Statistical Methods III: Spring 2013

Jonathan Wand

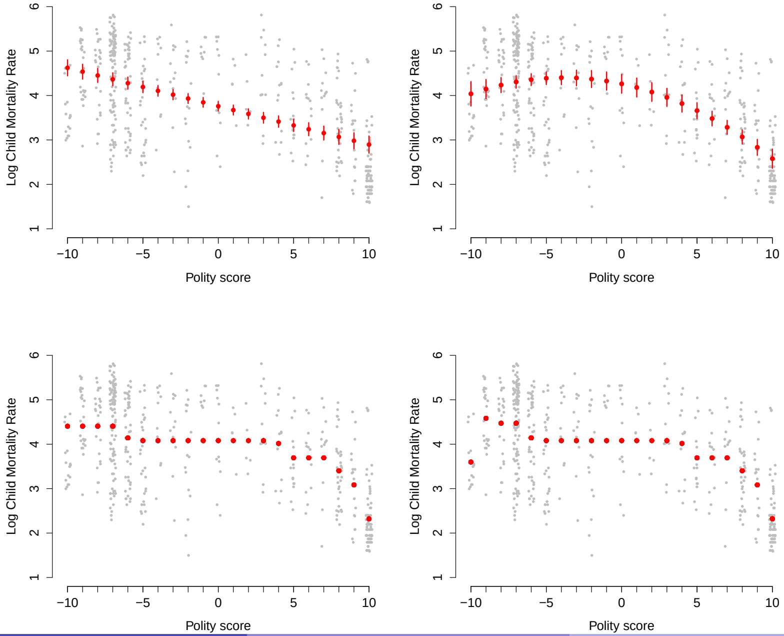
Stanford University

Ordered means + inference

Outline

1. Introduction
   - Ordered means
   - Regression/ML/Trinity of tests
2. Review
   - Normal, χ^2, t + 2-tail
3. Testing equality constraints

Polity scores and child mortality



Bartels (1996)

Table 1. Probit Parameter Estimates for Republican Vote Propensity, 1992

	Fully Informed Preferences	Uninformed Preferences	Information Effect (Difference)
Intercept	−1.542	−.348	−1.194
	(.766)	(1.112)	(1.673)
Age (years)	−.0435	.0000	−.0436
	(.0278)	(.0389)	(.0594)
Age squared (years)	.000429	−.000045	.000474
	(.000278)	(.00384)	(.000590)
Education (years)	.0962	.0017	.0945
	(.0337)	(.0536)	(.0779)
Income (percentile)	.399	.828	−.428
	(.329)	(.563)	(.802)
Black	−1.063	−2.285	1.222
	(.319)	(.479)	(.717)
Female	−.420	.326	−.746
	(.153)	(.269)	(.381)

Normal Distribution

If $X \sim N(\mu, \sigma^2)$, the pdf is

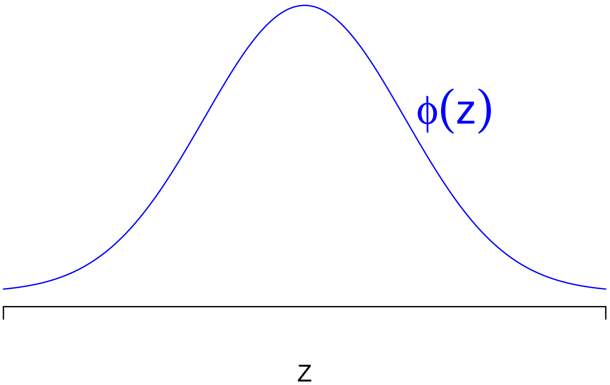
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Important: If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

