Victoria Lombardi

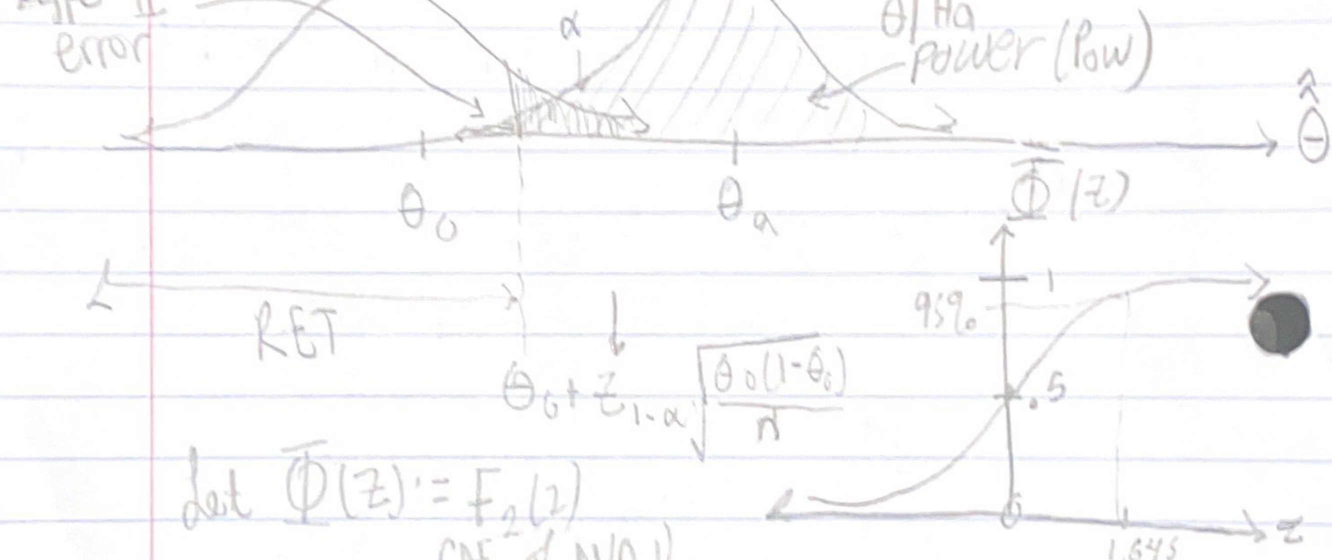
Math 369

Lecture 5

Lets look at power more generally (beyond 2 point hypothesis)

$H_0: \theta \leq \theta_0, \quad H_a: \theta = \theta_a > \theta_0$ at size α right tailed test

Under iid Bern(θ) and the normal approximation



Let $\Phi(z) := F_z(z)$
CDF of $N(0,1)$

$\alpha = 5\% \Rightarrow z_{1-\alpha} = 1.645$

$$\Phi(z_{1-\alpha}) = 1-\alpha$$

$$Pow = P\left(\hat{\theta} | H_a > \theta_0 + z_{1-\alpha}\sqrt{\frac{\theta_0(1-\theta_0)}{n}}\right)$$

$$= P\left(\frac{\hat{\theta} | H_a - \theta_a}{\sqrt{\frac{\theta_a(1-\theta_a)}{n}}} > \frac{\theta_0 + z_{1-\alpha}\sqrt{\frac{\theta_0(1-\theta_0)}{n}} - \theta_a}{\sqrt{\frac{\theta_a(1-\theta_a)}{n}}}\right)$$

$$= P\left(z > \frac{-\sqrt{n}(\theta_a - \theta_0) + z_{1-\alpha}\sqrt{\theta_0(1-\theta_0)}}{\sqrt{\theta_a(1-\theta_a)}}\right)$$

$$= 1-\Phi\left(\frac{-\sqrt{n}(\theta_a-\theta_0)+z_{1-\alpha}\sqrt{\theta_0(1-\theta_0)}}{\sqrt{\theta_a(1-\theta_a)}}\right)$$ Power function $= \text{Pow}(\theta_a, \theta_0, n, \alpha)$

(2)

Observations about the power function

If $n \to \infty \Rightarrow \text{Pow} \to 1$

As n increases

If $\theta_a \to \infty \Rightarrow \text{Pow} \to 1$

As $\alpha \to 0 \Rightarrow \text{Pow} \to 0$

RET

New type of survey. We ask "how tall are you (in inches)"? for men only. I'll ask 10 male students and get $x_1, \ldots, x_{10}$ (i.e. my data). The data is now continuous (no longer zeroes and ones). Height for a gender is known to be normally distributed

DGP: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$. Assume σ^2 is known and $= 4^2$

How can we estimate θ? θ is the mean of the r.v.s

Recall $\hat{\theta} = \bar{X}$ is unbiased. Lets use this estimator

$x_1 = (70, 72, 73, 68, 69, 70, 67, 72, 71, 73)$

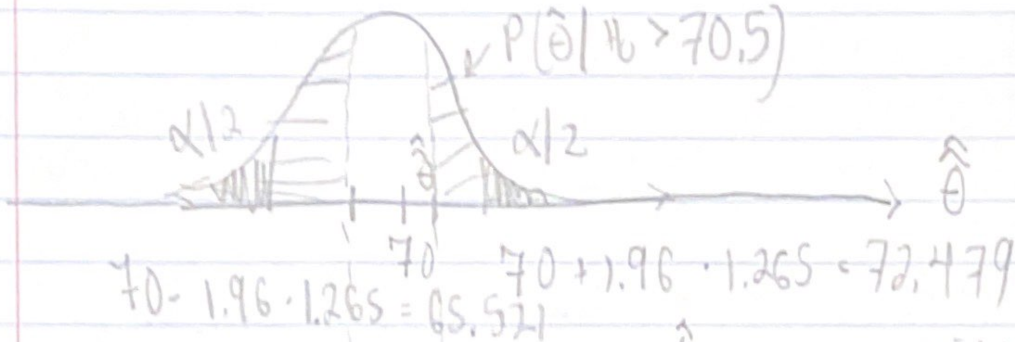
$\hat{\hat{\theta}} = \bar{x} = 70.5$

The American mean male adult height is 70"
Lets test if the mean of the population where this class is drawn from is different than 70". ③

