11 The t-test

Statisticians do it with two-tail T tests.

11.1 Confidence interval for the mean, unknown variance

Suppose $X_1, \dots, X_n$ IID $N(\mu, \sigma^2)$, but now σ^2 is unknown. Recall

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}} \sim t_{n-1}\,.$$

where $\hat{\sigma}^2 = S_{XX}/(n-1)$. A $100(1-\alpha)\%$ confidence interval for μ follows from

$$1 - \alpha = \mathbb{P}\left(-t^{(n-1)}_{\alpha/2} \le \frac{\sqrt{n}(\bar{X} - \mu)}{\hat{\sigma}} \le t^{(n-1)}_{\alpha/2}\right) = \mathbb{P}\left(\bar{X} - \frac{t_{\alpha/2}\hat{\sigma}}{\sqrt{n}} \le \mu \le \bar{X} + \frac{t_{\alpha/2}\hat{\sigma}}{\sqrt{n}}\right)$$

where $t^{(n-1)}_{\alpha/2}$ is the 'upper $\alpha/2$ point of a t-distribution on $n-1$ degrees of freedom', i.e., such that $\mathbb{P}\left(T > t^{(n-1)}_{\alpha/2}\right) = \alpha/2$.

Example 11.1 *In 'Sexual activity and the lifespan of male fruitflies',* Nature*, 1981, Partridge and Farquhar report experiments which examined the cost of increased reproduction in terms of reduced longevity for male fruitflies. They kept numbers of male flies under different conditions. 25 males in one group were each kept with 1 receptive virgin female. 25 males in another group were each kept with 1 female who had recently mated. Such females will refuse to remate for several days. These served as a control for any effect of competition with the male for food or space. The groups were treated identically in number of anaesthetizations (using CO2) and provision of fresh food.*

To verify 'compliance' two days per week throughout the life of each experimental male, the females that had been supplied as virgins to that male were kept and examined for fertile eggs. The insemination rate declined from approximately 1 per day at age one week to about 0.6 per day at age eight weeks.

The data was as follows

Groups of 25 males kept with	mean life (days)	s.e.
1 uninterested female	64.80	15.6525
1 interested female	56.76	14.9284

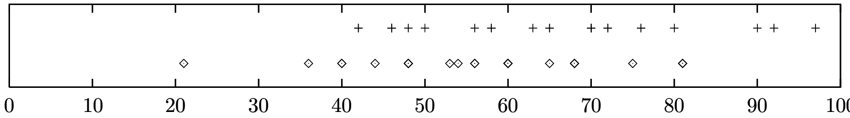
Here s.e. is an abbreviation for **standard error**, i.e. the value of $\hat{\sigma} = \sqrt{S_{xx}/(n-1)}$. Here $n = 25$. The mean life, $\bar{x}$ and the s.e., $\hat{\sigma}$, are sufficient statistics for (μ, σ^2), so there is nothing else we need to know about the individual values of the the longevities of these 50 flies in order to compute confidence intervals or test statistics.

From these summary statistics we can compute 95% confidence intervals for the mean lives of the control and test groups to be

$$[64.80 - 2.06(15.6526)/\sqrt{25}, 64.80 + 2.06(15.6526)/\sqrt{25}] = [58.35, 71.25]$$
$$[56.76 - 2.06(14.9284)/\sqrt{25}, 56.76 + 2.06(14.9284)/\sqrt{25}] = [50.61, 62.91]$$

It is interesting to look at the data, and doing so helps us check that lifespan is normally distributed about a mean. The longevities for control and test groups were

42 42 46 46 46 48 50 56 58 58 63 65 65 70 70 70 70 72 72 76 76 80 90 92 97
21 36 40 40 44 48 48 48 48 53 54 56 56 60 60 60 60 65 68 68 68 75 81 81 81



11.2 Single sample: testing a given mean, unknown variance (t-test)

Suppose that with the same assumptions as above it is required to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$.

Adopting the paradigm of the generalized likelihood ratio test we consider

$$\begin{aligned}
L_x(H_0, H_1) &= \frac{\max_{\mu,\sigma^2} f\left(x \mid \mu, \sigma^2\right)}{\max_{\sigma^2} f\left(x \mid \mu_0, \sigma^2\right)} \\
&= \frac{\left[2\pi \sum_i (x_i - \bar{x})^2/n\right]^{-n/2} \exp\left[-n/2\right]}{\left[2\pi \sum_i (x_i - \mu_0)^2/n\right]^{-n/2} \exp\left[-n/2\right]} \\
&= \left[\frac{\sum_i (x_i - \mu_0)^2}{\sum_i (x_i - \bar{x})^2}\right]^{n/2} \\
&= \left[\frac{\sum_i (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2}{\sum_i (x_i - \bar{x})^2}\right]^{n/2} \\
&= \left[1 + \frac{n(\bar{x} - \mu_0)^2}{\sum_i (x_i - \bar{x})^2}\right]^{n/2}\,.
\end{aligned}$$

This is large when $T^2 := n(n-1)(\bar{x} - \mu_0)^2 \,/\, \sum_i (x_i - \bar{x})^2$ is large, equivalently when $|T|$ is large. Under H_0 we have $T \sim t_{n-1}$. So a size α test is the two-tailed test which rejects H_0 if $t > t^{(n-1)}_{\alpha/2}$ or if $t < -t^{(n-1)}_{\alpha/2}$.

