University of Leicester

Comparing Groups (Continuous Variables)

MSc Medical Statistics

University of Leicester

Methods Module 2007

Comparing Groups

- Often in Medical Statistics we are interested in comparing two or more groups. E.g.
  - Two treatments in a clinical trial
  - Exposed versus not-exposed in observational studies
  - Differences between male/females, ethnic groups, areas etc.
- In this lecture we will look at comparing groups when we can assume that the population distribution the data are sampled from can assumed to be Normal.
- Interest mainly lies in comparing the means of two (or more) populations.

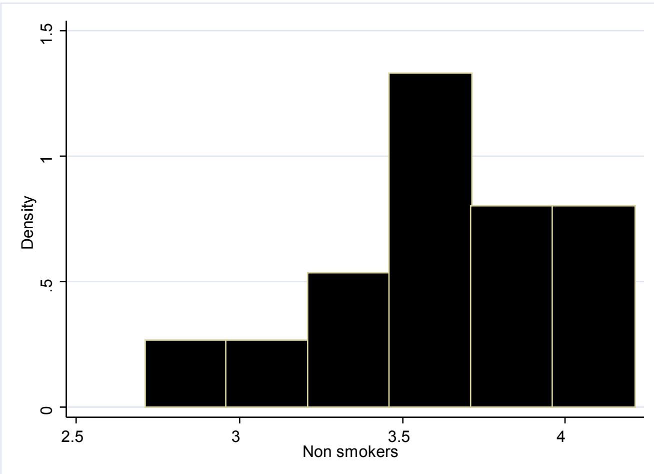
Example

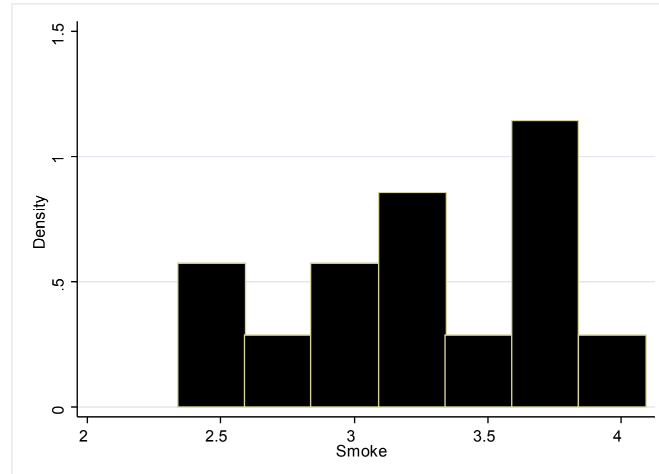
- Simple example comparing birth weights (kg) of children who have non-smoking mothers with mothers who are heavy smokers

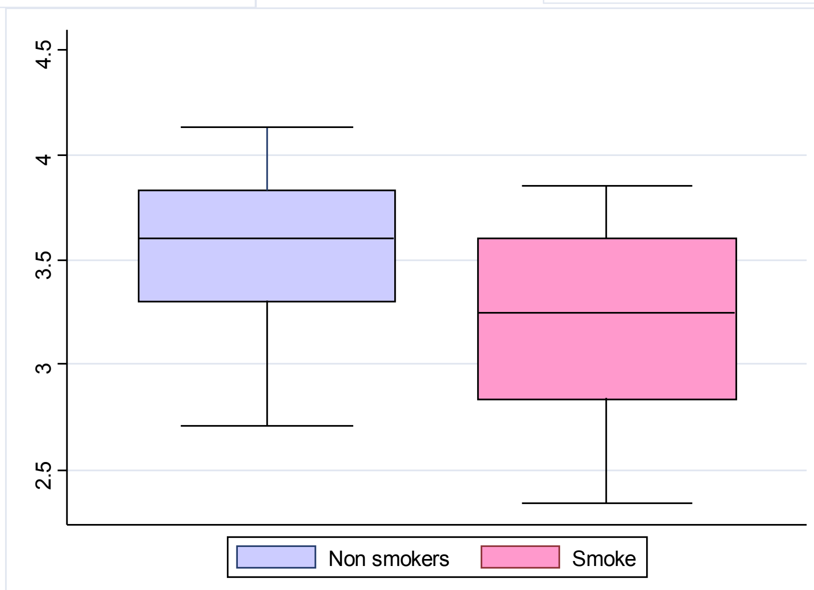
Variable	Obs	Mean	Std. Dev.	Min	Max
Nonsmoke	15	3.593333	.3707457	2.71	4.13
Smoke	14	3.202857	.4926916	2.34	3.85

