

# **BME 1132**

# **Probability and Biostatistics**

**Instructor:** Ali AJDER, *Ph.D.*

# Week-12

- Introduction
- Interval Estimation
- Confidence Interval When the Population Variance Is Unknown
- t Distribution
- Margin of Error
- Sample size Estimation
- Hypothesis Testing

# Interval Estimation

We have assumed previously that the distribution of birthweights was normal with mean  $\mu$  and variance  $\sigma^2$ . It follows from our previous discussion of the properties of the sample mean that  $\bar{X} \sim N(\mu, \sigma^2/n)$ . Thus, if  $\mu$  and  $\sigma^2$  were known then the behavior of the set of sample means over a large number of samples of size  $n$  would be precisely known.

In particular, 95% of all such sample means will fall within the interval  $(\mu - 1.96\sigma/\sqrt{n}, \mu + 1.96\sigma/\sqrt{n})$

Alternatively, if we re-express  $\bar{X}$  in standardized form by

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

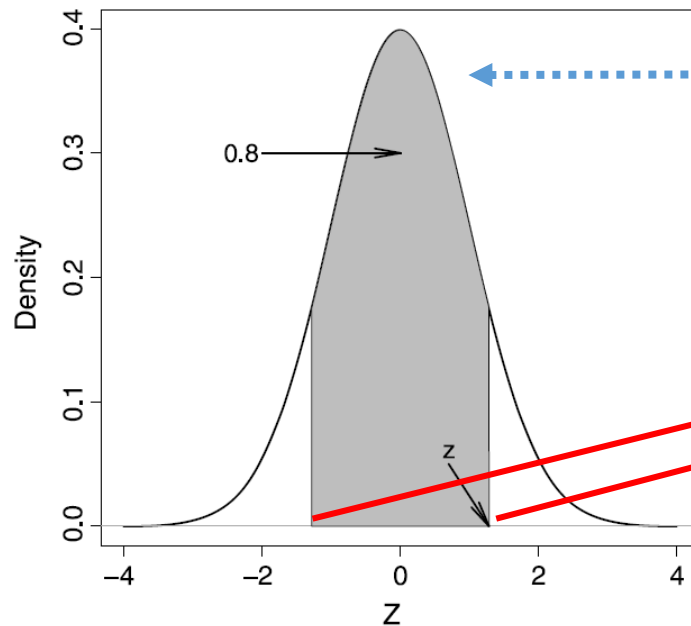
then  $Z$  should follow a standard normal distribution. Hence, 95% of the  $Z$  values from repeated samples of size  $n$  will fall between  $-1.96$  and  $+1.96$  because these values correspond to the 2.5th and 97.5th percentiles from a standard normal distribution. However, the assumption that  $\sigma$  is known is somewhat artificial, because  $\sigma$  is rarely known in practice.

# Confidence Intervals for the Population Mean

In particular, 95% of all such sample means will fall within the interval  $(\mu - 1.96\sigma/\sqrt{n}, \mu + 1.96\sigma/\sqrt{n})$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow \left[ \bar{X} - z_c * \sigma/\sqrt{n} \leq \mu \leq \bar{X} + z_c * \sigma/\sqrt{n} \right]$$

In general, for a given confidence level,  $c$ , we use the standard normal distribution to find the value whose upper tail probability is  $(1 - c)/2$ . We refer to this value as the  $z_c$  value for the confidence level of  $c$ .

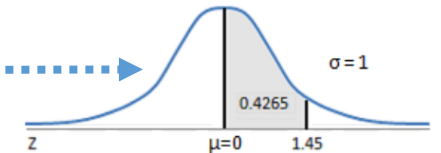


*Note the difference between the figures!*

Finding the  $z_c$  value for 0.8 confidence level. Approximately 80% of the values of the random variable fall between  $-1.28$  and  $1.28$

Areas Under the One-Tailed Standard Normal Curve

This table provides the area between the mean and some Z score. For example, when Z score = 1.45 the area = 0.4265.



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545

# Confidence Interval When the Population Variance Is Unknown

So far, we have assumed the population variance,  $\sigma^2$ , of the random variable is known.

Hence, we assumed that  $\sigma/\sqrt{n}$ , i.e., the standard deviation of the sample mean, is known. *This is an unrealistic assumption.*

Almost always, we need to estimate  $\sigma^2$  along with the population mean  $\mu$ . For this, we use our sample of  $n$  observations to obtain the sample variance  $s^2$  and sample standard deviation  $s$ .

As a result, the standard deviation for  $\bar{X}$  is estimated to be  $s/\sqrt{n}$ .

We refer to  $s/\sqrt{n}$  as the **standard error (SE)** of the sample mean  $\bar{X}$  to distinguish it from  $\sigma/\sqrt{n}$ .

To find confidence intervals for the population mean when the population variance is unknown,

$$\left[ \bar{X} - t_c * s/\sqrt{n} \leq \mu \leq \bar{X} + t_c * s/\sqrt{n} \right]$$

$t_c$  is the value with an upper tail probability of  $(1 - c)/2$  based on a ***t – distribution*** with  $n - 1$  degrees of freedom.

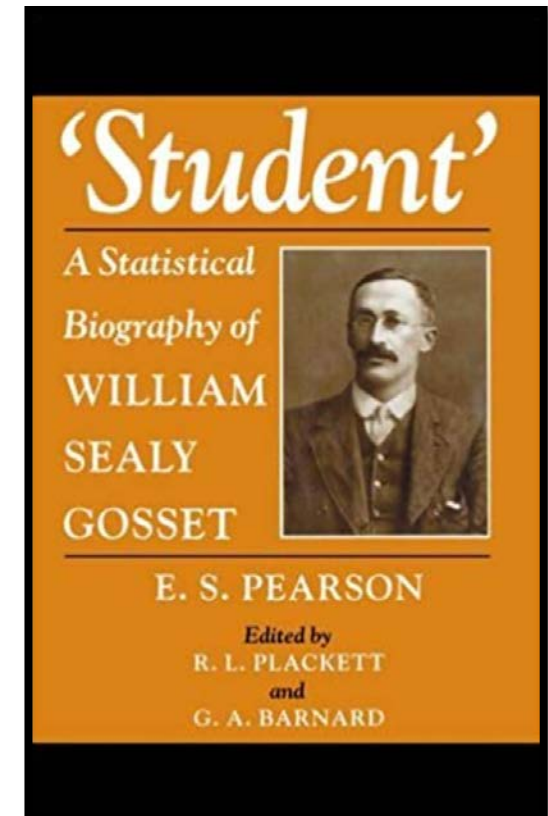


# t Distribution

Because  $\sigma$  is unknown, it is reasonable to estimate  $\sigma$  by the sample standard deviation  $s$  and to try to construct **CI**s using the quantity  $(\bar{X} - \mu)/(S/\sqrt{n})$ .

