LOGIC - HYPOTESIS TESTING

Null hypotesis, H_0 - NOTHING GOING ON
Alternative ", H_1 - SOMETHING EXTRA GOING ON

Eg.

$H_0 : P(T) = \frac{1}{2}$ (FAIR COIN) T = Tails

$H_a : P(T) \neq$ (UNFAIR COIN)

TEST STATISTIC -

Measures how far away the data are from the expected if Ho were true

The most common test statistic is the **z-statistic**

$$z = \frac{\text{observed} - \text{expected}}{SE}$$

Eg. COIN TOS

EXPECTED $= 10 \cdot \frac{1}{2} = 5$ $\quad SE = \sqrt{10} \cdot \sqrt{\frac{1}{2} \cdot \frac{1}{2}} = 1.58$

$$z = \frac{7 - 5}{1.58} = 1.27$$

P-Values as Measures of Evidence

Large values of $|z|$ are evidence against H_0

The strenght of the evidence is the

p-value (observed significance level)

Los valores de p por debajo del 5% se suelen considerar "statistically significant"

$$z = \frac{\text{observed} - \text{expected}}{SE}$$

- If p-value is larger than 5% we will not reject the null hypotesis.

T-Test

- We use it when the sample is small.

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\bar{x} \pm t_{n-1} SE$$

Statistical Significant vs Importance

TWO SAMPLE Z-TEST

<u>Rating:</u>

$p_1 \to n = 1000$ & 55%

$p_2 \to n = 1500$ & 58%

$H_0 \to p_1 = p_2$ (nothing inused coming up)

$$z = \frac{\text{observed difference} - \text{expected difference}}{\text{SE of difference}}$$

$\hat{p}_1 = 55\%$ $\quad p_1 - p_2 = 0$

$\hat{p}_2 = 58\%$

$$z = \frac{(\hat{p}_2 - \hat{p}_1) - (p_2 - p_1)}{\left[\sqrt{(SE(p_1))^2 + (SE(p_2))^2}\right]} \Rightarrow \frac{(0.58 - 0.55) - 0}{0.0202} = 1.48$$

$\to \sqrt{\sqrt{\frac{p_1(1-p_1)}{n_1}}^2 + \sqrt{\frac{p_2(1-p_2)}{n_2}}^2}$

Miramos Tabla

p-value = 2 (7%)

14% > 5% so we cannot reject null hypotesis

Also confidence interval $(\hat{p}_2 - \hat{p}_1) \pm z\, SE(\hat{p}_2 - \hat{p}_1)$

in case of 95% $z = 2$ $\quad [-1\%, 7\%]$

$0.33 \cdot 1000 = 330$
$0.58 \cdot 1500 = 870$ $\Big\} \rightarrow$ 1420 out of 2500

pooled estimate $\frac{1420}{2500} = 56.8$

$$SE(\hat{p}_1 - \hat{p}_2) \sqrt{\frac{0.568 \cdot (1-0.568)}{1000} + \frac{0.568(1-0.568)}{1500}} = 0.02022$$

which gives the same answer

TO COMPARE TWO MEANS

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{(SE(\bar{x}_1))^2 + (SE(\bar{x}_2))^2}$$

$SE(\bar{x}_1) = \frac{\sigma_1}{\sqrt{n_1}}$ estimated $\frac{s_1}{\sqrt{n_1}}$ t-test

Paired Difference Test

paired t-test

$$t = \frac{\bar{d} - 0}{SE(\bar{d})}$$

$$SE(\bar{d}) = \frac{\sigma}{\sqrt{n}} \to s_d = 0.55$$

$$t = \frac{1.4 - 0}{0.55/\sqrt{15}} = 5.69$$

COMPUTER SIMULATIONS IN PLACE OF CALCULATIONS

Recall - confidence interval

$$\boxed{\bar{x} \pm z SE(\bar{x})}$$

If we instead of $\bar{x}$ are interested in other estimator: for example $\hat{\theta}$ for a parameter θ and the normal aproximation is not valid for that estimator $\hat{\theta}$

