Formelsamling (Matematisk statistik för K)

Kombinatorik: k element kan väljas bland n på

- n^k sätt med återläggning och ordningshänsyn
- $n(n-1)\cdots(n-k+1)$ sätt utan återläggning och med ordningshänsyn
- $\binom{n}{k}$ sätt utan återläggning och ordningshänsyn

Varians: $\sigma^2 = Var(X) = \mathbf{E}[(X-\mu)^2] = \mathbf{E}[X^2] - \mu^2$, där $\mu = \mathbf{E}[X]$.

Binomialfördelningen, $Bin(n,p)$**:**

- $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n$
- $\mathbf{E}[X] = np, \quad Var(X) = np(1-p)$

Normalfördelningen, $N(\mu, \sigma)$**:**

- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \quad -\infty < x < \infty$
- $\mathbf{E}[X] = \mu, \quad Var(X) = \sigma^2$

Summor och medelvärden:

- $\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$, $\mathbf{E}[aX] = a\mathbf{E}[X]$, där a är en konstant.
- Kovarians $Cov(X,Y) = \mathbf{E}[(X-\mu_X)(Y-\mu_Y)]$, där $\mu_X = \mathbf{E}[X]$ och $\mu_Y = \mathbf{E}[Y]$.
- $Var(aX) = a^2 Var(X)$, där a är en konstant. Om X och Y är oberoende är $Var(X+Y) = Var(X) + Var(Y)$,
- Om $X_1, ..., X_n$ är ett stickprov från fördelningen av X med $\mathbf{E}[X] = \mu$ och $Var(X) = \sigma^2$ gäller för medelvärdet $\bar{X}$
  - $\mathbf{E}[\bar{X}] = \mu$, $Var(\bar{X}) = \sigma^2/n$
  - $\bar{X}$ är $N(\mu, \sigma/\sqrt{n})$ om X är $N(\mu, \sigma)$
  - $\bar{X}$ är approximativt $N(\mu, \sigma/\sqrt{n})$ om $n \geq 25$.
- Summor av oberoende normalfördelade variabler är normalfördelade.

- Skillnaden mellan två oberoende normalfördelade variabler är normalfördelad.

Ett stickprov: Låt $X_1, ..., X_n$ vara ett stickprov från fördelningen av X med $\mathbf{E}[X] = \mu$ och $Var(X) = \sigma^2$. Då är $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ och $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ väntevärdesriktiga skattningar av μ respektive σ^2. Om speciellt är $X \sim N(\mu, \sigma)$ gäller

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ och } \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim T_{n-1}$$

Två oberoende stickprov: Låt $X_1, ..., X_{n_1}$ vara ett stickprov från fördelningen av X som är $N(\mu_X, \sigma)$ och $Y_1, ..., Y_{n_2}$ vara ett stickprov från fördelningen av Y som är $N(\mu_Y, \sigma)$. Då är

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$T_{n_1+n_2-2}$-fördelad, där $S = \sqrt{S^2}$ och

$$S^2 = \frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}.$$

Regression:

- $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma)$ och oberoende, $i = 1, ..., n$.
- $\hat{\beta}_1 = \frac{n\sum x_i Y_i - \sum x_i \sum Y_i}{n\sum x_i^2 - (\sum x_i)^2}$ och $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$
- σ^2 skattas med $S^2 = \frac{\sum (Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{n-2}$
- $\frac{\hat{\beta}_0 - \beta_0}{S\sqrt{\frac{\sum x_i^2}{n\sum (x_i - \bar{x})^2}}}$ är T_{n-2}-fördelad.
- $\frac{\hat{\beta}_1 - \beta_1}{S/\sqrt{\sum (x_i - \bar{x})^2}}$ är T_{n-2}-fördelad.

Korrelation: Låt $(X_1, Y_1), ..., (X_n, Y_n)$ vara ett stickprov på (X, Y) ($\mathbf{E}[X] = \mu_X$, $Var(X) = \sigma_X^2$, $\mathbf{E}[Y] = \mu_Y$, $Var(Y) = \sigma_Y^2$). Korrelationen mellan X och Y

$$\rho = \frac{\mathbf{E}[(X - \mu_x)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

skattas med korrelationskoefficienten

$$r = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n\sum x_i^2 - (\sum x_i)^2)(n\sum y_i^2 - (\sum y_i)^2)}}.$$