BIM 105
Probability and Statistics for Biomedical Engineers

David M. Rocke

Department of Biomedical Engineering

Interval Estimates

The idea of an interval estimate is to provide an interval $\left[a,b\right]$ that is "likely" to contain the unknown parameter value θ. Suppose we are interested in the calcium content of a cell-growth medium, and in particular, we want to know that the mean content is across batches of the medium. Suppose we take a sample of size 100 and the sample mean calcium content is 36 gm/L with a sample standard deviation of 14 gm/L.

We know the sample mean has standard deviation (also call standard error of the mean)

$\sigma / \sqrt{100} \doteq 14 / 10 = 1.4$.

What values of the population mean μ are consistent with this evidence?

$\overline{x} = 36$

$SD(\overline{x}) \approx 1.4$

$\mu = 38$ or $\mu = 34$ are consistent with the findings.

$\mu = 42$ or $\mu = 30$ are not consistent with the findings.

Why?

Confidence Intervals

- A 95% confidence interval is an interval such that the probability that the true parameter value is in the interval is 0.95.
- We can also define 99% confidence intervals or, in general, the $100(1 - \alpha)\%$ confidence interval for any value of $0 < \alpha < 1$.
- The usual interpretation of this statement is based on the idea that the parameter has a fixed value that we do not know.
- We take a sample (which is random) and calculate the interval (which is random) and if we repeated this procedure, some of the intervals contain the true parameter value and some don't.

- In such repeated sampling, 95% of the resulting intervals are supposed to contain the true parameter value.
- This is called the *frequentist* approach since it is based on hypothetical repeats of the experiment and a statement about how frequently the interval will contain the true parameter value.
- An alternative approach is to treat the parameter value as itself a random variable (which we do not observe). In this *Bayesian* approach, the interval is based on the data, and then the statement that the parameter value is in the interval is a probability statement about the random parameter value.
- Much of the time, the two approaches yield very similar results.
- We will concentrate on the first approach, which is the more usual in science and engineering.

The calcium content across batches of a cell-growth medium has mean μ and variance σ^2. The sample mean $\bar{x}$ of a sample of size n has

$$E(\bar{x}) = \mu$$

$$V(\bar{x}) = \sigma^2 / n$$

From the central limit theorem, $\bar{x}$ is approximately normally distributed so that

$$\bar{x} \sim N(\mu, \sigma^2 / n)$$

$$\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

$$P(-z_{.025} < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} < z_{.025}) = 0.95$$

$$P(\mu - z_{0.025}\sigma / \sqrt{n} < \bar{x} < \mu + z_{0.025}\sigma / \sqrt{n}) = 0.95$$

$$P(\mu - z_{0.025}\sigma/\sqrt{n} < \overline{x} < \mu + z_{0.025}\sigma/\sqrt{n}) = 0.95$$

The event above is that both the following are true:

$\overline{x} < \mu + z_{0.025}\sigma/\sqrt{n}$ and

$\mu - z_{0.025}\sigma/\sqrt{n} < \overline{x}$ but

$$P(\overline{x} < \mu + z_{0.025}\sigma/\sqrt{n}) = P(-\mu < -\overline{x} + z_{0.025}\sigma/\sqrt{n}) = P(\mu > \overline{x} - z_{0.025}\sigma/\sqrt{n})$$

$$P(\mu - z_{0.025}\sigma/\sqrt{n} < \overline{x}) = P(-\overline{x} - z_{0.025}\sigma/\sqrt{n} < -\mu) = P(\overline{x} + z_{0.025}\sigma/\sqrt{n} > \mu)$$

The left-hand side is a statement about $\overline{x}$ that is true 97.5% of the time by definition.