The problem is that this quantity is no longer normally distributed.

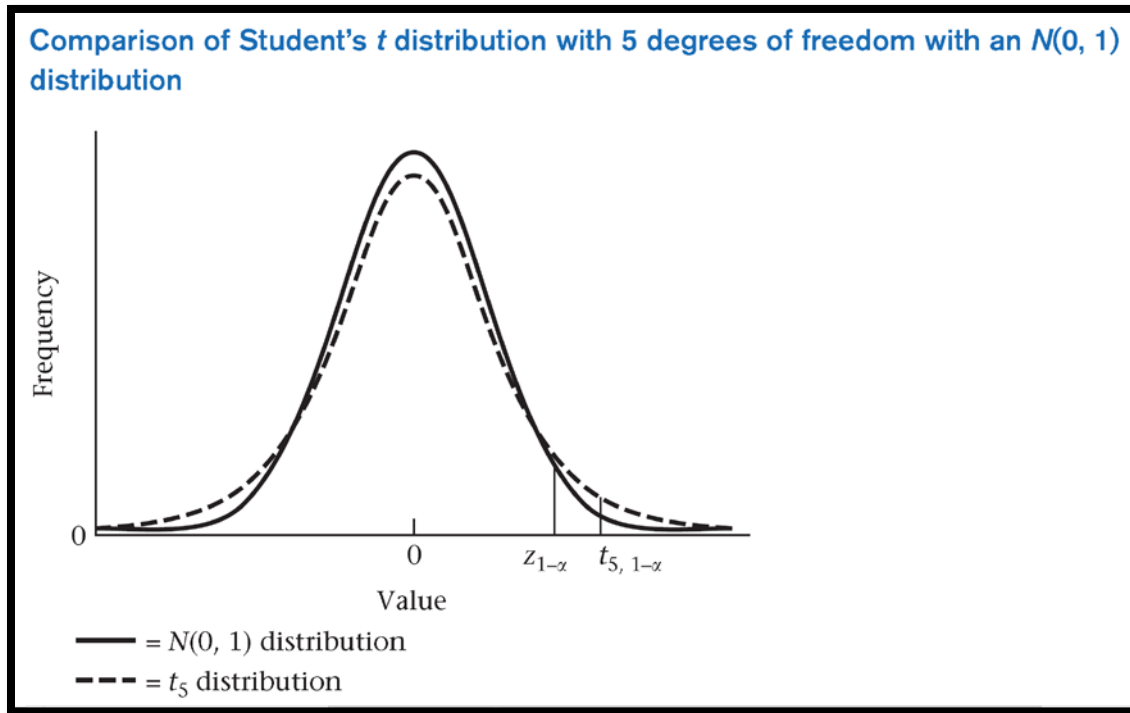
This problem was first solved in 1908 by a statistician named **William Gossett**. For his entire professional life, Gossett worked for the Guinness Brewery in Ireland. He chose to identify himself by the pseudonym “Student,” and thus the distribution of  $(\bar{X} - \mu)/(S/\sqrt{n})$  is usually referred to as **Student’s t distribution**. Gossett found that the shape of the distribution depends on the sample size  $n$ . Thus, the  $t$  distribution is not a unique distribution but is instead a family of distributions indexed by a parameter referred to as the **degrees of freedom** ( $df$ ) of the distribution.



If  $X_1, \dots, X_n \sim N(\mu, \sigma^2)$  and are independent, then  $(\bar{X} - \mu)/(S/\sqrt{n})$  is distributed as a  $t$  distribution with  $(n - 1)$   $df$ .

# t Distribution

It is interesting to compare a ***t* distribution** with *d* degrees of freedom with an  $N(0, 1)$  distribution. The density functions corresponding to these distributions are depicted in Figure for the special case where  $d = 5$ .



The two distributions thus get more and more alike as  $n$  increases in size. The upper 2.5th percentile of the *t* distribution for various degrees of freedom and the corresponding percentile for the normal distribution are given in Table.

Comparison of the 97.5th percentile of the *t* distribution and the normal distribution

$d$	$t_{d, .975}$	$z_{.975}$	$d$	$t_{d, .975}$	$z_{.975}$
4	2.776	1.960	60	2.000	1.960
9	2.262	1.960	$\infty$	1.960	1.960
29	2.045	1.960			

# t Distribution- Example

Compute a 95% CI for the following 10 birthweights.

2749.9 3543.7 1757.7 3401.9 3742.1  
3827.2 3345.2 3883.9 3572 3345.2

$$[\bar{X} - t_c * s/\sqrt{n} \leq \mu \leq \bar{X} + t_c * s/\sqrt{n}]$$

$$\bar{X} = 3316.88 \quad \frac{s}{\sqrt{n}} = 190.8$$

Subtract the confidence level from 1, then divide by two.

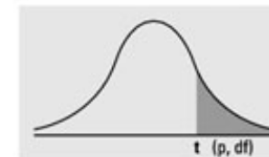
$$\alpha = (1 - 0.95)/2 = 0.025 \quad df = n - 1 = 10 - 1 = 9 \rightarrow$$

Look up in the t-distribution table for 9 degrees of freedom ( $df$ ) and  $\alpha = 0.025$ .

$$[3316.88 - 2.26 * 190.8 \leq \mu \leq 3316.88 + 2.26 * 190.8]$$

$$[2885.67 \leq \mu \leq 3748.9]$$

Numbers in each row of the table are values on a  $t$ -distribution with ( $df$ ) degrees of freedom for selected right-tail (greater-than) probabilities ( $p$ ).



