lecture5: sample proportion

13 Mart 2025 Perşembe 10:48

Sampling Distributions of Sample Proportions

P = the proportion of the population having some characteristic

- Sample proportion ($\hat{p}$) provides an estimate of P:

$$\hat{p} = \frac{X}{n} = \frac{\text{number of items in the sample having the characteristic of interest}}{\text{sample size}}$$

- $0 \le \hat{p} \le 1$
- $\hat{p}$ has a binomial distribution, but can be approximated by a normal distribution when $nP(1 - P) > 5$

→discrete

↳ higher sample size makes it closer to a normal distribution.

$\hat{p} = \frac{X}{n}$, $E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{1}{n}E[X] = \frac{1}{n}nP = P \Rightarrow \boxed{E(\hat{p}) = P}$

$\sigma^2_{\hat{p}} = var(\hat{p}) = var\left[\frac{X}{n}\right] = \frac{1}{n^2}var(X) = \frac{1}{n^2}nP(1-P)$

$= \frac{P(1-P)}{n} \Rightarrow \boxed{\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}}$

$$\boxed{Z = \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}}}$$

$nP(1-P) > 5 \rightarrow \hat{p} \sim \text{normal}$

$\mu_{\hat{p}} = P$, $\sigma_{\hat{p}} = \sqrt{\frac{P(1-P)}{n}}$

Example

- If the true proportion of voters who support Proposition A is P = .4, what is the probability that a sample of size 200 yields a sample proportion between .40 and .45?
- i.e.: if P = .4 and n = 200, what is P(.40 ≤ $\hat{p}$ ≤ .45) ?

$n = 200$, $P = 0.4$ } $200(0.4)(0.6) = 48 > 5$ ✓ normal

$P(0.4 < \hat{p} < 0.45) = ?$

convert to z:

$$P\left(\frac{0.4 - 0.4}{\sqrt{\frac{0.4(0.6)}{200}}} < \frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}} < \frac{0.45 - 0.4}{\sqrt{\frac{0.4(0.6)}{200}}}\right)$$

$=$

$P(0 < z < 1.47) = F(1.47) - F(0)$

$= 0.929219 - 0.5 \simeq 0.43$

Example

According to the US Census Bureau's American Community Survey, 87%, percent of Americans over the age of 25 have earned a high school diploma. Suppose we are going to take a random sample of 200 Americans in this age group and calculate what proportion of the sample has a high school diploma.

What is the probability that the proportion of people in the sample with a high school diploma is less than 85 percent?

$P = 0.87$, $n = 200$ } $P(\hat{p} < 0.85) = ?$

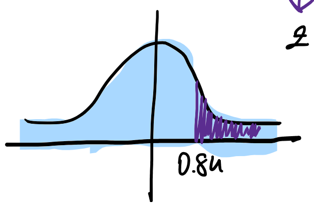
$\hat{p} \sim N\left(P, \frac{P(1-P)}{n}\right)$

$E(\hat{p}) = \mu_{\hat{p}}$ $var(\hat{p}) = \sigma^2_{\hat{p}}$

$\frac{\hat{p} - P}{\sqrt{\frac{P(1-P)}{n}}}$

$$P\left(z < \frac{0.85 - 0.87}{\sqrt{\frac{0.87(0.13)}{200}}}\right) = P(z < -0.84) = P(z > 0.84)$$

z score must be expressed by "less than" a positive number.



whole area − 0.84 $= 1 - P(z < 0.84)$

$= 1 - F(0.84)$

$= 1 - 0.7996$

$E(\hat{p}_x - \hat{p}_y) = E(\hat{p}_x) - E(\hat{p}_y)$

$= P_x - P_y$

$var(\hat{p}_x - \hat{p}_y) = \sigma^2_{\hat{p}_x - \hat{p}_y} = var(\hat{p}_x) + var(\hat{p}_y)$

$= \frac{P_x(1-P_x)}{n_x} + \frac{P_y(1-P_y)}{n_y}$

$\sigma_{\hat{p}_x - \hat{p}_y} = \sqrt{\frac{P_x(1-P_x)}{n_x} + \frac{P_y(1-P_y)}{n_y}}$

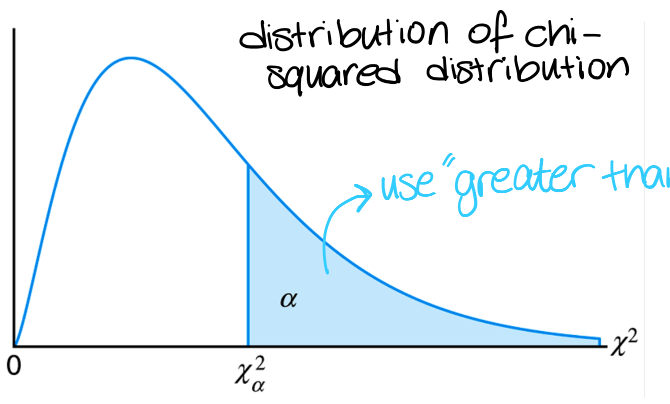
$$\boxed{z = \frac{(\hat{p}_x - \hat{p}_y) - (P_x - P_y)}{\sqrt{\frac{P_x(1-P_x)}{n_x} + \frac{P_y(1-P_y)}{n_y}}}}$$

If S^2 is the variance of a random sample of size n taken from a normal population having the variance σ^2, then the statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}$$

has a chi-squared distribution with $v = n - 1$ degrees of freedom.

