūtu x_r un x_{n-r+1}, izvēlas $k = 5$ un kā $I_1, I_2, \ldots, I_k$ tiek ņemti intervāli

$$(-\infty, x_r), (x_r, x_r+dx_r), (x_r+dx_r, x_{n-r+1}), (x_{n-r+1}, x_{n-r+1}+dx_{n-r+1}), (x_{n-r+1}+dx_{n-r+1}, \infty).$$

Vērtības $p_1, p_2, \ldots, p_k$ ir integrāļi pieciem intervāliem no $f(x)dx$, un attiecīgās $n_1, n_2, \ldots, n_k$ vērtības ir $r-1, 1, n-2r, 1, r-1$. Šīs iegūtās vērtības tiek ievietotas izteiksmē (B0.2), iegūst

$$\begin{aligned} &\frac{n!}{[(r-1)!]^2(n-2r)!} \left(\int_{-\infty}^{x_r} f(x)dx\right)^{r-1} \left(\int_{x_{n-r+1}}^{\infty} f(x)dx\right)^{r-1} \\ &\cdot \left(\int_{x_r}^{x_{n-r+1}} f(x)dx\right)^{n-2r} f(x_r) f(x_{n-r+1}) dx_r dx_{n-r+1}. \end{aligned} \qquad \text{(B0.3)}$$

Tālāk apzīmē $\left(\int_{-\infty}^{x_r} f(x)dx\right) = u$, $\left(\int_{x_{n-r+1}}^{\infty} f(x)dx\right) = v$, ja $du = f(x_r)$ un $dv = -f(x_{n-r+1})dx_{n-r+1}$, tad (B0.1) var pierakstīt, izmantojot u un v

$$\frac{\Gamma(n+1)}{\Gamma^2(r)\Gamma(n-2r+1)} u^{r-1} v^{r-1} (1-u-v)^{n-2r} du\, dv. \qquad \text{(B0.4)}$$

Nenulles u un v varbūtības apgabals ir ierobežots ar u un v asīm, kā arī ar taisni $u + v = 1$. Izdarot izmaiņas mainīgajos $1 - u - v = P$ un $u = Q$, integrējot pēc Q, un nosakot $r = (1/2)(1-a)(n+1)$, tiek atrasts proporcijas P, kas iekļaujas starp x_r un x_{n-r+1}, sadalījums izteiksmē (B0.1). Jāatzīmē, ka pat, ja L_1 un L_2 ir iegūtas, asimetriski nošķeļot izlasi, tas ir, ņemot $L_1 = x_s$ un $L_2 = x_t$, kur $t-s = n-2r+1$, tad $P = \int_{x_s}^{x_s} f(x)dx$ sadalījums paliek nemainīgs. Tādā veidā, pie dota p, ņemot $L_1 = x_s$ un $L_2 = x_t$, kur $t-s = n-2r+1 = a(n+1)$, un izvēloties mazāko no n vērtībām , kurai $\int_b^c g(P)dP \geq p$ un, tā ka $(1-a)(n+1)$ ir pozitīvs vesels skaitlis, esam atraduši īsto veidu L_1 un L_2 noteikšanai. Šī metode var tikt pielietota jebkuras nezināmas nepārtrauktas funkcijas $f(x)$ noteikšanai.

Neliels piemērs:

Pieņemsim, ka $a = 0.99$, $b = 0.985$, $c = 0.995$ un $p = 0.99$. Nepieciešamais izlases izmērs ir 1000(konkrētāk 999). Šajā gadījumā varbūtība, ka P atradīsies starp vērtībām 0.985 un 0.995 ir 0.992. Šajā piemērā varam uzskatīt, ka x ir nepārtraukts mainīgais, un, ja tiek ņemtas izlases apjomā 1000, tad tolerances robežas L_1 un L_2 tiek ņematas kā piektā mazākā un piektā lielākā vērtība no x attiecīgajā izlasē. Vidēji 99% no telpas iekļausies starp šīm robežām, turklāt, ja tolerances robežas tiek rēķinātas izlasēm ar apjomu 1000, tad apmēram 99.2% no izlasēm saturēs starp šim tolerances robežām ap 98.5% līdz 99.5% no visām vērtībām.