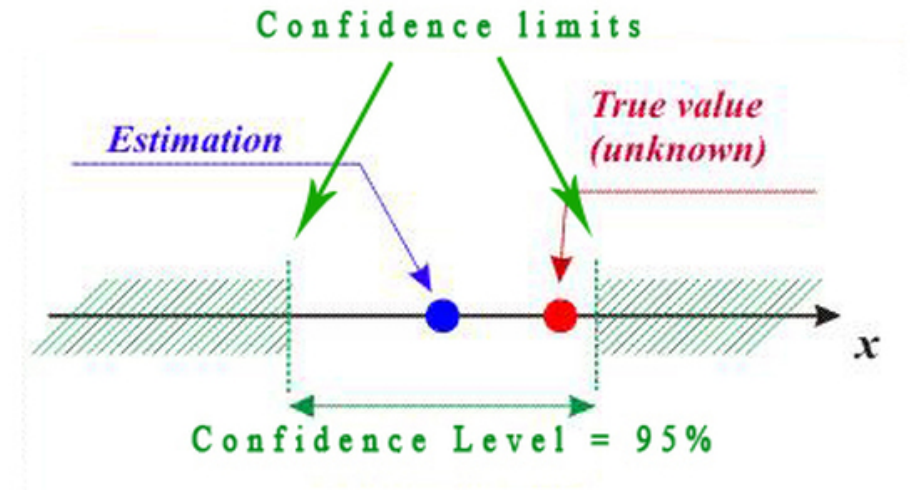
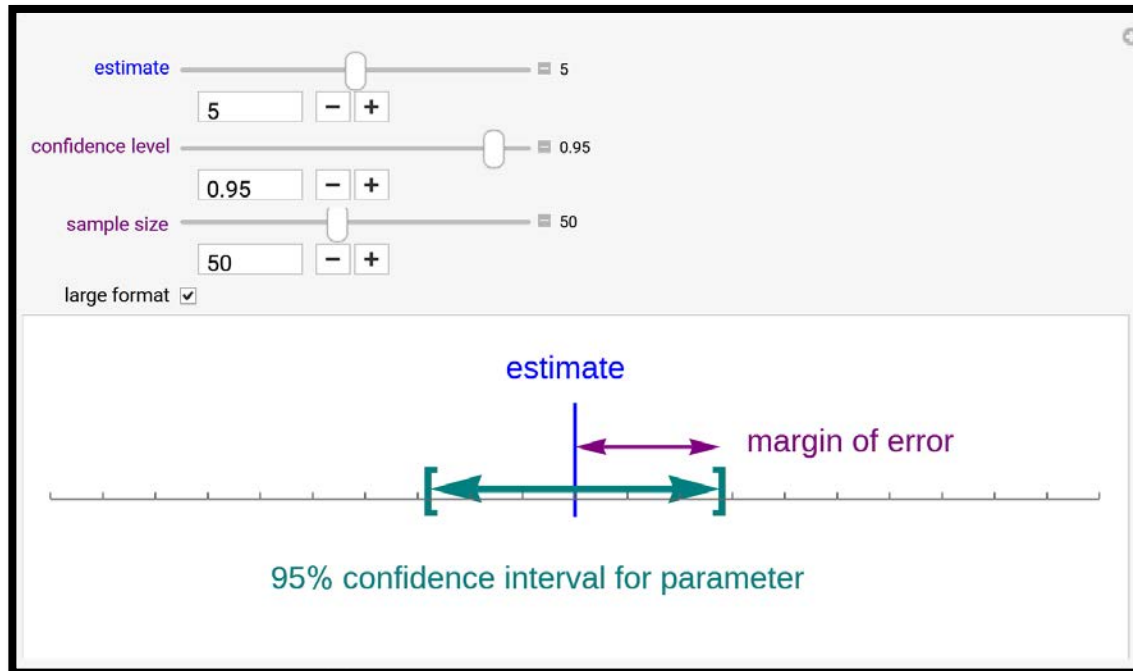
df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	4.3178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728



# Margin of error

In general, it is common to present interval estimates for  $c$  confidence level as

$$\text{Point estimate} \pm \text{Margin of error}$$



When the population variance  $\sigma^2$  is known, the margin of error  $e$  is calculated as follows:

$$e = z_c * \sigma / \sqrt{n}$$

When the population variance is not known, the margin of error is calculated as follows:

$$e = t_c * s / \sqrt{n}$$

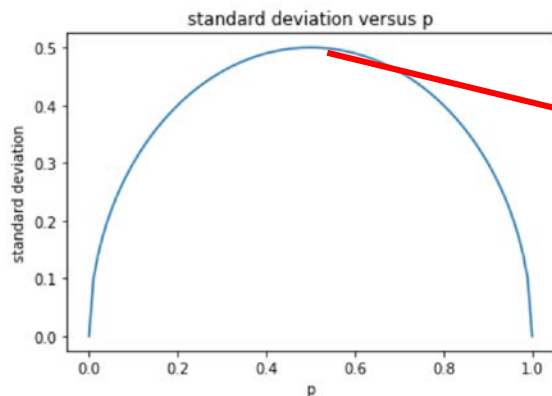
# Sample Size Estimation

Using the following equation for the margin of error:  $e = z_c * \frac{\sigma}{\sqrt{n}} \rightarrow n = \left( \frac{z_c * \sigma}{e} \right)^2$

**Example:** Suppose that we want to test a new drug for breast cancer, and we would like to estimate the 5-year survival mean (proportion) with the margin of error of 0.1 at 0.8 confidence level.  
(We want the true survival rate to fall within 10 % from its point estimate.)

