BIST P8130: Biostatistics Methods I

Recitation 04 – ANOVA, One and Two-sample proportions test in SAS and Sample Size Determination for two-proportions

Zilan Chai

Fall 2017

This recitation's big ideas:

- Use PROC GLM in SAS 9.4 to perform Analysis of Variance (ANOVA)
  1. Check model assumptions
  2. One-way ANOVA
  3. Multiple comparisons adjustments
- Use PROC FREQ to
  1. Summarize categorical data (frequency tables)
  2. Perform a one-sample test for proportion
- Estimation of Power and Sample size for comparing two binomial proportions
- Use PROC POWER in SAS to perform power and sample size calculation

Analysis of Variance (ANOVA) in SAS

Example From Lecture 8. A study is examining the effect of glucose on insulin release. Specimens of pancreatic tissue from experimental animals were treated with five different stimulants. Later, the amounts of insulin released were recorded.

Obs	insulin	trtgroup	mouse
1	1.53	1	1
2	1.61	1	2
…	…	…	…
31	7.49	5	7
32	8.98	5	8

1. Model Assumptions: (Will be covered in depth in the context of linear regression)

- Independent samples
- Responses within the groups are independent and identically distributed (i.i.d.)
- Residuals are normally distributed
- Equality of variances across groups (constant variance)

Constant variance (homoscedasticity)

- Plot residuals vs fitted values (Look for a random pattern)
- Levene's or Brown-Forsythe's tests for equality of variances
- Remedial measures:
  - Transformation of Y (dependent variable)
  - Square root, log, inverse, or arcsin if the response is a proportion or a percentage

Normality

- QQplot of the residuals (Look for a linear trend)
- Remedial measures:
  - Most of the ANOVA procedures are robust to minor departures from normality
  - Transformation of Y (dependent variable)
  - Use non-parametric Kruskal-Wallis

2. One-way ANOVA

Simplest SAS syntax (only one categorical factor):

```
PROC GLM data = data-name;
  class categorical_factor;
  model continuous_response = categorical_factor;
run;
```

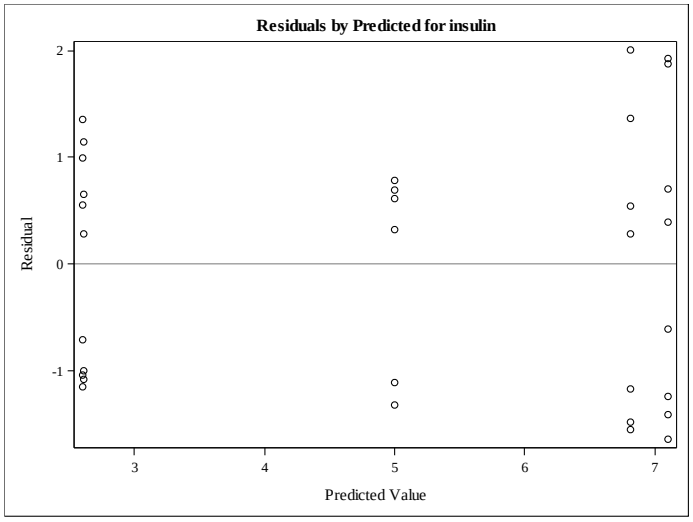
Example SAS code:

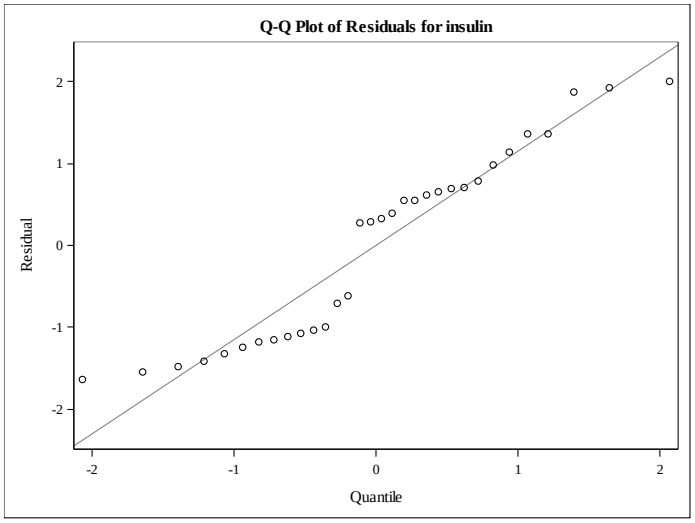
```
proc glm data = insulin plots=diagnostics(unpack);
class trtgroup;
model insulin=trtgroup;
means trtgroup/ HOVTEST=BF;
run;

* Plots = diagnostic(unpack) used for checking model assumptions:
Residual vs fitted values plot - constant variance check (look for random pattern)
QQplot - normality check (look for linear trend);

* HOVTEST=BF option generats the Brown-Forsythe test (with P-value)
used to assess constant variance assumption (homoscedasticity)
Available only for One-Way ANOVA;
```

