z-Test Discussion

C. Durso

z-Test

Set-up

- $x_1, x_2, \ldots x_n$ is a sample from $Normal(\mu, \sigma^2)$
- σ^2 is known
- Null hypothesis: $\mu = \mu_0$

Some Application Types

- Compare value estimated as mean of measurements with known Normal error to reference value.
- Give range (confidence interval) for value estimated by mean of measurements with known Normal error.
- Approximate large sample tests.

Distribution of $\bar{x}$

- $x_1, x_2, \ldots x_n$ can be viewed as values drawn from n independent Normal distributions, $X_1, X_2, \ldots X_n \sim Normal(\mu, \sigma^2)$
- $\bar{x}$ is a sample from the random variable $\frac{1}{n}\Sigma_{i=1}^{n} X_i$
- The mean of $\frac{1}{n}\Sigma_{i=1}^{n} X_i$ is μ
- The variance of $\frac{1}{n}\Sigma_{i=1}^{n} X_i$ is $\frac{\sigma^2}{n}$
- $\frac{1}{n}\Sigma_{i=1}^{n} X_i$ is Normally distributed, hence $\frac{1}{n}\Sigma_{i=1}^{n} X_i \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$

z-Statistic Transformation

Under the null hypothesis, $z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$ is a draw from $Normal(0,1)$.

p-value

- Two sided p-value for the null hypothesis is probability of the event $(Z \geq |z|) \cup (Z \leq -|z|)$.
- One sided p-value for the null hypothesis is probability of the event $(Z \geq |z|)$.

Confidence Interval

- P% confidence interval for the true mean is

$$\left(\bar{x} - a\frac{\sigma}{\sqrt{n}}, \bar{x} + a\frac{\sigma}{\sqrt{n}}\right)$$

  - a chosen with P% of the area under the standard Normal density in $(-a, a)$
  - $P = Pr\big(Z \in (-a, a)\big) = Pr\left(-a \le \frac{\frac{1}{n}\sum X_i - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} \le a\right)$

t-Test Discussion

C. Durso

t-Test

Set-up

- $x_1, x_2, \ldots x_n$ is a sample from $Normal(\mu, \sigma^2)$
- σ^2 is unknown
- Null hypothesis: $\mu = \mu_0$

Distribution of $\bar{x}$

- $x_1, x_2, \ldots x_n$ can be viewed as values drawn from n independent normal distributions, $x_1, x_2, \ldots x_n \sim Normal(\mu, \sigma^2)$.
- $\bar{x}$ is a sample from the random variable $\frac{1}{n}\Sigma_{i=1}^{n} X_i$.
- The mean of $\frac{1}{n}\Sigma_{i=1}^{n} X_i$ is μ.
- The variance of $\frac{1}{n}\Sigma_{i=1}^{n} X_i$ is $\frac{\sigma^2}{n}$.
- $\frac{1}{n}\Sigma_{i=1}^{n} X_i$ is Normally distributed, hence $\frac{1}{n}\Sigma_{i=1}^{n} X_i \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$.

t-Statistic

- Set $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$
- For null hypothesis $\mu = \mu_0$, set $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$, the sample mean, minus the hypothesized mean, divided by the sample standard deviation.

Student's t-Distribution
n -1 Degrees of Freedom

- Student's t-distribution with v degrees of freedom defined by density function: $\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}$
  - search "Gamma function," if you want
- Let $Z_1, Z_2, \ldots Z_n$ be independent standard normal distributions
- Set $\bar{Z} = \frac{1}{n}\Sigma_{i=1}^{n} Z_i$
- Random variable $\bar{Z}/\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(Z_i - \bar{Z})^2}$ has Student's t-distribution with $n-1$ degrees of freedom
- (Trust me.)