**Note:** Remember the  $\sigma = \sqrt{p * (1 - p)}$   
for binary RV.

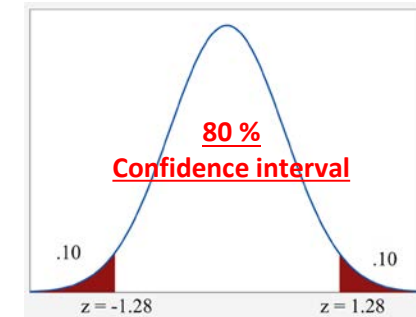
```
p = np.linspace(start = 0, stop = 1, num = 100)
sig = np.sqrt(p*(1-p))
plt.plot(p, sig);
plt.title('standard deviation versus p')
plt.xlabel('p')
plt.ylabel('standard deviation')
```



$$n = \left( \frac{1.28 * 0.5}{0.1} \right)^2 = 41$$

*We need to test the drug on 41 people to achieve the required margin of error at the given confidence level.*

$\sigma$  to 0.5 to be conservative since 0.5 is the highest possible standard deviation for binary random variable, and the above equation results in the highest value of  $n$  for the given margin of error and confidence level.



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545

# Hypothesis Testing- Introduction-1

We focused on **estimating parameters** such as the population mean and variance in the previous lecture. And now, we rely on estimators, their sampling distributions, and their specific values from observed data to evaluate **hypotheses**.

In general, many scientific investigations start by expressing a hypothesis.

For example, *Mackowiak et al.*<sup>1</sup> hypothesized that the average normal (i.e., for healthy people) body temperature is less than the widely accepted value of 37 °C.

If we denote the population mean of normal body temperature as  $\mu$ , then we can express this hypothesis as  $\mu < 37$ .

Typically, we can find another explanation, also expressed as a hypothesis, that invalidates (annuls) our proposed hypothesis. For this example, one might hypothesize that  $\mu \geq 37$ . We refer to this hypothesis as the **null hypothesis** and denote it as  $H_0$ . The null hypothesis usually reflects the “status quo” or “nothing of interest”.

In contrast, we refer to our hypothesis (i.e., the hypothesis we are investigating through a scientific study) as the **alternative hypothesis** and denote it as  $H_A$ .

<sup>1</sup>Mackowiak, P.A., Wasserman, S.S., Levine, M.M.: A critical appraisal of 98.6°F, the upper limit of the normal body temperature, and other legacies of Carl Reinhold August Wunderlich. *JAMA* 268, 1578–1580 (1992)

# Hypothesis Testing- Introduction-2

When we state our hypothesis, we are mainly proposing an explanation for an **observed phenomenon**.

For the above example, we might have observed that the body temperature of many healthy people is less than 37 °C. It is common to express the null hypothesis in the **simplest form** possible.

For the above example, to annul the alternative hypothesis,

$H_A: \mu < 37$ , it suffices to show that

$H_0: \mu = 37$ . (*This makes the task of evaluating a hypothesis easier.*)

The procedure for evaluating a hypothesis is called **hypothesis testing**, and it rises in many scientific problems.

A common approach for hypothesis testing is to focus on the **null hypothesis**, which is usually simpler than the alternative hypothesis, and **decide whether or not to reject** it.

If the evidence against  $H_0$  is strong, we **reject**  $H_0$ .

If not, we state that the evidence provided by the data is not strong enough to reject  $H_0$ , and we **fail to reject** it.

# Types of Errors

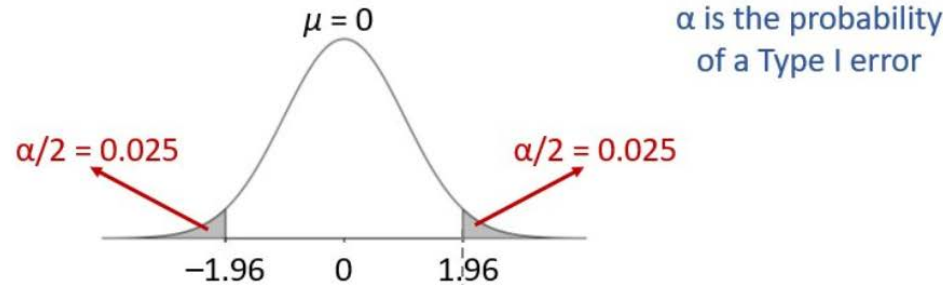
With respect to our decision regarding the null hypothesis  $H_0$ , we might make two types of errors:

**Type I error** : We reject  $H_0$  when it is true and should not be rejected.

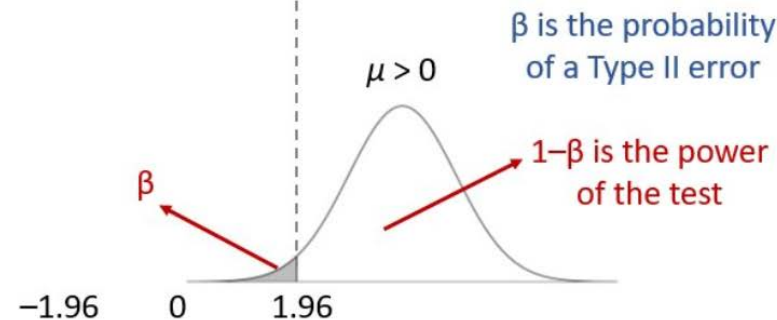
**Type II error** : We fail to reject  $H_0$  when it is false and should be rejected.

We denote the probability of making type I error as  $\alpha$  and the probability of making type II error as  $\beta$ .

If null hypothesis is true, this is the distribution of Z values



If the alternative hypothesis is true, this is the distribution of Z values



- ✓ We hope to avoid both type I and type II errors as much as possible. However, there is a trade-off between them.
- ✓ Now suppose that we have a hypothesis testing procedure that fails to reject the null hypothesis when it should be rejected with probability  $\beta$ . This means that our test correctly rejects the null hypothesis with probability  $1 - \beta$ .  
(Note that the two events are complementary.)
- ✓ We refer to this probability (i.e.,  $1 - \beta$ ) as the power of the test.
- ✓ In practice, it is common to first agree on a tolerable type I error rate  $\alpha$ , such as 0.01, 0.05, and 0.1.
- ✓ Then try to find a test procedure with the highest power among all reasonable testing procedures.

# Hypothesis Testing for the Population Mean-1

To decide whether we should reject the null hypothesis, we quantify the empirical support (provided by the observed data) against the null hypothesis using some statistics. Since these statistics are used to evaluate our hypotheses, we refer to them as **test statistics**.

