ORDLISTA

Beskrivande statistik	Presentera och sammanfatta data på ett överskådligt sätt
Binomialfördelning	$X \sim Bin(n,p)$, $\mathrm{E}(X) = np, \mathrm{var}(X) = np(1-p)$
Centrala gränsvärdessatsen	En summa av ”tillräckligt” många likafördelade stokastiska variabler är approximativt normalfördelad
Dummy-variabel	$x = 1$ eller 0, används för att separera kategorier vid tex. regressionsanalys
Empirisk korrelation	$r = \frac{\sum_1^n (x_j-\bar{x})(y_j-\bar{y})/(n-1)}{\sqrt{[\sum_1^n (x_j-\bar{x})^2/(n-1)][\sum_1^n (y_j-\bar{y})^2/(n-1)]}}$
Exponentialfördelning	$X \sim Exp(\mu)$, $\mathrm{E}(X) = \mu, \mathrm{var}(X) = \mu^2$
Flerdimesionell normalförd.	$\boldsymbol{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{C})$, med väntevärdesvektor $\boldsymbol{\mu}$ och kovariansmatris $\boldsymbol{C}$
F-fördelning	$Y_1 \sim \chi^2(r_1)$ och $Y_2 \sim \chi^2(r_2)$ oberoende, då blir $V = \frac{Y_1/r_1}{Y_2/r_2} \sim F(r_1, r_2)$
F-test i regr.analys	$V = \frac{Q_{REGR}/k}{Q_{RES}/(n-k-1)}$ testar H_0 : alla $\beta_i = 0$ $W = \frac{(Q_{RES}^{(1)} - Q_{RES}^{(2)})/p}{Q_{RES}^{(2)}/(n-k-p-1)}$ testar $H_0 : \beta_{k+1} 0 \ldots \beta_{k+p} = 0$
Histogram	Emperisk sannolikhets- eller täthetsfunktion
Hypotesprövning	Tex. $H_0 : \mu = \mu_0$ mot $H_1 : \mu \neq \mu_0$ på nivan α
Konfidensgrad	$1-\alpha$, är sannolikeheten att konfidensintervallet I_θ ”täcker över” det verkliga värdet på θ
Konfidensintervall vid N-förd.	$I_\mu, I_{\mu_i - \mu_j}, I_\sigma$ eller I_{σ^2}
Konfidensintervall vid Bin.förd.	$I_p, I_{p_1-p_2}$
Konfidensintervall via CGS	
Konsistentskattning	$\widehat{\Theta}_n$ är definierad för varje n. Om för varje $\varepsilon > 0$ gäller att $\mathrm{P}(\lvert\widehat{\Theta}_n - \theta\rvert > \varepsilon) \to 0$ då $n \to \infty$, så sägs $\widehat{\Theta}_n$ vara en konsistent skattning av θ.
