Appendix C.

Formelsamling

Andelar, medelvärde, standardavvikelse, varians, median	
Stickprovsandel	$p = \dfrac{antal\ enheter\ i\ stickprovet\ med\ studerad\ egenskap}{stickprovsstorlek}$
Populationsandel	$\pi = \dfrac{antal\ enheter\ i\ populationen\ med\ studerad\ egenskap}{populationsstorlek}$
Stickprovsmedelvärde beräknat på rådata	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$
Populationsmedelvärde beräknat på rådata	$\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$
Stickprovsmedelvärde beräknat på frekvenstabell	$\bar{x} = \dfrac{\sum_{i=1}^{g} f_i \cdot x_i}{n}$
Populationsmedelvärde beräknat på frekvenstabell	$\mu = \dfrac{\sum_{i=1}^{g} f_i \cdot x_i}{N}$
Stickprovsstandardav-vikelse beräknat på rådata	$s = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\dfrac{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$

Populationsstandard-avvikelse beräknat på rådata	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}{N}}$
Stickprovsstandard-avvikelse beräknat på frekvenstabell	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{g} f_i(x_i - \overline{x})^2} = \sqrt{\frac{\sum_{i=1}^{g} f_i \cdot x_i^2 - \frac{\left(\sum_{i=1}^{g} f_i \cdot x_i\right)^2}{n}}{n-1}}$
Populationsstandardav-vikelse beräknat på frekvenstabell	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{g} f_i(x_i - \mu)^2} = \sqrt{\frac{\sum_{i=1}^{g} f_i \cdot x_i^2 - \frac{\left(\sum_{i=1}^{g} f_i \cdot x_i\right)^2}{N}}{N}}$
Stickprovsvarians	s^2
Populationsvarians	σ^2

Kombinatorik	
Multiplikationsprincipen	$n_1 \cdot n_2 \cdot \ldots \cdot n_k$
Kombinationer utan återläggning	$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$
Kombinationer med återläggning	$C'^k_n = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$

Permutationer utan återläggning	$P_n^k = \frac{n!}{(n-k)!}$
Permutationer utan återläggning när vissa element är lika	$P_n^{k_1,k_2,\ldots} = \frac{n!}{k_1! \cdot k_2! \cdot \ldots}$
Permutationer med återläggning	$P'^k_n = n^k$

Sannolikhetslära	
Odds	$O(A) = \frac{\Pr(A)}{\Pr(A^C)}$
Additionssatsen för disjunkta händelser	$\Pr(A \cup B) = \Pr(A) + \Pr(B)$
Additionssatsen för icke disjunkta händelser	$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
Multiplikationssatsen	$\Pr(A \cap B) = \begin{cases} 0 \text{ om } A \text{ och } B \text{ är disjunkta} \\ \Pr(A) \cdot \Pr(B) \text{ om } A \text{ och } B \text{ är oberoende} \\ \Pr(A) \cdot \Pr(B \mid A) \text{ i övrigt} \end{cases}$
Betingad sannolikhet	$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$
Satsen om total sannolikhet	$\Pr(B) = \sum_{i=1}^{g} \Pr(A_i) \cdot \Pr(B \mid A_i)$
Bayes sats	$\Pr(A_j \mid B) = \frac{\Pr(A_j) \cdot \Pr(B \mid A_j)}{\sum_{i=1}^{g} \Pr(A_i) \cdot \Pr(B \mid A_i)}$

Slumpvariabler	
Väntevärde	$E(X) = \mu = \sum_{i=1}^{g} x_i \cdot p(x_i)$
Varians	$Var(X) = \sigma^2 = \sum_{i=1}^{g} p(x_i) \cdot (x_i - \mu)^2 = \sum_{i=1}^{g} x_i^2 \cdot p(x_i) - \mu^2$

Standardavvikelse	$\sigma = \sqrt{Var(X)}$
Linjär variabeltransformation	Om $Y = a + b \cdot X$ och X har väntevärde $E(X) = \mu_X$ och varians $Var(X) = \sigma_X^2$ gäller $E(Y) = \mu_Y = E(a + b \cdot X) = a + b \cdot \mu_X$ $Var(Y) = \sigma_Y^2 = Var(a + b \cdot X) = b^2 \cdot \sigma_X^2$