To evaluate hypotheses regarding the population mean, we use the sample mean  $\bar{X}$  as the test statistic.

For a statistic to be considered as a test statistic, its sampling distribution must be fully known (exactly or approximately) under the null hypothesis. That is, we should know the distribution of the test statistic if we assume that the null hypothesis is true. For the sample mean, the CLT states that the sampling distribution is approximately normal when the sample size is large.

Consider the body temperature example,

$$H_0: \mu = 37$$

$$H_A: \mu < 37$$

Suppose that  $\sigma^2 = 1$  is known and we have randomly selected a sample of 25 healthy people from the population and measured their body temperature.

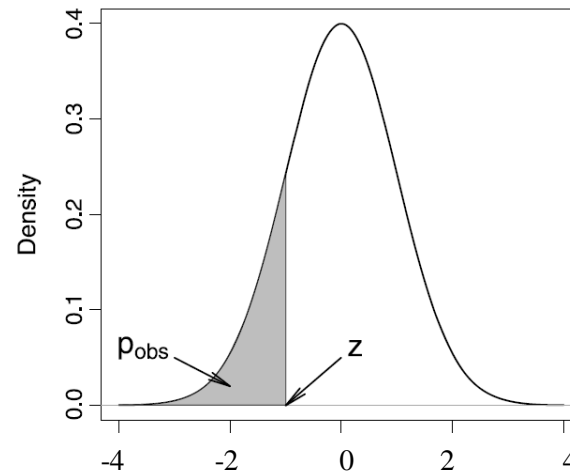
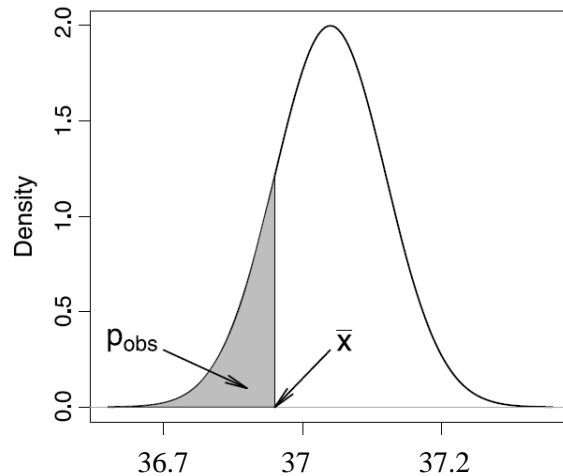
# Hypothesis Testing for the Population Mean-2

$$\left. \begin{array}{l} H_0: \mu = 37 \\ H_A: \mu < 37 \\ \sigma^2 = 1, n = 25 \end{array} \right\}$$

$\bar{X} \sim N(\mu, \sigma^2/n)$  The distribution is exactly normal if the variable itself is normal and the population variance is known

Now, suppose that the null hypothesis is true and the population mean is  $\mu = 37$ . By setting  $\mu$  to 37, the sampling distribution of  $\bar{X}$  becomes

$$\bar{X}|H_0 \sim N(37, 0.04)$$



The shaded area shows the lower-tail probability of the observed sample mean,  $\bar{x} = 36.8$ . This is the **observed significance level**, **p-value**, which is denoted as  $p_{obs}$ .

After standardizing, the p-value corresponds to the lower tail probability of  $z = -1$   $\left(z = \frac{36.8 - 37}{\sqrt{0.04}}\right)$  based on the standard normal distribution.

# z-Tests of the Population Mean

In practice, it is more common to use the standardized version of the sample mean as our test statistic,

$$Z = \frac{\bar{X} - 37}{0.02} \quad \{Z \sim N(0,1)\}$$

$\bar{x} = 36.8$ , to find the corresponding value for the random variable  $Z$ , we standardize  $\bar{x}$  in the same way we standardized  $\bar{X}$ . We denote this value as  $z$ :

$$z = \frac{36.8 - 37}{0.02} = -1$$

Now, instead of finding the  $p - value$  based on the lower tail probability of  $\bar{x} = 36.8$ , we can find it based on the lower tail probability of  $z = -1$ :

$$p_{obs.} = P(Z \leq -1)$$

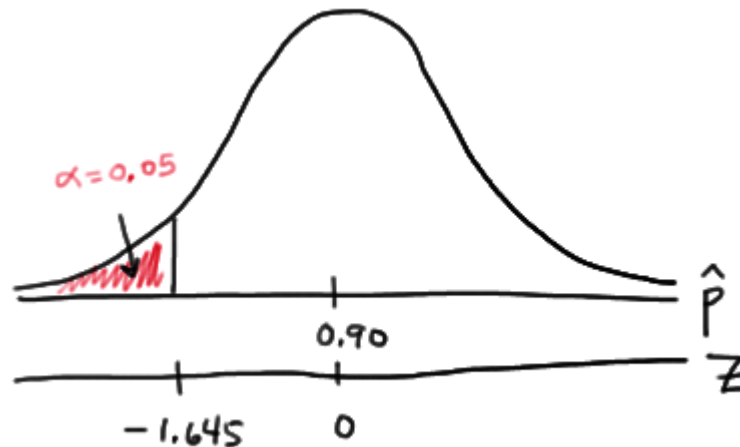
The  $p - value$  obtained based on the standardized test statistic  $Z$  is exactly the same as the  $p - value$  obtained based on the original test statistic  $\bar{X}$ .



# Interpretation of $p$ – *value*

The  $p$  – *value* is the conditional probability of extreme values (as or more extreme than what has been observed) of the test statistic assuming that the null hypothesis is true.

Based on the description of the  $p$  – *value*, we can interpret it as a measure of agreement between the observed data and the null hypothesis. Smaller  $p$ -values mean less agreement and provide stronger evidence against the null hypothesis.



# Interpretation of $p - value$

In order to use the  $p - value$  to decide whether we should reject the null hypothesis, a convenient approach is to prespecify a **cutoff** for the p-value and reject the null hypothesis if  $p_{obs.}$  is below the cutoff (i.e., when the measure of agreement between the null hypothesis and observed data is less than an acceptable level).