The right-hand side is a statement about μ that is equivalent

and so also true 97.5% of the time

$P(\overline{x} < \mu + z_{0.025}\sigma / \sqrt{n})$ Statement about $\overline{x}$

$= P(-\mu < -\overline{x} + z_{0.025}\sigma / \sqrt{n})$

$= P(\mu > \overline{x} - z_{0.025}\sigma / \sqrt{n})$ Statement about μ

$P(\mu - z_{0.025}\sigma / \sqrt{n} < \overline{x})$ Statement about $\overline{x}$

$= P(-\overline{x} - z_{0.025}\sigma / \sqrt{n} < -\mu)$

$= P(\overline{x} + z_{0.025}\sigma / \sqrt{n} > \mu)$ Statement about μ

$0.95 = P(\mu - z_{0.025}\sigma / \sqrt{n} < \overline{x} < \mu + z_{0.025}\sigma / \sqrt{n})$ Statement about $\overline{x}$

$= P(\overline{x} - z_{0.025}\sigma / \sqrt{n} < \mu < \overline{x} + z_{0.025}\sigma / \sqrt{n})$ Statement about μ

$$P(\overline{x} - z_{0.025}\sigma / \sqrt{n} < \mu < \overline{x} + z_{0.025}\sigma / \sqrt{n}) \approx P(\overline{x} - z_{0.025}s / \sqrt{n} < \mu < \overline{x} + z_{0.025}s / \sqrt{n})$$

$$P(\overline{x} - z_{0.025}\sigma / \sqrt{n} < \mu < \overline{x} + z_{0.025}\sigma / \sqrt{n}) \approx P(\overline{x} - z_{0.025}s / \sqrt{n} < \mu < \overline{x} + z_{0.025}s / \sqrt{n})$$

Suppose we take a sample of size 100 and the sample mean calcium content is 36 gm/L with a sample standard deviation of 14 gm/L. Then a 95% confidence interval is

$$\overline{x} \pm z_{\alpha/2}s / \sqrt{n}$$

$$36 \pm (1.960)(14) / \sqrt{100} = 36 \pm (1.960)(1.40) = 36 \pm 2.744$$

$$(33.256, 38.744)$$

Large-Sample Intervals for the Mean

A large-sample $100(1-\alpha)\%$ confidence interval for the mean μ

based on the sample mean $\overline{x}$ and sample standard deviation s of a sample of size n is

$$\overline{x} \pm z_{\alpha/2} s / \sqrt{n}$$

The statistic $s/\sqrt{n}$ is called the *standard error* of $\overline{x}$.

As against the statistic s which is the *sample standard deviation.*

If we repeat the process of taking a sample of size n and computing the confidence interval, then the fraction of the time that the interval contains μ is $(1-\alpha)$.

CI	**50%**	**90%**	**95%**	**99%**	**99.9%**
$z_{\alpha/2}$	0.6745	1.645	1.960	2.576	3.291

