Applied Stat Formula Sheet

Counting Formulas: $n!$ $\quad \frac{n!}{(n-k)!}$ $\quad \binom{n}{k} = \frac{n!}{(n-k)!k!}$

$$P(A^c) = 1 - P(A)$$

DeMorgan's $A' \cap B' = (A \cup B)'$

$$P(B) = P(B \cap A) + P(B \cap A')$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Law of total prob: $P(S) = P(S|A)P(A) + P(S|A')P(A')$

Discrete Distributions

	pmf	CDF	$E(X)$	$V(X)$
Binomial (n, p)	$\binom{n}{x}p^x(1-p)^{n-x}$ dbinom(x,n,p)	$F_B(x; n, p)$ pbinom(x,n,p)	np	$np(1-p)$
Negative Binomial (r, p)	$\binom{r+x-1}{r-1}(1-p)^x p^r$ dnbinom(x,r,p)	$F_{NB}(x; r, p)$ pnbinom(x,r,p)	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometric (n, m, N)	$\frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$ dhyper(x,m,N-m,n)	$F_{HG}(x; n, m, N)$ phyper(x,m,N-m,n)	$n\frac{m}{N}$	$n(\frac{m}{N})(1-\frac{m}{N})\frac{N-n}{N-1}$
Poisson (λ)	$\frac{e^{-\lambda}\lambda^x}{x!}$ dpois(x,λ)	$F_{POI}(x; \lambda)$ ppois(x,λ)	λ	λ

Continuos Distributions

	domain	f(x)	CDF	$E(X)$	$V(X)$
Normal (μ, σ^2)	$(-\infty, \infty)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x-\mu}{2\sigma^2}}$ dnorm(x,μ, σ)	$\Phi(\frac{x-\mu}{\sigma})$ pnorm(x,μ, σ)	μ	σ^2
Uniform (a, b)	$[a, b]$	$\frac{1}{b-a}$ dunif(x,a, b)	$\frac{x-a}{b-a}$ punif(x,a, b)	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Exponential (λ)	$[0, \infty)$	$\lambda e^{-\lambda x}$ dexp(x,λ)	$1 - e^{-\lambda x}$ pexp(x,λ)	$1/\lambda$	$1/\lambda^2$
Gamma (α, β)	$[0, \infty)$	$\frac{1}{\beta^\alpha \Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}$ dgamma(x,α,scale=β)	$F_{GAM}(x; \alpha, \beta)$ pgamma(x,α,scale=β)	$\alpha\beta$	$\alpha\beta^2$
$\chi^2(\nu)$	$[0, \infty)$	$\frac{1}{\beta^{k/2}\Gamma(k/2)}x^{k/2-1}e^{-x/2}$ dchisq(x,ν)	$F_{CHI}(x; \nu)$ pchisq(x,ν)	ν	2ν

(if $X_i \sim \text{Exp}(\lambda)$)

$$\text{CDF of } \max(X_1, \ldots, X_n) = [F(x)]^n = [1 - e^{-\lambda x}]^n$$

$$\text{CDF of } \min(X_1, \ldots, X_n) = 1 - [1 - F(x)]^n = 1 - e^{-n\lambda x}$$

Z-test

$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$\bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\bar{X} - \bar{Y} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}}$	$\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$	$\frac{\bar{X}-\bar{Y}-\Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

H_A	rejection region	p-value	power
upper-tailed	$z > z_\alpha$	$1 - \Phi(z)$	$1 - \Phi(z_\alpha - \mu_A)$
lower-tailed	$z < -z_\alpha$	$\Phi(z)$	$\Phi(-z_\alpha - \mu_A)$
Two-tailed	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$	$2\big(1 - \Phi(\lvert z \rvert)\big)$	$1 - \Phi(z_{\frac{\alpha}{2}} - \mu_A) + \Phi(-z_{\frac{\alpha}{2}} - \mu_A)$

$$\mu_A = (\mu - \mu_0)/\frac{\sigma}{\sqrt{n}}$$

$$\mu_A = (\mu_1 - \mu_2 - \Delta_0)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

For two-sample t-test, the df can be calculated by letting $a = S_1^2/n_1, b = S_2^2/n_2$ and

$$\nu = \frac{(a+b)^2}{\frac{a^2}{n_1-1} + \frac{b^2}{n_2-1}}$$

One-Sample Variance

$$\sigma^2 \in \Big(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \quad \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\Big), \qquad \frac{(n-1)S^2}{\sigma_0} \sim \chi^2(\text{n-1}) \text{ under } H_0$$

Two-Sample Variance

$$\frac{\sigma_2^2}{\sigma_1^2} \in \Big(\frac{S_2^2}{S_1^2}\mathcal{F}_{1-\frac{\alpha}{2},m-1,n-1}, \quad \frac{S_2^2}{S_1^2}\mathcal{F}_{\frac{\alpha}{2},m-1,n-1}\Big), \qquad \frac{S_1^2}{S_2^2} \sim F(\text{m-1, n-1}) \text{ under } H_0$$

Prediction Interval

$$\overline{X} \pm Z_{\frac{\alpha}{2}}\sigma\sqrt{\frac{1}{n} + 1}$$

ANOVA

Source	DF	Sum of Squares	MS	F
Group (between)	k-1	SSG $= \sum_{group} n_i(\bar{x}_i - \bar{x})^2$	MSG = SSG/k-1	F=MSG/ MSE
Error (within)	N-k	SSE $= \sum_{group}(n_i - 1)S_i^2$	MSE = SSE / N-k	
Total	N-1	SST		

$F \sim F(k-1, N-k)$ under the null hypothesis. $SST = SSG + SSE$

$R^2 = \frac{SSG}{SST}$ $\hat{\sigma}^2 =$ MSE

Least Square Regression Regression line:

$$y_i = b_0 + b_1 x_i + \epsilon_i$$

Fitted line:

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 X \qquad \hat{b}_1 = r\frac{S_2}{S_1} \qquad \hat{b}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\sigma}^2 = \text{ MSE}$$

$100(1-\alpha)\%$ (two-sided) CI for b_1 can be derived as

$$\left(\hat{b}_1 \pm t_{\frac{\alpha}{2}, n-2}\, \text{SE}_{\hat{b}_1}\right)$$

$100(1-\alpha)\%$ (two-sided) CI for mean response given $x = x^*$ is

$$(\hat{b}_0 + \hat{b}_1 x^*) \pm t_{\frac{\alpha}{2}, n-2}\sqrt{\text{SE}_{\hat{b}_0} + \text{SE}_{\hat{b}_1}(x - \bar{x})^2}$$

$100(1-\alpha)\%$ (two-sided) prediction interval for response given $x = x^*$ is

$$(\hat{b}_0 + \hat{b}_1 x^*) \pm t_{\frac{\alpha}{2}, n-2}\sqrt{\text{SE}_{\hat{b}_0} + \text{SE}_{\hat{b}_1}(x - \bar{x})^2 + \hat{\sigma}^2}$$