This cutoff is called the **significance level** or the size of the test. This is the acceptable type I error probability, i.e., the probability of rejecting the null hypothesis when it is true. As mentioned, we denote this probability  $\alpha$ .

The common significance levels are **0.01**, **0.05**, and **0.1**.

If  $p_{obs.}$  is less than the assumed cutoff, we say that the data provides statistically significant evidence against  $H_0$ , and we call the results statistically significant; that is, the difference between the observed value of the test statistic (here, 36.8) and the value specified by the null hypothesis (here, 37) is statistically significant.

For the body temperature example where  $p_{obs.} = 0.16$ , if we set the significance level at 0.05, we say that there is **not significant** evidence against the null hypothesis  $H_0: \mu = 37$  at the 0.05 significance level, so we **do not reject** the null hypothesis.

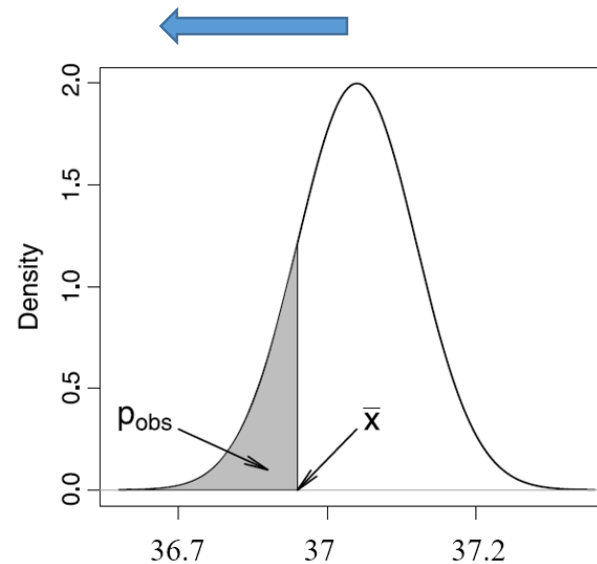
# One-Sided Hypothesis Testing-1

For the body temperature example, we tested the null hypothesis:

$H_0: \mu = 37$  against the alternative

$H_A: \mu < 37$ .

We refer to such tests as **one-sided hypothesis testing**, where the departure from the null is in one direction (here, in the direction of lower values than 37).



# One-Sided Hypothesis Testing-2

Let us denote the population mean according to the null hypothesis as  $\mu_0$ . Then, for the previous examples, we can express our alternative hypothesis that the population mean is less than a certain value as

$H_A: \mu < \mu_0$ . Likewise, our null hypothesis is  
 $H_0: \mu = \mu_0$ .

In this case, we quantified the support for the null hypothesis by finding the probability of test statistic values as small or smaller than the observed value if the null hypothesis is true.

The values more extreme than  $\bar{x}$  (here, smaller than the observed mean 36.8) represent a larger departure from  $\mu_0$  and provide stronger evidence against the null.

In some situations, we might hypothesize that the population mean is greater than a specific value and express our hypothesis as

$H_A: \mu > \mu_0$ . Our null hypothesis is still  
 $H_0: \mu = \mu_0$ .

This is also a one-sided test since the departure from the null is still in **one direction**: toward values larger than  $\mu_0$ .

# One-Sided Hypothesis Testing- Example

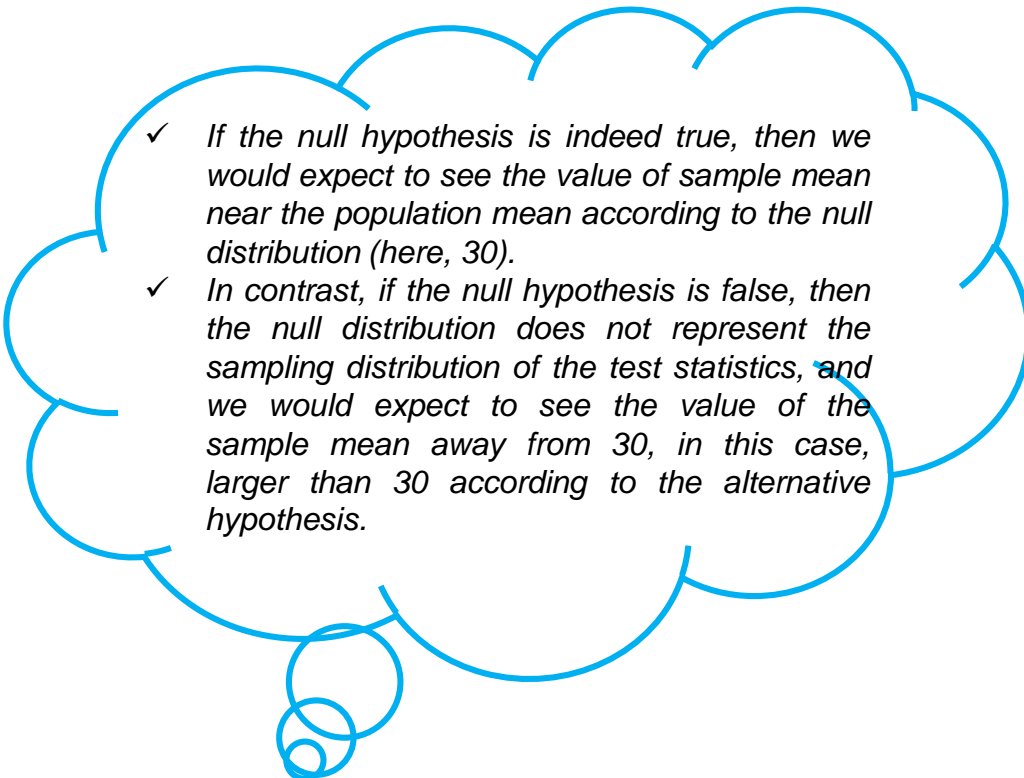
Suppose that we have observed that many Pima Indian women suffer from diabetes. We know that obesity and diabetes are related; we might therefore hypothesize that this population is obese on average, where obesity is defined as BMI higher than 30.