Ja L_1 un L_2 tiek ņemtas kā mazākā un lielākā x vērtība izlasē (attiecīgi $r = 1$, tas ir, izlase nav nošķelta), tad izlasēm apjomā 1000 tolerances robežas vidēji iekļaus 99.8% no populācijas, un ar varbūtību 0.996 L_1 un L_2 iekļaus vismaz 99.5% no populācijas vērtībām. Ja lielākā un mazākā x vērtība izlasēs tiek uzskatītas par tolerances robežām, un ja gribam uzskatīt, ka ar varbūtību 0.99 vismaz 99% no populācijas iekļausies starp tolerances robežam, tad nepieciešmais izlases apjoms ir 660. Ja varbūtību pazemina līdz 0.95, ka vismaz 99% iekļausies starp tolerances robežām, tad izlasei jābūt apjomā 130. Inženieri - statistiķi, praktisku pētījumu rezultātā, ir secinājuši, ka visobjektīvāk ir izmantot izlases apjomā no 100 līdz 1000, lai konstruētu tolerances robežas, kuras iekļautu vismaz 99% no populācijas vērtībām ar pietiekoši lielu ticamību. Piemēri par izlasēm apjomā 100, 660 un 130 atspoguļo stabilitāti, kas tiek gaidīta no šādu izmēru izlasēm tolerances apgabala noteikšanai. Stabilitātes rādītājs tolerances robežām izlasēm izmēros 500 līdz 1000 ir tik precīzs, kā to pieprasa inženieri - statistiķi.

Dažos gadījumos ir svarīgi noteikt izlases izmēru, kā arī noteikt kāda procentuāla populācijas daļa ar varbūtību vismaz p atradīsies sadalījuma astēs, kas nošķeltas ar L_1

un L_2, uzskatot, ka šī daļa atradīsies starp diviem skaitļiem d un e. Šajā gadījumā ir jānosaka mazāko vērtību no n tā, ka

$$\int_d^e \int_d^e h(u,v)du\,dv \geq p, \tag{B0.5}$$

kur $h(u,v)du\,dv$ apzīmē funkciju, kas dota (B0.4). Piemēram uzskata, ka $p = 0.99$, $d = 0$, $e = 0.005$ un $r = 1$. Nepieciešamais izlases apjoms ir 1060. Tas ir, izlasē ar apjomu 1060 varbūtība ir 0.99 , ka tolerances robežas L_1 un L_2 ņemtas kā izlases mazākā un lielākā vērtība nošķels populācijas vērtībām astes, kas nesaturēs vairāk kā 0.5% populācijas vērtību.

Ja ir jākonstruē tikai viena tolerances robeža, teiksim L_1, tad tiek izmantots u sadalījums. Tas var tikt atrasts, integrējot (B0.4) pēc v. Sadalījums ir

$$\frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)}u^{r-1}(1-u)^{n-r}du. \tag{B0.6}$$

Varbūtība p, ka populācijas proporcija astē, kas nošķelta ar tolerances robežu L_1, atrodas starp d un e, var tikt iegūta integrējot izteiksmi (B0.6).

