$$\begin{cases} \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \\ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \end{cases}$$

$$\begin{cases} E[\bar{X}] = \mu \\ E[S^2] = \sigma^2 \end{cases}, \text{ unbiased}$$

- ~~negat~~ Hypergeometric distribution

$$\mu = n \cdot \frac{m}{N}$$

sampling without replacement.

- binomial distri~~but~~ution.

$$P(X=x) = C_n^x \, p^x (1-p)^{n-x}$$

$$\mu = np, \ \sigma^2 = np(1-p)$$

• moment generating function

$$M(t) = E[e^{tX}] = \sum_{x>0} e^{tx} f(x)$$

$$\mu = M'(0) \quad \boxed{M^{(r)}(0) = E[X^r]}$$

$$\sigma^2 = E[X^2] - E[X]^2 = M''(0) - M'(0)^2$$

- geometric distribution

~~first~~ number of trials until first success.

$$F(x) = 1-(1-p)^x$$

$$f(x) = (1-p)^{x-1} p.$$

$$\mu = \frac{1}{p}, \ \sigma^2 = \frac{1-p}{p^2}$$

- negative ~~binomial~~ binomial distribution

$$f(x) = P(X=x) = \binom{x-1}{r-1}(1-p)^{x-r} p^r$$

$$= C_{x-1}^{r-1} (1-p)^{x-r} p^r$$

$$\mu = \frac{r}{p}, \ \sigma^2 = \frac{r(1-p)}{p^2}$$

- Poisson distribution

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\mu = \sigma^2 = \lambda.$$

- exponential distribution

$$f(x) = \frac{1}{\theta} e^{-x/\theta} \qquad \theta = \frac{1}{\lambda}$$

mean waiting time

$$\mu = \theta, \ \sigma^2 = \theta^2$$

$\boxed{\text{memoryless}}$

- gamma distribution

average waiting time until αth event occurs.

• $\mu = \alpha\theta, \ \sigma^2 = \alpha\theta^2$

- gamma function

$$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$$

$$\Gamma(t) = (t-1)\Gamma(t-1), \ \Gamma(n) = (n-1)!$$

- chi-square distribution.
  - a special case of Gamma distribution with

    $\theta = 2$, $\alpha = \frac{r}{2}$

    $\mu = r$, $\sigma^2 = 2r$

  r - degrees of freedom.

- Normal distribution.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} \sim N(\mu, \sigma^2)$$

Statistics   6&7

ISLR

- $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$

$$T = \frac{Z}{\sqrt{U/r}}, \quad Z \sim N(0,1), \quad U \sim \chi^2_{(r)} \text{ independent}$$

$\sim T(r-1)$

One sample $\boxed{\text{mean}}$

- use t when the data is normal
- not normal, $n \geq 30$, t and z are similar.
- not normal, n is small
  use t or nonparametric method.

$\boxed{\begin{array}{l} z \text{ (know population variance)} \\ t \text{ (do not know ...)} \end{array}}$

- Two samples.

① independent ..

② normal   ③ same variance σ^2.

$$(\bar{x}-\bar{y}) \pm 2\, t_{\frac{\alpha}{2}, n+m-2}\, S_p \sqrt{\frac{1}{n}+\frac{1}{m}}$$

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2} \quad \text{pooled variance.}$$

Welch's t $(\sigma_x^2 \neq \sigma_y^2)$

$$(\bar{x}-\bar{y}) \pm 2\, t_{\frac{\alpha}{2}, r} \sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}$$

$$r = \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\frac{(S_x/n)^2}{n-1} + \frac{(S_y/m)^2}{m-1}}$$

paired t-test $\rightarrow$ One sample.

- One variance:

$$\underset{a}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \leq \frac{(n-1)S^2}{\sigma^2} \leq \underset{b}{\chi^2_{\frac{\alpha}{2}, n-1}}$$

$$\Rightarrow \frac{(n-1)S^2}{b} \leq \sigma^2 \leq \frac{(n-1)S^2}{a}$$

- F distribution: (2 sample variances)

$$F = \frac{U/r_1}{V/r_2} \quad U \sim \chi^2(r_1), \quad V \sim \chi^2(r_2)$$

$\sim F(r_1, r_2)$

- Sample Proportion

$$\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Two proportions

$$\hat{\sigma} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}$$

- χ^2 test: goodness of fit.

$$\chi^2 \sim z^2 = \frac{(Y_1 - np_1)^2}{p_1(1-p_1)} \cdot (1-p_1+p_1)$$

$$= \frac{(Y_1-np_1)^2}{np_1} + \frac{(Y_2-np_2)^2}{np_2} = \sum_i \frac{(O_i - E_i)^2}{\text{Expected}}$$

Applications $\begin{cases} \text{homogeneity} \\ \text{independence.} \end{cases}$ (how data were collected)

- Wilcoxon tests (signed ranks)

$W = \sum_i z_i R_i$ z, score

For ~~two~~ paired samples not normal, $\boxed{H_0: \text{median } m = m_0}$

medians

= ANOVA . x_{ij} and $\bar{x}_{grand}$

SST - total ~~square~~ sum of squares

SSB - between groups $\bar{x}_{group} - \bar{x}_{grand}$

SSE - within group $x_{ij} - \bar{x}_{group}$

$$F = \frac{SST/(m-1)}{SSE/(n-m)} \sim F(m-1, n-m)$$

- order statistics

$$P(Y_1 < m < Y_5) = \sum_{k=1}^{4} P(W=k)$$

$$P(W=k) = C_n^k (0.5)^k (0.5)^{n-k}$$

- Run test

• test if two samples follow the same distribution

• test randomness

- K-S test (goodness of fit)

$F_n(x) \sim F_0(x)$

- Sample size

$$n = \frac{2\sigma^2 (z_\beta + z_{\alpha/2})}{diff^2}$$

- sufficient statistics

$f(x_1, x_2 \cdots x_n | Y)$ does not depend on θ.

$\cdot f(x_1, x_2 \cdots x_n | Y) = \phi[u(x_1, x_2 \cdots x_n); \theta] * h[x_1, x_2 \cdots x_n]$

or $f(x;\theta) = \exp[k(x)p(\theta) + S(x) + q(\theta)]$.