关系： μ，σ^2，$\frac{\sigma_1^2}{\sigma_2^2}$，线性回归

$\{x_i\}$

$\bar{X} = \frac{\sum x_i}{n}$ 样本均值

$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ 样本方差

Recall:

$z = \frac{x-\mu}{\sigma}$

$z^2 \sim \chi^2(1)$

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

样本均值

$$\bar{x} = \frac{\sum x_i}{n}$$

如果 $\{X_i\}$ iid $N(\mu, \sigma^2)$

$$E[\bar{x}] = \frac{\sum x_i}{n} = \mu$$

$$var(\bar{x}) = \frac{1}{n^2} var(\sum X_i) = \frac{1}{n^2} \sum var(X_i) = \frac{\sigma^2}{n}$$

那么 $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

在特殊情况下，样本均值服从正态分布

$X \sim N(\mu_X, \sigma_X^2)$

$Y \sim N(\mu_Y, \sigma_Y^2)$

$X+Y \sim N$?

当 X 和 Y 独立时 X+Y 服从正态分布.

独立 v. 相关

① 独立 $P(X|Y) = P(X)$

已知 Y，对于知道 X 的信息没有帮助

② 相关： 线性相关性.

不相关时 有可能是有非线性关系

独立 → 不相关

不独立 $\not\to$ 相关

样本方差 服从卡方分布

$$S = \frac{1}{n-1}\sum(x_i-\bar{x})^2$$

Start with $Z = \dfrac{x_i-\mu}{\sigma}$

$$Z^2 \sim \chi^2(1)$$

$$\sum^n Z^2 \sim \chi^2(n)$$

$$\sum Z^2 = \sum\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$= \sum\frac{(x_i-\bar{x}+\bar{x}-\mu)^2}{\sigma^2}$$

$$= \frac{\sum(x_i-\bar{x})^2+\sum(\bar{x}-\mu)^2+\underbrace{2\sum(x_i-\bar{x})(\bar{x}-\mu)}}{\sigma^2}$$

$$= \frac{\sum(x_i-\bar{x})^2}{\sigma^2} + n\left(\frac{\bar{x}-\mu}{\sigma}\right)^2$$

$$2(\bar{x}-\mu)\sum(x_i-\bar{x})$$

$$= 2(\bar{x}-\mu)\left(\sum x_i - n\cdot\bar{x}\right)$$

$$= 0$$

Recall: $S^2 = \dfrac{\sum(x_i-\bar{x})^2}{n-1}$

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\Sigma Z^2 = \frac{(n-1)\frac{\Sigma(X_i-\bar{X})^2}{n-1}}{\sigma^2} + \left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)^2$$

$$\underset{\chi^2(n)}{\underset{\wr}{\Sigma Z^2}} = \frac{(n-1)S^2}{\sigma^2} + \underset{\chi^2(1)}{\underset{\wr}{Z^2}}$$

$$\therefore \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

样本方差　除以 $n-1$

$$S^2 = \frac{1}{n-1}\sum (x_i-\bar{x})^2$$

除以 $n-1$ 可以得出 unbiased estimate.

证明: $E[S^2] = \sigma^2$

Recall: $x_i \sim N(\mu, \sigma^2)$

$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

$E[(x_i-\mu)^2] = \sigma^2$

$E[(\bar{x}-\mu)^2] = \frac{\sigma^2}{n}$

$$S^2 = \frac{1}{n-1}\sum (x_i-\mu+\mu-\bar{x})^2$$

$$= \frac{1}{n-1}\sum [\underbrace{(x_i-\mu)^2}_{\sigma^2} - \underbrace{(\bar{x}-\mu)^2}_{\frac{\sigma^2}{n}} - \underbrace{2(x_i-\mu)(\bar{x}-\mu)}_{0}]$$

求期望

$$E[S^2] = \frac{1}{n-1}\sum (\sigma^2 - \frac{\sigma^2}{n})$$

$$= \frac{n}{n-1}(\sigma^2 - \frac{\sigma^2}{n})$$

$$= \sigma^2$$

检验 μ

$\{x_i\}$, $x_i \sim N(\mu, \sigma^2)$

$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

$$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

如果已知总体方差 σ.

$\bar{x}$, σ, n 已知

检验 μ.

例: $n=126$, $\bar{x}=29.2$, $\sigma=7.5$, $\mu=18.2$.

问给定样本情况，
能否说 $\bar{x}$ 与 μ 有足够大的差别

如果不知总体方差σ^2，
用S^2近似

$\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ 服从什么分布？

Recall: $T = \dfrac{Z}{\sqrt{\frac{V}{r}}}$

$Z = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

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

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{\bar{X}-\mu}{\frac{1}{\sqrt{n}}}}{S} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\frac{S}{\sigma}}$$

$$= \frac{Z}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{Z}{\sqrt{\frac{(n-1)\frac{S^2}{\sigma^2}}{n-1}}} = \frac{Z}{\sqrt{\frac{\chi^2(n-1)}{n-1}}}$$

$\sim T(n-1)$

线性回归 t检验

$$y = a + bx + \varepsilon$$

$$b = \frac{\mathrm{cov}(x,y)}{\mathrm{Var}(x)} = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum (x_i-\bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

稍微改写下

$$\begin{aligned} y &= a + bx + \varepsilon \\ &= \bar{y} - b\bar{x} + bx + \varepsilon \\ &= \bar{y} + b(x-\bar{x}) + \varepsilon \end{aligned}$$

即 x 居了个中.

$$a = \bar{y}$$

$$b = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum (x_i-\bar{x})^2}$$

注意: X 是已知的 $\varepsilon \sim N(0, \sigma^2)$

$$y \sim N(a + b(x-\bar{x}), \sigma^2)$$

线性回归 t 检验：截距

$a = \bar{y}$ 样本均值

$y_i \sim N(a + b(x_i - \bar{x}), \sigma^2)$

$a \sim N(\alpha, \frac{\sigma^2}{n})$

斜率

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$var(b) = \left(\frac{1}{\sum (x_i - \bar{x})^2}\right)^2 var\left(\sum (x_i - \bar{x})(y_i - \bar{y})\right)$$

$$= \left(\frac{1}{\sum (x_i - \bar{x})^2}\right)^2 \cdot \sum \left((x_i - \bar{x})^2 var(y_i)\right)$$

$$= \sigma^2 \cdot \frac{1}{\sum (x_i - \bar{x})^2} \cdot \frac{1}{\sum (x_i - \bar{x})^2} \cdot \sum (x_i - \bar{x})^2$$

$$= \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$