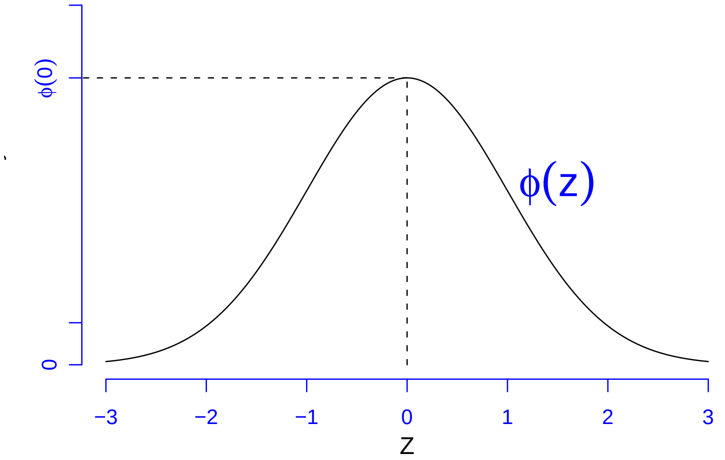
and conversely, we can start with $Z \sim N(0, 1)$,

$$X = Z\sigma + \mu \sim N(\mu, \sigma^2).$$



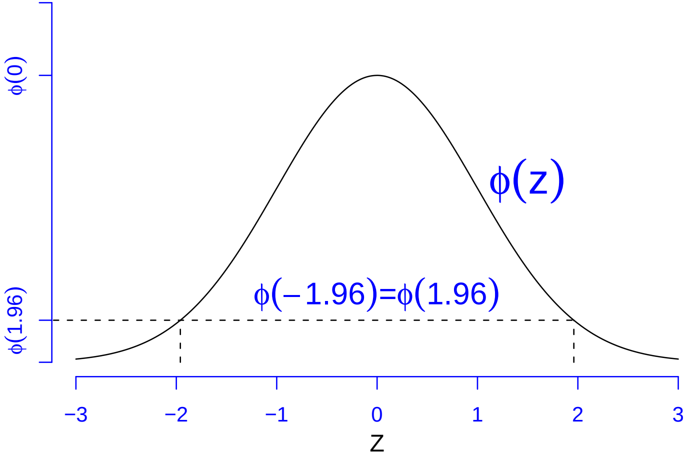
Let,

$$Z \sim N(0, 1)$$



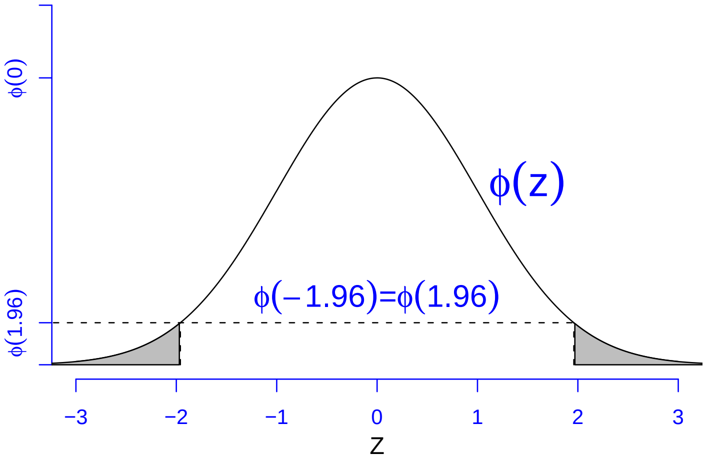
Let,

$$Z \sim N(0,1)$$



Let,

$$Z \sim N(0, 1)$$



Let,

$$Z \sim N(0,1)$$

Standard Normal Distribution

Memorize the following for quick calculations and seminars:

$$
\begin{aligned}
P(|Z| &> 1 &) = 2[1 - \Phi(& 1 &)] &\approx 0.32 \\
P(|Z| &> 1.5 &) = 2[1 - \Phi(& 1.5 &)] &\approx 0.14 \\
P(|Z| &> 1.96 &) = 2[1 - \Phi(& 1.96 &)] &\approx 0.05 \\
P(|Z| &> 2 &) = 2[1 - \Phi(& 2 &)] &\approx 0.044 \\
P(|Z| &> 2.5 &) = 2[1 - \Phi(& 2.5 &)] &\approx 0.01 \\
P(|Z| &> 3 &) = 2[1 - \Phi(& 3 &)] &\approx 0.003
\end{aligned}
$$

noting that the representation in second term is by symmetry, if $z > 0$ then

$$
\begin{aligned}
P(|Z| > z) &= \Phi(-z) + (1 - \Phi(z)) \\
&= (1 - \Phi(z)) + (1 - \Phi(z)) \\
&= 2(1 - \Phi(z))
\end{aligned}
$$

