Class 18 in-class problems, 18.05, Spring 2022

Concept questions

Concept question 1. NHST

You collect data from an experiment and do a left-sided z-test with significance 0.1. You find the z-value is 1.8

(i) *Which of the following computes the critical value for the rejection region?*

(a) `pnorm(0.1, 0, 1)` *(b)* `pnorm(0.9, 0, 1)`
(c) `pnorm(0.95, 0, 1)` *(d)* `pnorm(1.8, 0, 1)`
(e) `1 - pnorm(1.8, 0, 1)` *(f)* `qnorm(0.05, 0, 1)`
(g) `qnorm(0.1, 0, 1)` *(h)* `qnorm(0.9, 0, 1)`
(i) `qnorm(0.95, 0, 1)`

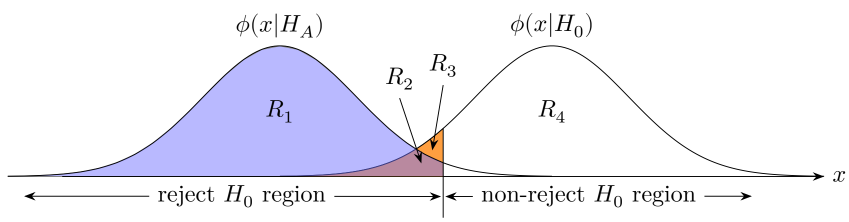
(ii) *Which of the above computes the p-value for this experiment?*

(iii) *Should you reject the null hypothesis?*

Solution: **(i)** g. **(ii)** d. **(iii)** No. (Draw a picture!)

Concept question 2. Power

The power of the test in the graph is given by the area of

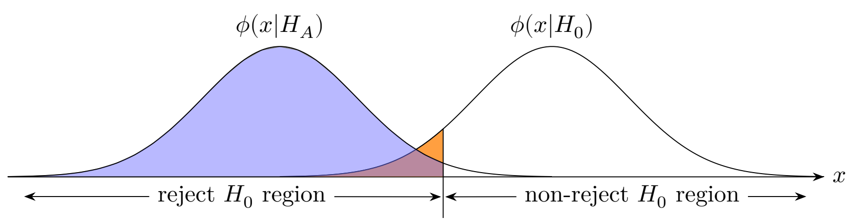


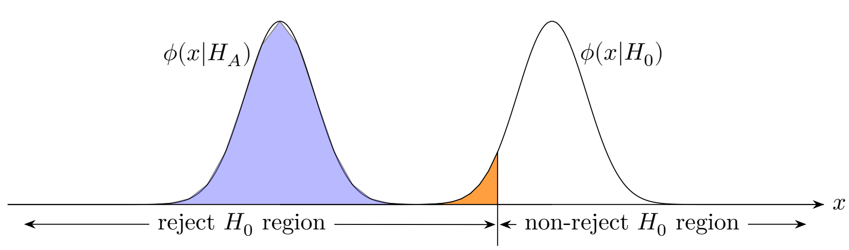
(a) R_1 *(b)* R_2 *(c)* $R_1 + R_2$ *(d)* $R_1 + R_2 + R_3$

Solution: (c) $R_1 + R_2$. Power $= P(\text{rejection region} \,|\, H_A) =$ area $R_1 + R_2$.

Concept question 3. Higher power

Which of the tests below has higher power?





(1) Top graph *(2) Bottom graph*

Solution: (2) The bottom graph.

Power $= P(x \text{ in rejection region} \mid H_A)$. In the bottom graph almost all the probability of H_A is in the rejection region, so the power is close to 1.

Board questions

Problem 1. Significance level and power

Our data x follows a binomial(θ, 10) distribution with θ unknown.

The rejection region is boxed in orange. The corresponding probabilities for different hypotheses are shaded below it.

x	0	1	2	3	4	5	6	7	8	9	10
$H_0: p(x\|\theta = 0.5)$	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	.001
$H_A: p(x\|\theta = 0.6)$	0.000	0.002	0.011	0.042	0.111	0.201	0.251	0.215	0.121	0.040	0.006
$H_A: p(x\|\theta = 0.7)$	0.000	0.000	0.001	0.009	0.037	0.103	0.200	0.267	0.233	0.121	0.028

***(a)** Find the significance level of the test.*

***(b)** Find the power of the test for each of the two alternative hypotheses.*

***(c)** What is the probability of a type I error? type II?*

Solution: (a) Significance level $= P(x \text{ in rejection region} \mid H_0) = 0.11$

(b) $\theta = 0.6$: power $= P(x \text{ in rejection region} \mid H_A) = 0.18$

$\theta = 0.7$: power $= P(x \text{ in rejection region} \mid H_A) = 0.383$

(c) The probability of a type I error is the significance: 0.11.

The probability of a type II error depends on H_A, it is 1 - power computed in problem 2.

Problem 2. z and one-sample t-test

For both problems use significance level $\alpha = 0.05$.

Assume the data 2, 4, 4, 10 are independently drawn from a $N(\mu, \sigma^2)$.

The hypotheses are: H_0: $\mu = 0$ and H_A: $\mu \neq 0$.

***(a)** Is the test one or two-sided? If one-sided, which side?*

***(b)** Assume $\sigma^2 = 16$ is known and test H_0 against H_A.*

***(c)** Now assume σ^2 is unknown and test H_0 against H_A.*

Solution: We have $\bar{x} = 5$, $\quad s^2 = \frac{9+1+1+25}{3} = 12$

(a) Two-sided. A standardized sample mean far above or below 0 is evidence against H_0, and consistent with H_A.