Korrelation	$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$
Kovarians	$\mathrm{cov}(X,Y) = \mathrm{E}((X - \mathrm{E}(X))(Y - \mathrm{E}(Y)))$
Likelihoodfunktionen	$L(\theta) = \prod_{i=1}^n f(x_i; \theta)$
Medelvärde av ober. stok. var.	$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$
MK-metoden	Minsta-Kvadrat-metoden, välj det värde $\hat{\boldsymbol{\theta}}$ som minimerar $Q(\boldsymbol{\theta}) = \sum_1^n (x_i - \mu_i(\boldsymbol{\theta}))^2$
ML-metoden	Maximum-Likelihood-metoden, välj det värde $\hat{\boldsymbol{\theta}}$ som maximerar likelihoodfunktionen $L(\theta)$
Momentmetoden	Om $E(X_i) = \mu(\theta)$ så ges momentskattningen av ekvationen $\mu(\hat{\theta}) = \bar{x}$. Vid fler parametrar så måste även högre moment beräknas
Observation av stok. var.	Observationer är givna värden som är resultatet från ett slumpmässigt försök
P-värde	Kan användas då vi genomför hypotesprövningar. Anger den nivån som nollhypotesen kan förkastas på
Parvisa mätningar	(i) mätt på samma enhet före och efter behandling/åtgärd (ii) mätt med två olika metoder inom par med likvärda enheter
Population	Samtliga möjliga observationer
Prediktionsintervall	Prediktera en framtida enskild mätning
Poissonfördelning	$X \sim Po(\mu)$, $\mathrm{E}(X) = \mu, \mathrm{var}(X) = \mu$
Punktskattning	Om $x_1, ..., x_n$ är observationer av oberoende s.v. X_i, vars sannolikhetsfunktion $p(k;\theta)$ eller täthetsfunktion $f(x;\theta)$ innehåller en okänd parameter θ så är en gissning av . Punktskattningen är en gissning på θ
R^2=R-sq	$R^2 = \frac{Q_{REGR}}{Q_{RES}}$, anger hur mycket av den totala variationen som kan förklaras av väntevärdesmodellen

Regressionsanalys	$Y_j = \beta_0 + \beta_1 x_{j1} + \ldots + \beta_k x_{jk} + \varepsilon_j = \mu_j + \varepsilon_j$, där $\varepsilon_j \sim N(0, \sigma)$. Skatta β_i och σ^2
Regressionskvadratsumma	$Q_{REGR} = \sum_{i=1}^{n} (\hat{\mu}_j - \bar{y})^2$, k frihetsgrader
Residualanalys	Studera residualerna $\hat{e}_j = y_j - \hat{\mu}_j$ vilka är observatioer från $\varepsilon_j \sim N(0, \sigma)$
Residualkvadratsumma	$Q_{RES} = \sum_{i=1}^{n} (y_j - \hat{\mu}_j)^2$, $n - k - 1$ frihetsgrader. Skattar σ^2 med $s^2 = Q_{RES}/(n - k - 1)$
Räknelagar för väntevärden	$\mathrm{E}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i \, \mathrm{E}\left(X_i\right)$
Räknelagar för varianser	Om $X_1, ..., X_n$ är oberoende, så gäller att $\mathrm{var}\left(\sum_1^n c_i X_i\right) = \sum_1^n c_i^2 \, \mathrm{var}\left(X_i\right)$
Signifikansnivå	Sannolikeheten att förkasta H_0 då H_0 är sann, dvs. $\alpha = \mathrm{P}(H_0$ förkastas om H_0 är sann)
Signifikanstest	Tex. $H_0 : \mu = \mu_0$ mot $H_1 : \mu \neq \mu_0$ på nivan α
Standardiserad normalfördeln.	$X \sim N(\mu, \sigma)$ då är $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$
Stickprov	Delmängd av en population, $x_1, \ldots, x_n$.
Stickprovsmedelvärde	$\bar{x} = \frac{1}{n} \sum_1^n x_i$
Stickprovsstandardavvikelse	$s = \sqrt{\frac{1}{n-1} \sum_1^n (x_i - \bar{x})^2}$
Stokastiska vektorer	$\boldsymbol{X} = (X_1, \ldots, X_n)' : (n \times 1)$ där X_i är stokastiska variabler
Styrka	Sannoliketen att förkasta H_0 då då H_0 är falsk, dvs. $h(\theta) = \mathrm{P}(H_0$ förkastas om θ är det sanna värdet)
t-fördelning	$X \sim N(0, 1)$ och $Y \sim \chi^2(f)$ oberoende, då är $\frac{X}{\sqrt{Y/f}} \sim t(f)$
Varians	$\mathrm{var}(X) = \mathrm{E}((X - \mathrm{E}(X))^2)$
Väntevärdesriktighet	Om $\hat{\theta}$ är en skattning av θ och $\mathrm{E}(\widehat{\Theta}) = \theta$ så är skattningen$\hat{\theta}$ en vvr skattning av θ
χ^2-fördelning	$X \sim \chi^2(f)$
χ^2-test	Tex. test av given fördelning eller Homogenitetstest