1. Podatci prikazuju broj stradalih od tornada u SAD-u u razdoblju od 1990. do 2000. godine:

$$53, 39, 39, 33, 69, 30, 25, 67, 130, 94, 40.$$

(a) Napravite dijagram stabljika-i-list (engl. *stem and leaf*).

2	5
3	0, 3, 9, 9
4	0
5	3
6	7, 9
7	
8	
9	4
10	
11	
12	
13	0

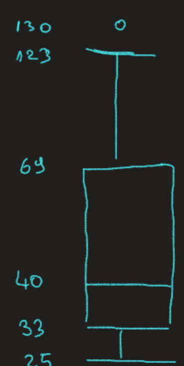
(b) Nađite srednju vrijednost, medijan, kvartile, interkvartilni rang (IQR), minimalni i maksimalni podatak te skiciraj pravokutni dijagram s izdancima (engl. *box and whiskers plot*).

$\bar{X} = 56.27$

25, 30, 33, 39, 39, 40, 53, 67, 69, 94, 130

min = 25, $Q_1 = 33$, $Q_2 = 40$, $Q_3 = 69$, max = 130

$IQR = Q_3 - Q_1 = 36$



(c) Ako je područje nestršećih vrijednosti definirano izdancima (engl. *whiskers*) veličine 1.5 IQR-a, utvrdite ima li stršećih vrijednosti u ovom skupu. Kako biste to komentirali? (Napomena: izdanci ne mogu prelaziti minimalni, odnosno maksimalni podatak).

stršeća vrijednost : 130

Moguće da je 1998. g. bila godina s jako puno tornadou, ili da je podatak pogrešan.

(d) Komentirajte oblik distribucije te navedite koju biste mjeru centra distribucije u ovom slučaju koristili i zašto.

Distribucija se čini relativno nepravilna i ukošena, te stršeći podaci poput 130 i 94 mogu previše utjecat na sredinu, stoga se medijan čini kao bolja mjera centra jer ukošenost manje utječe na njega.

2. Uređaj proizvodi metalne valjke. Izmjereni su promjeri $n = 9$ proizvedenih valjaka. Sredina tog uzorka jest 1.01 cm, a standardna devijacija uzorka jest 0.025.

$n = 9$

$\bar{x} = 1.01$

$s = 0.025$

(a) Izračunajte 99%-tni interval pouzdanosti za očekivani promjer valjaka koje proizvodi navedeni uređaj uz pretpostavku da podatci dolaze iz normalne distribucije.

$T = \frac{\bar{x} - \mu}{s}\sqrt{n}$ $\quad v = n-1 = 8 \quad t_{0.005} = 3.355$

$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$

$P\left(-t_{0.005} < \frac{\bar{x} - \mu}{s}\sqrt{n} < t_{0.005}\right) = 99\%$

$P\left(\bar{x} - t_{0.005}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{0.005}\frac{s}{\sqrt{n}}\right) = 99\%$

$\mu \in \langle 0.982, 1.038 \rangle$

(b) Bi li na gornji izračun utjecala spoznaja o varijanci populacije? Obrazložite odgovor.

Da, tada bismo mogli koristiti z-test koji je

precizniji jer normalna distribucija ima manje „repove"

od t-distribucija. Kao rezultat bismo dobili uži interval.

(c) Interpretirajte 99%-tni interval pouzdanosti.

Dobili smo 99%-tni interval pouzdanosti $\langle 0.982, 1.038 \rangle$.

To znači da smo 99% sigurni da se pravi μ nalazi

u tom intervalu.

Drugim riječima, kada bismo 100 ponovili isti test s novim podacima,

čime bismo dobili 100 različitih intervala,

99 od tih 100 intervala bi nam trebalo doista sadržavati

pravu vrijednost μ.

