Lecture 16. Linear model examples, and 'rules of thumb'

Two samples: testing equality of means, unknown common variance.

- Suppose we have two independent samples, $X_1, \ldots, X_m$ iid $N(\mu_X, \sigma^2)$, and $Y_1, \ldots Y_n$ iid $N(\mu_Y, \sigma^2)$, with σ^2 unknown.
- We wish to test $H_0 : \mu_X = \mu_Y = \mu$ against $H_1 : \mu_X \neq \mu_Y$.

Using the generalised likelihood ratio test

- $L_{\mathbf{x},\mathbf{y}}(H_0) = \sup_{\mu,\sigma^2} f_{\mathbf{X}}(\mathbf{x} \,|\, \mu, \sigma^2) f_{\mathbf{Y}}(\mathbf{y} \,|\, \mu, \sigma^2)$.
- Under H_0 the mle's are

$$\begin{aligned} \hat{\mu} &= (m\bar{x} + n\bar{y})/(m+n) \\ \hat{\sigma}_0^2 &= \tfrac{1}{m+n}\left(\sum (x_i - \hat{\mu})^2 + \sum (y_i - \hat{\mu})^2\right) = \tfrac{1}{m+n}\left(S_{xx} + S_{yy} + \tfrac{mn}{m+n}(\bar{x} - \bar{y})^2\right) \end{aligned}$$

so

$$\begin{aligned} L_{\mathbf{x},\mathbf{y}}(H_0) &= (2\pi\hat{\sigma}_0^2)^{-(m+n)/2} e^{-\frac{1}{2\hat{\sigma}_0^2}\left(\sum (x_i - \hat{\mu})^2 + \sum (y_i - \hat{\mu})^2\right)} \\ &= (2\pi\hat{\sigma}_0^2)^{-(m+n)/2} e^{-\frac{m+n}{2}}. \end{aligned}$$

- Similarly

$$L_{\mathbf{x},\mathbf{y}}(H_1) = \sup_{\mu_X,\mu_Y,\sigma^2} f_{\mathbf{X}}(\mathbf{x} \mid \mu_X, \sigma^2) f_{\mathbf{Y}}(\mathbf{y} \mid \mu_Y, \sigma^2) = (2\pi\hat{\sigma}_1^2)^{-(m+n)/2} e^{-\frac{m+n}{2}},$$

achieved by $\hat{\mu}_X = \bar{x}$, $\hat{\mu}_Y = \bar{y}$ and $\hat{\sigma}_1^2 = (S_{xx} + S_{yy})/(m+n)$.

- Hence

$$\Lambda_{\mathbf{x},\mathbf{y}}(H_0; H_1) = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_1^2}\right)^{(m+n)/2} = \left(1 + \frac{mn(\bar{x}-\bar{y})^2}{(m+n)(S_{xx}+S_{yy})}\right)^{(m+n)/2}.$$

- We reject H_0 if $mn(\bar{x}-\bar{y})^2/\big((S_{xx}+S_{yy})(m+n)\big)$ is large, or equivalently if

$$|t| = \frac{|\bar{x}-\bar{y}|}{\sqrt{\frac{S_{xx}+S_{yy}}{n+m-2}\left(\frac{1}{m}+\frac{1}{n}\right)}}$$

is large.

- Under H_0, $\bar{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{m})$, $\bar{Y} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$ and so

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

- From Theorem 16.3 we know $S_{XX}/\sigma^2 \sim \chi^2_{m-1}$ independently of $\bar{X}$ and $S_{YY}/\sigma^2 \sim \chi^2_{n-1}$ independently of $\bar{Y}$.
- Hence $(S_{XX} + S_{YY})/\sigma^2 \sim \chi^2_{n+m-2}$, from additivity of independent χ^2 distributions.
- Since our two random samples are independent, we have $\bar{X} - \bar{Y}$ and $S_{XX} + S_{YY}$ are independent.
- This means that under H_0,

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

- A size α test is to reject H_0 if $|t| > t_{n+m-2}(\alpha/2)$.

Example 16.1

Seeds of a particular variety of plant were randomly assigned either to a nutritionally rich environment (the treatment) or to the standard conditions (the control). After a predetermined period, all plants were harvested, dried and weighed, with weights as shown below in grams.

