IMPORTANT :- Final Exam is on Tuesday December 28^{th}, 2021 9:30 AM – 1:30pm

Recall : Hypothesis Testing

Z-test : Testing for sample mean when σ is known.

$H_0 : \mu = c$ (null) $\quad$ $H_A : \mu > c$ (alternative)

- sample $X_1, \dots, X_n$ from population X — mean μ, variance σ^2

- Compute $\frac{\sqrt{n}\,(\bar{X} - c)}{\sigma}$; Fix $\alpha \in (0,1)$
  Test statistic $\quad$ Level of significance

- Find $z_{\alpha/2}$: $\mathbb{P}(Z \geq z_{\alpha/2}) = \frac{\alpha}{2}$
  (Normal tables)

- Reject H_0 (null hypothesis) if

$$\frac{\sqrt{n}\,(\bar{x} - c)}{\sigma} > z_{\alpha/2}$$

$\boxed{\text{Critical value approach}}$

Recall:- Reject null hypothesis if

$$\mathbb{P}\left(Z \geq \frac{\sqrt{n}(\bar{x}-c)}{\sigma}\right) < \alpha$$

t-test : Test for sample mean when σ is unknown

$H_0: \mu = c \qquad H_A: \mu \geq c$

- Sampled $X_1, \ldots, X_n$ from X ; Fix $\alpha \in (0,1)$
- Compute

$$\frac{\sqrt{n}(\bar{x}-c)}{S}$$

S - sample variance

$\bar{x}$ - sample mean

- $T \sim t_{n-1}$ (t- distribution with $n-1$ degrees of freedom)

If $\mathbb{P}\left(T > \frac{\sqrt{n}(\bar{x}-c)}{S}\right) < \alpha$

then reject the null hypothesis.

Ex:- Frame it is a critical value approach.

χ^2-distribution

Thm 8.5.9 in Book

Let $n \geq 1$, $X_1, X_2, \ldots, X_n$ be i.i.d. Normal(μ,σ^2). Consider the sample mean and variance

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Then

- $\overline{X}_n$ is a Normal random variable with mean μ and variance $\frac{\sigma^2}{n}$.
- $\frac{(n-1)}{\sigma^2} s_n^2$ has the χ^2_{n-1} distribution.
- $\overline{X}_n$ and s_n^2 are independent.

Observation

Let $n \geq 1$, $X_1, X_2, \ldots, X_n$ be i.i.d. Normal(0,σ^2).

Then $Z_i = \frac{X_i}{\sigma}, i = 1, 2, \ldots, n$ are i.i.d. Normal(0,1).

$$U_X = \frac{\bar{X}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}} \qquad \text{and} \qquad U_Z = \frac{\bar{Z}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2}}$$

What can you say about U_X and U_Y ?

[Substitute: $Z_i = \frac{X_i}{\sigma}$ in U_Z and verify]

$U_X = U_Z$

Observation

Let $n \geq 1$, $X_1, X_2, \ldots, X_n$ be i.i.d. Normal(0,σ^2).

Then $Z_i = \frac{X_i}{\sigma}, i = 1, 2, \ldots, n$ are i.i.d. Normal(0,1).

$$U_X = \frac{\bar{X}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2}} \qquad \text{and} \qquad U_Z = \frac{\bar{Z}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (Z_i - \bar{Z})^2}}$$

$U_X = U_Z$ and U_Z **has distribution independent of** σ

Probability facts :- X and Y are independent random variables with p.d.f $f_X(\cdot)$ and $f_Y(\cdot)$

$$\Rightarrow \quad f(x,y) = f_X(x)\, f_Y(y)$$

↳ Joint density of X and Y.

(a) $Z = X+Y$ (Recall)

$$\mathbb{P}(Z \le z) = \iint_{x+y \le z} f_X(x)\, f_Y(y)\, dy\, dx$$

$$= \cdots$$

$$= \int_{-\infty}^{z} \left[\int_{-\infty}^{\infty} f_X(x)\, f_Y(u-x)\, dx \right] du$$

$$\Rightarrow \quad f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx$$

(b) $W = \dfrac{X}{Y}$; ______ $P(Y=0)=0$

$$F(w) = \mathbb{P}(W \le w)$$

$$= \mathbb{P}\left(\frac{X}{Y} \le w\right)$$

$$= \iint_{\{\frac{x}{y} \le w\}} f_X(x) f_Y(y)\, dx\, dy$$

$= \dots$ Ex: Proposition 5.5.8 in Book

$$= \int_{-\infty}^{w} \left[\int_{-\infty}^{\infty} |y|\, f_X(yu) f_Y(y)\, dy \right] du$$

$$f_W(w) = \int_{-\infty}^{\infty} |y|\, f_X(yw) f_Y(y)\, dy \quad -(*)$$

$X_1, \dots X_n$ i.i.d Normal $(0, \sigma^2)$

Recall: $\dfrac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$ [Use theorem and above facts to derive the same]

t-distribution

Let X_1 be a Normal random variable with mean 0 and variance 1. Let X_2 be an independent χ_n^2 random variable. Let

$$Z = \frac{X_1}{\sqrt{\frac{X_2}{n}}}.$$

Z is said to have t-distribution with n-degrees of freedom if its p.d.f. is given by

$$f_Z(z) \quad = \quad \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{z^2}{n}\right)^{-\frac{n+1}{2}}$$

for $z \in \mathbb{R}$.

verified by [illegible]

t-distribution

Let $n \geq 1$, $X_1, X_2, \ldots, X_n$ be i.i.d. Normal(μ,σ^2). Consider the sample mean and variance