3. Menadžer taksi-kompanije treba odlučiti treba li jedan tip guma zamijeniti drugim u svrhu manje potrošnje benzina. Dvanaest automobila ispitano je sa starom i novom vrstom guma bez promjene vozača i na istoj ruti. Podatci su dani u tablici (kilometri po litri):

Automobil	Nove gume	Stare gume	Automobil	Nove gume	Stare gume
1	4.2	4.1	7	5.7	5.7
2	4.7	4.9	8	6.0	5.8
3	6.6	6.2	9	7.4	6.9
4	7.0	6.9	10	4.9	4.7
5	6.7	6.8	11	6.1	6.0
6	4.5	4.4	12	5.2	4.9

(a) Možemo li na razini značajnosti $\alpha = 0.05$ zaključiti da su nove gume bolje od starih? Obrazložite.

i	D_i
1	-0.1
2	0.2
3	-0.4
4	-0.1
5	0.1
6	-0.1
7	0
8	-0.2
9	-0.5
10	-0.2
11	-0.1
12	-0.3

$\alpha = 0.05 \quad \nu = n-1 = 11 \qquad t_{0.05} = 1.796$

$n = 12$

$\overline{D}_n = -0.142$

$S_D = 0.198$

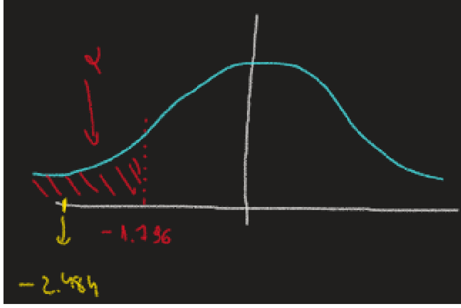
$H_0: \mu_D = \mu_0 = 0$ (gume su jednako dobre)

$H_1: \mu_D < 0$ (nove gume više km/l)

$$T = \frac{\overline{D}_n - \mu_0}{S_n}\sqrt{n} = -2.484$$

$T < -t_{0.05} \longrightarrow$ prihvaćamo H_1,

nove gume su bolje



(b) Koju ste pretpostavku testa ovdje iskoristili?

Pretpostavili smo da varijabla D ima normalnu razdiobu.

4. Izvagan je slučajan uzorak od 64 vrećice kokica i dobivena je prosječna težina od 5.23 dkg uz standarnu devijaciju 0.24.

   (a) Testirajte hipotezu da je prosječna težina vrećice kokica 5.5 dkg nasuprot alternative da je manja na razini značajnosti $\alpha = 0.05$.

$\overline{X} = 5.23$ $\qquad$ $\alpha = 0.05$

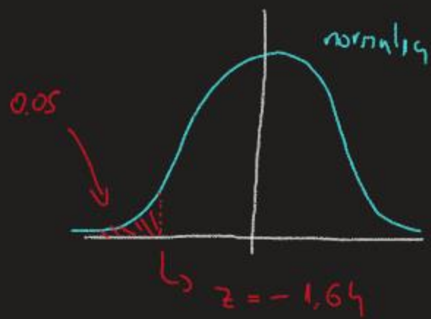
$s = 0.24$

$n = 64$

$H_0: \quad \mu = \mu_0 = 5.5$

$H_1: \quad \mu < 5.5$

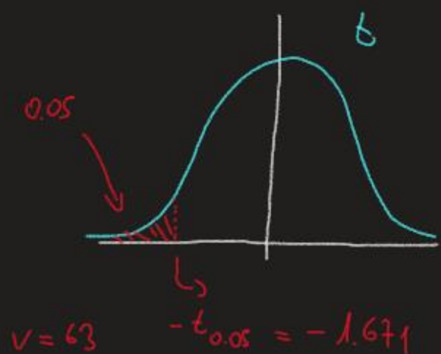
1. preko normalne :



$$Z = \frac{\bar{X} - \mu}{S}\sqrt{n} = -9$$

$Z < z \Rightarrow$ prihvaćamo H_1

2. preko t :



$$T = \frac{\bar{X} - \mu}{S}\sqrt{n} = -9$$

$T < t \Rightarrow$ prihvaćamo H_1

(b) Izračunajte snagu testa kada je alternativa 5.3 dkg (uz pretpostavku jednostranog testa) i veličina uzorka je 10, uz pretpostavku da je devijacija populacije poznata i iznosi 0.24.

