Stat 231 - Statistics **Spring 2017**

Lecture 19, 20, 21, 22, 23, 24: June 12th - June 23nd, 2017

Lecturer: Suryapratim Banerjee *Notes By: Harsh Mistry*

19.1 T-Distribution Recap

T_n is a continuous random variable in $(-\infty, \infty)$ that is said to follow a student's T distribution with n degrees of freedom if T is a ratio of two independent random variables.

$$T = \frac{Z}{W}$$

Where $Z \sim G(0,1)$ and $W = \sqrt{x^2(n)/n}$

Properties

- T_n is symmetric around zero for all n
- T_n looks like the Z-distribution but with a higher kurtosis.
- n = parameter of the T-distribution
- As ns approach infinity and $T \to Z$, The T looks like the z

19.2 T Table

Theorem 19.1 *Let $y_1 \dots y_n$ be independent Gaussian random variables with mean μ and σ^2 Define*

$$\bar{Y} = \frac{1}{n}\sum Y_i \text{ (estimator)}$$

$$S^2 = \frac{1}{n-1}\sum (Y_i - \bar{Y})^2 \text{ (estimator corresponding to its sample variance)}$$

Then

$$\frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim T_{n-1}$$

$$\frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$$

Confidence Interval for μ

1. $Y_1 \sim G(\mu, r)$

2. $\hat{\mu} = \bar{y} \rightarrow$ Estimate
3. $\bar{Y} \rightarrow$ estimator
4. Construct the pivotal distribution
$$\frac{\bar{y} - \mu}{S/\sqrt{n}} \sim T_{n-1}$$
5. Find the end points of the pivot (Look in row (n-1) column = percentage)
6. Construct the coverage Interval
$$P(-c < \frac{\bar{Y} - mu}{\frac{S}{\sqrt{n}}} < c) = percentage$$
$$P(\bar{Y} - c\frac{S}{\sqrt{n}} \leq \mu \leq \bar{Y} + c\frac{S}{\sqrt{n}}) = percentage$$
7. Replace $\bar{Y}$ with $\bar{y}$

Confidence Interval for σ^2

1. Find s^2
2. Estimator = S^2
3. Pivotal
$$\frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$$
4. Find Coverage interval from x^2 table
5. Find Confidence Interval

19.3 Other Distributions

Poison Problem

- Find Estimate for θ
$$\hat{\theta} = \bar{y}$$
- Estimator of $\bar{y} \rightarrow \bar{Y}$
- By the CLT
$$\frac{\bar{Y} - \theta}{\sqrt{\bar{Y}/n}} \sim G(0, 1)$$
- Construct the coverage Interval
$$P(-c < \frac{\bar{Y} - \theta}{\sqrt{\bar{Y}/n}} < c) = percentage$$
$$\bar{Y} \pm c\sqrt{\bar{Y}/n}$$

Exponential Problem

Mean $= \theta$; Variance θ^2

$$\bar{Y} \sim G(\theta, \theta/\sqrt{n})$$
$$\frac{\bar{Y} - \theta}{\theta/\sqrt{n}} = z$$

Exact Interval for the Exponential Problem

If $Y_i \sim Exp(\theta)$ then

$$\frac{2Y_i}{\theta} \sim Exp(2) \sim X^2(2)$$

19.4 Relationship between Likelihood Intervals and Confidence Intervals

Assume N is large

Theorem 19.2 *If θ is the true value of the parameter; $\hat{\theta}$ is the MLE; Then*

$$\delta(\theta) = -2log\frac{L(\theta)}{L(\tilde{\theta})} \sim X^2(1)$$

δ = *Likelihood Ratio Test Statistic*

19.5 Hypothesis Testing

Definition 19.3 *A hypothesis is a statement made about some attribute of the population*

$$H_0 : \theta = \theta_0$$

Two claims are tested against each other

- H_0 : Null Hypothesis = Current conventional wisdom
- H_1 : Alternate Hypothesis = Challenging view

Definition 19.4 *a **p-value** is the probability of observing your evidence (or worse) given that the null hypothesis is true.*

Notes

H_0 and H_1 are not treated symmetrically unless we have "overwhelming evidence against H_0, we do not reject it. The burden of proof is on H_1

- $p > 0.1 \implies$ No evidence against H_0
- $0.05 < p \leq 0.1 \implies$ Weak evidence against H_0
- $0.01 < p \leq 0.05 \implies$ Strong evidence against H_0
- $p \leq 0.01 \implies$ Very string evidence against H_0

"Statistically significant" $\rightarrow$ p-value of the test ≤ 0.05

- If $p < 0.05 \implies$ Reject H_0 at 5% level of significance
- If $p > 0.05 \implies$ Do not reject H_0

Type Errors

- Type I error : Rejecting H_0 when its actually true
- Type II error : Do not reject H_0 when it is actually false

These Two errors may conflict with each other.

Statistical Tests

- Discrepancy measure : A random variable that measures the level of disagreement of the data with the null hypothesis.
  - The distribution of D is known
  - $D \geq 0$ and $D = 0$, is the best evidence for the H_0
  - p-value : $P(D \geq d; H_0$ is true $)$ where d - value of D in your sample