8.6 Comparing experiments: unpaired data

MATHEMATICAL TECHNIQUES

♠ Recall the data in exercises 8.5.1 and 8.5.2. Suppose that plants in the experimental plot are being compared with those in equal sized populations in a control plot. Compare using both one and two-tailed tests, and compare the p-values with those found in the earlier problems.

- EXERCISE **8.6.1**

The 10 plants in the control plot have mean weight 10.0, and the variance for weight in both populations is known to be 9.0.

- EXERCISE **8.6.2**

The 10 plants in the control plot have mean height 38.0, and the variance for height in both populations is known to be 16.0.

♠ Test the null hypothesis that the means from two populations are equal in the following cases. $\hat{\mu}_1$ and $\hat{\mu}_2$ are sample means found from samples with size n_1 and n_2 drawn from normal distributions with known variances σ_1^2 and σ_2^2. State the significance level of the test. Use a two-tailed test.

- EXERCISE **8.6.3**

$\hat{\mu}_1 = 20.0$, $\hat{\mu}_2 = 25.0$, $n_1 = 25$, $n_2 = 25$, $\sigma_1^2 = 25.0$, $\sigma_2^2 = 25.0$.

- EXERCISE **8.6.4**

$\hat{\mu}_1 = 20.0$, $\hat{\mu}_2 = 21.0$, $n_1 = 25$, $n_2 = 50$, $\sigma_1^2 = 25.0$, $\sigma_2^2 = 25.0$.

- EXERCISE **8.6.5**

$\hat{\mu}_1 = 20.0$, $\hat{\mu}_2 = 21.0$, $n_1 = 50$, $n_2 = 100$, $\sigma_1^2 = 16.0$, $\sigma_2^2 = 16.0$.

- EXERCISE **8.6.6**

$\hat{\mu}_1 = 20.0$, $\hat{\mu}_2 = 21.0$, $n_1 = 200$, $n_2 = 100$, $\sigma_1^2 = 16.0$, $\sigma_2^2 = 9.0$.

♠ Use p-values to test the null hypothesis of equal means against an alternative that $\mu_2 > \mu_1$ when sample means of $\hat{\mu}_1$ and $\hat{\mu}_2$ are found from samples of size n_1 and n_2 with sample variances s_1^2 and s_2^2. Use a one-tailed test. State the significance level of the test.

- EXERCISE **8.6.7**

$\hat{\mu}_1 = 70.0$, $\hat{\mu}_2 = 74.0$, $n_1 = 50$, $n_2 = 50$, $s_1^2 = 25.0$, $s_2^2 = 25.0$.

- EXERCISE **8.6.8**

$\hat{\mu}_1 = 70.0$, $\hat{\mu}_2 = 70.4$, $n_1 = 500$, $n_2 = 500$, $s_1^2 = 25.0$, $s_2^2 = 25.0$ (the difference between means is 10 times smaller, but the populations are 10 times larger). Why is the significance level worse than in exercise 8.6.7?

- EXERCISE **8.6.9**

$\hat{\mu}_1 = 70.0$, $\hat{\mu}_2 = 74.0$, $n_1 = 500$, $n_2 = 500$, $s_1^2 = 250.0$, $s_2^2 = 250.0$. Why does this match the answer to exercise 8.6.7?

- EXERCISE **8.6.10**

$\hat{\mu}_1 = 70.0$, $\hat{\mu}_2 = 70.4$, $n_1 = 500$, $n_2 = 500$, $s_1^2 = 2.5$, $s_2^2 = 2.5$. Why does this match the answer to exercise 8.6.7?

♠ Use the normal approximation to test the null hypothesis that men and women have the same opinions in the following cases. State the significance level of a two-tailed test.

- EXERCISE **8.6.11**

35 out of 50 men believe that if dolphins were so smart they could find their way out of those nets. 40 out of 50 women believe this.

- EXERCISE **8.6.12**

350 out of 500 men and 400 out of 500 women.

- EXERCISE **8.6.13**

35 out of 50 men and 400 out of 500 women.

- EXERCISE **8.6.14**

70 out of 1000 men and 40 out of 500 women. Why do you think the difference is not significant even though the samples are very large?

♠ Use the method of support to test whether the following samples differ.

- EXERCISE **8.6.15**

One player makes 5 out of 10 shots, another makes 9 out of 10.

• EXERCISE **8.6.16**
One player makes 5 out of 10 shots, another makes 16 out of 20.