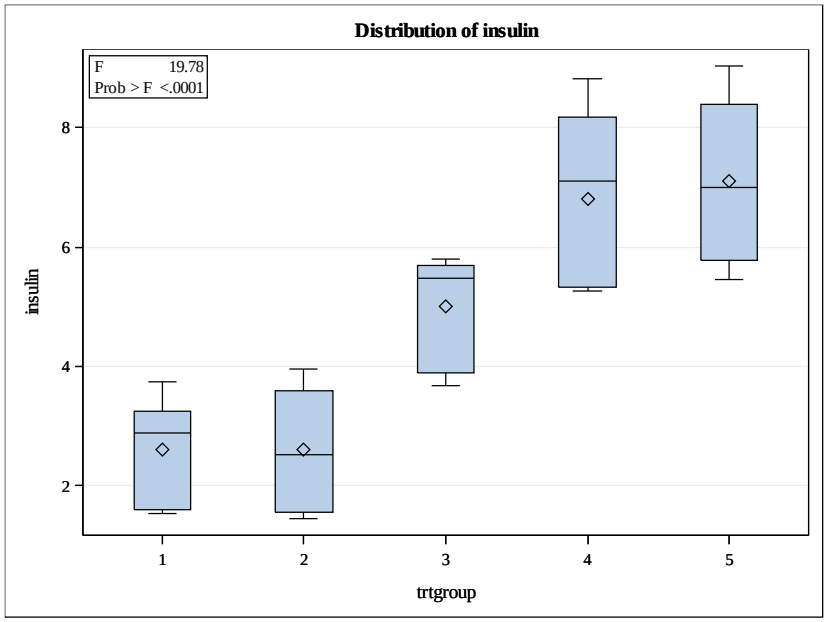
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	121.1854282	30.2963570	19.78	<.0001
Error	27	41.3573937	1.5317553		
Corrected Total	31	162.5428219			

R-Square	Coeff Var	Root MSE	insulin Mean
0.745560	24.27491	1.237641	5.098437

Source	DF	Type I SS	Mean Square	F Value	Pr > F
trtgroup	4	121.1854282	30.2963570	19.78	<.0001







Brown and Forsythe's Test for Homogeneity of insulin Variance ANOVA of Absolute Deviations from Group Medians					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
trtgroup	4	1.3870	0.3467	0.87	0.4958
Error	27	10.7859	0.3995		

3. ANOVA – Multiple Comparisons (MC)

- Bonferroni (bon): simplest and most conservative
- Tukey's HSD (tukey): most powerful method for pairwise comparisons, based on a studentized range distribution (q)
- Dunnett's (dunnett): most powerful test for comparisons with a control
- Scheffe's (scheffe): most powerful test for 'unplanned comparisons' – takes into account all contrasts, very conservative (not recommended for pairwise comparisons)

SAS syntax:

```
proc glm data = data_name;
class categorical_factor;
model continuous_response = categorical_factor;
means categorical_factor / bon tukey dunnett("B") scheffe;
run;
                *For Dunnett's, specify the reference group ("B");
```

Example SAS code:

```
proc glm data = insulin;
class trtgroup;
model insulin = trtgroup;
means trtgroup/ bon; *tukey dunnett("B") scheffe;
run;
```

