Some Recipes and Exercises With Tests, Type I and Type II errors, Size and Power, and Their Relation to Severity

(Complete over spring break)

We are sampling from a population of fish, or more properly, of fish lengths. A fish's length, in inches, may be represented by variable X; that is, to each fish a value of X (like a little badge) is attached.

Suppose X varies according to a pattern closely resembling a Normal distribution with an average (mean) value of μ (which is in question) and a known standard deviation σ equal to 2.

Let hypothesis $\mathcal{H}$ assert that μ = 12 and alternative $\mathcal{J}$ assert that μ > 12.
In order to test these claims, we observe a random sample of n fish (fish lengths), and average their lengths to get $\bar{X}$. We perform **a one-sided test**.

If H is true, then $\bar{X}$ follows the Normal distribution with mean μ, and the standard deviation of $\bar{X}$, written, $\sigma_{\bar{X}}$ equals σ $(1/n^{1/2})$. So $\sigma_{\bar{X}}$ equals 2 $(1/n^{1/2})$.

Null (or test) hypothesis H: $\bar{X}$ is Normal (μ, $\sigma_{\bar{X}}$), and μ = 12.

and

Alternative hypothesis J: $\bar{X}$ is Normal (μ, $\sigma_{\bar{X}}$), and μ >12.

where $\sigma_{\bar{X}}$ = 2 $(1/n^{1/2})$.

$\bar{X}_{obs}$ = the observed mean.

1. To calculate the **statistical significance level** of $\bar{X}_{obs}$

a. Calculate D_{obs} $= \bar{X}_{obs} - \bar{X}$ (expected assuming H)

$= \bar{X}_{obs} - 12$

b. Put D_{obs} in standard deviation units: $D_{obs}/\sigma_{\bar{X}}$

c. Find the area between 0 and $D_{obs}/\sigma_{\bar{X}}$ on the Standard Normal distribution chart.

d. .5 minus the value you get in c is the **statistical significance level** of the observed difference.

2. The **<u>severity</u>** with which J passes this test with outcome $\bar{X}_{obs}$ equals 1 minus the statistical significance level of D_{obs} . See #6. below.

Problem Set #1: Find the statistical significance level of each of the following outcomes with the indicated sample size n:

a. n = 25: (i) $\bar{X}_{obs} = 12.6$ (ii) $\bar{X}_{obs} = 12.8$ (iii) $\bar{X}_{obs} = 13$.

b. n = 100: (i) $\bar{X}_{obs} = 12.2$ (ii) $\bar{X}_{obs} = 12.4$ (iii) $\bar{X}_{obs} = 12.5$

c. n = 1600: (i) $\bar{X}_{obs} = 12.1$ (ii) $\bar{X}_{obs} = 12.2$ (iii) $\bar{X}_{obs} = 12.6$

3. Tests (e.g., in Neyman-Pearson theory) are specified by giving a rule for when outcomes should be taken to reject H and accept J, and when not. A typical rule would be to reject H whenever the difference between the observed mean and the mean expected assuming H were true reaches a given **small** statistical significance level, e.g., .05, or .03, or .01. When the significance level that will lead to rejecting H is set out ahead of time at some fixed value, say .03, this value is called the **<u>size</u>** of the test.

4. Suppose the hypotheses being tested are as above:

Null (or test) hypothesis H: $\bar{X}$ is Normal (μ, $\sigma_{\bar{X}}$), and $\mu = 12$.
Alternative hypothesis J: $\bar{X}$ is Normal (μ, $\sigma_{\bar{X}}$), and $\mu > 12$.
where $\sigma_{\bar{X}} = 2\,(1/n^{1/2})$.

From now on we will just write this as H: $\mu = 12$ vs. J: $\mu > 12$.

<u>Define</u> **<u>(one-sided) test T+ (with size = .03):</u>** Reject H and accept J whenever $\bar{X}_{obs}$ is statistically significantly different from 12 (in the positive direction) at level .03.

