Matematisk Statistik och Disktret Matematik, MVE051/MSG810, VT20

Föreläsning 11

Nancy Abdallah

Chalmers - Göteborgs Universitet

Comparing two means

In a similar way as for the proportions, we can compare the means of two different populations.

- Suppose we have two populations 1 and 2 with mean μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively.
- From each population we take a random sample such that the samples are independent from each other.
- For each sample we compute the point estimator for the mean: $\hat{\mu}_1 = \overline{X}_1$ and $\hat{\mu}_2 = \overline{X}_2$.
- A point estimator for $\mu_1 - \mu_2$ is $\hat{\mu}_1 - \hat{\mu}_2 = \overline{X}_1 - \overline{X}_2$.
- When sampling from a normal distribution, or if the sample sizes are large enough, $\overline{X}_1 - \overline{X}_2$ is normally distributed with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2/n_1 + \sigma_2^2/n_2$.

C.I. on the difference between two means

Populations normally distributed (or *n* large) with known variances

- A $100(1-\alpha)\%$ C.I. on $\mu_1 - \mu_2$ is given by

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$$

- If the population variances are equal to each other ($\sigma_1^2 = \sigma_2^2$) then the C.I. is

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sigma\sqrt{1/n_1 + 1/n_2}$$

C.I. on the difference between two means

Populations normally distributed with same unknown variance

Suppose $\sigma_1 = \sigma_2 = \sigma$, both unknown.

- **Pooled variance**

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

- The pooled variance is an unbiased estimator for σ_2^2.
- The random variable

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2(1/n_1 + 1/n_2)}}$$

  follows the T-distribution with $n_1 + n_2 - 2$ degrees of freedom.

C.I. and hypothesis test on $\mu_1 - \mu_2$

Populations normally distributed (or n large), unknown variances, equal variances

- A $100(1-\alpha)\%$ C.I. on $\mu_1 - \mu_2$ when the populations are normally distributed with $\sigma_1 = \sigma_2 = \sigma$ unknown is

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2}\sqrt{S_p^2(1/n_1 + 1/n_2)}$$

- To compare two means we can also do a hypothesis test. The alternative hypothesis can be one of the following

$$\mu_1 > \mu_2, \quad \mu_1 < \mu_2 \ \text{ or } \ \mu_1 \neq \mu_2$$

and we use the same procedure as for hypothesis testing on one parameter.

Example (Exercise 10.14)

To decide whether or not to purchase a new hand-held laser scanner for use in inventorying stock, tests are conducted on the scanner currently in use and on the new scanner. There data are obtained on the number of 7-inch bar codes that can be scanned per second:

New	**Old**
$n_1 = 61$	$n_2 = 61$
$\bar{x}_1 = 40$	$\bar{x}_2 = 29$
$s_1^2 = 24.9$	$s_2^2 = 22.7$

1. Find the pooled variance.
2. Find a 90% confidence interval on $\mu_1 - \mu_2$.
3. Does the new laser appear to read more bar codes per second on the average?

Solution of Exercise 10.14

1. $s_2^p = \frac{60(24.9)+60(22.7)}{120} = 23.8$
2. T-distribution with 120 degrees of freedom. $t_{(\alpha/2} = t_{0.05} = 1.658$ (*note that the table does not give the values for degrees of freedom greater than 100, use then an approximation*). A 90% C.I. is therefore

$$(40 - 29 \pm 1.658\sqrt{23.8(1/61 + 1/61)}) = (9.54, 12.45)$$

3. Yes, since the interval does not contain 0 and is positive-valued.

Can we find the same result using hypothesis testing?
$H_0 : \mu_1 - \mu_2 = 0$ and $H_1 : \mu_1 - \mu_2 > 0$ The test statistic is

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_p^2(1/n_1 + 1/n_2)}}$$

which follows a T-distribution with n_1+n_2-2 degrees of freedom. It is a right-tailed test with critical value $t_\alpha = t_{0.1} = 1.289$. $T = \frac{40-29}{\sqrt{23.8(1/61+1/61)}} = 12.45 > t_{0.1}$. Hence, we reject the hypothesis and we are 90% confident that H_1 is true.