If we denote the population mean of BMI for Pima Indian women, we can then express our hypothesis as  $\mu > 30$ . In this case, the null hypothesis is,

$H_0: \mu = 30$ ; that is,  $\mu_0 = 30$ .

For illustrative purposes, suppose that we have obtained a sample of size  $n = 100$  from the population of Pima Indian women. Further, suppose we know that the population variance is  $\sigma^2 = 6^2$ .

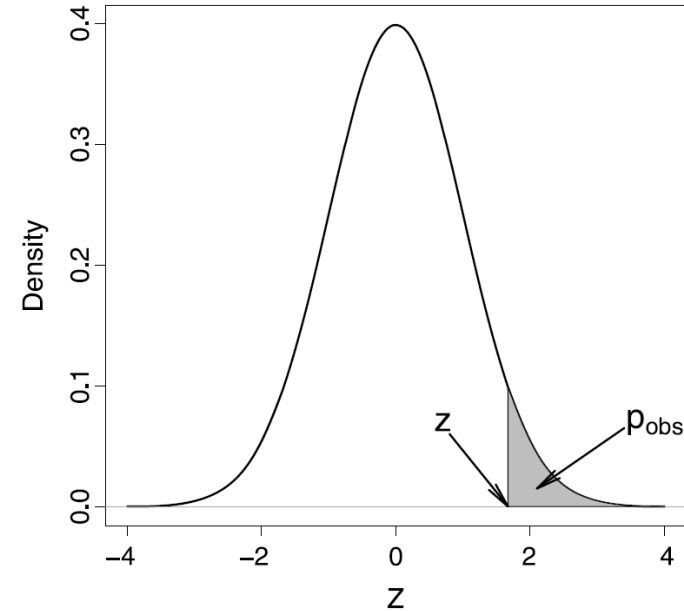
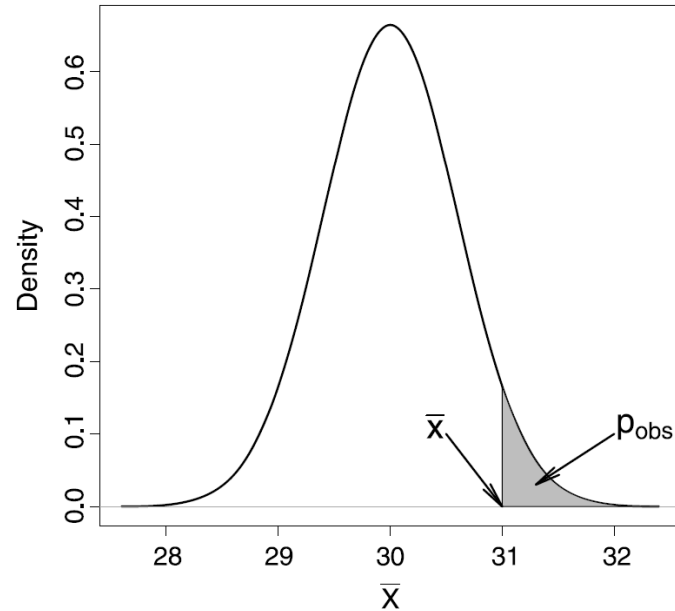
Suppose that from our sample of 100 Pima Indian women we find that the sample mean is  $\bar{x} = 31$ .

- 
- ✓ If the null hypothesis is indeed true, then we would expect to see the value of sample mean near the population mean according to the null distribution (here, 30).
  - ✓ In contrast, if the null hypothesis is false, then the null distribution does not represent the sampling distribution of the test statistics, and we would expect to see the value of the sample mean away from 30, in this case, larger than 30 according to the alternative hypothesis.

# One-Sided Hypothesis Testing- Example

$$H_0: \mu = 30,$$
$$H_A: \mu > 30.$$

$$p_{obs.} = P(\bar{X} \geq 31)$$



As before, we can standardize the test statistic by subtracting the mean and dividing the result by the standard deviation:

$$Z = \frac{\bar{X} - 30}{6/\sqrt{100}} \quad \{Z \sim N(0,1)\}$$

$$z - score = \frac{\bar{x} - 30}{6/\sqrt{100}} = \frac{31 - 30}{0.6} = 1.67$$

$$p_{obs.} = P(Z \geq 1.67)$$

# One-Sided Hypothesis Testing- Example

$$H_0: \mu = 30,$$

$$H_A: \mu > 30.$$

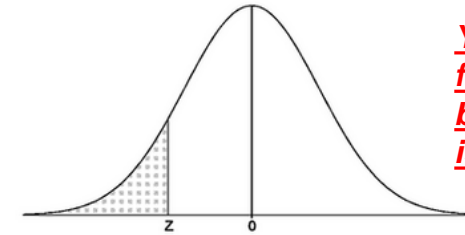
$$p_{obs.} = P(Z \geq 1.67)$$

For continuous random variables,  $P(Z \geq 1.67) = P(Z > 1.67)$  since the probability of any specific value (here, 1.67) is zero.

$$p_{obs.} = 0.0475$$

We can reject the null hypothesis at 0.05 level, we can conclude that the population mean of BMI for Pima Indian women is higher than 30 and the difference is statistically significant.

Cumulative Standard Normal Distribution Table



*You can also use this table for this type of problems because this distribution is symmetrical!*

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.00	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
-0.10	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.20	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.30	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.40	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.50	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.60	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.70	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.80	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.90	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-1.00	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-1.10	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.20	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.30	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.40	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.50	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.60	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.70	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.80	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.90	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-2.00	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183

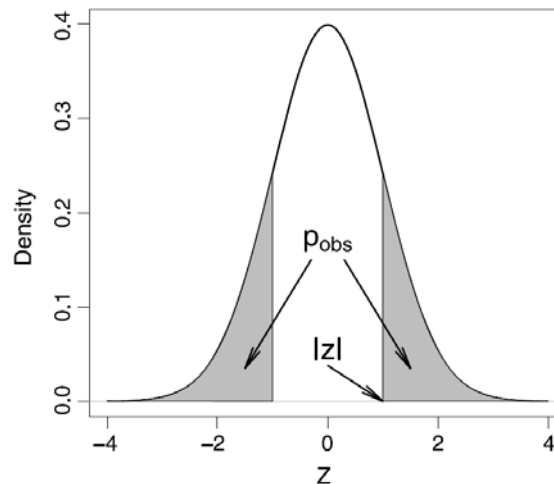
# Two-Sided Hypothesis Testing

For many hypothesis testing problems, we might be indifferent to the direction of departure from the null value. In such cases, we can express the null and alternative hypotheses as,

$$H_0: \mu = \mu_0.$$

$$H_A: \mu \neq \mu_0, \text{ respectively.}$$

Then we consider both large positive values and small negative values of z-score as evidence against the null hypothesis, and our alternative hypothesis is referred to as **two-sided**.