Useful facts: χ^2

CB Lemma 5.4.1 (simplified)

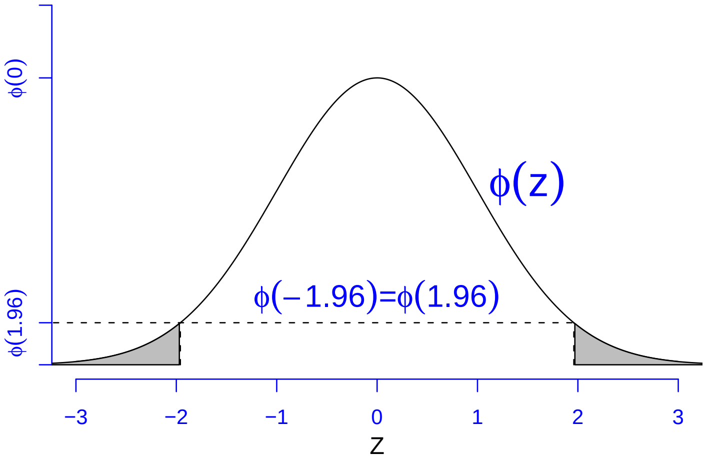
- If $Z \sim N(0,1)$ then $Z^2 \sim \chi^2_1$
- If $Z_1, \ldots, Z_n$ are iid, and $Z_i \sim \chi^2_1$ then $\sum Z_i \sim \chi^2_n$

Properties restated

- Sums of squared normals are distrib χ^2
- If $X \sim \chi^2_{k_1}$ and $Y \sim \chi^2_{k_2}$ then $X + Y \sim \chi^2_{k_1+k_2}$

