STATISTICS FOR BUSINESS (JOSEMARI SARASOLA, DONOSTIA): FORMULA SHEET

$B(n,p): p^x(1-p)^{n-x}\frac{n!}{x!(n-x)!}$

$\mu_{B(n,p)} = np, \sigma^2_{B(n,p)} = npq$

$G(p): (1-p)^x p$

$\mu_{G(p)} = \frac{q}{p}$

$BN(r,p): (1-p)^x p^{r-1}\frac{[x+(r-1)]!}{x!(r-1)!}p$

$\mu_{BN(r,p)} = \frac{rq}{p}$

p-value for sign test: $p = P[X \leq x] = 2\sum_{i=1}^{x} 0.5^n \frac{n!}{i!(n-i)!}$

$$P[X = x; H(k,N,n)] = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$H(k,N,n): \mu = \frac{nk}{N}; \sigma^2 = \frac{nk}{N} \times \frac{N-k}{N} \times \frac{N-n}{N-1}$$

$P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!}$

$B(n,p) \to P(\lambda = np)$

$F(x) = 1 - e^{-\lambda x}$

$\mu_{Exp(\lambda)} = \frac{1}{\lambda}$

$\sigma^2_{Exp(\lambda)} = \frac{1}{\lambda^2}$

Discrete uniform

$$P[X_{max} \leq x_i] = \left(\frac{i}{N}\right)^n$$

$$P[X_{max} = x_i] = \left(\frac{i}{N}\right)^n - \left(\frac{i-1}{N}\right)^n$$

$$P[X_{min} = x_i] = P[X_{min} \geq x_i] - P[X_{min} \geq x_{i+1}] = \left(\frac{N-(i-1)}{N}\right)^n - \left(\frac{N-i}{N}\right)^n$$

$$\hat{N} = x_{max} + \frac{x_{max} - n}{n}$$

Continuous uniform

$$X \sim U(a,b) \begin{cases} \mu = \frac{a+b}{2} \\ \sigma^2 = \frac{(b-a)^2}{12} \end{cases}$$

$F(x) = \frac{x-a}{b-a}$

$$F(M = x) = P[M < x] = \left(\frac{x-a}{b-a}\right)^n$$

$$E[M] = a + \frac{n(b-a)}{n+1}$$

$$F(m = x) = P[m < x] = 1 - \left(\frac{b-x}{b-a}\right)^n$$

$$E[m] = a + \frac{b-a}{n+1}$$

$$F(R = x) = P[R < x] = nx^{n-1}(1-x) + x^n$$

$$E[R] = (b-a)\frac{n-1}{n+1}$$

$B(n,p) \to N(\mu = np, \sigma = \sqrt{npq})$

$P(\lambda) \to N(\mu = \lambda, \sigma = \sqrt{\lambda})$

Validation

$X^2 = \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{\alpha,k-1}$

$X^2 = \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{\alpha,k-z-1}$

$$R \sim N\left(\mu = \frac{2n_1n_2}{n_1+n_2} + 1, \sigma = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}\right)$$

$$W_1 \sim N\left(\mu = \frac{n_1(n_1+n_2+1)}{2},\ \sigma = \sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}\right)$$

Parametric testing

Normal pop, σ known:

$\overline{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Normal pop, σ unknown:

$$t = \frac{\overline{x} - \mu}{\frac{\hat{s}}{\sqrt{n}}} \sim t_{n-1}$$

Non normal pop, σ known ($n > 30$):

$\overline{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Non normal pop, σ unknown ($n > 30$):

$\overline{x} \sim N\left(\mu, \frac{\hat{s}}{\sqrt{n}}\right)$

Proportion:

$\hat{p} \sim N\left(p, \sqrt{\frac{pq}{n}}\right)$

Variance: $\dfrac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$

Sample deviations:

$s = \sqrt{\frac{\sum_i (x_i - \overline{x})^2}{n}} = \sqrt{\frac{\sum_i x_i^2}{n} - \overline{x}^2}$

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

$\hat{s}^2 = \frac{n}{n-1} s^2$

$s^2 = \frac{n-1}{n} \hat{s}^2$

Confidence intervals

$\mu : \overline{x} \pm t_{n-1,\alpha/2} \dfrac{\hat{s}}{\sqrt{n}}$

$p : \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$$n = \frac{z^2_{\frac{\alpha}{2}} \cdot pq}{\epsilon^2}$$