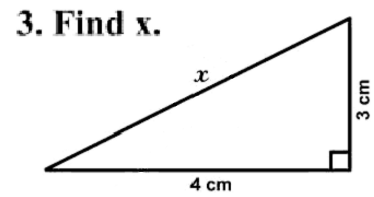
Genetics 300: Statistical Analysis of Biological Data

Lecture 3: Comparison of Means

- Two-sample t-test
- Analysis of variance
- Type I and Type II errors
- Power
- More R commands

September 23, 2010



Two-sample Test

- Previously, we compared a sample mean with a known value:

$$H_0 : \mu = \mu_0$$

- Now, we want to compare two unknown means:

$$H_0 : \mu_1 = \mu_2$$

- Are the data **paired** or **unpaired**?
- Two samples are paired if each data point in one sample is related to a unique data point in the other.
- Example: testing each patient before and after intervention

Paired Data

- For paired data, this is really just a one-sample test of the differences $d_i = X_{i2} - X_{i1}$ with mean Δ.

$$H_0 : \mu_1 = \mu_2 \longrightarrow H_0 : \Delta = 0$$

- So the test statistic $t = \dfrac{\bar{d}}{s_d / \sqrt{n}}$

where $\bar{d}$ is the sample average of the differences and s_d is the sample s.d. of the differences, follows a *t*-distribution with (n-1) degrees of freedom (d.o.f.) under H_0

Unpaired Two-Sample Test

- Two versions, depending on whether the variances in the two populations are equal or not

Unequal variances

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

t follows a *t*-distribution with some d.o.f. (the formula is complicated)

Equal variances

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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

t follows a *t*-distribution with (n_1+n_2-2) d.o.f.

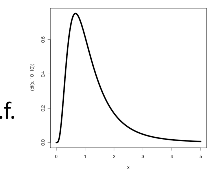
Unpaired Two-Sample Test

- How to decide whether the variances are the same or not?
- Using unequal variances should be fine in most cases

- F-test to test if the populations have equal variances
- For two normal populations with means μ_1 and μ_2 and a common variance σ^2, if you take two samples of size n_1 and n_2,

$$F = \frac{s_1^2}{s_2^2}$$

follows an F-distribution with $(n_1\text{-}1, n_2\text{-}1)$ d.o.f.



Example

- Is daily caloric consumption equal in two populations?
- We took two samples and found the following:

$$\bar{X}_1 = 2500, s_1 = 250, n = 13$$
$$\bar{X}_2 = 2000, s_2 = 200, n = 10$$

- Test the hypothesis that the mean consumption is the same in both populations.

Example

- Step 1: Test equality of variances

$$F = \frac{s_{max}^2}{s_{min}^2} = \frac{250^2}{200^2} = 1.56$$

$$F_{0.975}(13-1, 10-1) = 3.87$$

qf(.975,12,9)

- There is no evidence for different population variances → use equal variances method

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(13-1)250^2 + (10-1)200^2}{13+10-2} = 52857 = 229.9^2$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{500}{229.9\sqrt{\frac{1}{13} + \frac{1}{10}}} = 5.17 > 2.080 = t_{21, 0.975}$$

- We reject the null hypothesis and conclude that the two populations have different means.

Confidence Interval

- The 100(1-a)% confidence interval for the true mean differences
- Recall that we expected 95% of the data from a normal distribution to be contained in $\mu \pm 1.96\sigma$

Unequal variances

$$(\bar{x}_1 - \bar{x}_2) \pm t_{d',\,1-\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Equal variances

$$(\bar{x}_1 - \bar{x}_2) \pm t_{n_1+n_2-2,1-\frac{\alpha}{2}} \left(s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$$

- So in the current example, the CI is

$$(2500 - 2000) \pm t_{21,0.975} \left(229.9\sqrt{\frac{1}{13} + \frac{1}{10}} \right)$$
$$500 \pm 2.080(96.7) = 500 \pm 201$$

Other Methods for Two-group Comparisons

- Many methods have been devised for this problem
- "Signal-to-noise": $\dfrac{\bar{X}_1 - \bar{X}_2}{s_1 + s_2}$
- Where can the t-test go wrong?
  - With a small sample, *s* may be small by chance
  - "Regularize" $\dfrac{\bar{X}_1 - \bar{X}_2}{s_0 + s}$
  - Ref: "Significance analysis of microarrays", Tusher et al, PNAS 2001

ANOVA

- What if we have more than two groups?
- Suppose we have *k* populations, each roughly normal with common variance σ^2.
- How do we test for H_0: $\mu_1=\mu_2=\ldots=\mu_k$?
- The extension of the *t*-test to this case is known as one-way <u>**An**</u>**alysis** <u>**o**</u>**f** <u>**Va**</u>**riance**
- The name is deceptive: we need to analyze variances to test for a difference in means
- What is $\mathbf{H_A}$? **That at least one of the population means differs from one of the others**

ANOVA

- Can we just test all possible pairs using the two-sample *t*-test?
- For 3 groups, we need 3 tests ("3 choose 2")
- This becomes more complicated with more groups
- This is also likely to lead to an incorrect conclusion
- Suppose you have 3 groups and H_0 is true.