$H_0 : \mu = \mu_0 = 5.5$

$H_1 : \mu = \mu_1 = 5.3$

$n = 10$

$\sigma = 0.24$

$\alpha = 0.05$

$$\text{Snaga} = 1 - \beta = 1 - P(\text{nisam odbacio } H_0 \mid H_1) =$$

$$= 1 - P\left(\frac{\overline{X}_n - \mu_0}{\sigma}\sqrt{n} > z_{0.05} \mid H_1\right) =$$

$$= 1 - P\left(\frac{\overline{X}_n - \mu_0}{\sigma}\sqrt{n} > -1.64 \mid H_1\right) =$$

$$= P\left(\frac{\overline{X}_n - \mu_0}{\sigma}\sqrt{n} < -1.64 \mid H_1\right) =$$

$$= P\left(\overline{X}_n < -1.64\frac{\sigma}{\sqrt{n}} + \mu_0\right) =$$

$$= P\left(\overline{X}_n < 5.376\right) =$$

$$= P\left(\frac{\overline{X}_n - \mu_1}{\sigma}\sqrt{n} < \frac{5.376 - \mu_1}{\sigma}\sqrt{n}\right) =$$

$$= P\left(\mathcal{N}(0,1) < 1.001\right) =$$

$$= 0.8438$$

(c) Ako želimo da je snaga testa 0.9 kada je srednja vrijednost 5.2 dkg (uz pretpostavku jednostranog testa), koliko velik uzorak moramo uzeti? I u ovom slučaju koristite da je standarna devijacija populacije poznata i iznosi 0.24.

$1-\beta = 0.9$ $\qquad \alpha = 0.05$

$\mu_1 = 5.2$ $\qquad \sigma = 0.24$

$\mu_0 = 5.5$

$1-\beta = 1 - P(\text{nisam odbacio } H_0 \mid H_1)$

$$1-\beta = 1 - P\left(\frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n} > -1.64 \mid H_1\right)$$

$$0.9 = 1 - P\left(\bar{X}_n > -1.64\frac{\sigma}{\sqrt{n}} + \mu_0 \mid H_1\right)$$

$$0.1 = P\left(\frac{\bar{X}_n - \mu_1}{\sigma}\sqrt{n} > \frac{-1.64\frac{\sigma}{\sqrt{n}} + \mu_0 - \mu_1}{\sigma}\sqrt{n} \mid H_1\right)$$

$$0.1 = P\left(\mathcal{N}(0,1) > -1.64 + \frac{\mu_0 - \mu_1}{\sigma}\sqrt{n}\right)$$

$$0.1 = P\left(\mathcal{N}(0,1) > -1.64 + 1.25\sqrt{n}\right)$$

$$-1.64 + 1.25\sqrt{n} = 1.28$$

$$n = \cancel{5.45}\ 6$$

5. Istraživana je veza između pojave povišenog krvnog tlaka i pušenja tako da je ispitano 180 osoba i rezultati su prikazani tablicom:

	Nepušač	Blagi pušač	Teški pušač
Visok tlak	21	36	30
Normalan tlak	48	26	19

Zanima nas je li pušenje povezano s povišenim krvnim tlakom, na razini značajnosti $\alpha = 0.05$.

(a) Obrazložite koji test ćete koristiti i zašto ga smijete koristiti.

Koristimo χ^2 - test nezavisnosti.

Smijemo ga koristit jer je broj stupnjeva slobode,

$$(r-1)(c-1) = (2-1)(3-1) = 2 > 1$$

Također, očekivana vrijednost svake ćelije je ≥ 5.

(b) Postavite jasno hipoteze H_0 i H_1.

H_0 : broj pušača i broj osoba s povišenim krvnim tlakom su nezavisne varijable

H_1 : -//- zavisne varijable

(c) Provedite test i interpretirajte rezultate testa u kontekstu zadatka.