• EXERCISE **8.6.17**
A one meter square region in Utah is hit by 4 cosmic rays in one year, and a one meter square region at the north pole is hit by 10 cosmic rays in one year.

• EXERCISE **8.6.18**
Two one meter square regions in Utah are hit by 3 and 5 cosmic rays in one year, and a one meter square region at the north pole is hit by 10 cosmic rays in one year.

♠ Algorithm 8.6 uses $\hat{p}_1$ and $\hat{p}_2$ to estimate the variance under the null hypothesis.

• EXERCISE **8.6.19**
Why might it make more sense to use $\hat{p}$, the proportion in the pooled sample? What is the pooled proportion if 96 out of 200 are found in the control and 54 out of 100 in the treatment?

• EXERCISE **8.6.20**
Redo the test in the test using $\hat{p}$. How different are the results? Under what circumstances might it make a larger difference which was used?

♠ Show that the two sample test turns into the one sample test as n_1 approaches infinity.

• EXERCISE **8.6.21**
What is the null hypothesis about the difference between means?

• EXERCISE **8.6.22**
What is the distribution of sample means in the treatment population under the null hypothesis?

APPLICATIONS

♠ A cell is placed in a medium with volume equal to that of the cell. 100 marked molecules are placed inside, and after a long time, 40 are found inside and 60 are found outside. In a control, the membrane has been modified and 50 are found inside and 50 are found outside. A scientist is attempting to determine whether the cell membrane acts as a filter.

• EXERCISE **8.6.23**
What is the null hypothesis if the treatment is compared with the control? What is the null hypothesis if the treatment is compared with the expectation that molecules end up inside and outside with equal probability?

• EXERCISE **8.6.24**
Find the p-value associated with the comparison of the treatment with the control, and the comparison of the treatment with the expectation that molecules end up inside and outside with equal probability. Why do the p-values differ as they do?

♠ One organism has 8 mutations in one million base pairs, a second has 18 in one million, and a third has 28. Use the normal approximation to test whether the following differences are significant.

• EXERCISE **8.6.25**
The difference between the first and second organisms.

• EXERCISE **8.6.26**
The difference between the second and third organisms. Why is the significance level different?.

♠ One organism has 8 mutations in one million base pairs, a second has 18 in one million, and a third has 28. Use the method of support to test the following differences.

• EXERCISE **8.6.27**
Check whether organisms 1 and 2 differ and compare with exercise 8.6.25.

• EXERCISE **8.6.28**
Check whether organisms 2 and 3 differ and compare with exercise 8.6.26.

♠ 8 mice are treated with a chemical and 5 get cancer. 6 mice were kept as controls, and 1 got cancer. 10 mice are given both the chemical and a putative anticancer drug, and only 1 of them gets cancer.

• EXERCISE **8.6.29**
Use the method of support on the first two sets of mice to check whether there is reason to think that the chemical is carcinogenic.

- EXERCISE **8.6.30**
  Use the method of support on the first and third sets of mice to check whether there is reason to think that the anticancer drug works. Is this result consistent with the previous exercise?

♠ It has been proposed that a particular salubrious bath extends cell lifespan. Suppose that cell mortality follows an exponential model. Use the method of support to evaluate the following cases.

- EXERCISE **8.6.31**
  A cell in the salubrious bath survives 30 minutes, and a cell in standard culture survives only 5. Is there reason to think that the salubrious bath lengthens cell life?
- EXERCISE **8.6.32**
  In a repeated experiment, the cell in the salubrious bath survives 60 minutes, and a cell in standard culture survives only 3. Is there reason to think that the salubrious bath lengthens cell life?
- EXERCISE **8.6.33**
  Combine the data from exercises 8.6.31 and 8.6.32, and evaluate the difference in support between the null and alternative hypotheses.
- EXERCISE **8.6.34**
  What would happen to the result in exercise 8.6.33 if a third experiment were done and both cells survived 10 minutes? Why the change?