Diskreta sannolikhetsfördelningar

Binomialfördelning	$\Pr(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}$ $E(X) = \mu = n\pi$ $Var(X) = \sigma^2 = n\pi(1 - \pi)$
Hypergeometrisk fördelning	$\Pr(X = k) = \dfrac{\binom{N\pi}{k} \cdot \binom{N - N\pi}{n - k}}{\binom{N}{n}}$ $E(X) = \mu = n\pi$ $Var(X) = \sigma^2 = n\pi(1 - \pi)\dfrac{N - n}{N - 1}$
Geometrisk fördelning	$\Pr(X = k) = (1 - \pi)^{k-1} \cdot \pi$ $E(X) = \mu = \dfrac{1}{\pi}$ $Var(X) = \sigma^2 = \dfrac{(1 - \pi)}{\pi^2}$

Kontinuerliga sannolikhetsfördelningar	
Standardisering	$z = \dfrac{x-\mu}{\sigma}$
Normalfördelningsapproximation av binomialfördelning	Om $n\pi(1-\pi) > 5$ gäller $X \approx N\left(\mu = n\pi ; \sigma = \sqrt{n\pi(1-\pi)}\right)$

Stickprovsteori	
Stickprovsmedelvärde	$E(\overline{X}) = \mu$ $Var(\overline{X}) = \dfrac{\sigma^2}{n}$
Stickprovssumma	$E\left(\sum X\right) = n \cdot \mu$ $Var\left(\sum X\right) = n \cdot \sigma^2$
Stickprovsandel	$E(P) = \pi$ $Var(P) = \dfrac{\pi(1-\pi)}{n}$

Inferens om en population	
Dubbelsidigt konfidensintervall för populationsmedelvärde, σ okänd	$\overline{x} \pm t_{n-1;1-\alpha/2} \dfrac{s}{\sqrt{n}}$
Nedåt begränsat konfidensintervall för populationsmedelvärde, σ okänd	$\mu > \overline{x} - t_{n-1;1-\alpha} \dfrac{s}{\sqrt{n}}$
Uppåt begränsat konfidensintervall för populationsmedelvärde, σ okänd	$\mu < \overline{x} + t_{n-1;1-\alpha} \dfrac{s}{\sqrt{n}}$
Dubbelsidigt konfidensintervall för populationsandel	$p \pm z_{1-\alpha/2} \sqrt{\dfrac{p(1-p)}{n}}$

Nedåt begränsat konfidensintervall för populationsandel	$\pi > p - z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}$
Uppåt begränsat konfidensintervall för populationsandel	$\pi < p + z_{1-\alpha}\sqrt{\frac{p(1-p)}{n}}$
Hypotesprövning för populations-medelvärde, σ okänd	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ med kritiskt område/områden enligt om $H_a: \mu < \mu_0$ vänster om $t_{n-1;\alpha}$ om $H_a: \mu > \mu_0$ höger om $t_{n-1;1-\alpha}$ om $H_a: \mu \neq \mu_0$ både till vänster och höger om $t_{n-1;\alpha/2}$ och $t_{n-1;1-\alpha/2}$
Hypotesprövning för populationsandel	$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$ med kritiskt område/områden enligt om $H_a: \pi < \pi_0$ vänster om z_α om $H_a: \pi > \pi_0$ höger om $z_{1-\alpha}$ om $H_a: \pi \neq \pi_0$ både till vänster och höger om $z_{\alpha/2}$ och $z_{1-\alpha/2}$
Dubbelsidigt konfidensintervall för populationsmedelvärde, σ känd	$\bar{x} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$
Nedåt begränsat konfidensintervall för populationsmedelvärde, σ känd	$\mu > \bar{x} - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$
Uppåt begränsat konfidensintervall för populationsmedelvärde, σ känd	$\mu < \bar{x} + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$
Hypotesprövning för populations-medelvärde, σ känd	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ med kritiskt område/områden enligt om $H_a: \mu < \mu_0$ vänster om z_α om $H_a: \mu > \mu_0$ höger om $z_{1-\alpha}$ om $H_a: \mu \neq \mu_0$ både till vänster och höger om $z_{\alpha/2}$ och $z_{1-\alpha/2}$

Jämförelse av två populationer	
Dubbelsidigt konfidensintervall för jämförelse av populations-medelvärden, σ_1 och σ_2 okända	$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{df;1-\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ där $df=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$
Nedåt begränsat konfidens-intervall för jämförelse av populationsmedelvärden, σ_1 och σ_2 okända	$\mu_1-\mu_2 > \left(\bar{x}_1 - \bar{x}_2\right) - t_{df;1-\alpha}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ där $df=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$
Uppåt begränsat konfidens-intervall för jämförelse av populationsmedelvärden, σ_1 och σ_2 okända	$\mu_1-\mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{df;1-\alpha}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$ där $df=\frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}}$
Dubbelsidigt konfidensintervall för jämförelse av populations-andelar	$\left(p_1-p_2\right) \pm z_{1-\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$
Nedåt begränsat konfidens-intervall för jämförelse av populationsandelar	$\pi_1-\pi_2 > \left(p_1-p_2\right) - z_{1-\alpha}\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$
Uppåt begränsat konfidens-intervall för jämförelse av populationsandelar	$\pi_1-\pi_2 < \left(p_1-p_2\right) + z_{1-\alpha}\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}$