- Birth weight appears lower for children with mothers who are heavy smokers.

Initial Plots

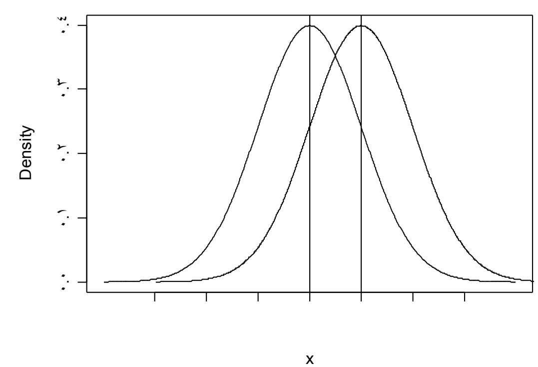






Sampling from two Normal Distributions

- Interested in comparing the sample means ($H_0: \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$) .
- Initially assume that the two population variances are the same, i.e.



- From the two samples of size n_1 and n_2, the two sample means, $\bar{x}_1 \text{ and } \bar{x}_2$ and the two variances, $s_1^2 \text{ and } s_2^2$ are calculated

Pooled variance and standard error

- For known variances or large samples

$$se(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- For small samples we use an estimate of the (common) variance which is a weighted average of the two sample variances (s_1^2 and s_2^2)

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- The standard error of the difference in the means can then be calculated as

$$se(\bar{x}_1 - \bar{x}_2) = s \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Confidence interval & t-test

For known variances (or large samples) the 95% confidence interval can be obtained

$$(\bar{x}_1 - \bar{x}_2) \pm 1.96 \times se(\bar{x}_1 - \bar{x}_2)$$

In small samples (sampled from a normal popn) it is better to use t distribution

- 95% confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.975, n_1 + n_2 - 2} \times se(\bar{x}_1 - \bar{x}_2)$$

- If you want a P-value use a **two sample t-test** where

$$t = \frac{\bar{x}_1 - \bar{x}_2}{se(\bar{x}_1 - \bar{x}_2)} \sim t_{n_1 + n_2 - 2}$$

(This will be normal for large samples)

Birth Weight Example.

- Assuming the two variances are the same the pooled estimate of the variance is

$$s^2 =$$

- The standard error of the difference in means is

$$se(\bar{x}_1 - \bar{x}_2) =$$

- 95% Confidence interval is kg (kg to kg)
- t-test

$$t = \quad , \quad df = \quad , \quad P =$$

Unequal variances (1)

- We have assumed that the two population variances are equal.
- t-test is 'robust' to departures from normality and unequal variances. However, an adjustment is possible in the case of unequal variances.
- The standard error of the mean difference is now calculated as

$$se(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- t-distribution with $n_1 + n_2 - 2$ df not appropriate. One solution is to adjust the df

$$df' = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}$$

Unequal variances (2)

- df` is then rounded to the nearest integer to get df``
- A 95% confidence interval can then be obtained by

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.975,df``} \times se\left(\bar{x}_1 - \bar{x}_2\right)$$

- The t-test is now

$$t = \frac{\bar{x}_1 - \bar{x}_2}{se\left(\bar{x}_1 - \bar{x}_2\right)} \sim t_{df``}$$

- Usually the two methods will give you very similar results.