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

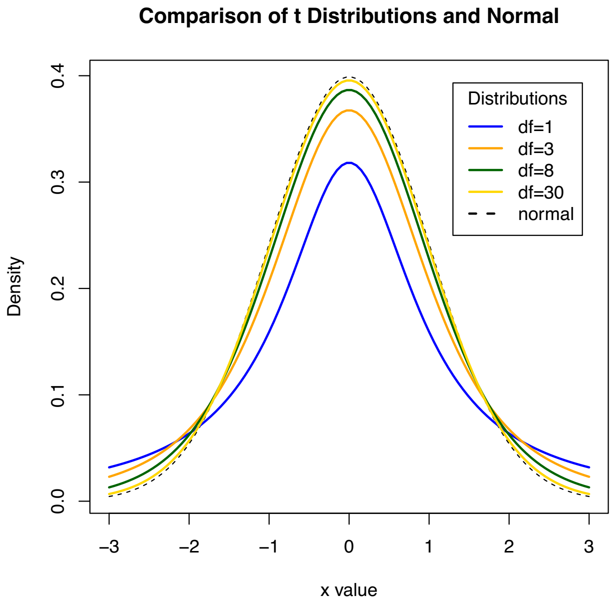
Then

$$\frac{\sqrt{n}(\overline{X} - \mu)}{S}$$

has the t_{n-1} distribution.

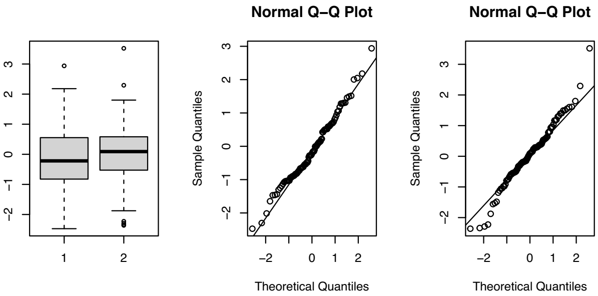
- hypothesis Testing

t-versus Normal



t-versus Normal

```
> x = rnorm(100); y = rt(100,30)
> par(mfrow=c(1,3))
> boxplot(x,y)
> qqnorm(x);qqline(x)
> qqnorm(y);qqline(y)
```



Hypothesis Testing– Proportions

Let $n \geq 1$, $X_1, X_2, \ldots, X_n$ be i.i.d. Bernoulli (p) random variables.

We want to test:

Null Hypothesis : $p = 0.5$

Alternative Hypothesis: $p \neq 0.5$

Normality assumption used in z-test is not necessary

Use Binomial Central Limit Theorem that

$$\frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}} \xrightarrow{d} Z,$$

One can use C.L.T.

where Z is standard Normal.

Hypothesis Testing– Proportions

Use `prop.test`. Suppose $n = 100, \bar{X} = 0.43$.

is built test.

```
> prop.test(43,100)

        1-sample proportions test with continuity correctio

data:  43 out of 100, null probability 0.5
X-squared = 1.69, df = 1, p-value = 0.1936
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.3326536 0.5327873
sample estimates:
   p
0.43
```

Hypothesis Testing– Proportions → Summarise

`prop.test` does the following:

- Computes $P(\mid Z - 0.5 \mid \geq \mid \frac{\sqrt{n}(\bar{X} - 0.5)}{0.5} - 0.5 \mid)$ towards p-value.
- Finds $100(1-\alpha)\%$- Confidence Interval by finding the region of p where

$$\mid \frac{\sqrt{n}(\bar{X} - p)}{\sqrt{p(1-p)}} \mid < z_{\frac{\alpha}{2}},$$

where $P(Z > z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

Hypothesis Testing: *t*-test – inbuilt in R

```
> t.test(x, mu=74)

        One Sample t-test

data:  x
t = 1.3571, df = 8, p-value = 0.2118
alternative hypothesis: true mean is not equal to 74
95 percent confidence interval:
 73.37848 76.39930
sample estimates:
mean of x
 74.88889
```

Hypothesis Testing: Two Sample-test

Suppose $X_1, X_2, \ldots, X_n$ i.i.d $X \sim \text{Normal}(\mu_1, \sigma_1^2)$

— independent of —

$Y_1, Y_2, \ldots, Y_n$ i.i.d $Y \sim \text{Normal}(\mu_2, \sigma_2^2)$

$H_0: \mu_1 = \mu_2$ $\quad$ $H_A: \mu_1 \neq \mu_2$

$\equiv$

Note: $\mu_1 - \mu_2 = 0 \longrightarrow \bar{X} - \bar{Y}$

if difference close to 0 then H_0 ✓

Two sample test if σ_1, σ_2 are known

If data is not normal then C.L.T can be used

Fix $\alpha \in (0,1)$

$$\bar{X} - \bar{Y} \sim \text{Normal}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}\right)$$

Test:- $Z \sim \text{Normal}(0,1)$

If
$$P\left(|Z| \geq \frac{|\bar{X} - \bar{Y}|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right) < \alpha$$

then reject H_0.

Hypothesis Testing: Two Sample Proportion-test

- Want to test if proportion of success $p_1 = p_2$ between two populations.
- Let $\hat{p}_1 = \overline{X^{(1)}}$ and $\hat{p}_2 = \overline{X^{(2)}}$
- The statistic is

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}, \quad \equiv \quad t_{n_1+n_2-1}$$

distribution

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

Large n_1, n_2 assume normality for Z

χ^2- test of independence

- Null Hypothesis: Variables are independent i.e

$$p_{ij} = p_i^R p_j^C \text{ for all } 1 \leq i \leq n_r \text{ and } 1 \leq j \leq n_c$$

- Alternate Hypothesis: Variables are not independent