Hypotesprövning för jämförelse av populationsmedelvärden, σ_1 och σ_2 okända	$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ om $H_a : \mu_1 - \mu_2 < d_0$ vänster om $t_{df;\alpha}$ om $H_a : \mu_1 - \mu_2 > d_0$ höger om $t_{df;1-\alpha}$ om $H_a : \mu_1 - \mu_2 \neq d_0$ både till vänster och höger om $t_{df;\alpha/2}$ och $t_{df;1-\alpha/2}$ där $df = \dfrac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$
Hypotesprövning för jämförelse av populationsandelar	$$z = \frac{(p_1 - p_2) - d_0}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$ med kritiskt område/områden enligt om $H_a : \pi_1 - \pi_2 < d_0$ vänster om z_α om $H_a : \pi_1 - \pi_2 > d_0$ höger om $z_{1-\alpha}$ om $H_a : \pi_1 - \pi_2 \neq d_0$ både till vänster och höger om $z_{\alpha/2}$ och $z_{1-\alpha/2}$
Dubbelsidigt konfidensintervall för jämförelse av populationsmedelvärden, σ_1 och σ_2 kända	$$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
Nedåt begränsat konfidensintervall för jämförelse av populationsmedelvärden, σ_1 och σ_2 kända	$$\mu_1 - \mu_2 > (\bar{x}_1 - \bar{x}_2) - z_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$
Uppåt begränsat konfidensintervall för jämförelse av populationsmedelvärden, σ_1 och σ_2 kända	$$\mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{1-\alpha}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Hypotesprövning för jämförelse av populationsmedelvärden, σ_1 och σ_2 kända	$z = \dfrac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ med kritiskt område/områden enligt om $H_a : \mu_1 - \mu_2 < d_0$ vänster om z_α om $H_a : \mu_1 - \mu_2 > d_0$ höger om $z_{1-\alpha}$ om $H_a : \mu_1 - \mu_2 \neq d_0$ både till vänster och höger om $z_{\alpha/2}$ och $z_{1-\alpha/2}$
Parvisa jämförelser	Bilda differens och använd metodik för inferens om en population

Inferens om en ändlig population

Dubbelsidigt konfidensintervall för populationsmedelvärde, σ okänd	$\bar{x} \pm t_{n-1;1-\alpha/2} \sqrt{\dfrac{s^2}{n}\left(1 - \dfrac{n}{N}\right)}$
Dubbelsidigt konfidensintervall för populationens totalmängd, σ okänd	$N \cdot \bar{x} \pm t_{n-1;1-\alpha/2} \cdot N \cdot \sqrt{\dfrac{s^2}{n}\left(1 - \dfrac{n}{N}\right)}$
Dubbelsidigt konfidensintervall för populationsandel	$p \pm z_{1-\alpha/2} \sqrt{\dfrac{p(1-p)}{n-1}\left(1 - \dfrac{n}{N}\right)}$
Dubbelsidigt konfidensintervall för totalt antal i populationen	$N \cdot p \pm z_{1-\alpha/2} \cdot N \cdot \sqrt{\dfrac{p(1-p)}{n-1}\left(1 - \dfrac{n}{N}\right)}$
Dubbelsidigt konfidensintervall för populationsmedelvärde vid stratifierat urval, $n_i \geq 30$	$\bar{x}_{STR} \pm z_{1-\alpha/2} \sqrt{\sum_{i=1}^{L} W_i^2 \cdot \dfrac{s_i^2}{n_i} \cdot \left(1 - \dfrac{n_i}{N_i}\right)}$ där $\bar{x}_{STR} = \sum_{i=1}^{L} W_i \cdot \bar{x}_i$ och $W_i = \dfrac{N_i}{N}$

