One population mean

One population proportion (Binary)

Two population means { Independence { Equal Variance / Unequal Variance ; Matched Pairs }

χ^2 - test.

χ^2 - goodness - of - fit test

One population proportion. p_1 = prob. improved

$p_A = 0.45$ $p_B = 0.40$ $p_0 = 0.15$

H_0: $p_A = 0.45$ $p_B = 0.4$ $p_0 = 0.15$

H_a: At least one proportion doesn't equal to the null value.

χ^2 - goodness - of - fit test

$n \to$ sample size

H_0 Prob.	Expected	Observed
0.45	$e_A = 90$	102
0.40	$e_B = 80$	82
0.15	$e_0 = 30$	16

$$e_i = n \cdot p_{i0}$$

$$\chi^2 = \sum^{n} \frac{(f_i - e_i)^2}{e_i}$$

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i}$$

Asymptotically under the null.

$$\chi^2 \sim \chi^2_{k-1}$$ k is the number of category

$$\text{P-value} = P(\chi^2_{k-1} > \chi^2)$$

$\chi^2 = 8.18$

$\text{P-value} = P(\chi^2_2 > 8.18) = 0.0167$

If $k=2$, χ^2 goodness-of-test is equivalent to the Z-test for one population proportion

6.5.2 Contingency Table

Subject, Two variables. Relationship

Voters
- Candidate $\rightarrow$ X
- gender $\rightarrow$ Y
- Age 20 – 100
- Education $\rightarrow$ Z

(1) Are Y and X independent?

(2) Are Y and Z independent?

H_0: two variables are independent

H_a: Two variables are not independent

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

$$e_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$$

	G	B	
M	109	138	247
F	152	112	264
	261	250	511

$$P(M \mid G) = P(M)$$

$$\frac{x}{261} = \frac{P(M \cap G)}{P(G)} = \frac{247}{511} \Rightarrow x = \frac{261 \times 247}{511}$$

Under H_0, asymptotically,

$$\chi^2 \sim \chi^2_{(k_1-1)(k_2-1)}$$

GLM Generalized linear Model

QQ Plot Quantile-Quantile Plot.

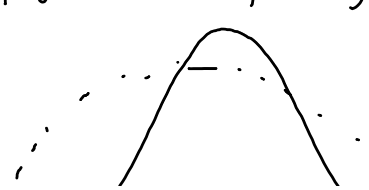
Model assumption $F(x)$

(i) Sort the observation $Y_1\ Y_2 \cdots Y_n$

$$\underset{\frac{1}{n}}{Y_{(1)}} \le \underset{\frac{2}{n}}{Y_{(2)}} \le \cdots \le \underset{\frac{3}{n}}{Y_{(n)}} \rightarrow \text{Empirical "Quantile"}$$

(ii) $q_i = F^{-1}\left(\frac{i-0.5}{n}\right)$ Theoretical Quantile.

(iv) Plot $y_{(i)}$ against the q_i



Kolmogorov - Smirnov Test (KS Test)

$\underline{H_0}$: $X \sim F_0(x)$

H_a: X doesn't follow distribution with $F_0(x)$

$$D_n = \sup_x \left| F_n(x) - F_0(x) \right|$$ D_n is the K-S test

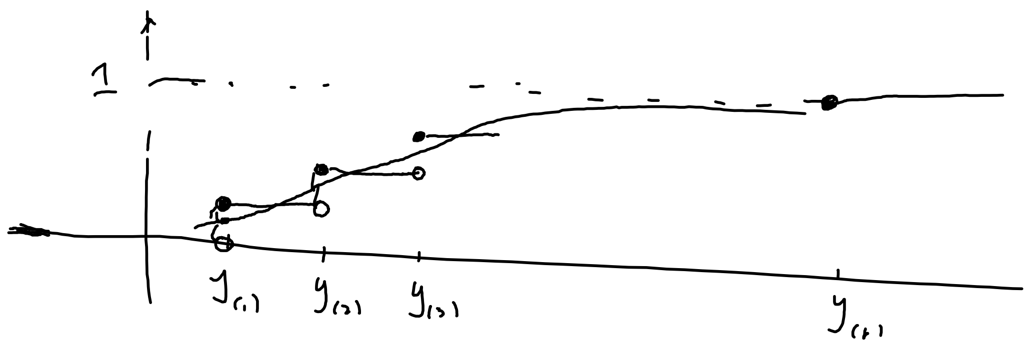
Under H_0, and $F_0(x)$ is continuous.

$$\sqrt{n} D_n \longmapsto K \quad \text{as} \quad n \to \infty$$

$$K = \sup_{t \in [0,1]} |B(t)|$$

Reject Region $= \{ \sqrt{n} D_n > k_\alpha \}$

$$\text{P-value} = P(K > \sqrt{n} D_n)$$ Kolmim



~~$P(Y \le y)$~~ $F_n(y) = \frac{1}{n} \sum \mathbb{1}(Y_i \le y)$

Permutation Test

Two ~~sample~~ population mean problems

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S_{\text{unpooled}}}$$

If the data doesn't follow normal, what is the distribution of T under the H_0?

Permutation Tests.

(i)

(ii)

(iii)

(iv) Rearrange the observations. compute the test statistic

(v) P-value $= 2P(\text{null distribution} > |T|)$

$$\approx 2 \cdot \frac{\Sigma \Delta (T_i > |T|)}{\text{number of permutations}}$$