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Comparisons significant at the 0.05 level are indicated by ***.				
trtgroup Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
5 - 4	0.2884	-1.6694	2.2462	
5 - 3	2.0996	0.0566	4.1426	***
5 - 1	4.4933	2.3367	6.6498	***
5 - 2	4.5013	2.4583	6.5442	***
4 - 5	-0.2884	-2.2462	1.6694	
4 - 3	1.8112	-0.2934	3.9158	
4 - 1	4.2049	1.9898	6.4199	***
4 - 2	4.2129	2.1083	6.3175	***
3 - 5	-2.0996	-4.1426	-0.0566	***
3 - 4	-1.8112	-3.9158	0.2934	
3 - 1	2.3937	0.1030	4.6843	***
3 - 2	2.4017	0.2176	4.5857	***
1 - 5	-4.4933	-6.6498	-2.3367	***
1 - 4	-4.2049	-6.4199	-1.9898	***
1 - 3	-2.3937	-4.6843	-0.1030	***
1 - 2	0.0080	-2.2826	2.2986	
2 - 5	-4.5013	-6.5442	-2.4583	***
2 - 4	-4.2129	-6.3175	-2.1083	***
2 - 3	-2.4017	-4.5857	-0.2176	***
2 - 1	-0.0080	-2.2986	2.2826	

PROC FREQ: summary statistics for categorical variables

Not presented in class. Please see the SAS code

One-Sample Test for Binomial Proportion

Example: Researchers want to test if the proportion of patients who does not have self-cough is different from 25%.

```
proc freq data=demo14;
  exact binomial;
  tables selfcough / binomial (p=.25);
run;
```

Selfcough	Frequency	Percent	Cumulative Frequency	Cumulative Percent
FALSE	71	83.53	71	83.53
TRUE	14	16.47	85	100.00
Frequency Missing = 6				

Binomial Proportion	
Selfcough = FALSE	
Proportion (P)	0.8353
ASE	0.0402
95% Lower Conf Limit	0.7564
95% Upper Conf Limit	0.9141
Exact Conf Limits	
95% Lower Conf Limit	0.7391
95% Upper Conf Limit	0.9069

Test of H0: Proportion = 0.75	
ASE under H0	0.0470
Z	1.8161
One-sided Pr > Z	0.0347
Two-sided Pr > |Z|	0.0694
Exact Test	
One-sided Pr >= P	0.0410
Two-sided = 2 * One-sided	0.0821

Two-Sample Test for Binomial Proportions

Example: Researchers want to test if the proportion of patients who does not have self-cough among the male is different from proportion of patients who does not have self-cough among female.

```
proc freq data=demo14 order = data;
  exact binomial;
    tables gender*selfcough / riskdiff(equal var=null cl=wald);
run;
```

Table of GENDER by Selfcough			
GENDER	**Selfcough**		
Frequency Percent Row Pct Col Pct	**FALSE**	**TRUE**	**Total**
Male	27 32.14 79.41 38.57	7 8.33 20.59 50.00	34 40.48
Female	43 51.19 86.00 61.43	7 8.33 14.00 50.00	50 59.52
Total	70 83.33	14 16.67	84 100.00
Frequency Missing = 7			

Proportion (Risk) Difference Test	
H0: P1 - P2 = 0	
Proportion Difference	-0.0659
ASE (H0)	0.0828
Z	-0.7953
One-sided Pr < Z	0.2132
Two-sided Pr > |Z|	0.4265

Power and Sample Size Determination (Balanced Design)

Suppose we want to test the null hypothesis: H_0: $p_1 = p_2$. Suppose that the sample size in each group is n (assuming equal sample size). The test statistics is:

$$z = \frac{p_1 - p_2}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n}+\frac{1}{n}\right)}} = \frac{p_1 - p_2}{\sqrt{2\,\frac{\bar{p}(1-\bar{p})}{n}}} \sim N(0,1) \quad under\ H_0$$

where

$$\bar{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{np_1 + np_2}{2n} = \frac{1}{2}(p_1 + p_2)$$

Also recall that the decision rule is: reject H_0 if $Z > z_{1-\alpha/2}$ or $Z < -z_{1-\alpha/2}$ (i.e. $|Z| > z_{1-\alpha/2}$).

