List of Formulæ
ST370 - Probability and Statistics for Engineers

(bring this list for all exams)

I. Exploring statistical data and regression analysis.

1. $X_1, X_2, \cdots, X_n$.: Sample data consisting of n observations.
2. $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.: Sample **mean**.
3. $Q_d = X_{(k)} + [(n+1)d - k](X_{(k+1)} - X_{(k)})$.: Sample d-th **fractile/quantile**, where k =largest integer$\leq (n+1)d$ and $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ represents data sorted in ascending order. Notice that the **median**, $m = Q_{0.5}$.
4. $s^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \frac{\sum X_i^2 - n\overline{X}^2}{n-1}$.: Sample **variance**.
5. $s = \sqrt{s^2}$: sample **standard deviation** (sd).
6. $R = X_{(n)} - X_{(1)}$.: Sample **range**.
7. $IQR = Q_{0.75} - Q_{0.25}$.: **Inter-quartile range**.
8. $\nu = \frac{s}{\overline{X}}$.: **Coefficient of variation**.
9. $SK = \frac{3(\overline{X} - m)}{s}$.: **Skewness coefficient**. $SK = 0 \Rightarrow$ *symmetric*, $SK > 0 \Rightarrow$ +vely skewed and $SK < 0 \Rightarrow -$vely skewed.
10. $r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{(n-1)s_X s_Y} = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{(n-1)s_X s_Y}$.: sample **correlation coefficient**.
11. $b = rs_Y/s_X = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{(n-1)s_X^2}$.: The **slope** when Y is regressed on X (i.e. $Y \approx a + bX$).
12. $a = \overline{Y} - b\overline{X}$: The vertical **intercept** when Y is regressed on X.
13. $\widehat{Y(x)} = a + bx$: Predicted value of Y when X=x is observed.
14. $s_{Y.X} = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{n-2}}$: **Standard error of the estimate** when Y is regressed on X.
15. $R^2 = 1 - \frac{(n-2)s_{Y.X}^2}{(n-1)s_Y^2}$.: **Multiple R-squared** (Coefficient of determination). Note $R^2 = r^2$ only for *Linear* regression.
16. $e_i = Y_i - \widehat{Y(X_i)}$.: **Residuals** when Y is regressed on X.

II. Probability and random variables (r.v.).

1. $\Pr[\text{not } A] = 1 - \Pr(A)$.: Probability of a complementary event.
2. $\Pr[A \text{ and } B] = 0$.: Implies A and B are **mutually exclusive** or disjoint events.
3. $\Pr[A \text{ or } B] = \Pr[A] + \Pr[B] - \Pr[A \text{ and } B]$.: General **Addition law** of probability.
4. $\Pr[A \mid B] = \Pr[A \text{ and } B] / \Pr[B]$.: **Conditional probability** of A given B.
5. $\Pr[A \mid B] = \Pr[A]$ or $\Pr[A \text{ and } B] = \Pr[A]\Pr[B]$.: Implies A and B are statistically **independent events**.
6. $\Pr[A \text{ and } B] = \Pr[A \mid B]\Pr[B] = \Pr[B \mid A]\Pr[A]$: General **Multiplication law** of probability.
7. $p(y) = \Pr[Y = y], \sum_{y=0}^{\infty} p(y) = 1$.: **Probability mass function** of a <u>discrete</u> random variable Y. Note $0 \leq p(y) \leq 1$, where $y = 0, 1, \ldots$.
8. $E(X) = \sum xp(x)$.: **Expected value** of a <u>discrete</u> random variable X.
9. $Var(X) = \sum [x - E(X)]^2 p(x)$. : **Variance** of a <u>discrete</u> random variable X. $SD(X) = \sqrt{Var(X)}$; **standard deviation** of X.
10. $f(x) \geq 0, \int_{-\infty}^{\infty} f(x)dx = 1$.: **Probability density function** of <u>continuous</u> random variable X.
11. $E(X) = \int_{-\infty}^{\infty} xf(x)\ dx$.: **Expected value** of a <u>continuous</u> random variable X.
12. $Var(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x)\ dx$.: **Variance** of a <u>continuous</u> random variable X. $SD(X) = \sqrt{Var(X)}$; **standard deviation** of X.
13. $E(a + bX) = a + bE(X)$.: Property of expected value (holds for both discrete and continuous r.v.).
14. $Var(a + bX) = b^2 Var(X)$.: Property of variance (holds for both discrete and continuous r.v.). Note that $Var(X) = E(X^2) - [E(X)]^2$.
15. $F(x) = \Pr[X \leq x]$.: Cumulative probability distribution function of X. Notice that $\Pr[a < X \leq b] = F(b) - F(a)$.