En estes situaciones las simulaciones pueden ser usados para estimar

LAW OF LARGE NUMBERS
TO APPROXIMATE QUANTITIES OF INTEREST

Monte Carlo Method

For the explanation we will use the average height of people living in USA

Sample $n = 100$

We are interested in a parameter θ of a population that we estimate with $\hat{\theta}$

Our statistic (estimator) $\hat{\theta}$ is the average so

$\hat{\theta}$ = average of sample $= \frac{1}{n} \sum_{i=1}^{n} X_i$

Monte Carlo Method or Simulation:

Approximation to a fixed quantity θ by the average of independent random variables with expected value seen θ. the larger the sample the smaller the SE of the Statistic

$$SE(\hat{\theta}) = \sqrt{E(\hat{\theta} - E(\hat{\theta}))^2}$$

E → variance

Example:

1. 1000 samples of 100 observations
2. Compute $\hat{\theta}$ for the samples $\hat{\theta}_1 \dots \hat{\theta}_{1000}$
3. Compute Standard Dev

$$s(\hat{\theta}_1 \dots \hat{\theta}_{1000}) = \sqrt{\frac{1}{1000-1} \sum_{i=1}^{1000} \left(\hat{\theta}_i - \text{average}(\hat{\theta}_i)\right)^2}$$

$$\bar{\hat{\theta}} = \text{average}(\hat{\theta})$$

Sumary

- Random Sampling
- The more Samples the less SE
- We can compute the SE with it

Plug-In / Bootstrap principle

The bootstrap principle uses the plug-in principle and the montecarlo method to aproximate quantities such as $SE(\hat{\theta})$

1. Draw sample $X_1 \dots X_n$ to compute $\hat{\theta}$
2. Repeat B times (Eg $B = 1000$) to get $\hat{\theta}_1 \dots \hat{\theta}_B$
3. If we have only 1 sample then the bootstrap simulate from the sample since we don't have access to the population.

   Basically it interact with the sample like if the sample was the population

Non-parametric bootstrap

Sometimes we can known or supose characteristics of the data. For example we might known that it follows a normal distribution but not its SE or mean

BOOTSTRAP CONFIDENCE INTERVALS

$$\hat{\theta} \pm z_{\alpha/2} SE(\hat{\theta})$$

We can estimate the sampling distribution of $\hat{\theta}$ not just $SE(\hat{\theta})$

Making a histogram of the bootstrap copies

Also we can do $\hat{\theta} - \theta$

bootstrap pivotal interval

$$\left[2\hat{\theta} - \hat{\theta}^*_{(1-\alpha/2)},\ 2\hat{\theta} - \hat{\theta}^*_{(\alpha/2)}\right]$$

Bootstraping in regression

We have data $(X_1, Y_1), \dots, (X_n, Y_n)$ from simple linear regresion model

$$Y_i = a + b X_i + e_i$$

From the data we can compute estimators $\hat{a}$, $\hat{b}$.

And compute residuals:

$$\hat{e}_i = Y_i - \hat{a} - \hat{b} X_i$$

* Remember: residuals are the difference between the observed values and the values predicted by the model/regression line

Residual = observed - predicted

Steps

1. Compute residuals $\hat{e}_i = Y_i - \hat{a} - \hat{b} X_i$
2. Resample from those residuals to get $e_1^* \dots e_n$
3. Compute the bootstrap responses $Y_i^* = \hat{a} + \hat{b} X_i + e_i^*$

This will give us a bootstrap sample $(X_1, Y_1^*) \dots (X_n, Y_n^*)$ in which one we can estimate the parameters

CATEGORICAL DATA

Did ticket class affect in your % of survival in the titanic

	First	Second	Third	Crew
Survived	202	118	178	215
Died	123	167	528	698

2x4 table (same as bidimensional arrays)

Also called contingency table (shows survival counts for each category class)