Dubbelsidigt konfidensintervall för populationsandel vid stratifierat urval	$p_{STR} \pm z_{1-\alpha/2}\sqrt{\sum_{i=1}^{L} W_i^2 \cdot \frac{p_i(1-p_i)}{n_i - 1} \cdot \left(1-\frac{n_i}{N_i}\right)}$ där $p_{STR} = \sum_{i=1}^{L} W_i \cdot p_i$
Proportionell allokering	$n_i = n \cdot \frac{N_i}{N}$
Neymanallokering för medelvärden	$n_i = n \cdot \frac{N_i \cdot \sigma_i}{\sum_{j=1}^{L} N_j \cdot \sigma_j}$
Neymanallokering för andelar	$n_i = n \cdot \frac{N_i \cdot \sqrt{\pi_i(1-\pi_i)}}{\sum_{j=1}^{L} N_j \cdot \sqrt{\pi_j(1-\pi_j)}}$
Optimal allokering för medelvärden	$n_i = n \cdot \frac{N_i \cdot \sigma_i / \sqrt{c_i}}{\sum_{j=1}^{L} N_j \cdot \sigma_j / \sqrt{c_j}}$
Optimal allokering för andelar	$n_i = n \cdot \frac{N_i \cdot \sqrt{\pi_i(1-\pi_i)} / \sqrt{c_i}}{\sum_{j=1}^{L} N_j \cdot \sqrt{\pi_j(1-\pi_j)} / \sqrt{c_j}}$

Samband mellan kvalitativa variabler	
Chitvåtest för frekvenstabell	$\chi^2 = \sum_{i=1}^{V} \frac{(O_i - E_i)^2}{E_i}$ med kritiskt område till höger om $\chi^2_{V-1;\alpha}$
Chitvåtest för korstabell	$\chi^2 = \sum_{i=1}^{W} \frac{(O_i - E_i)^2}{E_i}$ med kritiskt område till höger om $\chi^2_{(r-1)(c-1);\alpha}$

Samband mellan kvantitativa variabler	
Korrelationskoefficienten	$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}$
Enkel linjär regression	$y_i = b_0 + b_1 \cdot x_i$ där $b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ och $b_0 = \bar{y} - b_1 \cdot \bar{x}$
Förklaringsgrad	r^2
Residualer	$e_i = y_i - \hat{y}_i$

Regressionsmodellens standardavvikelse	$s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
Hypotesprövning av β_1	$t = \frac{b_1}{\left(\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right)}$ med kritiskt område/områden enligt om $H_a: \beta_1 < 0$ vänster om $t_{n-2;\alpha}$ om $H_a: \beta_1 > 0$ höger om $t_{n-2;1-\alpha}$ om $H_a: \beta_1 \neq 0$ både till vänster och höger om $t_{n-2;\alpha/2}$ och $t_{n-2;1-\alpha/2}$
Dubbelsidigt konfidensintervall för β_1	$b_1 \pm t_{n-2;1-\alpha/2}\frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$
Prognosticering	$\hat{y}_{x^*} = b_0 + b_1 \cdot x^*$
Dubbelsidigt konfidensintervall för prognosticering	$\hat{y}_{x^*} \pm t_{n-2;1-\alpha/2} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$
Dubbelsidigt prognosintervall för prognosticering	$\hat{y}_{x^*} \pm t_{n-2;1-\alpha/2} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$

Ickeparametriska metoder	
Mann-Whitneys test	Välj som population 1 och 2 enligt $n_1 \leq n_2$. Rangordna de $n_1 + n_2$ observationerna. Bestäm rangsummorna R_1 och R_2 i vardera stickprovet. $U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$ $U' = n_2 n_1 + \frac{n_2(n_2+1)}{2} - R_2$ Vid dubbelsidig mothypotes: om den största av U eller U' är större än eller lika med $U_{n_1;n_2;\alpha/2}$ förkastas H_0 Vid enkelsidig mothypotes: H_a: Population 1 har lägre värden än population 2 använd U som testvariabel H_a: Population 1 har högre värden än population 2 använd U' som testvariabel Om testvariabeln är större än eller lika med kritiskt värde $U_{n_1;n_2;\alpha}$ förkastas H_0
Teckentest	Undersök tecken på differensen av varje observationspar. Låt X vara det tecken som förekommer minst antal gånger varpå $X \sim bin\left(n; \pi = 0.5\right)$ Bestäm sannolikheten för mindre än eller lika med det observerade värdet på X och multiplicera med 2 för *p*-värdet
Wilcoxons teckenrangtest	Bestäm differensen för varje observationspar och rangordna absolutvärdet av dessa. Bestäm rangsummorna T_+ för de positiva differenserna och T_- för de negativa differenserna. Välj som testvariabel T den minsta av T_+ och T_-. Om testvariabeln är mindre än eller lika med $T_{n^*;\alpha/2}$ där n^* är stickprovsstorleken minus antalet observationspar med nolldifferens förkastas H_0
Spearmans rangkorrelation	Rangordna observationerna inom respektive variabel och beräkna differensen d_i för varje observationspar. $r_S = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)}$