Testing for unequal variances

- You will often see an **F-test** for comparing the two variances in a t-test.
- With the assumption that the data are from a Normal distribution and the two samples are independent.
- It can be shown that $$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$
- Also $$\frac{\chi^2_m / m}{\chi^2_n / n} \sim F_{m,n}$$
- so

$$\frac{s_1^2}{s_2^2} \sim F_{n_1-1, n_2-1}$$

Example

- For the birth weight and smoking data

$$F = \qquad , \quad df = 13{,}14, \quad P = 0.31$$

- Therefore, would not expect confidence interval of difference or results of t-test to change when not assuming equal variances.
- 95% CI 0.39kg (0.05 to 0.73)

$$t = \frac{0.39}{0.161} = 2.42, \quad df'' = 24, \quad P = 0.024$$

- Even when unequal variances often little difference in the results.

Assumptions

- Normality
- Check normality in **each sample** by histograms and normal probability plots.
- Generally robust to departures from normality, but look out for outliers.
- Transforming the data may help the sample to look approximately normal, but interpretation is much more difficult.
  - Log transformation best

Transformations

- Although transformation may lead to the assumption of normality, it is harder to interpret the data.
  - E.g. the differences in the square root of blood pressure was ??
- However, log transformations are interpretable,

$$y_1 = \ln(x_1), \quad y_2 = \ln(x_2)$$

$$y_1 - y_2 = \ln(x_1) - \ln(x_2) = \ln\left(\frac{x_1}{x_2}\right)$$

- If we exponentiate the mean difference in logs we get an estimate of the geometric mean of the ratio of the variables
- By exponentiating the (log) confidence interval limits we get a confidence interval for this geometric mean.

Paired Data

- There are situations when there appears to be two groups and the two sample t-test is often wrongly applied. This occurs when the groups are *paired.* Paired data will occur if
  - the same subjects are studied on two occasions (before and after)
  - different subjects may have been matched (e.g. for age)
- Interested in the mean *difference* between observations.
- This reduces the data to a one sample problem
- We assume that the *differences* are Normal and not necessarily the two original samples.

Confidence Intervals

- Obtain the differences, d_i, and obtain confidence interval as in one sample case.

$$\bar{d} \pm t_{0.975,n-1} \times s_d / \sqrt{n}$$

- For hypothesis test, test to see if difference is different from 0

$$\frac{\bar{d}}{s_d / \sqrt{n}} \sim t_{n-1}$$

Example

Patient	Drug	Placebo	difference
1	6.1	5.2	.90
2	7	7.9	-.90
3	8.2	3.9	4.30
4	7.6	4.7	2.90
5	6.5	5.3	1.20
6	8.4	5.4	3.00
7	6.9	4.2	2.70
8	6.7	6.1	.60
9	7.4	3.8	3.60
10	5.8	6.3	-.50

- The data displayed in the table shows the hours of sleep for 10 patients, once when using a sleeping drug and once when using a placebo (one week later).
- The difference between the the two treatments has been calculated.
- It can be seen that for 7 of the patients use of the drug led to a longer sleep.

Example continued

- Mean difference is 1.78 hours
- Standard error of the difference is 0.559 hours
- 95% confidence interval (0.52 hours to 3.04 hours)
- Wide CI as only small amount of data.
- P-value=0.011.
- Again mean difference with CI much more informative.
- Note that we are assuming normality which is very hard to assess with such a small sample.

Sample Size Calculations

- Assuming normality

$$x_{1i} \sim N[\mu_1, \sigma^2] \ \& \ x_{2i} \sim N[\mu_2, \sigma^2]$$

$$\text{then} \quad \bar{x}_1 - \bar{x}_2 \sim N\left[\delta, \sigma^2\left(1/n_1 + 1/n_2\right)\right]$$

where $\delta = \mu_1 - \mu_2$ for large samples and equal variances

- Assume H_0 $\quad \bar{x}_1 - \bar{x}_2 \sim N\left[0, \sigma^2\left(1/n_1 + 1/n_2\right)\right]$