Chapter 9

Answers

8.6.1. The distribution of the sample means from experimental plot is $N(\mu_1, 0.9)$ and for the control is $N(\mu_2, 0.9)$, because the standard error is $3.0/\sqrt{10} = 0.949$ in both cases. The distribution of the difference D is then $D \sim N(\mu_1 - \mu_2, 1.8)$. The difference in sample means is 0.99, so we must find the probability that $D \geq 0.99$ (for a one-tailed test) or that $D \geq 0.99$ or $D \leq -0.99$ (for a two-tailed test). Transforming into a standard normal distribution, the two-tailed test gives

$$\begin{aligned}\Pr(D \geq 0.99 \text{ or } D \leq -0.99) &= \Pr(\frac{D}{\sqrt{1.8}} \geq \frac{0.99}{\sqrt{1.8}} \text{ or } \frac{D}{\sqrt{1.8}} \leq \frac{-0.99}{\sqrt{1.8}}) \\ &= \Pr(Z \geq 0.738) + \Pr(Z \leq -0.738) \\ &= 1 - \Phi(0.738) + \Phi(-0.738) = 0.46.\end{aligned}$$

The difference is not significant with a two-tailed test. The p-value with a one-tailed test is half of this, 0.23, which is also not significant. These p-values are larger because of the uncertainty about the true mean of the control population.

8.6.3. The distribution of the sample means from the first sample is $N(\mu_1, 1.0)$ and for the second is $N(\mu_2, 1.0)$, because the standard error is $\sigma/\sqrt{n} = 1.0$ in both cases. The distribution of the difference D is then $D \sim N(\mu_1 - \mu_2, 2.0)$. We want to find the probability that $D \geq 5.0$ or $D \leq -5.0$. Transforming into a standard normal distribution,

$$\begin{aligned}\Pr(D \geq 5.0 \text{ or } D \leq -5.0) &= \Pr(\frac{D}{\sqrt{2.0}} \geq \frac{5.0}{\sqrt{2.0}} \text{ or } \frac{D}{\sqrt{2.0}} \leq \frac{-5.0}{\sqrt{2.0}}) \\ &= \Pr(Z \geq 3.53) + \Pr(Z \leq -3.53) \\ &= 1 - \Phi(3.53) + \Phi(-3.53) = 0.0046.\end{aligned}$$

The difference is very highly significant.

8.6.5. The distribution of the sample means from the first sample is $N(\mu_1, 0.32)$ and for the second is $N(\mu_2, 0.16)$, because the standard error is $4.0/\sqrt{50} = 0.566$ for the first population and $4.0/\sqrt{100} = 0.4$ for the second. The distribution of the difference D is then $D \sim N(\mu_1 - \mu_2, 0.48)$. We want to find the probability that $D \geq 1.0$ or $D \leq -1.0$. Transforming into a standard normal distribution, we get

$$\begin{aligned}\Pr(D \geq 1.0 \text{ or } D \leq -1.0) &= \Pr(\frac{D}{\sqrt{0.48}} \geq \frac{1.0}{\sqrt{0.48}} \text{ or } \frac{D}{\sqrt{0.48}} \leq \frac{-1.0}{\sqrt{0.48}}) \\ &= \Pr(Z \geq 1.443) + \Pr(Z \leq -1.443) \\ &= 1 - \Phi(1.443) + \Phi(-1.443) = 0.149.\end{aligned}$$

The difference is not significant.

8.6.7. The distribution of the sample means from the first sample is $N(\mu_1, 0.5)$ and for the second is $N(\mu_2, 0.5)$, because the standard error is $5/\sqrt{50} = \sqrt{1/2}$ for both populations. The distribution of the difference D is then

$D \sim N(\mu_1 - \mu_2, 1.0)$. We want to find the probability that $D \geq 4.0$. Transforming into a standard normal,

$$\Pr(D \geq 4.0) = \Pr\left(\frac{D}{\sqrt{1.0}} \geq \frac{4.0}{\sqrt{1.0}}\right) = \Pr(Z \geq 4.0) = 1 - \Phi(4.0) = 0.00003.$$

The difference is very highly significant.

8.6.9. The distribution of the sample means from the first sample is $N(\mu_1, 0.5)$ and for the second is $N(\mu_2, 0.5)$, because the standard error is $\sqrt{250/500} = \sqrt{1/2}$ for both populations. The distribution of the difference D is then $D \sim N(\mu_1 - \mu_2, 1.0)$. We want to find the probability that $D \geq 4.0$. Transforming into a standard normal,

$$\Pr(D \geq 4.0) = \Pr\left(\frac{D}{\sqrt{1.0}} \geq \frac{4.0}{\sqrt{1.0}}\right) = \Pr(Z \geq 4.0) = 1 - \Phi(4.0) = 0.00003.$$

The difference is very highly significant, and matches the earlier problem because the samples are 10 times larger, and which cancels the variances that are 10 times larger.