Control	4.17	5.58	5.18	6.11	4.50	4.61	5.17	4.53	5.33	5.14
Treatment	4.81	4.17	4.41	3.59	5.87	3.83	6.03	4.89	4.32	4.69

- Control observations are realisations of $X_1, \ldots, X_{10}$ iid $N(\mu_X, \sigma^2)$, and for the treatment we have $Y_1, \ldots, Y_{10}$ iid $N(\mu_Y, \sigma^2)$.
- We test $H_0 : \mu_X = \mu_Y$ vs $H_1 : \mu_X \neq \mu_Y$.
- Here $m = n = 10$, $\bar{x} = 5.032$, $S_{xx} = 3.060$, $\bar{y} = 4.661$ and $S_{yy} = 5.669$, so $\tilde{\sigma}^2 = (S_{xx} + S_{yy})/(m + n - 2) = 0.485$.
- Then $|t| = |\bar{x} - \bar{y}|/\sqrt{\tilde{\sigma}^2(\frac{1}{m} + \frac{1}{n})} = 1.19$.
- From tables $t_{18}(0.025) = 2.101$, so we do not reject H_0. We conclude that there is no evidence for a difference between the mean weights due to the environmental conditions.

Arranged as analysis of variance:

Source of variation	d.f.	sum of squares	mean square	F statistic
Fitted model	1	$\frac{mn}{m+n}(\bar{x}-\bar{y})^2$	$\frac{mn}{m+n}(\bar{x}-\bar{y})^2$	$F = \frac{mn}{m+n}(\bar{x}-\bar{y})^2/\tilde{\sigma}^2$
Residual	$m+n-2$	$S_{xx}+S_{yy}$	$\tilde{\sigma}^2$	
	$m+n-1$			

Seeing if $F > F_{1,m+n-2}(\alpha)$ is exactly the same as checking if $|t| > t_{n+m-2}(\alpha/2)$.

Notice that although we have equal size samples here, they are not paired; there is nothing to connect the first plant in the control sample with the first plant in the treatment sample.

Paired observations

- Suppose the observations *were* paired: say because pairs of plants were randomised.
- We can introduce a parameter γ_i for the ith pair, where $\sum_i \gamma_i = 0$, so that we assume

$$X_i \sim \mathrm{N}(\mu_X + \gamma_i, \sigma^2), \quad Y_i \sim \mathrm{N}(\mu_Y + \gamma_i, \sigma^2), \; i = 1, .., n,$$

  and all independent.
- Working through the generalised likelihood ratio test, or expressing in matrix form, leads to the intuitive conclusion that we should work with the differences $D_i = X_i - Y_i, i = 1, .., n$, where

$$D_i \sim \mathrm{N}(\mu_X - \mu_Y, \phi^2), \text{ where } \phi^2 = 2\sigma^2.$$

- Thus $\overline{D} \sim \mathrm{N}(\mu_X - \mu_Y, \frac{\phi^2}{n})$, and we test $H_0 : \mu_X - \mu_Y = 0$ by the t statistic

$$t = \frac{\overline{D}}{\tilde{\phi}/\sqrt{n}},$$

  where $\tilde{\phi}^2 = S_{DD}/(n-1) = \sum_i (D_i - \overline{D})^2/(n-1)$, and $t \sim t_{n-1}$ distribution under H_0.

Example 16.2

Pairs of seeds of a particular variety of plant were sampled, and then one of each pair randomly assigned either to a nutritionally rich environment (the treatment) or to the standard conditions (the control).

Pair	1	2	3	4	5	6	7	8	9	10
Control	4.17	5.58	5.18	6.11	4.50	4.61	5.17	4.53	5.33	5.14
Treatment	4.81	4.17	4.41	3.59	5.87	3.83	6.03	4.89	4.32	4.69
Difference	- 0.64	1.41	0.77	2.52	-1.37	0.78	-0.86	-0.36	1.01	0.45

- Observed statistics are $\overline{d} = 0.37, S_{dd} = 12.54, n = 10$, so that $\tilde{\phi} = \sqrt{S_{dd}/(n-1)} = \sqrt{2.33/9} = 1.18$.
- Thus $t = \frac{\overline{d}}{\tilde{\phi}/\sqrt{n}} = \frac{0.37}{1.18/\sqrt{10}} = 0.99$.
- This can be compared to $t_{18}(0.025) = 2.262$ to show that we cannot reject $H_0 : \mathbb{E}(D) = 0$, i.e. that there is no effect of the treatment.
- Alternatively, we see that the observed p-value is the probability of getting such an extreme result, under H_0, i.e.