Confidence Interval for the IgM Data

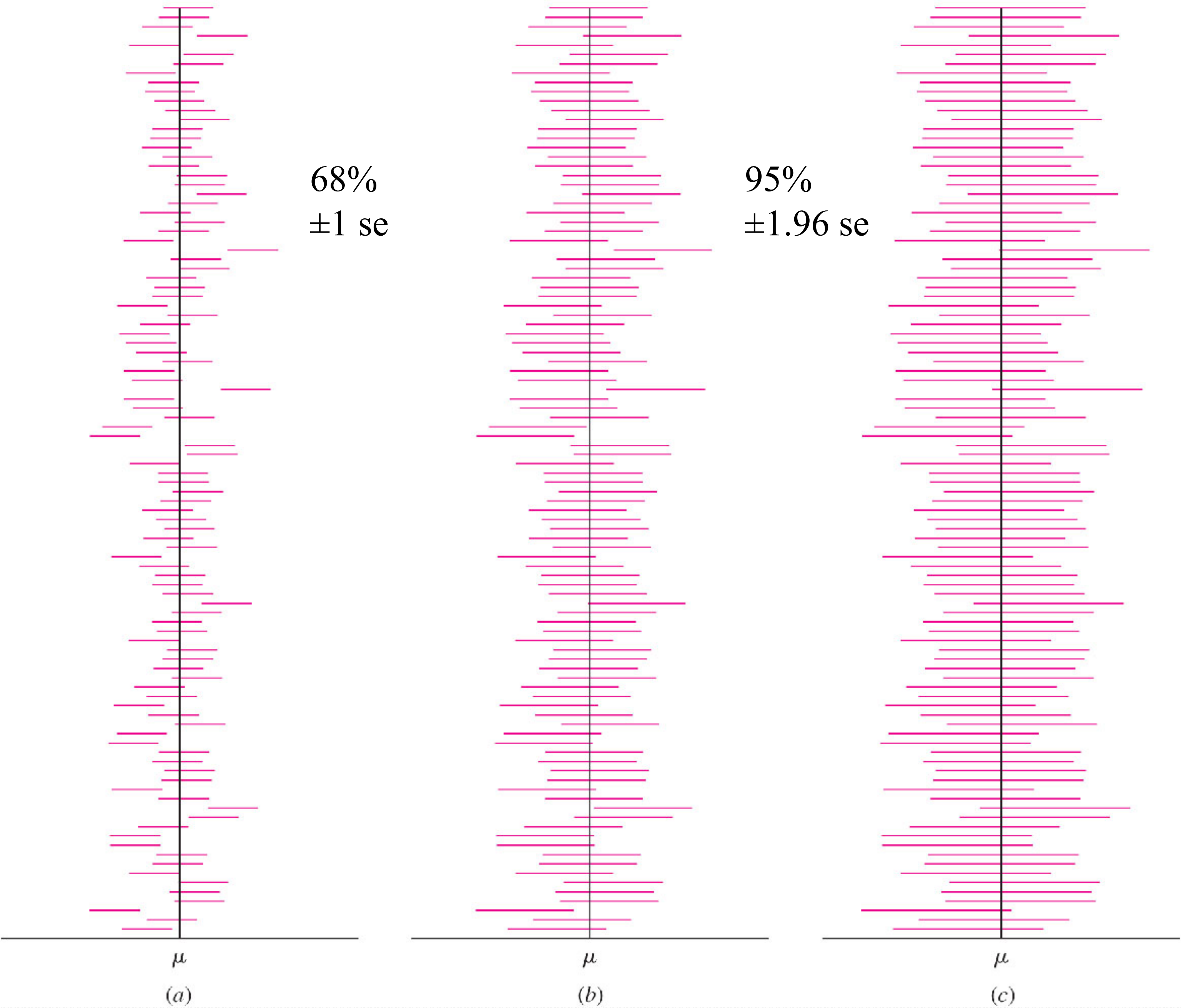
- There are 298 observations in the IgM data set (concentrations of IgM in g/L), so this is a large sample.
- Since the distribution looks more normal on the log scale, we do the analysis on that scale.
- This gives us an interval for the mean of ln(IgM), which is the geometric mean of IgM on the original scale.
- The mean is –0.3632 and the standard deviation is 0.5469

Confidence Interval for the IgM Data

- The mean is -0.3632, the standard deviation is 0.5469, and the sample size is $n = 298$
- A 95% CI on the natural log scale is
  $-0.3632 \pm (1.960)(0.5469)/\sqrt{298}$
  $-0.3632 \pm (1.960)(0.0317)$
  -0.3632 ± 0.0621
  $(-0.4253, -0.3011)$
- Or (0.653, 0.740) on the original scale of g/L, which we obtain by exponentiating each end of the CI.

Behavior of Confidence Intervals

- A 95% confidence interval will cover the true value 95% of the times that the procedure is run.
- A particular 95% confidence interval either covers the true value or not, though we don’t know which in an particular case.
- So (from the frequentist point of view) it is incorrect to say that the probability that the interval covers the true value is 95%.
- That is why we call it 95% confidence and not 95% probability.



68%
±1 se

95%
±1.96 se

99.7%
±3 se

Confidence Intervals in MATLAB

```
>> [h p ci stats] =ztest(ligm,0,std(ligm))

% ignore h and p. These are for a hypothesis test that the
% mean of the log IgM is 0, which is not meaningful
                                   %
h =
     1
p =

   2.0190e-30

ci =
   -0.4253
   -0.3011

stats =

  -11.4632        %this is the z-statistic testing mu=0, not relevant here
```

Confidence Intervals in MATLAB

```
Confidence interval for mean log IgM

ci =
   -0.4253
   -0.3011

Confidence interval for geometric mean IgM

>> exp(ci)

ans =

    0.6536
    0.7400
```

An Interpretation of CI's

Another perspective on confidence intervals is to ask the following question: given a sample mean $\overline{x}$, sample variance s^2, and a possible population mean μ, is the possible value μ for the mean consistent with the data?

If μ is too far from $\overline{x}$, then we would say no.

The distance from $\overline{x}$ to μ is $|\overline{x}-\mu|$

If that distance is bigger than $1.96s/\sqrt{n}$, then we would perhaps say no.