C Tabulas izlases apjomam

C1. tabula: n vērtības, kur (a),(b) un (c) ir $(p, 1-\alpha)$ neparametriskie tolerances intervāli. (a) $\left(X_{(\frac{m}{2})}, X_{(n-\frac{m}{2})}\right)$, ja m ir pāra skaitlis; (b) $\left(X_{(\frac{m+1}{2})}, X_{(n-\frac{m+1}{2}+1)}\right)$, ja m ir nepāra skaitlis; (c) $(X_{(m)}, X_{(n)})$ jebkuram m

	$p = 0.50$			$p = 0.75$			$p = 0.80$		
	$1-\alpha$			$1-\alpha$			$1-\alpha$		
m	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
1	7	8	11	15	18	24	18	22	31
2	9	11	14	20	23	31	25	30	39
3	12	13	17	25	29	37	32	37	47
4	14	16	19	30	34	43	38	44	55
5	17	18	22	35	40	49	45	50	62
6	19	21	25	40	45	54	51	57	69
7	21	23	27	45	50	60	57	63	76
8	24	26	30	50	55	65	63	69	83
9	26	28	33	55	60	70	69	76	89
10	28	30	35	59	65	76	75	82	96
11	31	33	38	64	70	81	81	88	102
12	33	35	40	69	74	86	86	94	109
13	35	37	42	73	79	91	92	100	115
14	37	40	45	78	84	96	98	106	122
15	39	42	47	82	89	101	104	112	128
16	42	44	50	87	93	106	109	118	134
17	44	47	52	91	98	111	115	124	141
18	46	49	54	96	103	116	121	129	147
19	48	51	57	100	107	121	126	135	153
20	51	53	59	105	112	126	132	141	159

C2. tabula: n vērtības, kur (a),(b) un (c) ir $(p, 1-\alpha)$ neparametriskie tolerances intervāli. (a) $\left(X_{(\frac{m}{2})}, X_{(n-\frac{m}{2})}\right)$, ja m ir pāra skaitlis; (b) $\left(X_{(\frac{m+1}{2})}, X_{(n-\frac{m+1}{2}+1)}\right)$, ja m ir nepāra skaitlis; (c) $(X_{(m)}, X_{(n)})$ jebkuram m (turpinājums)

	$p = 0.90$			$p = 0.95$			$p = 0.99$		
	$1-\alpha$			$1-\alpha$			$1-\alpha$		
m	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
1	38	46	64	77	93	130	388	473	662
2	52	61	81	105	124	165	531	628	838
3	65	76	97	132	153	198	667	773	1001
4	78	89	113	158	181	229	798	913	1157
5	91	103	127	184	208	259	926	1049	1307
6	104	116	142	209	234	288	1051	1182	1453
7	116	129	156	234	260	316	1175	1312	1596
8	128	142	170	258	286	344	1297	1441	1736
9	140	154	183	282	311	371	1418	1568	1874
10	152	167	197	306	336	398	1538	1693	2010
11	164	179	210	330	361	425	1658	1818	2144
12	175	191	223	353	386	451	1776	1941	2277
13	187	203	236	377	410	478	1893	2064	2409
14	199	215	249	400	434	504	2010	2185	2539
15	210	227	262	423	458	529	2127	2306	2669
16	222	239	275	446	482	555	2242	2426	2798
17	233	251	287	469	506	580	2358	2546	2925
18	245	263	300	492	530	606	2473	2665	3052
19	256	275	312	515	554	631	2587	2784	3179
20	267	286	325	538	577	656	2701	2902	3304