16. $b(r; n, \pi) = C_r^n \pi^r (1-\pi)^{n-r} = \Pr[R = r]$.: **Binomial** probability mass function, where R is the number of *success* out of n trials and $\Pr[success] = \pi$.

17. $E(R) = n\pi, Var(R) = n\pi(1-\pi)$.: Expected value and variance for a Binomial Distribution.

18. $B(r; n, \pi) = \Pr[R \leq r] = \sum_{x=0}^{r} b(x; n, \pi)$.: Binomial probability distribution function. (see Table A)

19. $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}$.: Probability Density function of a **Normal** distribution.

20. $E(X) = \mu, Var(X) = \sigma^2$.: Expected value and variance for a normal distribution.

21. $Z = \frac{X-\mu}{\sigma}$.: **Normal deviate.** Notice that $E(Z) = 0$ and $Var(Z) = 1$.

22. $\Phi(z) = \Pr[Z \leq z] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy$.: **Standard normal** probability distribution function. (see Table D)

23. $p(x; \lambda, t) = \Pr[X = x] = \frac{e^{-\lambda t}(\lambda t)^x}{x!}; x = 0, 1, 2 \ldots$.: Probability mass function for **Poisson** distribution.

24. $P(x; \lambda, t) = \Pr[X \leq x] = \sum_{k=0}^{x} p(k; \lambda, t)$.: Poisson probability distribution function. (see Table C)

25. $E(X) = \lambda t = Var(X)$.: Expected value and variance for a Poisson Distribution.

26. $f(t) = \lambda e^{-\lambda t}; t \geq 0$.: Probability density function for the **Exponential** distribution.

27. $F(t) = \Pr[T \leq t] = 1 - e^{-\lambda t}$.: Exponential probability distribution function.

28. $E(T) = \frac{1}{\lambda}, Var(T) = \frac{1}{\lambda^2}$.: Expected value and variance for an Exponential Distribution.

III. Sampling distribution and statistical inference.

1. $E(\overline{X}) = \mu, SD(\overline{X}) = \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$: Property of **sample mean** for large population. Note for small population, $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.
2. **Central Limit Theorem (CLT)**: *The sampling distribution of $\overline{X}$ approaches a normal curve with expected value μ and variance $\frac{\sigma^2}{n}$ as n becomes large, regardless of the form of population frequency distribution.*
3. $\overline{X} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.: **Confidence interval** estimate of μ, when σ known. (see Table D or E for $z_{\alpha/2}$)
4. $\overline{X} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$.: **Confidence interval** estimate of μ, when σ unknown. (see Table G for $t_{\alpha/2}$ with $df = n - 1$)
5. $n = \frac{z_{\alpha/2}^2 \sigma^2}{d^2}$. : Required **sample size** for estimating the mean, where $d =$ desired precision, $z_{\alpha/2}$ =critical normal deviate and $\sigma =$ assumed population sd. If σ is unknown, replace it by its estimate say $R/6$ (recall, $R = X_{(n)} - X_{(1)}$).
6. $P \pm z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$.: **Confidence interval** estimate of the **proportion** when the population is large. This estimate is based on the assumption that $n\pi \geq 5$ and $n(1-\pi) \geq 5$.
7. $n = \frac{z_{\alpha/2}^2 \pi(1-\pi)}{d^2}$. : Required **sample size** for estimating the proportion, where $d =$ desired precision, $z_{\alpha/2}$ =critical normal deviate and $\pi =$ assumed population proportion. If π is unknown, replace it by a conservative estimate, $\pi = 0.5$.
8. $(\overline{X_A} - \overline{X_B}) \pm z_{\alpha/2}\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$.: Confidence interval estimate for $\mu_A - \mu_B$ using large independent samples ($\min(n_A, n_B) > 30$).
9. $(\overline{X_A} - \overline{X_B}) \pm t_{\alpha/2} s_D$.: Confidence interval estimate for $\mu_A - \mu_B$ using small independent samples ($\min(n_A, n_B) \leq 30$).

   When $\sigma_A = \sigma_B, s_D = \sqrt{\frac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A+n_B-2}}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$. (see Table G for $t_{\alpha/2}$ with $df = n_A + n_B - 2$)

   When $\sigma_A \neq \sigma_B, s_D = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$. (see Table G for $t_{\alpha/2}$ with $df = \frac{(s_A^2/n_A + s_B^2/n_B)^2}{(s_A^2/n_A)^2/(n_A-1) + (s_B^2/n_B)^2/(n_B-1)}$)
10. $\overline{d} \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$.: Confidence interval estimate of $\mu_A - \mu_B$ using **matched pairs**. (see Table G for $t_{\alpha/2}$ with $df = n - 1$)