$\dfrac{\overline{x}-\mu}{s/\sqrt{n}} > 1.96$ means that μ is not consistent with the data.

If true mean is μ, then $\overline{x} \sim N(\mu, \sigma/\sqrt{n})$

$\Pr(|\overline{x}-\mu| > 1.960\sigma/\sqrt{n}) = .05$

The 95% CI is the set of possible values of μ that are consistent with the data with a 5% criterion for not consistent.

Example

- The birth weight of 189 infants was collected at Baystate Medical Center in Springfield, MA.
- The mean weight was 2945 gm with a standard deviation of 729 gm.
- A 95% confidence interval for the mean birth weight is $2945 \pm (1.960)(729)/\sqrt{189} = 2945 \pm 104 = (2841, 3049)$

Sample Size Determination

A $(1-\alpha)100\%$ confidence interval for the mean in large samples is

$$\overline{x} \pm z_{\alpha/2} s / \sqrt{n}$$

If we want the confidence interval to have half-width *w* then

$$w = z_{\alpha/2} s / \sqrt{n}$$

$$n = \frac{z_{\alpha/2}^2 s^2}{w^2}$$

Hypothesis Tests

- We have seen that a confidence interval for a parameter based on a sample from a population can be thought of as the set of parameter values that is consistent with the data collected, at the specified level of confidence.
- For example, if a 95% confidence interval for the mean was 111 ± 12, or (99, 123), then values for the population mean consistent with the data include 100, 99, 120, but not 98 or 125.

Hypothesis Tests

- A hypothesis test is used when we have one specific value of the parameter that we wish to know if it is consistent with the data.
- We could just look to see if it is in the 95% confidence interval, but that does not distinguish between μ = 122.5, which is almost out of the CI, and μ = 114, which is near the sample mean.
- It also does not distinguish between 98, barely out of the CI, and 90, which is very far from the CI.

- The hypothesis we wish to test is called the *null hypothesis*, largely because we often want to test if the difference between two or more things is null, that is zero.
- We may or may not have a specific alternate hypothesis, but this is often not needed.
- We construct a statistic which under the null hypothesis has a known distribution; that is, we temporarily assume that the null hypothesis is true, and see what that would imply.
- If the value of this statistic is sufficiently unlikely under the null hypothesis, then we conclude that the null hypothesis probably is not true.
- **We can never conclude that the null hypothesis is true, only that it is not shown to be false.**

Example

The mean calcium content of a type of cell growth medium is supposed to be 30 g/L. A sample of 100 batches has a mean concentration of 36 g/L with a standard deviation of 14 g/L. Test the hypothesis that the population mean is actually 30 as specified.

$H_0 : \mu = 30$

Under the null hypothesis,

$$\frac{\bar{x} - 30}{14/\sqrt{100}} \sim N(0,1)$$

$$\frac{36-30}{1.4} = \frac{6}{1.4} = 4.286 \sim N(0,1)?$$

This is too large to have come from a standard normal random variable. Almost all values of a standard normal lie between -3 and 3.

Fixed Level Tests

- We compute a statistic that should be standard normal if the null hypothesis is true.
- If that statistics is either too large (like 3) or too small (like -3), then we will reject the null hypothesis, that is, decide that it is likely not to be true.
- For example, the chance that a standard normal random variable is larger than 1.960 in magnitude is 5%.
- If we choose 5% as the criterion level, then we will reject the null hypothesis whenever the z statistic is larger than 1.960 or smaller than –1.960.

P-Values

- Suppose we had a sample mean level of 33 g/L with a standard deviation of 14 g/L and $n = 100$. To test the null hypothesis that the population mean is 30, we calculate $z = (33 - 30)/1.4 = 2.143$. This is approximately standard normal if the true mean is 30.
- The chance that z exceeds 2.143 is 0.0161. We double this to account for the fact that we might have gotten a sample mean below 30. This is called a *two-sided test*.
- The (two-sided) p-value is then 2(0.0161) = 0.0322
- This means that if the null is true, we would get a value as extreme as this only 3.22% of the time.
- At this point, we may choose to reject the hypothesis based on a p-value threshold like 5%.
- We have two choices: either something unlikely has happened or else the null hypothesis is false.