D Izveidoto programmu kods

```
#tolerances faktora aproksimacija
n<-3
m<-n-1
p<-0.95
alpha<-0.1
sqrt(m*qchisq(p, 1, 1/n)/qchisq(alpha, m))

n<-10
m<-n-1
p<-0.95
alpha<-0.1
sqrt(m*qchisq(p, 1, 1/n)/qchisq(alpha, m))

#1.piemers
n<-15
d<-c(200,120,15,7,8,6,48,61,380,80,29,1000,350,1400,110)
l<-log(d)#logaritms no datiem
m<-mean(l)#videja vertiba
s<-sd(l)#standartnovirze
p<-0.95 #proporcija
alpha1<-0.1 #nozimibas limenis
b<-sqrt(n)
#rekina vienpusejo tolerances faktoru
k1<-qt(1-alpha1,n-1,qnorm(p)*sqrt(n))*(1/b)
#tolerances augš.robeža
t1<-m+k1*s
library(nortest)
lillie.test(l)

alpha2<-0.05
v<-b*(m-log(50))/s
```

```
#atrod kvantiles vertibu
f<-function(x)(qt(1-alpha2,n-1,x)-v)
uniroot(f,c(-300,0))
#rekina ticamibas intervalu
t.test(l)
#iebūvētie tolerances intervāli
library(tolerance)
normtol.int(d,alpha1,p,side=1,method="HE",log.norm=TRUE)
out.p<-normtol.int(l,alpha1,p,side=1,method="HE",log.norm=FALSE)
nptol.int(d,alpha1,p,side=1,method="WILKS")
nptol.int(l,alpha1,p,side=1,method="WILKS")
#izlases apjoma nosacijuma parbaude
(n-1)*p^(n)-n*p^(n-1)+1

#2.piemers
n<-20
l<-c(0.968,0.982,1.030,1.003,1.046,1.020,0.997,1.010,1.027,1.010,0.973,
1.000,1.044,0.995,1.020,0.993,0.984,0.981,0.997,0.992)
v<-mean(l) #videja vertiba
s<-sd(l) #standartnovirze
m<-n-1
p<-0.99
alpha<-0.05
sqrt(m*qchisq(p,1,1/n)/qchisq(alpha, m)) #tolerances faktors
sqrt(m*qchisq(p, 1, 1/n)/qchisq(alpha, m))*s
l1<-v-sqrt(m*qchisq(p, 1, 1/n)/qchisq(alpha, m))*s #apakš.tol.robeža
l2<-v+sqrt(m*qchisq(p, 1, 1/n)/qchisq(alpha, m))*s #augš.tol.robeža
t.test(l)
#iebūvētie tolerances intervāli
library(tolerance)
out.p<-normtol.int(l,alpha,p,side=2,method="HE",log.norm=FALSE)
plottol(out.p,l,plot.type="control",x.lab=NULL,y.lab=NULL)
abline(0.9932563,0,lty=7,col=1)
abline(1.0139437,0,lty=7,col=1)
#neparametriskais
outn<-nptol.int(l,alpha,p,side=2,method="WILKS")
plottol(outn,l,plot.type="control")
#izlases apjoma nosacijuma parbaude
(n-1)*p^(n)-n*p^(n-1)+1
```

```
p11<-0.9
p12<-0.85
p13<-0.8
(19)*p13^(20)-20*p13^(19)+1

#3.piemers
as.x<-c(0.160,0.170,0.180,0.1,0.17,0.1,0.06,0.1,0.17,0.056,
        0.111,0.162,0.143,0.079,0.006)
elpa.Y<-c(0.145,0.156,0.181,0.108,0.180,0.112,0.081,0.104,
          0.176,0.048,0.092,0.144,0.121,0.065,0.0)
dati.alko<-data.frame(as.x,elpa.Y)
plot(as.x,elpa.Y)
lin.reg2<-lm(elpa.Y~as.x)#lineara regresija
abline(lin.reg2)
lin.reg2$coef
lin.reg2$resid#atlikumi

#parbauda datu normalitati
qqnorm(elpa.Y)
qqline(elpa.Y)
shapiro.test(elpa.Y)
lillie.test(elpa.Y)

anova(lin.reg2)
#rekina determinacijas koeficientu
0.035403/sum(0.035403,0.002426)
#atlikumu standartkļūda