Example 11.2 *Does jogging lead to a reduction in pulse rate? Eight non-jogging volunteers engaged in a one-month jogging programme. Their pulses were taken before and after the programme.*

pulse rate before	74	86	98	102	78	84	79	70
pulse rate after	70	85	90	110	71	80	69	74
decrease	4	1	8	-8	7	4	10	-4

Although there are two sets of data it is really just the changes that matter. Let the decreases in pulse rates be $x_1, \ldots, x_8$ and assume these are samples from $N(\mu, \sigma^2)$ for some unknown σ^2. To test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ we compute

$$\sum x_i = 22, \quad \bar{x} = 2.75, \quad \sum x_i^2 = 326, \quad S_{xx} = \sum x_i^2 - 8\bar{x}^2 = 265.5.$$

Hence the test statistic is

$$t = \frac{\sqrt{8}(2.75 - 0)}{\sqrt{265.5/(8-1)}} = 1.263,$$

which is to be compared to $t_{0.025}^{(7)} = 2.365$. Hence the data is not sufficient to reject H_0 at the 5% level. This may surprise you since 6 of the 8 subjects had lowered pulse rates. This sort of test is called a **paired samples t-test**.

11.3 Two samples: testing equality of means, unknown common variance (t-test)

We have the same samples as in 8.5, i.e., $X_1, \ldots, X_m$ are IID $N(\mu_1, \sigma^2)$ and $Y_1, \ldots, Y_n$ are IID $N(\mu_2, \sigma^2)$. These two samples are independent. It is required to test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$, but now σ^2 is unknown. Note how this differs from the pair t-test above: the samples are not paired, and can be of unequal sizes.

As above, a maximum likelihood approach could convince us that the test should be of the form 'reject H_0 if $(\bar{x} - \bar{y})^2/(S_{xx} + S_{yy})$ is large.

As in 8.5 we have that under H_0,

$$(\bar{X} - \bar{Y}) \left(\frac{1}{m} + \frac{1}{n}\right)^{-\frac{1}{2}} \frac{1}{\sigma} \sim N(0,1)\,.$$

If $S_{XX} = \sum_{i=1}^{m}(X_i - \bar{X})^2$, and $S_{YY} = \sum_{j=1}^{n}(Y_j - \bar{Y})^2$, then

$$(S_{XX} + S_{YY})/\sigma^2 \sim \chi^2_{m+n-2}$$

so that (since $\bar{X}$ is independent of S_{XX}, $\bar{Y}$ is independent of S_{YY} and the two samples are independent)

$$T = (\bar{X} - \bar{Y}) \Big/ \sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\left(\frac{S_{XX} + S_{YY}}{m+n-2}\right)} \sim t_{m+n-2}.$$

A test of size α rejects H_0 if $t > t_{\alpha/2}^{(m+n-2)}$ or if $t < -t_{\alpha/2}^{(m+n-2)}$.

Example 11.3 *For the fruitfly data we might test H_0 : that mean longevity is the same for males living with 8 interested females as with 8 uninterested females.* The test statistic is

$$t = (64.80 - 56.76) \Big/ \sqrt{\left(\frac{1}{25} + \frac{1}{25}\right)\left(\frac{24(15.6525) + 24(14.9284)}{25 + 25 - 2}\right)} = 1.859$$

which can be compared to $t_{0.025}^{(48)} = 2.01$, and therefore is not significant at the 0.05% level. H_0 is not rejected. (It is however, significant at the 10% level, since $t_{0.05}^{(48)} = 1.68$).

Similarly, we can give a 95% confidence interval for the difference of the means. This has endpoints

$$\begin{aligned}(64.80 - 56.76) \pm 2.01 &\sqrt{\left(\frac{1}{25} + \frac{1}{25}\right)\left(\frac{24(15.6525) + 24(14.9284)}{25 + 25 - 2}\right)} \\ &= 8.04 \pm 8.695.\end{aligned}$$

I.e., a 95% confidence interval for the extra longevity of celibate males is $[-0.655, 16.735]$ days. Notice again that finding we cannot reject $\mu_1 - \mu_2 = 0$ at the 5% level is equivalent to finding that the 95% confidence interval for the difference of the means contains 0.

In making the above test we have assumed that the variances for the two populations are the same. In the next lecture we will see how we might test that hypothesis.

11.4 Single sample: testing a given variance, unknown mean (χ^2-test)

Let $X_1, \ldots, X_n$ be IID $N(\mu, \sigma^2)$, and suppose we wish to test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$, where μ is unknown, and therefore a '**nuisance parameter**'.

Following Theorem 8.1, the likelihood ratio is

$$L_x(H_0, H_1) = \frac{\max_{\mu,\sigma^2} f\left(x \mid \mu, \sigma^2\right)}{\max_{\mu} f\left(x \mid \mu, \sigma_0^2\right)} = \frac{\left[2\pi \frac{\sum_i (x_i - \bar{x})^2}{n}\right]^{-n/2} \exp\left[-n/2\right]}{(2\pi\sigma_0^2)^{-n/2} \exp\left[-(1/2\sigma_0^2)\sum_i (x_i - \bar{x})^2\right]}$$

As in Section 8.4 this is large when $\sum_i (x_i - \bar{x})/n\sigma_0^2$ $(= S_{xx}/n\sigma_0^2)$ differs substantially from 1.

Under H_0, $S_{XX}/\sigma_0^2 \sim \chi^2_{n-1}$. Given the required size of test α, let a_1, a_2 be such that

$$\mathbb{P}(S_{XX}/\sigma_0^2 < a_1) + \mathbb{P}(S_{XX}/\sigma_0^2 > a_2) = \alpha$$

under H_0. Then a size α test is to reject H_0 if $S_{xx}/\sigma_0^2 < a_1$ or if $S_{xx}/\sigma_0^2 > a_2$.

Usually we would take $a_1 = F_{n-1}^{-1}(\alpha/2), a_2 = F_{n-1}^{-1}(1 - \alpha/2)$, where F_{n-1} is the distribution function of a χ^2_{n-1} random variable.