$\alpha = 0.05 \quad v = (r-1)(c-1) = 2 \quad \Rightarrow \quad \chi^2_{0.05,2} = 5.991$

	Nepušač	Blagi pušač	Teški pušač	
Visok tlak	21 (33.35)	36 (29.97)	30 (23.68)	87
Normalan tlak	48 (35.65)	26 (32.03)	19 (25.32)	93
	69	62	49	180

$$\chi^2 = \sum \frac{(o-e)^2}{e} = 4.573 + 1.213 + 1.687 + 4.278 + 1.135 + 1.578 = 14.464$$

$\chi^2 > \chi^2_{0.05,2} \Rightarrow$ zavisne

(b) Izračunajte jackknife procjenitelj i jackknife 95%-tni interval pouzdanosti varijance za uzorak $\{1, 2, 3, 4\}$. Procjene možete izračunati preko parcijalnih procjena ili preko pseudovrijednosti.

$$\hat{\theta} = s^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2$$

1,2,3,4 $\quad \hat{\Theta} = 1.667$

$$\left.\begin{array}{ll} 2,3,4 & \hat{\theta}_1 = 1 \\ 1,3,4 & \hat{\theta}_2 = 2.33 \\ 1,2,4 & \hat{\theta}_3 = 2.33 \\ 1,2,3 & \hat{\theta}_4 = 1 \end{array}\right\} \quad \hat{\theta}_{(\cdot)} = 1.665 \quad \Rightarrow \quad bias_{jack}(\hat{\theta}) = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}) =$$

$$= -0.006$$

$$\hat{\Theta}_{jack} = \hat{\theta} - bias_{jack}(\hat{\theta}) = \underline{\underline{1.673}}$$

$$SE(\hat{\theta})_{jack} = \sqrt{\frac{n-1}{n} \sum_i (\hat{\theta}_i - \hat{\theta}_{\cdot})^2} = 1.15$$

$$t_{1-\frac{\alpha}{2}, n-1} = t_{0.025, 3} = 3.182$$

95% $\quad \sigma^2 \in \langle \hat{\Theta}_{jack} - t_{1-\frac{\alpha}{2}, n-1} \cdot SE, \; \hat{\Theta}_{jack} + t_{1-\frac{\alpha}{2}, n-1} \cdot SE \rangle$

$$\underline{\langle -1.9863, \; 5.3323 \rangle}$$

Alternativno, preko pseudovrijednosti:

1,2,3,4 $\quad \hat{\Theta} = 1.667$

$$\begin{array}{llll} 2,3,4 & \hat{\theta}_1 = 1 & \Rightarrow & ps_1 = 3.668 \\ 1,3,4 & \hat{\theta}_2 = 2.33 & \Rightarrow & ps_2 = -0.322 \\ 1,2,4 & \hat{\theta}_3 = 2.33 & \Rightarrow & ps_3 = -0.322 \\ 1,2,3 & \hat{\theta}_4 = 1 & \Rightarrow & ps_4 = 3.668 \end{array} \left\} \quad \begin{array}{l} S_{ps_i} = 2.303 \\ SE = \frac{S_{ps_i}}{\sqrt{n}} = 1.15 \end{array}\right.$$

$$\hat{\Theta}_{jack} = \overline{ps} = \underline{\underline{1.673}}$$

95% $\quad \sigma^2 \in \langle \hat{\Theta}_{jack} - t_{1-\frac{\alpha}{2}, n-1} \cdot SE, \; \Theta_{jack} + t_{1-\frac{\alpha}{2}, n-1} \cdot SE \rangle \quad \leftarrow$ isto kao i prije

$$\underline{\langle -1.9863, \; 5.3323 \rangle}$$

(c) Ukratko objasnite koja je ideja procjene standardne pogreške preko pseudovrijednosti te koja je pritom uloga centralnog graničnog teorema.

Pseudovrijednost tretiramo kao novu slučajnu varijablu sa sredinom jednakom $\hat{\theta}_{jack}$, pa možemo koristiti CGT kako bismo računali SE i intervalne procjene.

Aux