(b) We'll use the standardized mean z for the test statistic (we could also use $\bar{x}$). The null distribution for z is $\mathrm{N}(0, 1)$. This is a two-sided test so the rejection region is

$$(z \leq z_{0.975} \text{ or } \ z \geq z_{0.025}) \ = \ (-\infty, -1.96] \cup [1.96, \infty)$$

Since $z = (\bar{x} - 0)/(4/2) = 2.5$ is in the rejection region we reject H_0 in favor of H_A.

Repeating the test using a p-value. Since $z > 0$ and the test is two-sided:

$$p = 2P(Z \geq 2.5 \mid H_0) = 2 * (1 - \texttt{pnorm}(2.5)) \approx 0.012$$

Since $p < \alpha$ we reject H_0 in favor of H_A.

(c) We'll use the Studentized $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ for the test statistic. The null distribution for t is t_3. For the data we have $t = 5/\sqrt{3}$. This is a two-sided test and t is positive, so the p-value is

$$p = 2P(T \geq 5/\sqrt{3} \,|\, H_0) = 2 * (1 - pt(5/sqrt(3), 3)) \approx 0.06318$$

Since $p > \alpha$ we do not reject H_0.

Problem 3. Two-sample t-test

Real data from 1408 women admitted to a maternity hospital for (i) medical reasons or through (ii) unbooked emergency admission. The duration of pregnancy is measured in complete weeks from the beginning of the last menstrual period.

Medical: 775 obs. with $\bar{x} = 39.08$ *and* $s^2 = 7.77$.

Emergency: 633 obs. with $\bar{x} = 39.60$ *and* $s^2 = 4.95$

(a) *Set up and run a two-sample t-test to investigate whether the duration differs for the two groups.*

(b) *What assumptions did you make?*

Solution: (a) The pooled variance for this data is

$$s_p^2 = \frac{774(7.77) + 632(4.95)}{1406}\left(\frac{1}{775} + \frac{1}{633}\right) = 0.0187$$

The t statistic for the null distribution is

$$\frac{\bar{x} - \bar{y}}{s_p} = -3.8064$$

Rather than compute the two-sided p-value using `2*tcdf(-3.8064,1406)` we simply note that with 1406 degrees of freedom the t distribution is essentially standard normal and 3.8064 is almost 4 standard deviations. So

$$P(|t| \geq 3.8064) \approx P(|z| \geq 3.8064)$$

which is very small, much smaller than $\alpha = 0.05$ or $\alpha = 0.01$. Therefore we reject the null hypothesis in favor of the alternative that there is a difference in the mean durations.

(b) We assumed the data was normal, independent and that the two groups had equal variances. Given the big difference in the sample variances this assumption might not be warranted.

Note: there are significance tests to see if the data is normal and to see if the two groups have the same variance.

Discussion questions

1. Significance and power

The null distribution for test statistic x is $N(4, 8^2)$. The rejection region is $\{x \geq 20\}$.

What is the significance level and power of this test?

Solution: 20 is two standard deviations above the mean of 4. Thus,

$$\text{significance} = P(x \geq 20|H_0) \approx 0.025$$

Trick question: we can't compute the power without an alternative distribution.

2. Type I errors Q1

Suppose a journal will only publish results that are statistically significant at the 0.05 level. What percentage of the papers it publishes contain type I errors?

Solution: This is asking for $P(H_0|\text{rejected } H_0)$. This is the probabilitiy of a hypothesis. Since we are not given a prior (baserates), we can't know this. **The percentage could be anywhere from 0 to 100!**

Remember: signifcance is the false positive rate, i.e $P(\text{rejection}|H_0)$. You need the base rate (prior) to know how often the test as a whole is wrong

3. Type I errors Q2

Jerry desperately wants to cure diseases but he is terrible at designing effective treatments. He is however a careful scientist and statistician, so he randomly divides his patients into control and treatment groups. The control group gets a placebo and the treatment group gets the experimental treatment. His null hypothesis H_0 is that the treatment is no better than the placebo. He uses a significance level of $\alpha = 0.05$. If his p-value is less than α he publishes a paper claiming the treatment is significantly better than a placebo.

***(a)** Since his treatments are never, in fact, effective what percentage of his experiments result in published papers?*

***(b)** What percentage of his published papers contain type I errors, i.e. describe treatments that are no better than placebo?*

Solution: **(a)** Since in all of his experiments H_0 is true, roughly 5%, i.e. the significance level, of his experiments will have $p < 0.05$ and be published.

(b) This is asking for $P(H_0|\text{rejected } H_0)$. This is the probabilitiy of a hypothesis. Since we are given the prior (baserate), that is, since all his treatments are no better than placebo, we can answer this: All of his published papers contain type I errors.

4. Type I errors Q3

Jen is a genius at designing treatments, so all of her proposed treatments are effective. She is also a careful scientist and statistician, so she too runs double-blind, placebo controlled, randomized studies. Her null hypothesis is always that the new treatment is no better than the placebo. She also uses a significance level of $\alpha = 0.05$ and publishes a paper if $p < \alpha$.

***(a)** How could you determine what percentage of her experiments result in publications?*

(b) *What percentage of her published papers contain type I errors, i.e. describe treatments that are, in fact, no better than placebo?*

Solution: **(a)** The percentage that get published depends on the power of her treatments. If they are only a tiny bit more effective than placebo then roughly 5% of her experiments will yield a publication. If they are a lot more effective than placebo then as many as 100% could be published.

(b) This is asking for $P(H_0|\text{rejected})$. Since we are given the prior (baserate), that is, since all her treatments are better than placebo, we can answer this: None of her published papers contain type I errors.