

Additional Statistics and Probability Practice Problems

1. You are studying fruit flies. You have a mutant fly strain which is homozygous for black body and white eyes, both recessive autosomal alleles. You cross it with a wild-type fly, with brown body and red eyes, producing many heterozygotes. Lastly, you cross these flies against black-bodied, white-eyed flies. The results of this cross are as follows.

- black-bodied, white-eyed: 28
- black-bodied, red-eyed: 20
- brown-bodied, white-eyed: 23
- brown-bodied, red-eyed: 29

Your question: Are the body color gene and eye color gene linked?

**(a) What statistical test should you use here? T-test or chi-squared? And which version?**

**(b) Perform the test. What is the null hypothesis, the p-value? What do you conclude?**

A researcher obtained mice with tumors and randomly divided them into two groups. She then injected one group of mice with the drug and used the second group as a control. After 2 weeks, she sacrificed the mice and weighed the tumors. The weight of tumors for each group of mice is below. The researcher is interested in learning if the drug reduces the growth of tumors.

<u>Treated with Drug</u>	<u>Control- Not Treated</u>
0.72	0.71
0.68	0.83
0.69	0.89
0.66	0.57
0.57	0.68
0.66	0.74
0.70	0.75
0.63	0.67
0.71	0.80
0.73	0.78

**(a) What statistical test should you use here? T-test or chi-squared? And which version?**

**(b) Perform the test. What is the null hypothesis, the p-value? What do you conclude?**

3. If HIV has a prevalence of 3% among adults in the Republic of the Congo, and a particular HIV test has a false positive rate of .001 and a false negative rate of .01, **what is the probability that a random adult in the Republic of the Congo who tests positive is actually infected?** (also known as “positive predictive value”)

## Answers

**(1a)** Chi-squared test for goodness of fit. You are testing whether the given data fits against a distribution which you already know. You have categorical data as opposed to numerical data.

**(1b)** Null hypothesis: The genes are unlinked. The expected distribution is 1:1:1:1 if the genes are completely unlinked. So that's an expected distribution of 25, 25, 25, 25.

Chi-squared is  $1/25[(9) + (25) + (4) + (16)] = 1/25(54) = 2.16$ . Degrees of freedom is 3.

Using the chi-squared value, that's a p-value  $> 0.10$ . This is not significant.

We fail to reject the null hypothesis. This data does not provide evidence that the genes are anything but unlinked, much though the proportions look skewed. (If you were to multiply these numbers by 10, then the difference would become significant enough to reject. Try it!)

**(2a)** Two sample T-test. You are dealing with numerical data as opposed to categorical data.

**(2b)** Null hypothesis: The drug has no effect on the tumors.

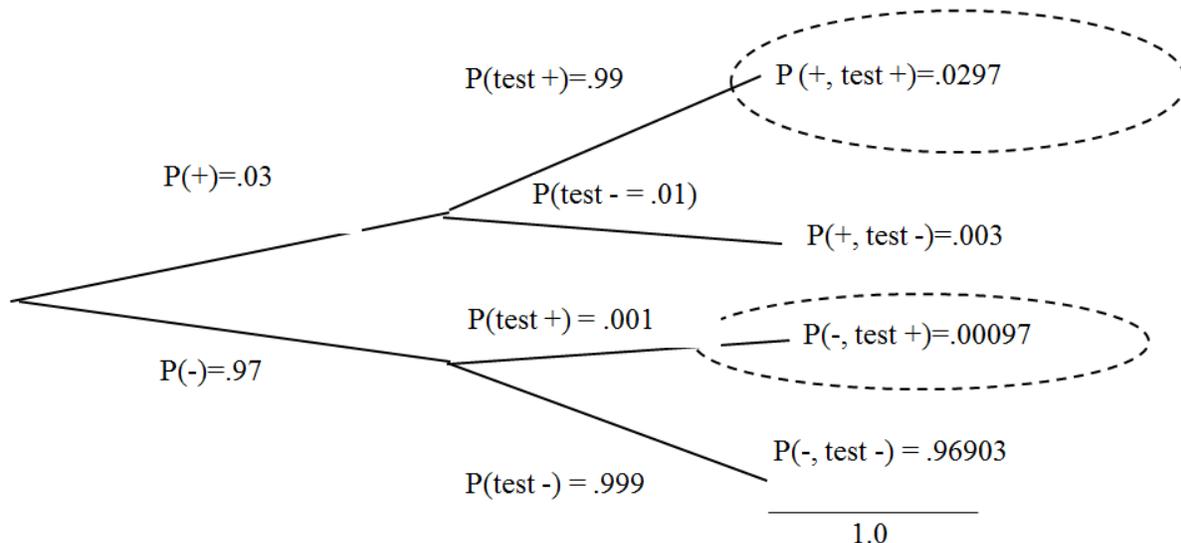
Alternate hypothesis: The drug reduces the tumors.

Treated with Drug – mean: 0.675, stdev: 0.0479. Control – mean: 0.742, stdev: 0.0908

T-statistic is  $(0.675 - 0.742) / \sqrt{1/10(0.0479^2 + 0.0908^2)} = -2.064$

Degrees of freedom is 18. p-value  $< 0.05$ . We reject the null hypothesis. The data are statistically significant evidence that the drug reduces the size of the tumors.

**(3)** Note: this test has a sensitivity of 0.99 and a specificity of 0.999. You can calculate these by subtracting the false positive and negative rates from 1. (A highly sensitive test, when negative, rules out disease “snout”. A highly specific test, when positive, rules 'in' disease “spin”.)



A positive test places one on either of the two “test +” branches. But only the top branch also fulfills the event “true infection.” Therefore, the probability of being infected is the probability of being on the top branch given that you are on one of the two circled branches above.

$$P(\text{true+} | \text{test+}) = \frac{P(\text{test+} \& \text{true+})}{P(\text{test+})} = \frac{.0297}{.0297 + .00097} = 96.8\%$$