- H_0 is rejected if $\quad |\bar{x}_1 - \bar{x}_2| \geq \varepsilon_{\alpha/2}\sigma\sqrt{1/n_1 + 1/n_2}$ [1]

where ε_α is standard normal deviate, i.e. $\Phi(\varepsilon_\alpha) = 1 - \alpha$

and the right hand side of [1] is the critical value

Sample Size Calculations

- Under H_1 $\quad \bar{x}_1 - \bar{x}_2 \sim N\left[\delta, \sigma^2\left(1/n_1 + 1/n_2\right)\right]$

  the critical value is $\delta - \varepsilon_\beta \sigma \sqrt{1/n_1 + 1/n_2}$ [2]

- Equating [1] and [2] and assuming that $n_1 = n_2 = n$

$$n = \frac{2\sigma^2}{\delta^2}\left(\varepsilon_{\alpha/2} + \varepsilon_\beta\right)^2$$

$$= \frac{2\sigma^2}{(\mu_1 - \mu_2)^2}\left(\varepsilon_{\alpha/2} + \varepsilon_\beta\right)^2$$

More than 2 groups

- May be more than two groups we want to compare
- Could compare each pair of groups using t-tests.
  - leads to multiple testing
- A method that simultaneously compares all groups means is the **one way analysis of variance (anova).**
- If only there are only two groups then it reduces to the t-test

One way ANOVA

- Assumes samples come from Normal distributions with the *same* variance.
- Therefore, obtain a *pooled* variance estimate of the population variance.
- Null hypothesis is that the population means in the *k* groups are the same.
- *Between-group variation* is contrasted with *within-group* variation. If there is larger variation between groups than one would expect by chance then there is evidence that the population means are different.

Calculations for ANOVA (1)

- Between Group, Within Group and Total Sums of squares are calculated. Let

$$\bar{y}_i = \text{ represent the mean in the } i^{th} \text{ group}$$

$$S_i = \text{ the sums of squares, } \sum_{j=1}^{n_i} \left(y_{ij} - \bar{y}_i\right)^2 \text{ in the } i^{th} \text{ group}$$

$$T = \text{ sum of all observations, } \sum_{\mathrm{i}=1}^{k} n_i \bar{y}_i$$

$$S = \text{ sums of squares of all observations } = \sum_{i=1}^{k} S_i$$

$$N = \text{ total number of observations } = \sum_{i=1}^{k} n_i$$

Calculations for ANOVA (2)

- Using these the following table can be obtained

Source of Variation	df	Sums of Squares	Mean Squares	$F_{k-1,N-k}$
Between Groups	$k-1$	$BSS = \sum_{i=1}^{k} n_i \bar{y}_i^2 - T^2/N$	$s_B^2 = \dfrac{BSS}{k-1}$	$\dfrac{s_B^2}{s_W^2}$
Within Groups	$N-k$	$WSS = S - \sum_{i=1}^{k} n_i \bar{y}_i^2$	$s_W^2 = \dfrac{WSS}{N-k}$	
Total	$N-1$	$TSS = S - T^2/N = (BSS + WSS)$		

- A P-value for H_0: the population means are equal can be obtained by looking up the appropriate F value. (p-value given in most statistical software)

Homoscedasticity and Heteroscedasticity

- **Homoscedasticity** – the variances in the *k* groups are the same.
- **Heteroscedasticity** – the variances in the *k* groups are not the same.

- An extension of the F-test for comparing variances is *Bartlett's test* (good if data is normal)
- Most packages will produce this, but very sensitive to non-normality (comparison of means is not) and will rarely detect differences in moderate samples. An alternative is *Levene's Test* (less sensitive to departures from normality).
- Unless differences in variances are obvious, unlikely to effect comparison of means

Example

- Another study investigating the relationship between birth weight and mother's smoking status.
  - 4 Groups (Non-Smoker, Ex-smoker (gave up smoking), <1 pack per day, >1 pack per day).

Group	**Mean birth weight (lb)**	**SD**
Non-smokers	7.56	0.96
Ex-smokers	7.24	0.91
<1 pack /day	6.33	1.14
>1 pack /day	6.01	0.72

  - Birth weight appears to decrease as smoking increases.