Equivalently we can define:
Test T+ (with size = .03) : Reject H and accept J
whenever $\bar{X}_{obs} \geq 12 + 2\,\sigma_{\bar{X}}$.
Or

Test T+ :Reject H and accept J whenever $D_{obs} \geq 2\ \sigma_{\bar{X}}$

Define **(one-sided) test T+ (with size = .001):** Reject H and accept J whenever $\bar{X}_{obs}$ is statistically significantly different from 12 (in the positive direction) at level .001.

Equivalently we can define:
Test T+ (with size = .001): Reject H and accept J
whenever $\bar{X}_{obs} \geq 12 + 3\ \sigma_{\bar{X}}$.
or
Test T+ :Reject H and accept J whenever $D_{obs} \geq 3\ \sigma_{\bar{X}}$

5. Let $\bar{X}^*$ be the **cut-off for rejection**. Then $\bar{X}^*$ for T+ with size .03 = 12 + 2 $\sigma_{\bar{X}}$.
Questions: What is $\bar{X}^*$ for this test when n = 100?____
(answer: 12.4).
What is $\bar{X}^*$ for Test T+ with size .001 and n= 100? (answer: 12.6.)

Problem Set #2:
1. Find $\bar{X}^*$ for Test T+ with size .03 and n= 25, and for n = 1600.
2. Which of the outcomes in Problem Set #1, a. and b., would lead to rejecting H at this level? The outcomes were:

a. n = 25: (i) $\bar{X}_{obs}$ = 12.6 (ii) $\bar{X}_{obs}$ = 12.8 (iii) $\bar{X}_{obs}$ = 13.

c. n = 1600: (i) $\bar{X}_{obs}$ = 12.1 (ii) $\bar{X}_{obs}$ = 12.2 (iii) $\bar{X}_{obs}$ = 12.6

5. **The Type I error:** This is the error of rejecting H when actually J is true.
The probability that test T+ with size .03 commits a Type I error, written as α = P(Test T+ rejects H ; H is true)
= P($\bar{X} \geq \bar{X}^*$; H is true)
= P($\bar{X}$ is statistically significant at level .03; H is true)
=.03

So, the fixed <u>size</u> of the test ensures that the probability of a Type I error is no more than .03.

In general, by fixing a small <u>size</u> α, the test is assured of having a small probability, namely α , of committing a Type I error.)

6. Suppose Test T+ with size α = .03 rejects H and "passes" J with an outcome that just reaches the cut-off for rejection, i.e., suppose $\bar{X}_{\text{obs}} = \bar{X}*$.
We can show that the severity with which J passes test T+ with α = .03 = 1 - α = 1 - .03 = .97.

By definition of severity, the severity with which a hypothesis J passes a test T with outcome e = P(Test T would <u>not</u> yield such a passing result ;J is false)
= 1 - P(Test T+ ***would*** yield such a passing result ; J is false)

Notice that "passing J" (that μ > 12) is the same as rejecting H in the case of test T+. And "J is false" = H is true. Therefore, this equals
= 1 - P(Test T+ reject H ; H is true)

and from #5 above, we have this is
= 1 - α = 1 - .03 = .97.

(Note: There is no difference in this test if we let H: μ ≤ 12)

---

7. **The Type II Error, β, and Power**
Once again, consider Test T+ (with size α = .03) and $\bar{X}*$ the cut-off for rejection of H where
H: μ = 12 vs. J: μ > 12.

<u>The Type II error</u> is the error of failing to reject H, even though H is false and J is true.
The probability a test T commits a Type II error, written as β = P(Test T commits a Type II error) = P(Test T accepts H ; J is true)

where "accept H" just means you do not reject it. (We will clarify the precise way to interpret "accept H" as we proceed.)