$$\mathbb{P}(|t_9| > |t||H_0) = 2\mathbb{P}(t_9 > |t|) = 2 \times 0.17 = 0.34.$$

In R code:

```
> t.test(x,y,paired=T)

t = 0.9938, df = 9, p-value = 0.3463
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4734609  1.2154609
sample estimates:
mean of the differences
                  0.371
```

Rules of thumb: the 'rule of three'*

Rules of Thumb 16.3

If there have been n opportunities for an event to occur, and yet it has not occurred yet, then we can be 95% confident that the chance of it occurring at the next opportunity is less than $3/n$.

- Let p be the chance of it occurring at each opportunity. Assume these are independent Bernoulli trials, so essentially we have $X \sim \text{Binom}(n, p)$, we have observed $X = 0$, and want a one-sided 95% CI for p.
- Base this on the set of values that cannot be rejected at the 5% level in a one-sided test.
- i.e. the 95% interval is $(0, p')$ where the one-sided p-value for p' is 0.05, so

$$0.05 = \mathbb{P}(X = 0|p') = (1 - p')^n.$$

- Hence

$$p' = 1 - e^{\log(0.05)/n} \approx \frac{-\log(0.05)}{n} \approx \frac{3}{n},$$

since $\log(0.05) = -2.9957$.

- For example, suppose we have given a drug to 100 people and none of them have had a serious adverse reaction.
- Then we can be 95% confident that the chance the next person has a serious reaction is less than 3%.
- The exact p' is $1 - e^{\log(0.05)/100} = 0.0295$.

Rules of thumb: the 'rule of root n'*

Rules of Thumb 16.4

After n observations, if the number of events differs from that expected under a null hypothesis H_0 *by more than* $\sqrt{n}$*, reject* H_0*.*

- We assume $X \sim \text{Binom}(n, p)$, and $H_0 : p = p_0$, so the expected number of events is $\mathbb{E}(X|H_0) = np_0$.
- Then the probability of the difference between observed and expected exceeding $\sqrt{n}$, given H_0 is true, is

$$
\begin{aligned}
\mathbb{P}(|X - np_0| > \sqrt{n}|H_0) &= \mathbb{P}\left(\frac{|X - np_0|}{\sqrt{np_0(1-p_0)}} > \frac{1}{\sqrt{p_0(1-p_0)}}\middle| H_0\right) \\
&< \mathbb{P}\left(\frac{|X - np_0|}{\sqrt{np_0(1-p_0)}} > 2\middle| H_0\right) \text{ since } \frac{1}{\sqrt{p_0(1-p_0)}} > 2 \\
&\approx \mathbb{P}(|Z| > 2) \\
&\approx 0.05
\end{aligned}
$$

- For example, suppose we flip a coin 1000 times and it comes up heads 550 times, do we think the coin is odd?
- We expect 500 heads, and observe 50 more. $\sqrt{n} = \sqrt{1000} \approx 32$, which is less than 50, so this suggests the coin is odd.
- The 2-sided p-value is actually $2 \times \mathbb{P}(X \geq 550) = 2 \times (1 - \mathbb{P}(X \leq 549))$, where $X \sim \text{Binom}(1000, 0.5)$, which according to R is

```
> 2 * (1 - pbinom(549,1000,0.5))

0.001730536
```

Rules of thumb: the 'rule of root 4 × expected'*

The 'rule of root *n*' is fine for chances around 0.5, but is too lenient for rarer events, in which case the following can be used.

Rules of Thumb 16.5

After n observations, if the number of rare events differs from that expected under a null hypothesis H_0 by more than $\sqrt{4 \times expected}$, reject H_0.