Example With M&Ms

OLD

BLUE	Orange	Green	Yellow	Red	Brown
24%	20%	16%	14%	13%	13%

ACTUAL

BLUE	Orange	Green	Yellow	Red	Brown	
85	79	56	64	58	68	410

Goodness of fit test

H_0 : nothing different going on (old %)

H_1 : Different

Under null hypotesis

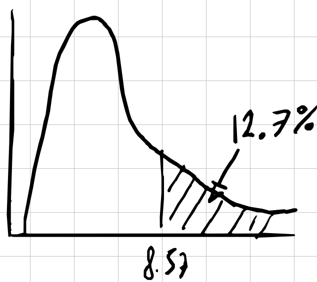
BLUE	Orange	Green	Yellow	Red	Brown
98.4	82	65.6	57.4	53.3	53.3

$$\chi^2 = \sum_{\text{categories}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$\frac{(85 - 98.4)^2}{98.4} + \dots + \frac{(68 - 53.3)^2}{53.3} = 8.57$$

Large values of the chi-square statistic χ^2 ARE EVIDENCE AGAINST H_0

The p-value is the right tail of χ^2 distribution with degrees of freedom = nºcategories - 1 = 5



- Not enought evidence to regeet the hypotesis

If we only want to test 1 category we use z-test. Esentially chi-square provides an extension of z-test testing several categories

If p-value is small we can regect null hypotesis

TESTING HOMOGENITY

Comparation of several population

* We will use the M&Ms' and Titanic's 'Gecple'

Test Of Homogenety - χ^2

- Assumes independency across and within populations

In this exomple (the titanic one) we are using the whole population

	First	Second	Third	Crew
Survived	202	118	178	215
Died	123	167	528	698
Total	325	285		

Eg. 325 observations (firs class)

285 observations (second class)

H_0: probability of survival is the same in all 4 histograms (clases)

	First	Second	Third	Crew
713 Survived	202	118	178	215
Died	123	167	528	698
Total	325	285		

total: 2229

1. First. Estimate the survival probability across all

$$\frac{713}{2229} = 32\%$$

2. Calculate expected values for that data

3. Calculate chi-squared - χ^2 over all cells

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

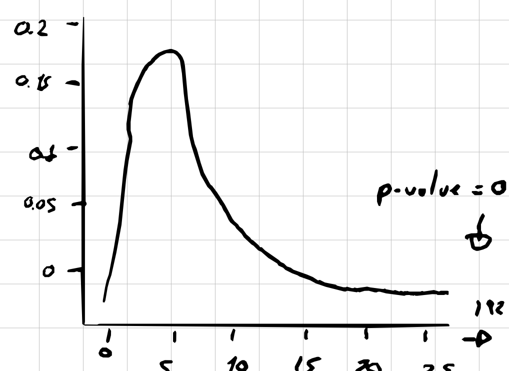
$$= \frac{(202-104)^2}{104} + \dots = 192.2$$

		FIRST	SECOND	THIRD	CREW
Survive	EXPECTED	104	91.2	225.8	292.1
	OBSERVED	202	118	178	215
Died	EXPECTED	221	193.8	480.1	620.8
	OBSERVED	123	167	528	698

4. Calculate degrees of freedom.

$$(n^\circ \text{ col} - 1) \cdot (n^\circ \text{ rows} - 1) = (4-1) \cdot (2-1) = 3$$

p-value for 192.2 is 0% so its a huge evidence that H_0 IS FALSE



ANOTHER EXAMPLE IS TESTING INDEPENDENCE

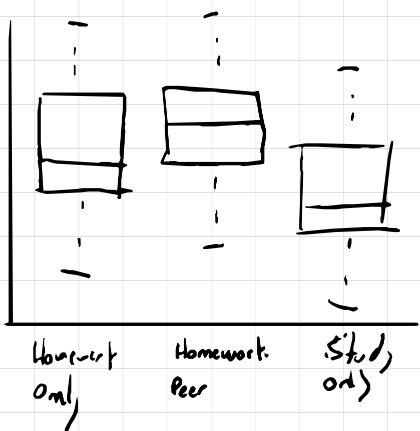
Gender (male/female) related to voting preference (Liberal/Conservative)