$$\text{Power} = P(\text{reject } H_0 \mid H_1 \text{ is true}) = P(|\text{test statistic}| > \text{critical value} \mid H_1 \text{ is true})$$

The formula is given by:

$$\text{Power} = P\left(Z < \frac{-z_{1-\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + \sqrt{n}\,|\,p_1 - p_2\,|}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}} \;\middle|\; \text{H}_1 \text{ true} \right)$$

Relationships:

- α decreases $\Rightarrow$
- sample size increases $\Rightarrow$
- effect size increases $\Rightarrow$

Recall that our goal is to find the value n such that the power is 1- β,

$$P\left(Z < z_{1-\beta}\right) = 1-\beta$$

$$\frac{-z_{1-\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + \sqrt{n}\,|\,p_1 - p_2\,|}{\sqrt{p_1(1-p_1) + p_2(1-p_2)}} = z_{1-\beta}$$

And so we can solve this equation for n:

$$n = \frac{\left(z_{1-\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + z_{1-\beta}\sqrt{p_1(1-p_1) + p_2(1-p_2)}\right)^2}{(p_1 - p_2)^2}$$

where n is the sample size **per group**. So the total sample size required is $\mathbf{2}\boldsymbol{n}$.

Relationships:

- α decreases $\Rightarrow$
- effect size increases $\Rightarrow$
- power increases $\Rightarrow$

Computing Sample Size Given Power for One-Sample test for Proportion

Suppose that we want to test a proportion p, is equal to a reference value, p_0, the hypotheses are: H_0: $p = p_0$ and H_1: $p \neq p_0$ or H_1: $p = p_1$

The formulas for sample size and power are shown below:

$$n = p_1(1-p_1)\left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{p_1 - p_0}\right)^2$$

$$1-\beta = \Phi\left(z - z_{1-\frac{\alpha}{2}}\right) + \Phi\left(-z - z_{1-\frac{\alpha}{2}}\right), \qquad z = \frac{p_1 - p_0}{\sqrt{p_1(1-p_1)/n}}$$

Example. It is known that a certain type of skin lesion will develop into cancer. There is a drug on the market, the expected response rate for this standard treatment is 25%. A pharmaceutical company is developing a new drug to treat skin lesions, the expected response rate for this standard treatment is 40%. Suppose the company recruited 100 patients per treatment group. What is the power of a test that is conducted at the 5% significance level with the sample size provided?

Example SAS Code:

```
proc power;
   twosamplefreq test = pchi
   alpha = 0.05
   grouppproportions = (0.25 0.40)
   nullproportiondiff = 0
   groupweights = (1 1)
   ntotal = 200
   power = .;
run;
```

Fixed Scenario Elements	
Distribution	Asymptotic normal
Method	Normal approximation
Null Proportion Difference	0
Alpha	0.05
Group 1 Proportion	0.25
Group 2 Proportion	0.4
Group 1 Weight	1
Group 2 Weight	1
Total Sample Size	200
Number of Sides	2

Computed Power
Power
0.621

What would happen to the power if the difference in response rate between the two treatment groups were smaller? (Suppose that the response rate for new treatment is only 35% and response rate is the same for the standard treatment group.)

What would happen to the power if the sample sizes were halved in each group?

Example. It is known that a certain type of skin lesion will develop into cancer. There is a drug on the market, the expected response rate for this standard treatment is 25%. A pharmaceutical company is developing a new drug to treat skin lesions, the expected response rate for this standard treatment is 40%. Suppose the company want to conduct a test at the 5% significance level, how many patients per group are needed in order to achieve an 80% power?

Example SAS Code:

```
proc power;
   twosamplefreq test = pchi
   alpha = 0.05
   grouppproportions = (0.25 0.40)
   nullproportiondiff = 0
   npergroup = .  /*could also do ntotal = . here*/
   power = 0.80;
run;
```

Fixed Scenario Elements	
Distribution	Asymptotic normal
Method	Normal approximation
Null Proportion Difference	0
Alpha	0.05
Group 1 Proportion	0.25
Group 2 Proportion	0.4
Nominal Power	0.8
Number of Sides	2

Computed N Per Group	
Actual Power	**N Per Group**
0.800	152