We Never "Accept the Null"

- If we have a null hypothesis that the population mean is 30, then we know from first principles that it is false at some decimal point. The true mean may be in 30 ± 1 or 30 ± 0.1 or 30 ± 0.01, but it is not in 30 ± 10^{-9}.
- So when we reject the null, it means that we know that the true mean is not 30 or very close to 30.
- If we do not reject the null, that does not mean that we know that the true mean is exactly 30.
- In fact, the **confidence interval contains all the values of the true mean consistent with the data**.

A sample of 100 batches has a mean concentration of 33 g/L with a standard deviation of 14 g/L. A 95% confidence interval for μ is

$$33 \pm (1.960)(14/\sqrt{100})$$

$$33 \pm 2.744$$

$$(30.256, 35.744)$$

The value 30.256 is the smallest value still in the 95% confidence interval.

A sample of 100 batches has a mean concentration of 33 g/L with a standard deviation of 14 g/L. A 95% confidence interval for μ is

$(30.256, 35.744)$

The value 30.256 is the smallest value still in the 95% confidence interval.

If we test the hypothesis that the population mean is actually 30.256, then

$H_0 : \mu = 30.256$

Under the null hypothesis,

$$\frac{\overline{x} - 30.256}{14/\sqrt{100}} \sim N(0,1)$$

$$\frac{33 - 30.256}{1.4} = \frac{2.744}{1.4} = 1.960$$

The p-value is 2(0.0250)=0.05 or 5%

The 95% confidence interval is the set of possible null hypotheses for μ such that the p-value is greater than 5%

We can make a 95% confidence interval by inverting the test because the 95% confidence interval is the set of possible null hypotheses for μ such that the p-value is greater than 5%. This happens when the test statistic is larger than 1.960 in absolute value.

$\frac{33-\mu}{14/\sqrt{100}} < 1.960$ z-statistic smaller than 1.960 = $z_{0.025}$

$33-\mu < (1.960)(14/\sqrt{100})$

$\mu > 33-(1.960)(14/\sqrt{100})$ μ larger than the lower end of the 95% CI

$\frac{33-\mu}{14/\sqrt{100}} > -1.960$ z-statistic larger than 1.960 = $z_{0.025}$

$33-\mu < -(1.960)(14/\sqrt{100})$

$\mu < 33+(1.960)(14/\sqrt{100})$ μ smaller than the upper end of the 95% CI

$33-(1.960)(14/\sqrt{100}) < \mu < 33+(1.960)(14/\sqrt{100})$

$30.256 < \mu < 35.744$

Hypothesis Tests in MATLAB

```
>> [h p ci stats] =ztest(ligm,-0.5,std(ligm))
Null hypothesis is that the mean of the log iGM values is -0.5

h =
     1                    # the hypothesis is rejected at p = 0.05
p =

   1.5657e-05             # this is the actual two-sided p-value


ci =                      # Confidence interval as before

   -0.4253
   -0.3011

stats =

    4.3192                # t-statistic
```

Type I and Type II Errors

- We make a Type I error when we reject a null hypothesis even though it is in fact true.
- By definition, we will always make some Type I errors, but we can explicitly control how many we make.
- If we use a 5% criterion, then we will reject a true null hypothesis 5% of the time.
- We make a Type II error when we fail to reject a false null. We can’t calculate this without specifying what false null we are talking about, and it cannot reasonably be so close to the null that it makes no difference.

Type I and Type II Errors

- For example, if the null hypothesis is that the population mean is 30 g/L, we may choose to calculate the Type II error for a concentration of 35 g/L.
- The value of the parameter that we use to calculate the Type II error is sometimes called the *alternate hypothesis*.
- For a given sample size, the *power* of a test is one minus the Type II error. This depends on the alternate hypothesis.
- We would like to have small Type I error and large power, but this may require very large values of n.

Why do we double the tail area?

- Suppose we want to test the hypothesis that the true mean is 30 g/L. A sample of 100 gives a mean of 33 g/L and a standard deviation of 14 g/L.
- If we construct the test statistic, we get $z = 2.143$, which has a tail area to the right of 0.0161. This happens 1.61% of the time when the null is true.
- The p-value is the fraction of time that we would reject the null if the null is true, and it looks like this would be 1.61%.
- But if the sample mean had been 27 g/L, then the test statistic would have been $z = -2.143$, which has a tail area of 1.61% to the left.
- So if we use the 1.61% criterion, we reject the null 1.61% of the time with positive z and 1.61% with negative z, for a total of 2(1.61%) = 3.22%

What about one-sided tests?