8.6.11. The distribution of the sample proportion from the first sample is $N(p_1, p_1(1-p_1)/n1)$ and for the second is $N(p_2, p_2(1-p_2)/n_2)$. Using $\hat{p}_1 = 0.7$, $\hat{p}_2 = 0.8$ and $n_1 = n_2 = 50$, we find that the distributions of the sample proportions are $N(p_1, 0.0042)$ and $N(p_2, 0.0032)$. The distribution of the difference D is then $D \sim N(p_1 - p_2, 0.0074)$. We want to find the probability that $D \geq 0.1$ or $D \leq -0.1$. Transforming into a standard normal distribution, we get

$$\begin{aligned}
\Pr(D \geq 0.1 \text{ or } D \leq -0.1) &= \Pr(\frac{D}{\sqrt{0.0074}} \geq \frac{0.1}{\sqrt{0.0074}} \text{ or } \frac{D}{\sqrt{0.0074}} \leq \frac{-0.1}{\sqrt{0.0074}}) \\
&= \Pr(Z \geq 1.16) + \Pr(Z \leq -1.16) \\
&= 1 - \Phi(1.16) + \Phi(-1.16) = 0.245.
\end{aligned}$$

The difference is not significant.

8.6.13. The distribution of the sample proportion from the first sample is $N(p_1, p_1(1-p_1)/n1)$ and for the second is $N(p_2, p_2(1-p_2)/n_2)$. Using $\hat{p}_1 = 0.7$, $\hat{p}_2 = 0.8$, $n_1 = 50$, and $n_2 = 500$, we find that the distributions of the sample proportions are $N(p_1, 0.0042)$ and $N(p_2, 0.00032)$. The distribution of the difference D is then $D \sim N(p_1 - p_2, 0.00452)$. We want to find the probability that $D \geq 0.1$ or $D \leq -0.1$. Transforming into a standard normal distribution, we get

$$\begin{aligned}
\Pr(D \geq 0.1 \text{ or } D \leq -0.1) &= \Pr(\frac{D}{\sqrt{0.00452}} \geq \frac{0.1}{\sqrt{0.00452}} \text{ or } \frac{D}{\sqrt{0.00452}} \leq \frac{-0.1}{\sqrt{0.00452}}) \\
&= \Pr(Z \geq 1.487) + \Pr(Z \leq -1.487) \\
&= 1 - \Phi(1.487) + \Phi(-1.487) = 0.137.
\end{aligned}$$

The difference is not significant.

8.6.15. The maximum likelihood estimator for the first player is $\hat{p}_1 = 0.5$, for the second is $\hat{p}_2 = 0.9$, and for the two players pooled is $\hat{p}_1 = 0.7$. The likelihood function is

$$L(p_1, p_2) = \left(\begin{array}{c} 10 \\ 5 \end{array}\right) p_1^5 (1-p_1)^5 \left(\begin{array}{c} 10 \\ 9 \end{array}\right) p_2^9 (1-p_2).$$

Because we are comparing hypotheses, we can ignore the constants in the support function $S(p_1, p_2) = 5\ln(p_1) + 5\ln(1-p_1) + 9\ln(p_2) + \ln(1-p_2)$. Using the individuals separately, we find $S(0.5, 0.9) = -10.18$. The support for the pooled hypothesis is $S(0.7, 0.7) = -12.22$. The difference is slightly greater than 2, giving reason to suspect that the second player is a better shot.

8.6.17. The maximum likelihood estimator for Utah is $\hat{\Lambda}_1 = 4$, for the North Pole is $\hat{\Lambda}_2 = 9$, and for the two sites pooled is $\hat{\Lambda} = 7$. The likelihood function is

$$L(\Lambda_1, \Lambda_2) = \frac{e^{-\Lambda_1} \Lambda_1^4}{4!} \frac{e^{-\Lambda_2} \Lambda_2^{10}}{10!}$$

Because we are comparing hypotheses, we can ignore the factorials in the support function

$$S(\Lambda_1, \Lambda_2) = 4\ln(\Lambda_1) - \Lambda_1 + 10\ln(\Lambda_2) - \Lambda_2$$

. Using the individuals separately, we find $S(4,10) = 14.57$ (the value can be positive because we left out constants). The support for the pooled hypothesis is $S(7,7) = 13.24$. The difference is less than 2, giving no reason to suspect that the North Pole gets more cosmic rays.

8.6.19. The null hypothesis is based on the idea that the treatment and control are really the same. Using the pooled proportion treats them together. There is a total of 150 out of 300, so $\hat{p} = 0.5$.