one sample z-test

$H_a: \theta \neq 70, \quad H_0: \theta = 70$

$$\hat{\theta} \mid H_0 \sim N\left(70, \frac{4^2}{10}\right) = N(70, 1.265^2)$$



Ret

← RET →

$\Rightarrow \hat{\theta} \in RET \Rightarrow$ Fail to reject

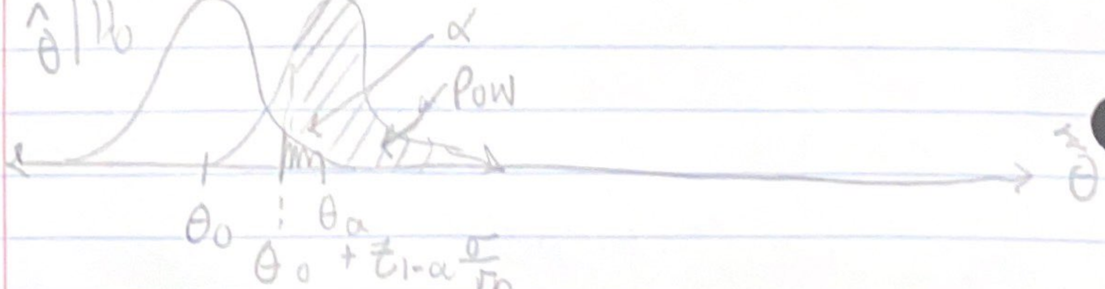
$P_{val} = P(\text{estimate is more extreme than observed} \mid H_0)$

$= P(|\hat{\theta} \mid H_0| > |\hat{\theta} - \theta|) = 2(P\,\hat{\theta} \mid H_0 > 70.5)$

$= 2P\left(Z > \frac{70.5 - 70}{1.265}\right) = 69.3\% \nless \alpha$

$\therefore$ statistically insignificant

$H_0: \theta \leq \theta_0, \quad H_a: \theta = \theta_a > \theta_0$, size α



$$Pow = P\left(\hat{\theta} \mid H_a > \theta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right)$$
$$= P\left(\frac{\hat{\theta} \mid H_a - \theta_a}{\frac{\sigma}{\sqrt{n}}} > \frac{\theta_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}} - \theta_a}{\frac{\sigma}{\sqrt{n}}}\right)$$
$$= 1 - \Phi\left(-\frac{\sqrt{n}}{\sigma}(\theta_a - \theta_0) + z_{1-\alpha}\right) = Pow(\theta_a, \theta_0, n, \alpha, \sigma)$$

More realistic: we don't know sigsq. But sigsq is a "nuisance parameter". It means we need to estimate it in order to estimate θ but we don't intrinsically care about it.

DGP: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ and both θ, σ^2 are unknown

How do we estimate sigsq? Recall... for a rv X

$$\sigma^2 := E\left[(X - \theta)^2\right] \qquad \theta = E[X],\ \hat{\theta} = \frac{1}{n}\sum x_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \theta)^2 \qquad \text{Problem: I need to know } \theta!$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \bar{X})^2 \text{ seems like a reasonable estimator}$$

Is this estimator unbiased? For any iid DGP....

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n}\sum (X_i - \bar{X})^2\right] = \frac{1}{n}\sum E(X_i - \bar{X})^2$$
$$\overset{iid}{=} \frac{1}{n} n E\left[(X_1 - \bar{X})^2\right] = E\left[X_1^2 - 2X_1\bar{X} + \bar{X}^2\right]$$
$$= E[X_1^2] - 2E\left[X_1 \cdot \frac{X_1 + \ldots + X_n}{n}\right] + E[\bar{X}^2]$$

Recall: $Var[X] = E[X^2] - E[X]^2$

$$= \sigma^2 + \theta^2 - \frac{2}{n}E\left[X_1^2 + X_1X_2 + \ldots + X_1X_n\right] + \frac{\sigma^2}{n} + \theta^2$$
$$= \frac{n+1}{n}\sigma^2 + 2\theta^2 - \frac{2}{n}\left(\sigma^2 + \underbrace{\theta^2 + \theta^2 + \ldots + \theta^2}_{n}\right) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

It's a little bit biased....

However, it is asymptotically unbiased" meaning...

$$\lim_{n\to\infty} E[\hat{\theta}] = \theta \quad \text{e.g. } \lim_{n\to\infty} E[\hat{\sigma}^2] = \lim \frac{n-1}{n}\sigma^2$$
$$= \sigma^2$$

Consider the following estimator.

$$S^2 := \frac{n}{n-1}\hat{\sigma}^2 = \frac{n}{n-1}\cdot\frac{1}{n}\Sigma(X_i - \bar{X})^2$$
$$= \frac{1}{n-1}\Sigma(X_i - \bar{X})^2$$

The beauty of this estimator is that

$$E[S^2] = E\left[\frac{n}{n-1}\hat{\sigma}^2\right] = \frac{n}{n-1}E[\hat{\sigma}^2]$$
$$= \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2 \quad \text{i.e. unbiased}$$

And its the default estimator for σ^2 (variances in DGP's) and its really important in normal theory