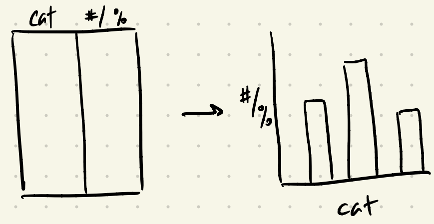
Chapter 2:

Categorical:



Continuous:

$$k = \lceil \log_2 n \rceil + 1$$

$$\bar{X} = \sum_{i=1}^{n} x_i / n$$

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / (n-1)$$

$$\tilde{X} = \frac{1}{2}\left(X_{(\lfloor \frac{n+1}{2} \rfloor)} + X_{(\lceil \frac{n+1}{2} \rceil)} \right)$$

$\bar{X}_{k\%}$ remove top/bot $k\%$

mode

$$CV = s/\bar{x}$$

IQR: Q3 - Q1

Q1 - 1.5 (IQR)

Q3 + 1.5 (IQR)

$$z = \frac{x - \bar{x}}{s}$$

Emp: 68/95/99.7

Chebyshev: $1 - \frac{1}{k^2}$ w/in k stdevs.

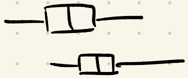
QQ plots

Box-Cox:

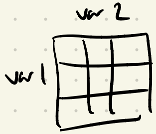
$$Y_\lambda = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(x) & \lambda = 0 \end{cases}$$

Chapter 3:

CQ:



CC:



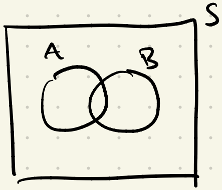
QQ:

Pearson correlation r: $$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)}} \in [-1,1]$$

sample covariance: $$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}$$

Chapter 4:

Prob:



$$(A \cup B)^c = A^c \cap B^c$$

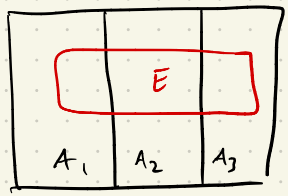
$$(A \cap B)^c = A^c \cup B^c$$

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

$$Pr(A \cap B) = Pr(A) \cdot Pr(B \mid A) = Pr(B) \cdot Pr(A \mid B)$$

Bayes': $Pr(A \mid B) = Pr(B \mid A)\, Pr(A) \,/\, Pr(B)$

LoTP:



$$Pr(E) = Pr(E \cap A_1) + Pr(E \cap A_2) + \cdots$$

$$= Pr(E \mid A_1)\, Pr(A_1) + \cdots$$

Sensitivity: $Pr(T^+ \mid D^+)$ PPV: $Pr(D^+ \mid T^+)$

Specificity: $Pr(T^- \mid D^-)$ NPV: $Pr(D^- \mid T^-)$

Comb:

$P(n, k)$

$\binom{n}{k}$

Stars and bars: $\binom{n+k-1}{k-1}$ w/ n objects, k bins, ≥ 0

Chapter 5:

Random variables: X, f_X, F_X

$$E(X) = \int x f(x)\, dx$$

$$Var(X) = E\left((X - E(X))^2\right) = E(X^2) - E(X)^2$$

$$Cov(X,Y) = E\left((X-E(X))(Y-E(Y))\right) = E(XY) - E(X)E(Y) \Leftarrow \sigma_{XY}$$

$$corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

$$Z = aX + bY \rightarrow E(Z) = a\mu_X + b\mu_Y$$

$$Var(Z) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\sigma_{XY}.$$

Discrete:

$Bern(p)$. $E(X) = p$, $Var(X) = p(1-p)$

$Bin(n,p)$ $E(X) = np$, $Var(X) = np(1-p)$

$Pois(\lambda)$ $E(X) = Var(X) = \lambda$. $Pr(X=x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

$Geom(p)$ $E(X) = 1/p$, $Var(X) = \dfrac{1-p}{p^2}$

Cont:

$Unif(a,b)$ $E(X) = \dfrac{a+b}{2}$, $Var(X) = \frac{1}{12}(b-a)^2$

$Exp(\lambda)$ $E(X) = 1/\lambda$, $Var(X) = 1/\lambda^2$

$N(\mu, \sigma)$

Sampling: $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$

CLT

Chapter 6:

CIs: Known σ^2: $\Pr\left(z_{\alpha/2} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le z_{1-\alpha/2}\right) = \alpha$ 1- vs. 2-sided

$$\Pr\left(\bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = \alpha$$

Hypothesis Tests

$\boxed{\mu}$

$\bar{x} \sim N(\ )$?

yes

populations

1 or paired

known σ^2 ?

yes

z-test

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$z \sim N(0,1)$$

no

t-test

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$df = n - 1$$

$$t \sim t_{n-1}$$

2

equal σ^2 ?

yes

known σ^2 ?

yes

2-sample z-test

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$z \sim N(0,1)$$

no

2-sample t-test

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 - 1 + n_2 - 1}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$df = n_1 - 1 + n_2 - 1$$

$$t \sim t_{n_1+n_2-2}$$

no

Welch's t-test

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right]}$$

$$t \sim t_\nu$$

no

populations

1 or paired

Wilcoxon signed-rank test

$$T = T^+ - T^-$$

$$\mu = 0$$

$$\sigma = \sqrt{\frac{n(n+1)(2n+1)}{6}}$$

2

W rank-sum test

$$W = \min(W_1, W_2)$$

$$\mu_W = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_W = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

$\boxed{\sigma^2}$

1 sample: $T = (n-1) \cdot \frac{s^2}{\sigma^2} \sim \chi^2_{n-1}$

2 samples: $F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim \frac{\chi^2_{n_1-1}/(n_1-1)}{\chi^2_{n_2-1}/(n_2-1)} = F_{n_1-1,\, n_2-1}$

$\boxed{p}$

1 sample: $\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

CIs: $p = \hat{p}$ (Wald)

$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$

H-tests: $p = p_0$

2 samples: $\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$

CIs: $p_1 = \hat{p}_1,\ p_2 = \hat{p}_2$

$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$

H-tests: $p_1 = p_2 = p = \frac{X_1 + X_2}{n_1 + n_2}$

$\boxed{\chi^2}$

GoF: $X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-1}$

χ^2 ToI: $X^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(r-1)(c-1)}$