Test for normal variance, $H_0 : \sigma^2 = \sigma_0^2$

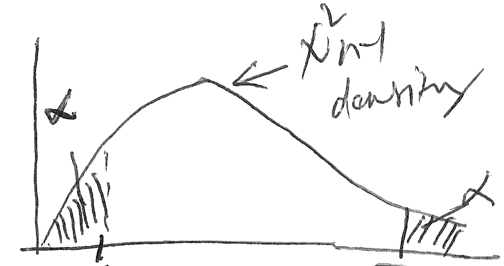
Data: $X_1, \ldots, X_n$

$X \sim N[\mu, \sigma^2]$ both unknown.

Test statistic:

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

$T = \frac{(n-1) s^2}{\sigma_0^2} \sim \chi^2_{n-1}$ when H_0 is true.



Critical point for the level α test:

One-sided alternative: $H_1: \sigma^2 > \sigma_0^2$. Reject when $T > \chi^2_{n-1,\alpha}$.

$\chi^2_{n-1,\alpha}$ = $(1-\alpha)$-th percentile of χ^2_{n-1} = qchisq$(1-\alpha, n-1)$.

$H_1: \sigma^2 < \sigma_0^2$ - Reject when $T < \chi^2_{n-1, 1-\alpha}$

$\chi^2_{n-1, 1-\alpha}$: α-th percentile of χ^2_{n-1} dist.

Two-sided alternative:

$H_1: \sigma^2 \neq \sigma_0^2$: Reject when $T < \chi^2_{n-1, 1-\alpha/2}$ or $T > \chi^2_{n-1,\alpha/2}$

- **Chi-square test**

p-value:

already saw.

H_1	reject when	p-value	computing p-value
$\sigma^2 > \sigma_0^2$		$P[\chi^2_{n-1} \geq T_{obs}]$	1 - pchisq(T_{obs}, n-1)
$\sigma^2 < \sigma_0^2$		$P[\chi^2_{n-1} \leq T_{obs}]$	pchisq(T_{obs}, n-1)
$\sigma^2 \neq \sigma_0^2$		$2 \min \{P[\chi^2_{n-1} \geq T_{obs}], P[\chi^2_{n-1} \leq T_{obs}]\}$	

Testing Hypotheses on ratio of two normal variances.

Set up: $X \sim N[\mu_X, \sigma_X^2]$ $\quad Y \sim N[\mu_Y, \sigma_Y^2]$

Data: $X_1, \ldots, X_m$ $\quad$ Data: $Y_1, \ldots, Y_n$

mutually indep.

$H_0: \sigma_X^2 = \sigma_Y^2$ $\quad H_1$: one of the three possibilities.

Test statistic:

Note: $$\frac{\frac{\cancel{(m-1)} S_X^2}{\sigma_X^2} \cdot \frac{1}{\cancel{m-1}}}{\frac{\cancel{(n-1)} S_Y^2}{\sigma_Y^2} \cdot \frac{1}{\cancel{n-1}}} \sim F_{m-1,\,n-1}.$$

correct pivot for CI for $\frac{\sigma_X^2}{\sigma_Y^2}$.

F-test

$$T = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \underset{H_0}{=} \frac{S_X^2}{S_Y^2} \sim F_{m-1,\,n-1} \text{ when } H_0 \text{ is true.}$$

$\rightarrow$ ~~Proc~~ Proceed as in χ^2-test to compute critical point and p-value.

Chi-square tests [Chapter 10].

Chi-square Goodness of Fit Test

Set up: Count data on a **categorical** (or qualitative.) **variable**. — Ex: Gender, Race, majors, etc. (Assume/ignore the ordering b/w categories, if it exists).

- There are k categories, labeled as $i = 1, \ldots, k$.
- Observed data: O_i, $i = 1, \ldots, k$, where $O_i = \#$ of observations in the i-th category. (observed frequencies)
- $n = \sum_{i=1}^{k} O_i$ is the total number of observations.

Hypotheses: H_0 : The data follow a given model, versus, H_1 : The data don't follow the given model.

- Let p_i = proportion of observations in the population that fall in the i-th category, $i = 1, \ldots, k$. We can also think of p_i as the probability that a randomly selected observation from the population falls in the i-th category.
- $H_0 : p_i = p_{i,0}$ (known), $i = 1, \ldots, k$, where $p_{i,0}$ are known proportions that add up to 1. [i.e., all the cell prob. are known.]
- Model is **completely known** under H_0.

Case 1: Two categories : p_1, p_2 , $p_1+p_2=1$, i.e., $p_2 = 1-p_1$

$H_0: p_1 = p_{1,0},\ p_2 = p_{2,0} = 1 - p_{1,0}$. Nothing new here as we can label one category as "success" and the other as "failure", and use test for proportions.

Case 2: More than two categories

Basic idea: Compare O_i's with E_i's — counts expected assuming H_0 is true — using a **chi-square statistic**

$$\chi^2 = \sum_{i=1}^{k} (O_i - E_i)^2 / E_i$$

- Large χ^2 : Large diff. in obs. and exp. counts.
- Reject H_0 when χ^2 is large.
- $E_i = n p_{i,0}$, $i = 1, 2, \ldots, k$; $\sum_{i=1}^{k} E_i = n$
- Null distribution: Approx. χ^2_{k-1} dist. if n is large.
- Rule of thumb: All $E_i \geq 5$. Collapse adjacent categories if this is not the case.

Ex: Suppose 60 independent rolls of a die lead to the following data.

category	1	2	3	4	5	6	total
observed count (O_i)	4	6	17	16	8	9	60
expected count (E_i)	10	10	10	10	10	10	60

Is the die fair? Answer this question by performing an appropriate test of hypothesis at 5% level of significance.

H_0: Die is fair, $p_{i,0} = \frac{1}{6}$, $i = 1, 2, \ldots, 6$.

H_1: Die is NOT fair, i.e., at least one prob. is $\neq \frac{1}{6}$.

$E_i = n\, p_{i,0} = (60)\left(\frac{1}{6}\right) = 10.$

$\chi^2_{obs} = \frac{(4-10)^2}{10} + \frac{(6-10)^2}{10} + \cdots + \frac{(9-10)^2}{10} = 14.2$

p-value $= P\left[\chi^2_{K-1} \geq \chi^2_{obs}\right] = 1 - \text{pchisq}(\chi^2_{obs}, K-1) = 0.014$

Conclusion: Reject H_0 since p-value $< \alpha = 0.05$. Die is NOT fair.

R code:

```
> x <- c(4, 6, 17, 16, 8, 9)
> sum(x)
[1] 60
> sum( (x-10)^2/10)
[1] 14.2 → χ²_obs
> 1-pchisq(14.2, 5)
[1] 0.01438768    ↑ K-1
>
```