8.6.21. The null hypothesis is

$$D \sim N(0.0, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}).$$

As n_1 becomes large, the null becomes

$$D \sim N(0.0, \frac{\sigma_2^2}{n_2}).$$

8.6.23. The null hypothesis is that the cell does not act as a filter. In comparison with the control, it means that the two proportions will be equal. In comparison with the expectation, it means that molecules end up inside with probability 0.5.

8.6.25. If the mutation rates are Λ_1 and Λ_2, the numbers M_1 and M_2 are approximately

$$M_1 \sim N(\Lambda_1, \Lambda_1), \quad M_2 \sim N(\Lambda_2, \Lambda_2).$$

The difference D is then $D \sim N(\Lambda_1 - \Lambda_2, \Lambda_1 + \Lambda_2)$. The null hypothesis has $\Lambda_1 = \Lambda_2$. For the sum, we have no better guess than the sum of the estimators $\hat{\Lambda}_1 + \hat{\Lambda}_2 = 8 + 18 = 26$, for a null hypothesis of $D \sim N(0, 26)$. The measured difference is 10, which happens to be exactly 1.96 standard deviations. With a two-tailed test, the significance level is 0.05, right on the borderline of significance.

8.6.27. The maximum likelihood estimator for the first organism is $\hat{\Lambda}_1 = 8$, for the second organism is $\hat{\Lambda}_2 = 18$, and for the two organisms pooled is $\hat{\Lambda} = 13$. The likelihood function is

$$L(\Lambda_1, \Lambda_2) = \frac{e^{-\Lambda_1}\Lambda_1^8}{8!}\frac{e^{-\Lambda_2}\Lambda_2^{18}}{18!}$$

Because we are comparing hypotheses, we can ignore the factorials in the support function

$$S(\Lambda_1, \Lambda_2) = 8\ln(\Lambda_1) - \Lambda_1 + 18\ln(\Lambda_2) - \Lambda_2$$

. Using the individuals separately, we find $S(8,18) = 42.66$ (the value can be positive because we left out constants). The support for the pooled hypothesis is $S(13,13) = 40.69$. The difference is very close to 2, consistent with the marginally significant difference in exercise 8.6.25.

8.6.29. The null hypothesis is that the chemical has no effect, with the alternative that the chemical causes cancer. The pooled proportion is $\hat{p} = \frac{6}{14} = 0.428$. The support for the null is

$$S_n(0.428) = \ln(b(5; 8, 0.428)) + \ln(b(1; 6, 0.428)) = -3.74.$$

The proportion in the treated group is 0.625, and in the control is 0.167. Therefore,

$$S_a(0.625, 0.167) = \ln(b(5; 8, 0.625)) + \ln(b(1; 6, 0.167)) = -2.18.$$

The difference is only 1.56, not enough to provide convincing evidence that

8.6.31. The null is that the two leave at the same rate λ, with an alternative that the rates are different. The average mortality rate is $1/17.5 = 0.057$ (see section 8.1). The support is

$$S_n(\lambda) = \ln(\lambda e^{-30\lambda}) + \ln(\lambda e^{-5\lambda}).$$

With $\lambda = 0.057$, this is -7.72. The best guess for λ_1 is $1/30 = 0.0333$ and and the best guess for λ_2 is $1/5 = 0.2$. The support is

$$S_a(\lambda_1, \lambda_2) = \ln(\lambda_1 e^{-30\lambda_1}) + \ln(\lambda_2 e^{-5\lambda_2}).$$

With the maximum likelihood estimators of λ_1 and λ_2, this is -7.01. The difference is far too small to be convincing.

8.6.33. The null is that the two leave at the same rate λ, with an alternative that the rates are different. The average mortality rate in the bath is $1/45 = 0.022$, the average mortality rate without the bath is $1/4 = 0.25$, and the overall average mortality rate is $1/24.5 = 0.041$. The support for the null is

$$S_n(\lambda) = \ln(\lambda e^{-30\lambda}) + \ln(\lambda e^{-5\lambda}) + \ln(\lambda e^{-60\lambda}) + \ln(\lambda e^{-3\lambda}).$$

With $\lambda = 0.022$, this is -16.79. The support for the alternative is

$$S_a(\lambda_1, \lambda_2) = \ln(\lambda_1 e^{-30\lambda_1}) + \ln(\lambda_2 e^{-5\lambda_2}) + \ln(\lambda_1 e^{-60\lambda_1}) + \ln(\lambda_2 e^{-3\lambda_2}).$$

With the maximum likelihood estimators of λ_1 and λ_2, this is -14.39. The difference is greater than 2, and there is reason to think it is significant.