MVE055/MSG810 2017 Lecture 10

Petter Mostad

Chalmers

November 29, 2017

Confidence interval for difference between expectations, unknown $\sigma_1 = \sigma_2 = \sigma$.

Assume we have a random sample $X_1, \dots, X_{n_1} \sim \text{Normal}(\mu_1, \sigma)$ and an independent random sample $Y_1, \dots, Y_{n_2} \sim \text{Normal}(\mu_2, \sigma)$.

- If
$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$
denotes the pooled variance estimate, if $t_{\alpha/2}$ is so that $[-t_{\alpha/2}, t_{\alpha/2}]$ contains $100(1-\alpha)\%$ of the probability in a t distribution with $n_1 + n_2 - 2$ degrees of freedom, and if
$$\begin{aligned} L_1 &= \overline{X} - \overline{Y} - t_{\alpha/2}\sqrt{S_p^2(1/n_1 + 1/n_2)} \\ L_2 &= \overline{X} - \overline{Y} + t_{\alpha/2}\sqrt{S_p^2(1/n_1 + 1/n_2)} \end{aligned}$$
then $[L_1, L_2]$ is a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$.
- The proof is based on first proving
$$\left(\overline{X} - \overline{Y} - (\mu_1 - \mu_2)\right) / \sqrt{S_p^2(1/n_1 + 1/n_2)} \sim \mathrm{T}(n_1 + n_2 - 2).$$

Confidence interval for difference between expectations, unknown σ_1 and σ_2.

Assume we have a random sample $X_1, \dots, X_{n_1} \sim \text{Normal}(\mu_1, \sigma_1)$ and an independent random sample $Y_1, \dots, Y_{n_2} \sim \text{Normal}(\mu_2, \sigma_2)$.

- If
$$\gamma = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1-1} + \frac{\left(S_2^2/n_2\right)^2}{n_2-1}},$$
if $t_{\alpha/2}$ is so that $[-t_{\alpha/2}, t_{\alpha/2}]$ contains $100(1-\alpha)\%$ of the probability in a t distribution with γ degrees of freedom, and if
$$\begin{aligned} L_1 &= \overline{X} - \overline{Y} - t_{\alpha/2}\sqrt{S_1^2/n_1 + S_2^2/n_2} \\ L_2 &= \overline{X} - \overline{Y} + t_{\alpha/2}\sqrt{S_1^2/n_1 + S_2^2/n_2} \end{aligned}$$
then $[L_1, L_2]$ is an approximate $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$.
- The proof is based on first proving that, approximately,
$$\left(\overline{X} - \overline{Y} - (\mu_1 - \mu_2)\right) / \sqrt{S_1^2/n_1 + S_2^2/n_2} \sim \text{T}(\gamma).$$

Hypothesis test for difference between expectations, unknown $\sigma_1 = \sigma_2 = \sigma$.

Assume we have a random sample $X_1, \dots, X_{n_1} \sim \text{Normal}(\mu_1, \sigma)$ and an independent random sample $Y_1, \dots, Y_{n_2} \sim \text{Normal}(\mu_2, \sigma)$.

- Assume we want to compare $H_0 : \mu_1 - \mu_2 = d_0$ with $H_1 : \mu_1 - \mu_2 \neq d_0$ for some known d_0 (often $d_0 = 0$).
- We choose as test statistic

  $$\frac{\overline{X} - \overline{Y} - d_0}{\sqrt{S_p^2(1/n_1 + 1/n_2)}}$$

  which has a $\text{T}(n_1 + n_2 - 2)$ distribution when H_0 is true.
- The rejection region are all values outside the interval $[-t_{\alpha/2}, t_{\alpha/2}]$, where $t_{\alpha/2}$ is so that $P[T > t_{\alpha/2}] = \alpha/2$ when T has a $\text{T}(n_1 + n_2 - 2)$ distribution.
- Correspondingly, one can make one-sided tests where we use t_α instead of $t_{\alpha/2}$, and significance tests, where we instead compute a p-value.

Hypothesis test for difference between expectations, unknown σ_1 and σ_2.

Assume we have a random sample $X_1, \dots, X_{n_1} \sim \text{Normal}(\mu_1, \sigma_1)$ and an independent random sample $Y_1, \dots, Y_{n_2} \sim \text{Normal}(\mu_2, \sigma_2)$.

- Assume we want to compare $H_0 : \mu_1 - \mu_2 = d_0$ with $H_1 : \mu_1 - \mu_2 \neq d_0$ for some known d_0 (often $d_0 = 0$).
- We choose as test statistic

$$\frac{\overline{X} - \overline{Y} - d_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

  which has a approximate $\text{T}(\gamma)$ distribution when H_0 is true (γ computed as before).
- The rejection region are all values outside the interval $[-t_{\alpha/2}, t_{\alpha/2}]$, where $t_{\alpha/2}$ is so that $P[T > t_{\alpha/2}] = \alpha/2$ when T has a $\text{T}(\gamma)$ distribution.
- Correspondingly, one can make one-sided tests where we use t_α instead of $t_{\alpha/2}$, and significance tests, where we instead compute a p-value.

Confidence interval for difference between expectations, known σ_1 and σ_2.

Assume we have a random sample $X_1, \dots, X_{n_1} \sim \text{Normal}(\mu_1, \sigma_1)$ and an independent random sample $Y_1, \dots, Y_{n_2} \sim \text{Normal}(\mu_2, \sigma_2)$.

