- Probability density functions PDF

$P(a \le x \le b) = \int_a^b f(x)\,dx \qquad 1 = \int_{-\infty}^{\infty} f(x)\,dx \qquad$ eg) $f(x) = \begin{cases} bx & 0 \le x < 3 \\ b(6-x) & 3 \le x \le 6 \\ 0 & \text{otherwise} \end{cases}$

expectation value: $E(g(x)) = \int g(x) f(x)$ eg) $E(3X+4)$ $g(x) = 3x+4$

variance : $Var(g(x)) = \int (g(x))^2 f(x)\,dx - (E(g(x)))^2$

- Cumulative distribution functions CDF

$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$ (swap variable for t since x in limit)

$f(x) = \frac{d}{dx} F(x)$

- median m : $\int_{-\infty}^{m} f(x)\,dx = \int_{m}^{\infty} f(x)\,dx = \frac{1}{2}$

- mode value for which $f(x)$ has maximum value

- to switch variables
1.) given $f(x)$ is PDF of X find PDF of $Y = X^2$
2.) find cumulative distribution function CDF first
3.) substitute to get CDF in terms of Y
4.) differentiate CDF to get PDF in terms of Y

- T test
  - random sampling (sample reflects pop well)
  - data can be assumed to follow normal distribution

- confidence intervals $CI = \bar{x} \pm \Upsilon \frac{s}{\sqrt{n}}$ $\quad V = n-1$

- test static $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \qquad t = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} \qquad t = \frac{\bar{x} - \bar{y} - \mu_0}{s\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$

(if $s_x \approx s_y$)

Ho: population mean $= \mu_0$

H_1: $\mu \neq \mu_0$ pop mean not μ_0 $-T_2 < t < T_2$ then accept Ho

H_1: $\mu > \mu_0$ pop mean greater than μ_0 $t < T_1$ then accept Ho

H_1: $\mu < \mu_0$ pop mean smaller than μ_0 $t > T_1$ then accept Ho

- if standard deviation known $V = \infty$ so normal distribution

---

- contingency table

f_0	✓	×	Total
✓			
×			
Total			

$V = (m-1)(n-1)$

- chi squared static $\chi^2 = \sum \frac{(f_0 - f_E)^2}{f_E}$

Ho: the two variables are independent ∴ not associated
H_1: the two variables are dependent ∴ associated

if $\chi^2 > \chi^2_{\text{critical}}$ reject Ho accept H1
$\chi^2 < \chi^2$ critical accept Ho reject H_1

- used for godness of fit tests
- if $f_E \geq 5$ is ok if $f_E < 5$ combine sets

$V = n -$ number of constraints

1 from given total for all entries
1 from approximating a parameter

$V = n - 1 - 1$
total given, approximated parameter

- Wilcoxon rank test

Ho: population median M equal to given value $M = M_0$
H_1: population median M not equal / lower / higher $M \neq M_0$

H_1: $M = M_0$ accept Ho if W_+ and $W_- >$ T critical

H_1: $M > M_0$ accept Ho if $W_- > T_{\text{critical}}$

H_1: $M < M_0$ accept Ho if $W_+ <$ T critical

- Wilcoxon rank sum test

  m - number of samples in smaller set of data

  n - number of samples in larger set of data

check $W_x + W_y = \frac{1}{2}(m+n)(m+n+1)$

H_0 : median of X and Y are equal $M_x = M_y$

H_1: lower tail test   W = rank sum of set with less samples (m)

H_1: upper tail test   test static $= m(m+n+1) - W$

H_1: two tailed test   use smaller so either W or $m(n+n+1) - W$

reject H_0 if test static $\leq$ critical value

- model as normal distribution $X \sim N(\mu, \sigma^2)$ use continuity correction

---

- probability generating function PGF

  $G(t) = E(t^X) = \sum_x t^x P(X=x)$

- sum of probabilities $= 1$ so $t=1$ $G(1) = 1$

- expectation value $E(X) = G'(1) = \mu$

- variance $Var(X) = G''(1) + G'(1) - (G'(1))^2$

- for 2 independent random variables X, Y

  $G_{X+Y}(t) = G_X(t)\, G_Y(t)$