# Two-Sided Hypothesis Testing

Suppose we believe that the average normal body temperature is different from the accepted value 37°C, but we are not sure whether it is higher or lower than 37.

Then the null hypothesis remains,

$H_0: \mu = 37$ , but the alternative hypothesis is expressed as

$H_A: \mu \neq 37$ .

As before, we calculate the sample mean  $\bar{x} = 36.8$  and standardize it to obtain the ***z – score***, which is  $-1$ .

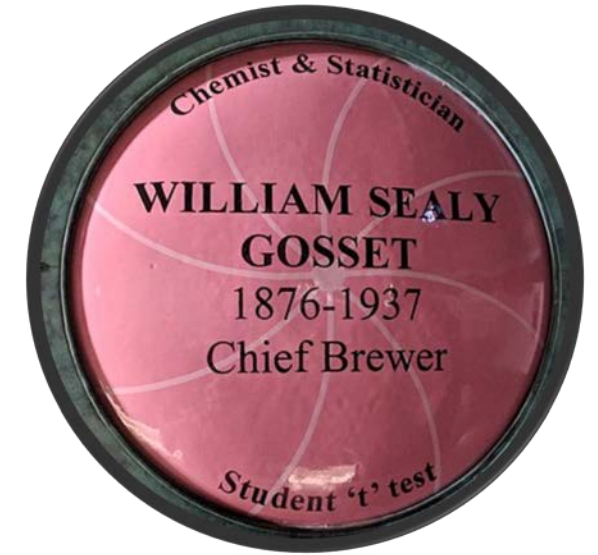
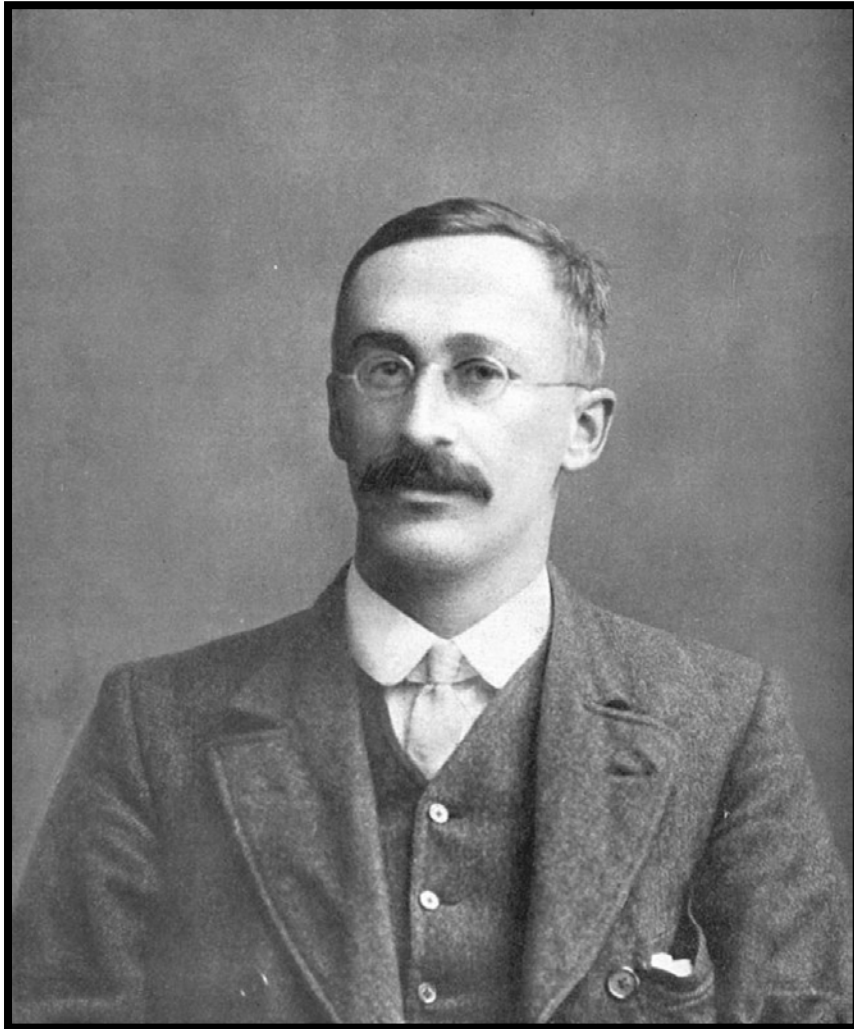
To find the ***p – value*** (observed significance level) we need to add the probabilities for  $Z \leq -1$  and  $Z \geq 1$ :

$$p_{obs.} = P(Z \leq -1) + P(Z \geq 1)$$

$$p_{obs.} = 2 * P(Z \geq 1) = 2 * 0.16 = 0.32.$$

The p-value is greater than typical significance levels such as 0.01, 0.05, and 0.1, so we cannot reject it at these levels. Therefore, we conclude that the observed difference is not statistically significant, and could be due to chance alone.

# Questions?



**William Sealy Gosset** (13 June 1876 – 16 October 1937) *was Head Brewer of Guinness, Head Experimental Brewer of Guinness, and father of modern British statistics. He pioneered small sample experimental design and analysis with an economic approach to the logic of uncertainty. Gosset published under the pen name "Student," and developed most famously **Student's t-distribution** - originally called Student's "z" - and "Student's test of statistical significance".*[11](#)