H_0: nothing happend

2x2 table

H_1: it effect

Comparing Several Means



H_0 : all means are equal

We will need to use t-test

$$t = \frac{\text{diff sample means}}{\text{SE of diff}}$$

Analysis of Variance ANOVA

1. Compare sample variance between the means and the groups

k groups of n_j observations

group1	group2	...	group k
y_{11}	y_{12}		y_{1k}
$\vdots$	$\vdots$		$\vdots$
$y_{n_1 1}$	$y_{n_2 2}$		$y_{n_k k}$

* We dont need the same n for each group

$$N = n_1 + n_2 \dots + n_k$$

$$\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$$

$$\bar{\bar{y}} = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} y_{ij}$$

Overall mean

Analisis of Variance 2

- TREATMENT SUM OF SQUARES

$$SST = \sum_j \sum_i (\bar{y}_j - \bar{\bar{y}})^2$$ has $k-1$ degrees of freedom

- TREAMENT MEAN SQUARES

$$MST = \frac{SST}{k-1}$$

- ERROR SUM OF SQUARES

$$SSE = \sum_j \sum_i (y_{ij} - \bar{y}_j)^2$$ has $N-k$ degrees of freedom

- ERROR MEAN SQUARE

$$MSE = \frac{SSE}{N-k}$$ Measure variability within the groups

F-Distribution to evaluate ANOVA

$F = \frac{MST}{MSE}$ F distribution with $k-1$ and $N-k$ degrees of freedom

We reject H_0 if p-value is smaller than 5%

ANOVA TABLE

Source	df	SUM of Squares	MEAN SQUARE	F	p-val
Treatment	$k-1$	SST	MST	MST/MSE	
Error	$N-k$	SSE	MSE		
Total	$N-1$	TSS			

where $TSS = \sum_j \sum_i (y_{ij} - \bar{\bar{y}})^2$

EXAMPLE OF HOMEWORK | ANOVA

Source	df	Sum of Squares	Mean Square	F	p-value
Treatment	2	98.4	47.2	2.49	0.097
Error	38	723.8	19.1		
Total	40	822.2			

$$\boxed{\begin{aligned} y_{ij} &= \mu_j + \epsilon_{ij} \\ y_{ij} &= \mu + \tau + \epsilon_{ij} \end{aligned}}$$

$\tau_j = \mu_j - \mu \Rightarrow y_j - \bar{\bar{y}}$ | Estimate

Estimate of ϵ_{ij} becomes the residual $y_{ij} - \bar{y}_j$

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_j - \bar{\bar{y}}) + (y_{ij} - \bar{y}_j)$$

$$TSS = SST + SSE$$

$$\sum_j \sum_i (y_{ij} - \bar{\bar{y}})^2 = \sum_j \sum_i (\bar{y}_j - \bar{\bar{y}})^2 + \sum_j \sum_i (y_{ij} - \bar{y}_j)^2$$

EXTRA OF ANOVA

- F-Test assumes all groups have the same variance σ^2

- The data within and across groups are independent

- If F-Test reject we can differ and compare all pairs of means with two sample t-test using $s_{pooled} = \sqrt{MSE}$

Data Snooping - Multiple Testing Fallacy

Smaller the p-value the highly significant and stronger.

Interpretation: If there is no effect then there is only a 1% chance to get such a highly significant result

Carefull with the multiple testing fallacy
(If we do a lot of test just by ? we will have high significant in some of them)

FDP

False discobery Proportion = $\frac{\text{number of false discoveries}}{\text{total number of discoveries}}$

$\frac{\text{false positive}}{\text{all}}$

FALSE DISCOVERY RATE [FDR]

- Control the expect proportion of false discoveries
  1. Sort p-val $p_1 \leq \dots \leq p_{(m)}$
  2. Find larger k such $p_{(k)} \leq \frac{k}{m} \alpha$
  3. Declare discoveries for all tests i from 1 to k