Expectations

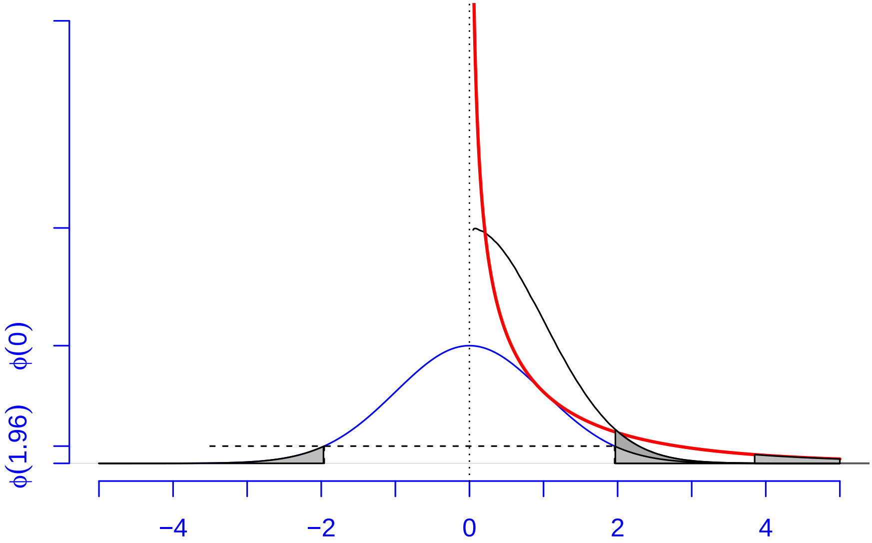
- $E\chi^2_k = k$
- $V\chi^2_k = 2k$

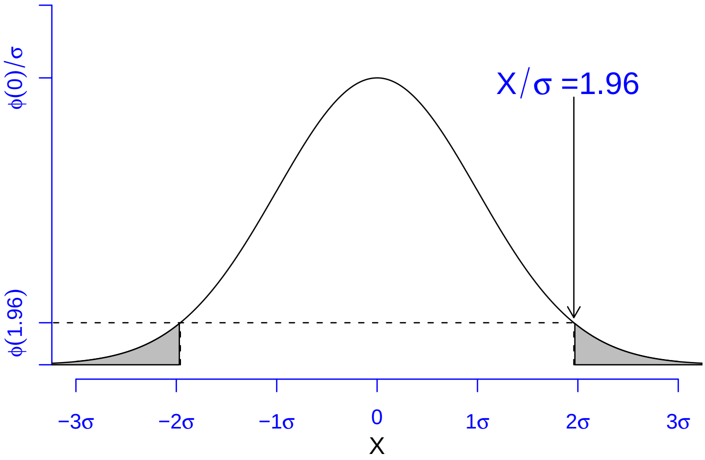


Let,

$$Z \sim N(0, 1)$$

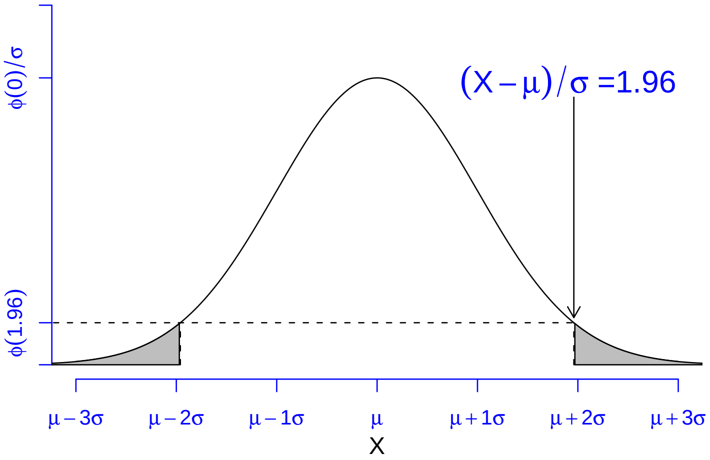
χ_1^2





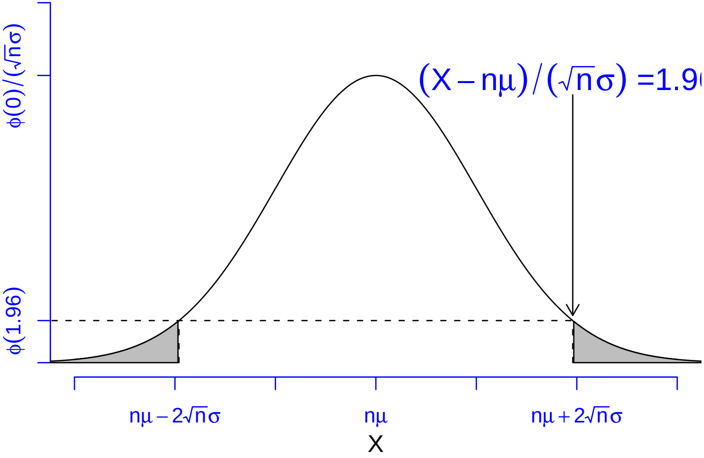
Let,

$$X = Z\sigma \quad \text{thus} \quad X \sim N(0, \sigma^2)$$



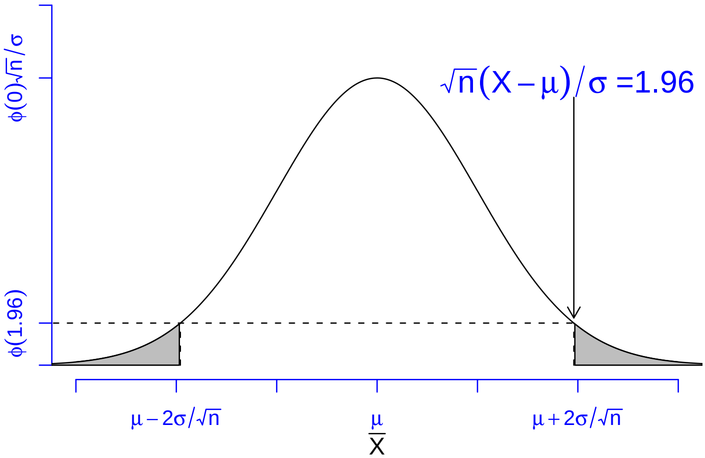
Let,

$$X = \mu + Z\sigma \quad \text{thus} \quad X \sim N(\mu, \sigma^2)$$



Let,

$$X = \sum_{i=1}^{n} (\mu + Z_i \sigma) \quad \text{thus} \quad X \sim N(n\mu, n\sigma^2)$$