  Type I error: P(reject in at least 1) = 1 - P(fail to reject in all 3)

  $= 1 - (1\text{-}0.05)^3 = 1 - 0.857 = 0.143$

Bonferroni Correction

- We want to specify α, the probability of Type I error out of all comparisons.
- We can do this by using, for each individual test $\boxed{\alpha^* = \frac{\alpha}{\binom{k}{2}}}$
- More group → more stringent threshold
- **Bonferroni correction** – simply divide the desired overall α by the number of tests
- This is **conservative**. The level of the test is guaranteed to be less than or equal to α.
- Controlling Type I error is fine if you're worried about the possibility of a single false positive. We will discuss **false discovery rate** later

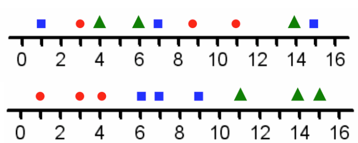
ANOVA

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

- We have to assume that the underlying population variances are identical ($\sigma_1=\sigma_2=\ldots=\sigma$)
- Variance consists of two components: the variation of the individual values around their population means, and the variation of the population means around the overall mean
- Is there more variation **within** groups or **between** groups?

ANOVA

- Compare the following scenarios:



- Does the variability in the data come mainly from the variation within groups, or is it mostly a result of the variation between groups?

ANOVA

$$\boxed{SS_{total} = SS_{between} + SS_{within}}$$

$$\boxed{SS_{total} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{\bar{X}})^2}$$

$\bar{\bar{X}}$ = mean of all observations

$\bar{X}_i$ = mean of group i

$$\boxed{\begin{aligned} SS_{between} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{X}_i - \bar{\bar{X}})^2 \\ SS_{within} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2 \end{aligned}}$$

- Some notations:
  - x_{ij} is the *jth* observation in the *i*th group
  - *i*th group has size n_i
  - Use sum of squares to measure variability

Sum of Squares

- We also define the following **mean squares** by dividing by the degrees of freedom

$$MS_{between} = \frac{SS_{between}}{k-1}$$
$$MS_{within} = \frac{SS_{within}}{n-k}$$

- We can now test for H_0 using $$F = \frac{MS_{between}}{MS_{within}}$$

- We reject if $F > F_{k-1,n-k,1-\alpha}$ and conclude that at least one mean is different. To know which one is different, we must do t-tests between groups

ANOVA Table

Source of variation	Sum of Squares	df	Mean Square	F statistic	p-value
Between	$SS_{between}$	$k-1$	$\frac{SS_{between}}{k-1}$	$F = \frac{MS_{between}}{MS_{within}}$	$P(F_{k-1,n-k} > F)$
Within	SS_{within}	$n-k$	$\frac{SS_{within}}{n-k}$		
Total	SS_{total}	$n-1$			

Errors in hypothesis testing

- Type I and Type II errors

	H_0 is true	H_1 is true
Reject H_0	Type I error	correct
Not reject H_0	correct	Type II error

- P(Type I error) = P(reject H_0 | H_0 is true) = α "false alarm"
- P(Type II error) = P(not reject H_0 | H_1 is true) = β "alarm failure"

- Power = P(reject H_0 | H_1 is true) = $1-\beta$
- We want our test to be **powerful** so that we can detect a different if it exists.

Statistical Power

- Consider $H_0 : \mu = \mu_0$ vs $H_0 : \mu = \mu_1$

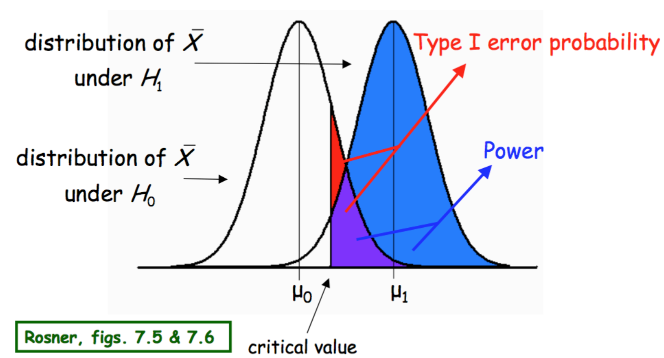


Figure from HST 190 notes (Betensky)

Power

- Trade-off between Type I and Type II errors
- Lower Type I error rate → higher Type II error rate & loss of power
- How do we compute power? (for one-sided test)

$$\boxed{1-\beta=\Phi\left(-z_{1-\alpha}+\left|\mu_1-\mu_0\right|\frac{\sqrt{n}}{\sigma}\right)}$$

$\Phi(c)=P(Z\le c)$ (area to the left of c for Z~N(0,1))

- What are the factors affecting power? n, σ, α, $|\mu_1-\mu_0|$

Example

- Researchers wish to test whether a drug is effective in reducing intraocular pressure by 5mm Hg (for glaucoma prevention). They will conclude effectiveness if they see a result significant at the 5% level. The std. dev. of the pressure change, on the basis of previous experiments, is believed to be 20. If they administer the drug to 100 patients, what is the probability they will detect a change **if the drug is truly effective?**
- n=100, σ=20, α=0.05, $|\mu_1-\mu_0|$=5

$$1-\beta=\Phi\left(-z_{1-\alpha}+\left|\mu_1-\mu_0\right|\frac{\sqrt{n}}{\sigma}\right)$$

$$=\Phi\left(-z_{0.95}+5\frac{\sqrt{100}}{20}\right)=\Phi\left(-1.68+\frac{5}{2}\right)$$

$$=\Phi(0.82)=0.794=\text{power}$$

Sample size

- Instead of fixing n and calculating power, we can select desired level of power and calculate n.

$$\boxed{n=\frac{\sigma^2\left(z_{1-\beta}+z_{1-\alpha}\right)^2}{\left(\mu_1-\mu_0\right)^2}}$$

- For a two-sided test, substitute $z_{1-\alpha/2}$ for $z_{1-\alpha}$ for all these formulas