rse2<-round(sqrt(deviance(lin.reg2)/df.residual(lin.reg2)),4)
summary(lin.reg2)#salīdzinam atlikumu standartkļūdu
abline(lin.reg2)
xprim2<-c(1,0.1)
X2<-model.matrix(lin.reg2)
H2<-solve(t(X2)%*%X2)#matricu reizinājums(inversais)
betah<-solve(t(X2)%*%X2)%*%t(X2)%*%elpa.Y #koeficientu novērtējums
xhb<-sum(betah*xprim2)
d<-sqrt(xprim2%*%H2%*%xprim2)
n<-15
m<-2
p<-0.9 #proporcija
```

```
alpha<-0.05 #nozimibas limenis
qnorm(p) #kvantile
kr<-qt(1-alpha,n-m,qnorm(p)*(1/d)) #kritiskā vērtība
kf<-d*kr
kopa<-xhb-kf* 0.01366 #tolerances robeža
#iebūvētās komandas
library(tolerance)
out1<-regtol.int(lin.reg1,new.x=NULL,side=1,alpha1, p1)

#4.piemers
temp.dati.x1<-c(80,93,100,82,90,99,81,96,94,93,97,95,100,85,86,87)
katal.dati.x2<-c(8,9,10,12,11,8,8,10,12,11,13,11,8,12,9,12)
viskozit.dati.Y<-c(2256,2340,2426,2293,2330,2368,2250,2409,2364,2379,
                   2440,2364,2404,2317,2309,2328)
kopa.dati<-data.frame(temp.dati.x1,katal.dati.x2,viskozit.dati.Y)
lin.reg1<-lm(viskozit.dati.Y~temp.dati.x1+katal.dati.x2)
anova(lin.reg1)
xprim1<-c(1,88,9)
X1<-model.matrix(lin.reg1)
H1<-solve(t(X1)%*%X1) #inversias matr.reizin.
betah1<-solve(t(X1)%*%X1)%*%t(X1)%*%viskozit.dati.Y #koeficientu novērt.
xhb1<-sum(betah1*xprim1)
d1<-sqrt(xprim1%*%H1%*%xprim1)
# rekina atlikumu standartkļūdu (residual standard error)
rse1<-round(sqrt(deviance(lin.reg1)/df.residual(lin.reg1)),2)
#rezultātu salīdzina ar kopsavilkumu (summary)
summary(lin.reg1)
#veicam datu normalitātes testu
shapiro.test(viskozit.dati.Y)
lillie.test(viskozit.dati.Y)
qqnorm(viskozit.dati.Y)
qqline(viskozit.dati.Y)
n1<-16
m1<-3
p1<-0.9 #proporcija
alpha1<-0.05 #nozīmības līmenis
qnorm(p1) #kvantile
kr1<-round(qt(1-alpha1,n1-m1,qnorm(p1)*(1/d1)),3)#tolerances faktors
kf1<-d1*kr1
```

```
kopa1<-xhb1+kf1*rse1 #tolerances robeža
#iebūvētās komandas
library(tolerance)
out2<-regtol.int(lin.reg2,new.x=NULL,side=1,alpha, p)

#5.piemers
sarm.dati<-c(28,32,39,rep(40,2),rep(42,3),49,rep(51,2),52,rep(54,2),
55,58,rep(59,2),60,63,66,70,79,82,89,96,118)
sarm.kub<-(sarm.dati)^(1/3) #rekina kubsakni no datiem
vid.sarm<-round(mean(sarm.kub),4)#videja vertiba datu kubsaknem
st.sarm<-round(sd(sarm.kub),4)#dispersija
#normalitates parbaude
library(nortest)
lillie.test(sarm.kub)

library(MASS)
f<-fitdistr(sarm.dati,"gamma") #novertejam parametrus
f$estimate #otrs veids ka novertēt parametrus gamma sad.(shape un rate)
scale<-round(1/0.1612490,3) #iegūstam mēroga parametru
xlab<-paste("Gamma(shape)=",round(f$estimate[[1]],3))
#kvantilu-kvantilu grafiks
qqplot(qgamma(ppoints(sarm.dati),shape=f$estimate[[1]],