Let,

$$\bar{X} = \frac{1}{n}\left(\sum_{i=1}^{n}(\mu + Z_i\sigma)\right) \quad \text{thus} \quad \bar{X} \sim N(\mu, \sigma^2/\sqrt{n})$$

Normal Distribution: critical values

Given $X_i \sim N(\mu, \sigma^2)$, and chosen n and α, and

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_A : \mu \neq \mu_0$$

We can often state test statistic in different, equivalent ways—choose to make your life easy:

$$\begin{aligned}
\alpha &= P(|\bar{X}_n - \mu_0| > c) \\
&= P(|\bar{X}_n - \mu_0|/(\sigma/\sqrt{n}) > c/(\sigma/\sqrt{n})) \\
&= P(|Z| > z) = \alpha
\end{aligned}$$

where $Z \sim N(0, 1)$; e.g., if $\alpha = .0455$, then $z = 2$

Normal Distribution: critical values (cont'd)

For $P(|Z| > z) = \alpha$,

- In terms of normed value, $C = \{(-\infty, -z) \cup (z, \infty)\}$
- If we observe $Z \in C$, or equivalently

  $$|\bar{X} - \mu_0|/(\sigma/\sqrt{n}) > z$$

  we would reject H_0.

For $P(|\bar{X}_n - \mu_0| > c) = \alpha$

- could also solve for $c = z\sigma/\sqrt{n}$
- In terms of deviation from mean

  $$C = (-\infty, -z\sigma/\sqrt{n}) \cup (z\sigma/\sqrt{n}, \infty)\}$$

- given $\alpha = .0455$

  $$C = (-\infty, -2\sigma/\sqrt{n}) \cup (2\sigma/\sqrt{n}, \infty)\}$$

Normal Distribution: critical values (example)

Let $X_i \sim N(\mu, \sigma^2)$, and $\sigma^2 = 9$

With a sample size $n = 16$, you observe $\bar{X} = 12$

Question: What decision would we make about $H_0 : \mu_0 = 10$?

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{12 - 10}{3/4} = 3/2$$

The observed test statistic is not in critical region region, $T \notin C$ where

$$C = (-\infty, -2) \cup (2, \infty)\}$$

so would fail to reject H_0 (or loosely, accept H_0)

Normal Distribution: critical values (cont'd)

Connection to confidence intervals

With respect to $P(|\bar{X}_n - \mu| > c) = \alpha$, we just solved for $c = z\sigma/\sqrt{n}$ given α

Using this c, we can construct an interval centered around $\bar{X}$,

$$(\bar{X}_n - c, \bar{X}_n + c) = (\bar{X}_n - z\sigma/\sqrt{n}, \bar{X}_n + z\sigma/\sqrt{n})$$

for specified values of z, σ and n

- this interval is refered to as a $(1 - \alpha) \times 100$ percent confidence interval
- $(1 - \alpha) \times 100$ percent of the time, this interval will include the true μ.

Normal Distribution: p-value

Let $X_i \sim N(\mu, \sigma^2)$, and $\sigma^2 = 9$

With a sample size $n = 16$, you observe $\bar{X} = 12$

Question: How confident are we that this sample could be produced with $\mu_0 = 10$?