t-Statistic Numerator Transformation

- Replace every x_i in $\bar{x}$ with $\frac{x_i - \mu_0}{\sigma}$ to get $\bar{x}' = \frac{1}{n}\Sigma_{i=1}^{n}\frac{x_i - \mu_0}{\sigma}$.
  - Note $\bar{x}' = \frac{\bar{x} - \mu_0}{\sigma}$.
  - Also, $\frac{x_1 - \mu_0}{\sigma}, \frac{x_2 - \mu_0}{\sigma}, \ldots \frac{x_n - \mu_0}{\sigma}$ are observations from $Z_1, Z_2, \ldots Z_n$ as above.
  - Thus $\bar{x}'$ is corresponding observation from $\bar{Z}$ above.

t-Statistic s Transformation

- Replace every x_i in s with $\frac{x_i-\mu_0}{\sigma}$ to get $s' = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i-\mu_0}{\sigma}-\bar{x}'\right)^2}{n-1}}$
- Note $s' = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i-\mu_0}{\sigma}-\frac{\bar{x}-\mu_0}{\sigma}\right)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i}{\sigma}-\frac{\bar{x}}{\sigma}\right)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}\frac{1}{\sigma^2}(x_i-\bar{x})^2}{n-1}} =$
  $\frac{1}{\sigma}\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \frac{s}{\sigma}$
- Also, with $\frac{x_1-\mu_0}{\sigma}, \frac{x_2-\mu_0}{\sigma}, \ldots \frac{x_n-\mu_0}{\sigma}$ observations from $Z_1, Z_2, \ldots Z_n$ as above, $s'/\sqrt{n}$ is the corresponding realization of $\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(Z_i - \bar{Z})^2}$
- Thus $\frac{\bar{x}'}{s'/\sqrt{n}}$ has a Student's t-distribution with $n-1$ degrees of freedom

t-Statistic Distribution

Conclude $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{\frac{\bar{x}-\mu_0}{\sigma}}{\frac{s}{\sigma}/\sqrt{n}} = \frac{\bar{x}'}{s'/\sqrt{n}}$ has a Student's t-distribution with $n-1$ degrees of freedom.

Theorem

If $x_1, x_2, \ldots x_n$ is a sample from $Normal(\mu, \sigma^2)$, $\bar{x} = \frac{1}{n}\Sigma_{i=1}^{n} x_i$, and $s = \sqrt{\frac{\Sigma_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$, then $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ has a Student's t-distribution with $n-1$ degrees of freedom.

t-Test

- If $x_1, x_2, \ldots x_n$ is a sample from $Normal(\mu, \sigma^2)$:
  - σ^2 is unknown
  - $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
- Then null hypothesis $\mu = \mu_0$ has p-value equal to area under the Student's t-distribution density for n-1 degrees of freedom above $(-\infty, -|t|] \cup [|t|, \infty)$.

Student's t Confidence Interval 100-p Percent

- 100-p% confidence interval (ex. p=5, 95% CI) for true mean μ is $\left(\bar{x} - \alpha\left(\frac{s}{\sqrt{n}}\right), \bar{x} + \alpha\left(\frac{s}{\sqrt{n}}\right)\right)$ where:
  - $x_1, x_2, \ldots x_n$ is a sample from *normal*(μ, σ^2)
  - μ, σ is unknown
  - $\bar{x}$ is a sample mean, s is a sample standard deviation
  - α satisfies $P(t \leq -\alpha) = \frac{1}{2}p/100$, t Student t-distributed with $n-1$ degrees of freedom

Student's t Confidence Interval Justification

Note $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ has Student's t-distribution with $n-1$ degrees of freedom.

$$\frac{(100-p)}{100} = P\left(-\alpha \le \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \le \alpha\right)$$

$$= P\left(-\alpha \le \frac{\mu-\bar{x}}{s/\sqrt{n}} \le \alpha\right)$$

$$= P\left(-\alpha\left(\frac{s}{\sqrt{n}}\right) \le \mu-\bar{x} \le \alpha\left(\frac{s}{\sqrt{n}}\right)\right)$$

$$= P\left(\bar{x}-\alpha\left(\frac{s}{\sqrt{n}}\right) \le \mu \le \bar{x}+\alpha\left(\frac{s}{\sqrt{n}}\right)\right)$$