Corr 5.5.4c: If $A \xrightarrow{p} a$, $B \xrightarrow{p} b$ then $cA + dB \xrightarrow{p} ca + db$ for constants a, b, c, d.

An example 2-prop z-test. $n_1 = 81$, $\sum X_{1i} = 27$, $n_2 = 79$, $\sum X_{2i} = 12$.

$H_0: \theta_1 = \theta_2$, $\alpha = 5\% \Rightarrow z_{97.5\%} = 1.96$

$$\left(\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2\right)_{STD} = \frac{\hat{\hat{\theta}}_1 - \hat{\hat{\theta}}_2}{\sqrt{\frac{\sum x_i + \sum x_{2i}}{n_1 + n_2}\left(1 - \frac{\sum x_i + \sum x_{2i}}{n_1 + n_2}\right)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{\frac{27}{81} - \frac{12}{79}}{\sqrt{\frac{27+12}{81+79}\left(1 - \frac{27+12}{81+79}\right)}\sqrt{\frac{1}{81} + \frac{1}{79}}}$$

$$= 2.66 \notin [-1.96, +1.96] \Rightarrow \text{Reject } H_0.$$

Let $X_1, \ldots, X_n \overset{iid}{\sim}$ DGP with mean θ and variance σ². We want to test theories of θ and σ² is a nuisance parameter.

$\hat{\theta} = \bar{X}$ unbiased, consistent

by CLT, $\frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1)$ but σ is unknown

If the DGP was iid normal, then we know the 1-sample t test:

$$\frac{\bar{X} - \theta}{\frac{S}{\sqrt{n}}} \sim T_{n-1}$$

But we don't know the DGP here. So what do we do? Use Wald test:

$$\frac{\bar{X} - \theta}{\frac{S}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

since $S \xrightarrow{p} \sigma$, we show this from Slutsky's thm.

Then you have an approx z-test!

Let $X_{11}, \ldots, X_{1n_1} \overset{iid}{\sim}$ DGP with mean θ_1 and variance σ_1^2 independent of
$X_{21}, \ldots, X_{2n_2} \overset{iid}{\sim}$ DGP with mean θ_2 and variance σ_2^2

And you wish to test differences in the two populations' means.
We use the Wald Test again:

$$\frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \xrightarrow{d} N(0,1)$$

by corr 5.5.4c and Slutsky's and the two-sample CLT since $S_1 \to \sigma_1$ and $S_2 \to \sigma_2$

In a basic stats class, they teach you to just use the T distribution here with Welch's T test. It's technically wrong. But some people still recommend this as it is more conservative than the Z distribution. It makes sense to be conservative since the CLT + Slutsky's thm may be slow to reach N(0,1).

Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Gumbel(θ, 1) := $e^{-\left((x-\theta) + e^{-(x-\theta)}\right)}$

The Gumbel model is a good model for "extreme" events e.g. wind speed in hurricanes, damage due to flooding, etc. We want to test against values of θ. At $\alpha = 5\%$ ~~test~~ $H_a: \theta > 2$.

We need a test statistic to measure departure from the null. We'll use the best estimator for θ we can find. Default: MLE if you can derive it. Then you get its asymptotic distribution from the monster thm:

$$\frac{\hat{\theta}^{MLE} - \theta}{\sqrt{\frac{I(\hat{\theta}^{MLE})^{-1}}{n}}} \xrightarrow{d} N(0,1)$$

Let's derive I(θ) first:

$$\ell(\theta; x) = -\left((X-\theta) + e^{-(X-\theta)}\right) = \theta - X - e^{-X}e^{\theta}$$
$$\ell'(\theta; x) = 1 - e^{-X}e^{\theta}$$
$$\ell''(\theta; x) = -e^{-X}e^{\theta} \Rightarrow -\ell''(\theta; x) = e^{-X}e^{\theta}$$
$$I(\theta) = E[-\ell''(\theta; x)] = E[e^{-X}e^{\theta}] = e^{\theta}E[e^{-X}] = e^{\theta}e^{-\theta} = 1$$

$$E[e^{-X}] = \int_{\mathbb{R}} e^{-x} f_X(x)\,dx \overset{\text{calculus}}{=} \ldots\ldots = e^{-\theta}$$

$$\hat{\theta}^{MLE}: \ell'(\theta; X_1, \ldots X_n) \overset{set}{=} 0$$
$$\Rightarrow 0 = \sum_{i=1}^{n} \ell'(\theta; X_i) = \sum_{i=1}^{n} 1 - e^{-X}e^{\theta} = n - e^{\theta}\sum_{i=1}^{n} e^{-X_i}$$
$$\Rightarrow e^{\theta}\sum e^{-X_i} = n \Rightarrow e^{\theta} = \frac{n}{\sum e^{-X_i}} \Rightarrow \hat{\theta}^{MLE} = \ln\left(\frac{n}{\sum e^{-X_i}}\right)$$
$$\Rightarrow \frac{\ln\left(\frac{n}{\sum e^{-X_i}}\right) - \theta}{\sqrt{\frac{1}{n}}} \xrightarrow{d} N(0,1)$$

Let's say your sample is: <2.15, 1.91, 3.66, 4.85, 3.03, 1.03, 3.58> which is n = 7. Run the test above. We calculate the z statistic and if it is > 1.645 => Reject the null.

$$z = \frac{2.26 - 2}{\sqrt{\frac{1}{7}}} = 0.688 \in (-\infty, 1.645] \Rightarrow \text{Retain } H_0.$$

The goals of statistic inference were threefold:

(1) Point estimation: provide the best guess $\hat{\hat{\theta}}$ for θ. Your guess is a realization from the estimator $\hat{\theta}$. So "good point estimates" really means that the estimator is good e.g. unbiased, consistent, low MSE (risk if you care about another loss function than L2).

(2) Testing: test a theory about θ. Our approach was "hypothesis testing". What makes a good test? (a) It is properly sized and (b) very powerful.

(3) Confidence sets: a set of values for θ that you are "confident about". Our approach here will be the "frequentist confidence interval" (CI).

Definition of an "interval estimate" is a lower and upper bound statistic:
$[w_L(x_1, \ldots, x_n), w_U(x_1, \ldots, x_n)]$

and the analogous "interval estimator" is:

$[w_L(X_1, \ldots X_n), w_U(X_1, \ldots, X_n)]$

which realizes random intervals e.g. [1.76, 2.48].

Def: the "coverage probability" of an interval estimator is:

$$P\left(\theta \in [w_L(X_1, \ldots, X_n), w_U(X_1, \ldots, X_n)] \mid \theta\right).$$