→



v	0.90	0.95	0.975	0.99	0.999
1	0.016	0.004	0.001	0.000	0.000
2	0.211	0.103	0.051	0.020	0.002
3	0.584	0.352	0.216	0.115	0.024
4	1.064	0.711	0.484	0.297	0.091
5	1.610	1.145	0.831	0.554	0.210
6	2.204	1.635	1.237	0.872	0.381
7	2.833	2.167	1.690	1.239	0.598
8	3.490	2.733	2.180	1.646	0.857
9	4.168	3.325	2.700	2.088	1.152
10	4.865	3.940	3.247	2.558	1.479
11	5.578	4.575	3.816	3.053	1.834
12	6.304	5.226	4.404	3.571	2.214
13	7.042	5.892	5.009	4.107	2.617
14	7.790	6.571	5.629	4.660	3.041
15	8.547	7.261	6.262	5.229	3.483
16	9.312	7.962	6.908	5.812	3.942
17	10.085	8.672	7.564	6.408	4.416
18	10.865	9.390	8.231	7.015	4.905

how to read chi-squared chart?

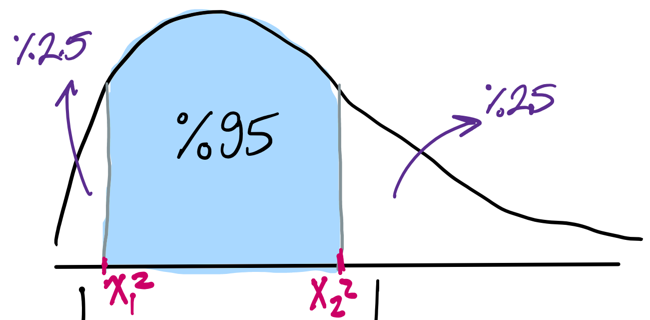
example:

$P(\chi^2 < 6.908) = ?$ $(n = 17)$

$v = n - 1 = 16$

$P(\chi^2 > 6.908) = 0.975$

$P(\chi^2 < 6.908) =$



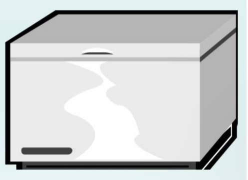
assuming they are symmetrical

$\chi^2_1 = \chi^2_{0.975, n-1}$

$\chi^2_2 = \chi^2_{0.025, n-1}$

Chi-square Example

- A commercial freezer must hold a selected temperature with little variation. Specifications call for a standard deviation of no more than 4 degrees (a variance of 16 degrees[2]).
- A sample of 14 freezers is to be tested
- What is the upper limit (K) for the sample variance such that the probability of exceeding this limit, given that the population standard deviation is 4, is less than 0.05?



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

$n = 14$, $v = 13$

$\sigma^2 = 16$

$P(S^2 > K) < 0.05$

convert $S^2 \rightarrow \chi^2$

$P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)K^2}{\sigma^2}\right) < 0.05$

$= P\left(\chi^2 > \frac{(n-1)K^2}{\sigma^2}\right) \le 0.05$, $n - 1 = 13$

v	0.10	0.05	0.025	0.01	0.001
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.124
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264
12	18.549	21.026	23.337	26.217	32.909
13	19.812	22.362	24.736	27.688	34.528
14	21.064	23.685	26.119	29.141	36.123

$\chi^2_{0.05} = \frac{(n-1)K}{\sigma^2}$

$22.362 = \frac{13K}{16} \Rightarrow K = 27.52$

Example

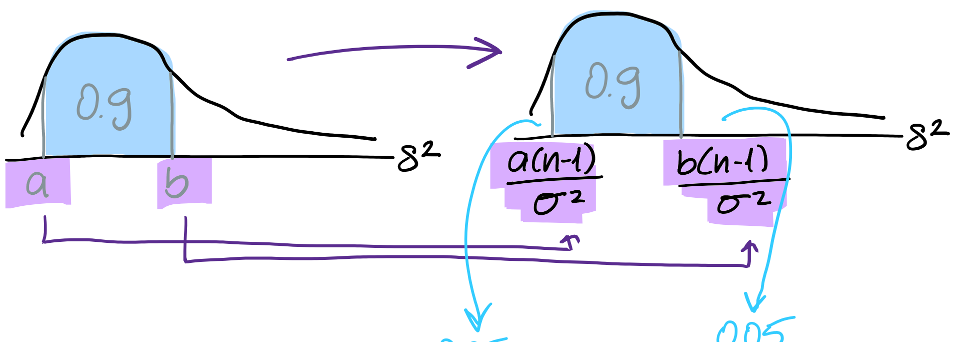
A particular type of vacuum-packed coffee packet contains an average of 16 oz. It has been observed that the number of ounces of coffee in these packets is normally distributed with σ = 1.41 oz. A random sample of 15 of these coffee packets is selected, and the observations are used to calculate s. Find the numbers a and b such that $P(a \le S^2 \le b) = 0.90$.

$n = 15$, $\sigma = 1.41$, $\sigma^2 = 1.988$

$P(a \le S^2 \le b) = 0.9$

$P\left(\frac{a(n-1)}{\sigma^2} < \chi^2 < \frac{b(n-1)}{\sigma^2}\right) = 0.9$

$= P\left(\frac{14a}{1.988} < \chi^2 < \frac{14b}{1.988}\right) = 0.9$



found from the table {

$\chi^2_{0.95} = \frac{14a}{1.988}$

$= 6.571$

$a = 0.93$

$\chi^2_{0.05} = \frac{14b}{1.988}$

$= 23.685$

$b = 3.36$