Now, J in test T+ is a **<u>composite</u>** or **<u>complex</u>** hypothesis. In order to calculate β, we have to consider particular point values of μ in the

alternative J (i.e., values of μ ≥ 12). Let J' abbreviate a specific value of μ in the alternative J. Then
β = P(Test T+ accepts H ; J') = 1 - P(Test T+ rejects H ; J')

= P($\bar{X} < \bar{X}*$; J') = 1 - P($\bar{X} \geq \bar{X}*$; J')

Case 1: J' < $\bar{X}*$

Draw a picture of the Normal curve, label J' right at the midpoint, H to the left of J' and $\bar{X}*$ somewhere to the right of J'. shade in the entire area under the curve to the left of $\bar{X}*$. This shaded area, which must clearly be <u>greater</u> than .5, equals β. So, the probability of a Type II error in this case is not low, it is greater than .5.

______H(12)_____J'_______$\bar{X}*$.____

Case 2: J' > $\bar{X}*$

Draw a picture of the Normal curve, label J' right at the midpoint, $\bar{X}*$ to the left of J' and H to the left of $\bar{X}*$. Shade in the area under the curve to the left of $\bar{X}*$. This shaded area, which must clearly be <u>less</u> than .5, equals β.

__H(12)____ $\bar{X}*$_____J'_________.____

To Calculate a particular value for β, i.e., P($\bar{X}* < \bar{X}$; J")

Note that it is always calculated using the cut-off value for rejecting H, namely, $\bar{X}*$. Again, you begin by calculating a difference, but now, since the assumption is that J' is true, you calculate:

1. Find D = $\bar{X}*$ - J'

2. Divide D by $\sigma\bar{X}$
(Note that in case 1, D is positive, in case 2, D is negative.)

3. Find the area to the <u>left</u> of the value you get in #2, under the standard Normal curve.

4. The area to the <u>right</u> of the value you get in #2 is the **<u>power</u>** of the test (to reject H and accept J'). (So power = 1 - β)

Example: For T+ (with α = .03) and n = 100, the cut-off for rejecting H, $\bar{X}*$, is 12.4. ($\sigma\bar{X}$ = .2) Say you want to find β for J': μ = 12.2.
You ask: What is the probability that Test T+ would accept H when in fact J' is true? That is: what is the probability that Test T+ would accept H: μ = 12 when in fact the true value of μ is 12.4?
To answer, you calculate D = 12. 4 - 12.2 = .2 which is equal to 1 standard deviation. (Note, this is an example of case 1.)
The area to the left of 1 on the standard Normal curve is about .84. So β = .84.

Problem Set #3: For Test T+ (with α = .03) and n = 100 with
H: μ = 12 vs. J: μ > 12 and σ = 2, so ($\sigma\bar{X}$ = .2)

1. Find the probability that Test T+ commits a Type II error when J': μ = 12.6.
A Useful Abbreviation: Since the probability of a Type II error, β, always varies with the particular point value of J', it will be useful to abbreviate the probability that Test T+ commits a Type II error when J': μ =μ' as simply **β(μ') or β(J').**
So problem 1 in set #3 is to calculate β(12.6).

I'll do this first one: D = 12.4 - 12.6 = -.2, which is -1 $\sigma\bar{X}$. The area to the left of -1 on the standard Normal curve = .5 - (the area between 0 and 1) = .5 - .34 = .16.
Note that this is an example of case 2: J > $\bar{X}$*.

2. Find β(12).3. Find β(12.2) 4. Find β(12.4). .
5. Find β(12.7). 6. Find β(12.8). 7. Find β(12.9). 8. Find β(13).

9. Plot these points on a curve on whose horizontal axis are values of J' and whose verticle axis are the corresponding values of β(J').

10. Explain why β(H) = 1 - α.
.
11. Graph the corresponding values (in 1-8) for the power. This yields a **<u>power curve</u>**.

12. Explain: If Test T+ accepts H (say it just misses the cut-off $\bar{X}$*) and β(μ') is very low, then "μ < μ'" passes a severe test.