- We assume $X \sim \text{Binom}(n, p)$, and $H_0 : p = p_0$, so the expected number of events is $\mathbb{E}(X|H_0) = np_0$.
- Under H_0, the critical difference is $\approx 2 \times \text{s.e.}(X - np_0) = \sqrt{4np_0(1 - p_0)}$, which is less than $\sqrt{n}$: this is the rule of root n.
- But $\sqrt{4np_0(1 - p_0)} < \sqrt{4np_0}$, which will be less than $\sqrt{n}$ if $p_0 < 0.25$.
- So for smaller p_0, a more powerful rule is to reject H_0 if the difference between observed and expected is greater than $\sqrt{4 \ \times \ \text{expected}}$.
- This is essentially a Poisson approximation.

- For example, suppose we throw a die 120 times and it comes up 'six' 30 times; is this 'significant'?
- We expect 20 sixes, and so the difference between observed and expected is 10.
- Since $\sqrt{n} = \sqrt{120} \approx 11$, which is more than 10, the 'rule of root n' does not suggest a significant difference.
- But since $\sqrt{4 \times \text{ expected}} = \sqrt{80} \approx 9$, the second rule does suggest significance.
- The 2-sided p-value is actually $2 \times \mathbb{P}(X \geq 30) = 2 \times (1 - \mathbb{P}(X \leq 29))$, where $X \sim \text{Binom}(120, \frac{1}{6})$, which according to R is

```
> 2 * (1 - pbinom(29,120, 1/6 ))

0.02576321
```

Rules of thumb: non-overlapping confidence intervals*

Rules of Thumb 16.6

Suppose we have 95% confidence intervals for μ_1 and μ_2 based on independent estimates $\overline{y}_1$ and $\overline{y}_2$. Let $H_0 : \mu_1 = \mu_2$.

(1) *If the confidence intervals* **do not** *overlap, then we can reject H_0 at $p < 0.05$.*

(2) *If the confidence intervals* **do** *overlap, then this does not necessarily imply that we cannot reject H_0 at $p < 0.05$.*

- Assume for simplicity that the confidence intervals are based on assuming $\overline{Y}_1 \sim \mathrm{N}(\mu_1, s_1^2), \overline{Y}_2 \sim \mathrm{N}(\mu_2, s_2^2)$, where s_1 and s_2 are known standard errors.
- Suppose wlg that $\overline{y}_1 > \overline{y}_2$. Then since $\overline{Y}_1 - \overline{Y}_2 \sim \mathrm{N}(\mu_1 - \mu_2, s_1^2 + s_2^2)$, we can reject H_0 at $\alpha = 0.05$ if

$$\overline{y}_1 - \overline{y}_2 > 1.96\sqrt{s_1^2 + s_2^2}.$$

- The two CIs will not overlap if

$$\overline{y}_1 - 1.96 s_1 > \overline{y}_2 + 1.96 s_2, \text{ i.e. } \overline{y}_1 - \overline{y}_2 > 1.96(s_1 + s_2).$$

- But since $s_1 + s_2 > \sqrt{s_1^2 + s_2^2}$ for positive s_1, s_2, we have the 'rule of thumb'.

- Non-overlapping CIs is a more stringent criterion: we cannot conclude 'not significantly different' just because CIs overlap.
- So, if 95% CIs just touch, what is the p-value?
- Suppose $s_1 = s_2 = s$. Then CIs just touch if $|\bar{y}_1 - \bar{y}_2| = 1.96 \times 2s = 3.92 \times s$.
- So p-value $=$

$$\begin{aligned} \mathbb{P}(|\overline{Y}_1 - \overline{Y}_2| > 3.92s) &= \mathbb{P}\left(\left|\frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{2}s}\right| > \frac{3.92}{\sqrt{2}}\right) \\ &= \mathbb{P}(|Z| > 2.77) = 2 \times \mathbb{P}(Z > 2.77) = 0.0055. \end{aligned}$$

- And if 'just not touching' $100(1-\alpha)\%$ CIs were to be equivalent to 'just rejecting H_0', then we would need to set α so that the critical difference between $\bar{y}_1 - \bar{y}_2$ was exactly the width of each of the CIs, and so

$$1.96 \times \sqrt{2} \times s = s \times \Phi^{-1}(1 - \tfrac{\alpha}{2}).$$

- Which means $\alpha = 2 \times \Phi(-1.96/\sqrt{2}) = 0.16$.
- So in these specific circumstances, we would need to use 84% intervals in order to make non-overlapping CIs the same as rejecting H_0 at the 5% level.