Example (continued)

- ANOVA Table

```
                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      11.6726896      3   3.89089655      4.41     0.0137
 Within groups       20.303607     23   .882765523
------------------------------------------------------------------------
    Total           31.9762967     26   1.22985756

Bartlett's test for equal variances:  chi2(3) = 1.2656  Prob>chi2 =
0.737
```

Comparing means (1)

- A confidence interval for the *means* in each of the groups constructed in standard way – except use pooled estimate of standard error. Thus,

$$se(\bar{y}_i) = \frac{s_W}{\sqrt{n_i}}$$

- Confidence interval obtained by

$$\bar{y}_i \pm t_{0.975, N-k} \times se(\bar{y}_i)$$

- Note df are N-k not n_i-1

Comparing means (2)

- Similarly when obtaining the difference between means,

$$se(\bar{y}_i - \bar{y}_j) = s_W \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- With 95% confidence interval,

$$\bar{y}_i - \bar{y}_j \pm t_{0.975,N-k} \times se(\bar{y}_i - \bar{y}_j)$$

- Note, that the df are N-k.

- Some problems with multiple testing, should we adjust?

Example

- From ANOVA table, s_w=0.94.
- Compare Non-smoking group to 3 other groups. Use difference in means and 95% confidence intervals
  - Ex-smokers -0.35lb (-1.48lb to 0.79lb)
  - <1 pack -1.26lb (-2.30lb to –0.22lb)
  - >1 pack -1.57lb (-2.58lb to –0.57lb)

- Easiest way to get these results is to fit a regression model with 3 dummy variables (linear models module).

Multiple Testing

- With more than 2 groups complex to describe differences.
  - E.g. A different from B and C, but B and C similar.
- Do we compare all differences?
- Specify differences of interest in advance (contrasts)?
- With many groups there are many 2-way comparisons which leads to multiple testing.
  - With 5 groups 9 comparisons.
  - If 5% level for each – overall level >5%
- Methods exist that attempt to keep overall significance level at 5%.
  - Tend to be conservative so could miss true difference.
- Best to decide in advance which groups to compare.

Bonferroni Adjustment

- Easiest adjustment is Bonferroni, which adjusts P-values

- If there are *k* comparisons then the P-value obtained from each test is multiplied by *k*. If the resulting P-value is greater than 1 then set it equal to 1.
  - Assumes comparisons are independent (which they are not)
  - Too conservative
- Good for making clinicians think about their research question.

Other methods

- Many methods that adjust the P-value in different ways.
- e.g. Scheffe,
- The t value is calculated as before, but then calculate

$$a_1 = -\sqrt{(k-1)F_{k-1,n-k,1-\alpha}} \qquad a_2 = \sqrt{(k-1)F_{k-1,n-k,1-\alpha}}$$

- Reject H_0 if
  - $t>a_2$ or $t<a_1$
- SAS will perform 10+ different multiple test adjustments (multiple comparisons tests).

Example

```
 Comparison of bwt by group
                                (Bonferroni)
Row Mean-|
Col Mean |        Non          Ex          <1
---------+------------------------------------
      Ex |   -.345714
         |      1.000
         |
      <1 |   -1.25714   -.911429
         |      0.119      0.667
         |
      >1 |   -1.57321    -1.2275   -.316071
         |      0.022      0.188      1.000
```

```
                         Comparison of bwt by group
                                  (Scheffe)
Row Mean-|
Col Mean |        Non          Ex          <1
---------+------------------------------------
      Ex |   -.345714
         |      0.940
         |
      <1 |   -1.25714   -.911429
         |      0.130      0.449
         |
      >1 |   -1.57321    -1.2275   -.316071
         |      0.032      0.185      0.934
```

Checking ANOVA assumptions using Residuals

- Often comparing small groups so normal prob. plots are difficult to interpret.
- Calculate residuals $\varepsilon_{ij} = y_{ij} - \bar{y}_i$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

where σ^2 can be estimated by s_w^2

- Produce a normal probability plot of the residuals
- The residuals can be plotted against the group means to investigate how similar the variances are and to see if there are any outliers.