- If we assume that σ_1 and σ_2 are known (so the distribution standard deviations, not the sample standard deviations), if $z_{\alpha/2}$ is so that $[-z_{\alpha/2}, z_{\alpha/2}]$ contains $100(1-\alpha)\%$ of the probaility in a standard normal distribution, and if

$$\begin{aligned} L_1 &= \overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \\ L_2 &= \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} \end{aligned}$$

  then $[L_1, L_2]$ is a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$.
- The proof is based on showing that

$$\overline{X} - \overline{Y} \sim \text{Normal}(\mu_1 - \mu_2, \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}).$$

General example of p-value computation

- Assume we are investigating a sequence of independent experiments which each result in success (1) or failure (0). Assume the probability for success is an unknown parameter p. Assume the data from 8 experiments is

$$0, 1, 0, 0, 1, 0, 0, 1$$

  Our null hypothesis for p is $H_0 : p \geq 0.6$, while the alternative hypothesis is $H_1 : p < 0.6$. What is the p value?
- To answer the question we must know which *test statistic* that should be used. Different test statistics give different results.
- Alternative 1: The procedure to obtain the test statistic is to do 8 experiments and let T be the number of successes.
- Alternative 2: The procedure to obtain the test statistic is to continue doing experimentns until 3 successes have been obtained, and let T be the number of experiments done.
- Using alternative 1, we get a p value of 0.174. If we use alternative 2 we get a p value of 0.095. So with a significance level of 0.1, we will reject H_0 using the second test statistic, but not using the first.

Exact computation of the p values in the example

- Alternative 1: If we assume $p = 0.6$ we get $T \sim \text{Binomial}(8, 0.6)$. The possible values for T and their probabilities are given in the table:

0	1	2	3	4	5	6	7	8
0.001	0.008	0.041	0.124	0.232	0.279	0.209	0.090	0.017

  The sum of the probabilities for $T = 0, 1, 2,$ or 3 is 0.174. This is the probability for the test statistic obtaining the observed value (3) or "something more extreme" in the direction of H_1. Thus it is the p values.

- Alternative 2: If we assume $p = 0.6$ we get $T \sim \text{Neg-Binomial}(3, 0.6)$. The possible values for T and their probabilities are given in the table:

3	4	5	6	7	8	9	10	11
0.216	0.259	0.207	0.138	0.083	0.046	0.025	0.013	0.006

12	13	14	15	16,17,...
0.003	0.001	0.001	0.000	totalt 0.000

  The sum of the probabilities that $T = 8, 9, 10, \ldots$ is 0.095. This is the probability for the test statistic obtaining the observed value (8) or "something more extreme" in the direction of H_1. Thus it is the p value.

Example: p values in development of new medicines

- Approval of new drugs is a firmly regulated process based on p values.
- Without regulation, it would be possible for drug companies to first perform medical testing and then choose between possible hypotheses, data, and test statistics to report the ones that give the best results for the company.
- To avoid this, companies must submit a detailed description of their experiments *and* how they will obtain their p value *before* the testing starts.
- One consequence is that it is possible for two companies to have performed exactly the same experiments and have obtained exactly the same results, while one gets approval and the other does not, if the first company has made a bet on a better way to compute the p value than the second company.

Usefulness of p values

- According to Milton, p values are "coming into widespread use" because of their "logical appeal".
- I would instead say that hypothesis testing in general, and p values in particular, are controversial.
- An example: The journal "Basic and Applied Social Psychology" decided on February 2015 to no longer accept papers with methodology depending on p values, as they are controversial.

The F distribution (extra material)

- If X_1 and X_2 are independent, if X_1 has a χ^2 distribution with γ_1 degrees of freedom, and if X_2 has a χ^2 distribution with γ_2 degrees of freedom, then
$$W = \frac{X_1/\gamma_1}{X_2/\gamma_2}$$
has an F distribution with γ_1 *and* γ_2 degrees of freedom, we write $W \sim \mathrm{F}(\gamma_1, \gamma_2)$.
- As $\mathbb{E}[X_1] = \gamma_1$ and $\mathbb{E}[X_2] = \gamma_2$ it is natural that the expected value for W is about 1; more exactly we have $\mathbb{E}[W] = \frac{\gamma_2}{\gamma_2 - 2}$.
- The F distribution is similar to a normal distribution when γ_1 and γ_2 are very large.
- There are tables for the F distribution combining various degrees of freedom.
- Note that if $W \sim \mathrm{F}(\gamma_1, \gamma_2)$ then $1/W \sim \mathrm{F}(\gamma_2, \gamma_1)$.

Hypthesis test for identical variances (extra material)

Assume we have a random sample $X_1, \ldots, X_{n_1} \sim \text{Normal}(\mu_1, \sigma_1)$ and an independent random sample $Y_1, \ldots, Y_{n_2} \sim \text{Normal}(\mu_2, \sigma_2)$.

- Assume we would like to compare $H_0 : \sigma_1 = \sigma_2$ with $H_1 : \sigma_1 \neq \sigma_2$.
- If S_1^2 and S_2^2 are the sample variances for the X's and the Y's, respectively, and if we choose the test statistic

  $$S_1^2 / S_2^2$$

  then it has an $\text{F}(n_1 - 1, n_2 - 1)$ distribution when H_0 is true.
- The rejection region are all values outside $[F_{1-\alpha/2}(n_1 - 1, n_2 - 1), F_{\alpha/2}(n_1 - 1, n_2 - 1)]$, where $F_{\alpha/2}(\gamma_1, \gamma_2)$ is so that $P[W > F_{\alpha/2}(\gamma_1, \gamma_2)] = \alpha/2$ when W has an $F(\gamma_1, \gamma_2)$ distribution.
- Note: When using the test, define what is X and what is Y so that $S_1^2 > S_2^2$ (to facilitate use of tables).