- The book discusses a one-sided test, used when one expects only deviations on one side of the null.
- This should usually not be used because can be a form of self deception.
- Suppose I think that vitamin C prevents cancer. I might advocate a hypothesis test with null hypothesis that vitamin C has no effect and alternate hypothesis that it helps. Say I use a 5% cutoff.
- I am then promising that no matter how much the evidence seems to show that vitamin C causes cancer, I will not make any conclusion from the data.
- This is obvious nonsense; if the evidence looks strong, I will conclude that vitamin C is harmful.
- But then my chance of being wrong is 5% plus the chance that the evidence of harm is sufficiently great, which might also be 5%.

What about one-sided tests?

- One-sided tests can be used in a decision context. If we test a new manufacturing process for a medical device to see if the defective rate can be reduced from 4% to a lower rate, then we only want to adopt or pursue the new process if the defective rate is lower than 4%. If the defective rate of the new process is the same or higher, then we don't want to use the new process.
- One-sided tests are appropriate for some decision analysis contexts. They are not usually appropriate for science.
- It is not sufficient to say that "we want to know if the mean is less than 12." It must be the case that departures on the other side lead to the same conclusions as if the sample mean is near the true mean.
- Some book problems will assume a one-sided test. You can tell if the alternate is phrased with "greater than" or "less than" instead of "not equal." Then you should do the one-sided test.

Non-inferiority Tests

- A type of stent placed in coronary arteries as therapy for atherosclerosis has a mean increase in blood flow of 32 ml/min.
- A new type of stent is proposed to be equivalent to the standard type.
- It is unfair to require that the new stent be shown to have a higher mean increase in blood flow than the standard because this would require a sample mean blood flow increase much larger than 32 (depending on the standard deviation and the sample size.

Non-inferiority Tests

- For example, if 100 patients had a mean increase of 35 ml/min with a standard deviation of 25ml/min, then the test statistic $(35-32)/(25/\sqrt{100}) = 1.2$. The area to the right of 1.2 is 0.1151, so not significant.
- Instead, the non-inferiority method says that the new device should be shown to be "equivalent" to the old one or better, where equivalent is ±20%. The band of 32 ± 6.4 is a sort of zone of equivalence. Thus we only need to show that the mean increase is at least $32 - 6.4 = 25.6$ ml/min.
- $(35-25.6)/(25/\sqrt{100}) = 3.76$ which has $p \sim 0$. Correctly one sided.
- This is a decision context (approve/don't approve drug)

Hypothesis Tests Summary

- To test a statistical hypothesis you use a statistic whose distribution is known if the null hypothesis is true.
- Compute the probability that the statistic would be as greater (if positive) or less than (if negative) and double it. This is the (two-sided) p-value.
- Reject the null hypothesis if this p-value is sufficiently small.
- A $100(1 - \alpha)\%$ confidence interval is the set of values of the parameter such that the null hypothesis two-sided p-value is less than α.

Hypothesis Tests for the Mean, Large Samples

If we have a sample $x_1, x_2, ..., x_n$ from a population with mean μ

and we have a possible value $\mu = \mu_0$ in mind

and n is large (say > 30)

and want to see if the data are consistent with that value of μ,

then we construct the test statistic

$$z = \frac{\overline{x} - \mu}{s / \sqrt{n}}$$

and we compare it to the normal distribution.

If z is positive and $\Pr(Z > z) = \alpha$,

then the p-value of the test is 2α.

If z is negative and $\Pr(Z < z) = \alpha$,

then the p-value of the test is 2α.

Confidence Intervals for the Mean, Large Samples

If we have a sample $x_1, x_2, ..., x_n$ from a population with mean μ and n is large (say > 30)

Let $\overline{x}$ and s be the sample mean and standard deviation

Then a $(1-\alpha)100\%$ confidence interval for the mean is

$$\overline{x} \pm z_{\alpha/2} s / \sqrt{n}$$

This is the set of possible values of μ consistent with the data at level α