Set $c = \bar{X} - \mu_0 = 12 - 10 = 2$

Significance level interpretation of α:

$$\begin{aligned}
\alpha &= P(|\bar{X}_n - \mu_0| > c) \\
&= P(|\bar{X}_n - \mu_0|/(\sigma/\sqrt{n}) > c/(\sigma/\sqrt{n})) \\
&= P(|\bar{X}_n - \mu_0|/(\sigma/\sqrt{n}) > 2/(3/4)) \\
&= P(|Z| > 3/2) \approx .14
\end{aligned}$$

Note: referred to as p-value

Useful facts: χ^2

CB Lemma 5.4.1 (simplified)

- If $Z \sim N(0,1)$ then $Z^2 \sim \chi^2_1$
- If $Z_1, \ldots, Z_n$ are iid, and $Z_i \sim \chi^2_1$ then $\sum Z_i \sim \chi^2_n$

Properties restated

- Sums of squared normals are distrib χ^2
- If $X \sim \chi^2_{k_1}$ and $Y \sim \chi^2_{k_2}$ then $X + Y \sim \chi^2_{k_1+k_2}$

Expectations

- $E\chi^2_k = k$
- $V\chi^2_k = 2k$

Hypothesis tests of r=Rank(R) restrictions

- Equality restriction versus unconstrained

  $H_0 : R\theta = 0$ versus $H_A : R\theta \neq 0$

Inference with two means

Let

$$\hat{\Delta} = (\hat{\Delta}_1, \hat{\Delta}_2) \sim N(\Delta, I_2)$$

where, possibly,

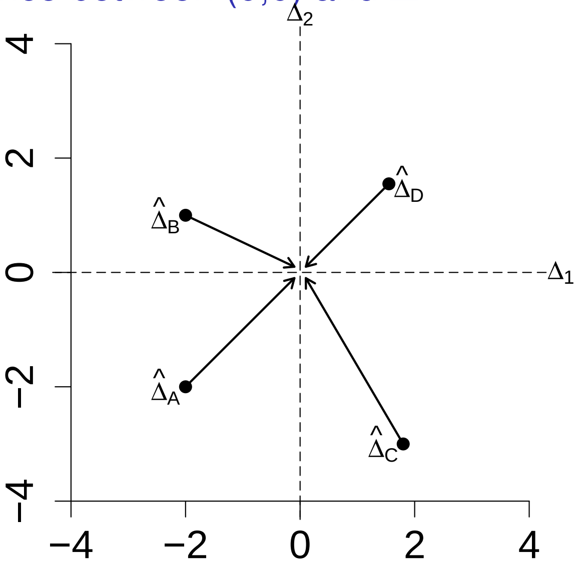
$$\hat{\Delta}_j = \hat{\mu}_j - \hat{\mu}_{j-1} \qquad j \in \{1, 2\}$$

and I_k is a $k \times k$ identity matrix.

Hypotheses/comparisons

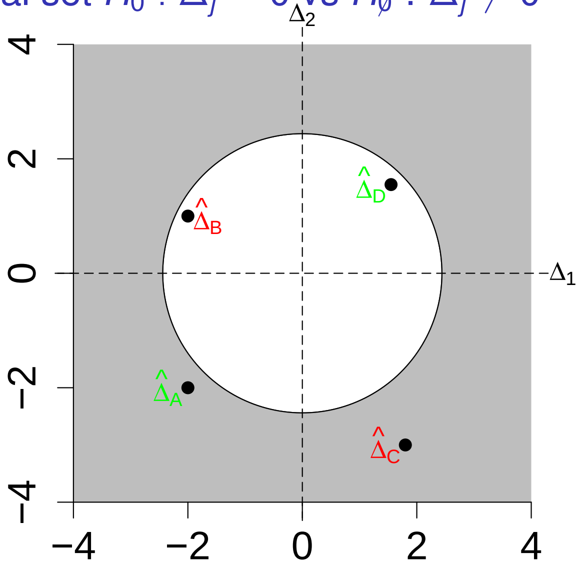
- $H_0 : \Delta = 0 \quad$ vs $\quad H_{\emptyset} : \Delta \neq 0$
- what does this imply about R,r?
- alternatives?

Distance between (0,0) and $\hat{\Delta}$



$\Delta_1^2 + \Delta_2^2 \sim \chi_2^2$

Critical set $H_0 : \Delta_j = 0$ vs $H_{\emptyset} : \Delta_j \neq 0$



Critical value $\chi^2_{2,\alpha=.05} = 6$, so $\sqrt{\Delta_1^2 + \Delta_2^2} = \sqrt{6} =$ radius