rate=f$estimate[[2]]),sarm.dati,main="QQplot(Gamma)",xlab=xlab)

#vienpusejie tolerances faktori(Wilson-Hilferty)
p1<-0.9
alpha1<-0.05
b<-1/sqrt(n)
k11<-round(qt(1-alpha1,n-1,qnorm(p1)*sqrt(n))*b,4)
p2<-0.95
k12<-round(qt(1-alpha1,n-1,qnorm(p2)*sqrt(n))*b,4)
p3<-0.99
k13<-round(qt(1-alpha1,n-1,qnorm(p3)*sqrt(n))*b,4)
#apakšejas tolerances robežas(Wilson-Hilferty)
L1<-round((vid.sarm-k11*st.sarm)^3,3)
L2<-round((vid.sarm-k12*st.sarm)^3,3)
L3<-round((vid.sarm-k13*st.sarm)^3,3)
#Augšējās tol.robežas(Wilson-Hilferty)
U1<-round((vid.sarm+k11*st.sarm)^3,3)
```

```
U2<-round((vid.sarm+k12*st.sarm)^3,3)
U3<-round((vid.sarm+k13*st.sarm)^3,3)
#divpusejais tolerances faktors(Wilson-Hilferty)
m<-n-1
k21<-round(sqrt(m*qchisq(p1,1,1/n)/qchisq(alpha1,m)),4)
k22<-round(sqrt(m*qchisq(p2,1,1/n)/qchisq(alpha1,m)),4)
k23<-round(sqrt(m*qchisq(p3,1,1/n)/qchisq(alpha1,m)),4)
#divpusejie tolerances intervali(W-H)
round((vid.sarm - k21*st.sarm)^3,3)
round((vid.sarm + k21*st.sarm)^3,3)
round((vid.sarm - k22*st.sarm)^3,3)
round((vid.sarm + k22*st.sarm)^3,3)
round((vid.sarm - k23*st.sarm)^3,3)
round((vid.sarm + k23*st.sarm)^3,3)
##parsniegsanas varbutiba
#apakšeja tol.robeža
ap.rob<-(vid.sarm-41^(1/3))/(st.sarm/sqrt(n))
b1<-sqrt(n)
#aprekinam kvantiles vertibu
f<-function(xx)(qt(1-alpha1,n-1,xx)-ap.rob)
k<-uniroot(f,c(0,300))
k[[1]]
#tīra kvantiles vertiba bez saknes
k[[1]]/b1

###neparametriskie tolerances intervali
n<-27
m<-n-1
p<-0.75
alpha<-0.05
#rekina divpusejo tolerances faktoru
round(sqrt(m*qchisq(p,1,1/n)/qchisq(alpha,m)),3)
#tolerances intervāli
round((3.8274 - 1.523*0.4298)^3,2)
round((3.8274 + 1.523*0.4298)^3,2)
b<-1/sqrt(n)
#vienpusejais tolerances faktors
round(qt(1-alpha,n-1,qnorm(p)*sqrt(n))*b,3)
#vienpusējās tolerances robežas
```

```
round((3.8274 - 1.083*0.4298)^3,2)

round((3.8274 + 1.083*0.4298)^3,2)

round(qbeta(0.05,22,6),3)#beta kvantilu funkcija
```

Diplomdarbs “Statistiskie tolerances intervāli” izstrādāts LU Fizikas un Matemātikas fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Margarita Fjodorova

________________ ______________

(paraksts) (datums)

Rekomendēju darbu aizstāvēšanai.

Vadītājs: doc. Dr.math. Jānis Valeinis

________________ ______________

(paraksts) (datums)

Recenzents: Edmunds Cers

________________ ______________

(paraksts) (datums)

Darbs iesniegts Matemātikas nodaļā ______________

(datums)

____

(darbu pieņēma)

Darbs aizstāvēts diplomdarbs valsts pārbaudījuma komisijas sēdē

______________ prot. Nr. ________, vērtējums________

(datums)

Komisijas sekretārs/-e: Ingrīda Uļjane ______________

(paraksts)