Statistical Inference: Introduction

Outline of presentation:

1) How to form confidence interval for population mean μ when population sd σ is known and population is normally distributed.

2) How to test hypothesis that pop'n mean is some specified number in same situation.

3) How to form confidence interval for population mean when σ is unknown but n is large.

4) How to test hypothesis that μ is some specified number in situation in 3).

5) How to find confidence interval for population proportion p when n is large.

6) How to test hypothesis about population proportion when n large.

7) Normal populations, small n, σ unknown.

Then back to why they work, how to interpret.

Confidence Intervals

Example: 25 precision measurements of the mass of an object.

Probability model or **statistical model** for results:

Each measured value is sum of three numbers:

1) true mass of object.

2) bias of measuring device.

3) random error.

In each measurement, 1) and 2) are the same.

Random errors are different each time: behave like sample with replacement from "hypothetical" population of possible errors.

Population mean of errors is 0.

Added assumption: no bias in measuring device.

If so: measurements like sample from population with mean μ which is the true mass of the object.

Added assumption: have used measuring device so often that we "know" the standard deviation, σ, of the measurement errors.

This means that our measurements are like a sample from a population with mean μ and standard deviation σ where we know σ.

Confidence interval method:

1) collect sample of n from population.

2) assume population mean μ unknown.

3) assume population sd σ known.

4) assume population distribution normal.

5) compute sample mean $\bar{x}$.

6) select desired confidence level (usually 0.95 or 95%).

7) find z so that area between $-z$ and z under normal curve is desired confidence level.

8) work out lower and upper limits:

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$$

9) these pair of numbers are a 95% (or other percentage) confidence interval for the population mean μ.

10) if we do many experiments of this sort and work out a 95% confidence interval each time then about 95% of the intervals work — i.e., the two numbers come out on either side of μ.

Actual mechanical steps are 1, 5, 6, 7, 8.

Specific example: weigh object 16 times.

Step 1: Get mean of 9.999594 g.

Long experience (SD of many mmnts of same object) on scale shows $\sigma = 25\mu g$ (25 micro-grams $= 25 \times 10^{-6} = 0.000025$ g).

Step 2: to find 80% confidence interval: must find z.

Need area from $-z$ to z to be 0.8.

So area to left of $-z$ plus area to right of z is 0.2.

Area to right of z is 0.1; area to left of z is 0.9.

From tables $z = 1.28$.

Step 3: confidence interval is

$$9.999594 \pm 1.28 \frac{25 \times 10^{-6}}{\sqrt{16}}$$
$$= 9.999594 \pm 0.000008$$

so runs from 9.999586 to 9.999602.

To be more confident: take bigger z.

For 95% confidence interval use $z = 1.96$.

Make sure you know why!

To be certain: IMPOSSIBLE.

Could true weight be below 9.999580? YES but this is unlikely.

How do I know?

Steps in hypothesis testing:

1) collect sample of n from population.

2) specify value of μ to be examined for credibility. Use notation μ_0.

(In weight example $\mu_0 = 9.999580$.)

3) assume population sd σ known.

4) assume population distribution normal.

5) work out sample mean $\bar{x}$.

6) compute z-statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma}$$

(In example

$$z = \frac{9.999594 - 9.999580}{0.000025/4} = 2.24$$

7) compute area to right of z or to left of z or outside range of $-z$ to z. Area is called P-value.

(In example with weight: area to right of z.)

Area is 0.0125.

8) Interpret P-value. Small values of P mean strong evidence against assumptions used to compute P. We assumed: true weight at or below 9.999580. So: "reject" that assumption.

More usual case: σ unknown.

Easiest case: large sample size n.

Get confidence interval for population mean:

1) compute $\bar{x}$ and s, sample mean and sd.

2) select confidence level as before.

3) find corresponding multiplier z so area from $-z$ to z is desired confidence level.

4) work out
$$\bar{x} \pm z\frac{s}{\sqrt{n}}$$
as your interval.

Notice: just substitute s for σ.

Example: want to estimate number of trees in some park infected by pine beetles. (All figures hypothetical.)

Setup: partition area into 10m by 10m plots. Imagine park is 10km by 10km so there are 1,000,000 plots.

Take sample of say 64 such plots.

Go to plot, count trees with pine beetles in plot (X_i).

Work out sample mean $\bar{X}$ and sample SD s;

Values: $\bar{X} = 2.5$, $s = 3$.

Get confidence interval for average number of infected trees per 10 by 10 plot in the park.

Mechanics: do 99% confidence interval.

Find z so that area to left is 0.995.

Find $z = 2.57$.

Confidence interval is

$$\bar{x} \pm z \frac{s}{\sqrt{n}} = 2.5 \pm 2.57 \frac{3}{\sqrt{64}}$$

Simplifies to 2.5 ± 0.96 or 1.54 to 3.46.

Note: units are trees per 10 by 10 plot. Likely want confidence interval for trees per hectare or per square km or for trees in the park.

To go from trees per plot to trees in park: multiply by number of plots in park.

We are 99% confident that there are between 1.54 million and 3.46 million infected trees in the park.

Hypothesis testing:

Sample of 144 salmon from 1 day's catch.

Concern: newly caught fish have higher dioxin / furan content than historically found.

Fish Tissue Action Level (FTAL): 1.5 parts per trillion for "protection of cancer related effects".

Measure d/f content in sample fish.

Find $\bar{X} = 1.3$, and $s = 0.8$ (ppt).

Trying to convince ourselves: average d/f level in catch below 1.5.

Mechanics: compute observed sample mean in standard units, assuming actual level is just at FTAL.
$$z = \frac{1.3 - 1.5}{0.8/\sqrt{144}} = -3$$

Then find area under normal curve beyond -3. Get 0.0013.

We call 0.0013 a P-value for a test of the hypothesis that the mean d/f concentration in today's catch is 1.5 (or more).

Conclusion: average d/f concentration in population (today's catch) is very likely below 1.5.

(If not then today's sample of 144 fish is one of a pretty rare group of samples – should only happen 1.3 times in 1000 trials.)

Confidence interval for a proportion.

Toss tack 100 times. It lands upright 46 times.

What is $p = Prob(U)$ on a single toss?

Solution:
$$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Meaning: $\hat{p} = 46/100 = 0.46$.

Value of z from normal curve as always: 1.96 for 95%.

CI is
$$0.46 \pm 1.96\frac{\sqrt{.46 \times .54}}{10} = 0.46 \pm 0.098$$

Notice that $n = 100$ trials gives wide range of credible values for p.

Comment: better approximation − page 508 in text.

Could $p = 1/2$?

Measure discrepancy using

$$z = \frac{\widehat{p} - 1/2}{\sqrt{0.5(1 - 0.5)/100}}$$

Get $z = -0.8$. Look up area outside $-z$ to z. Get

$$P = 0.4237$$

Not small, so no real evidence against $p = 1/2$.

Notice use of "two tails". More to come later.

Final pair of examples: n small, σ unknown.

Fifteen pairs of plants: each pair off spring of same parent plant. In each pair one cross-fertilized, one self-fertilized.

Record difference in heights (in eights of an inch): cross minus self.

The data

$$
\begin{array}{cccccccc}
49 & -67 & 8 & 16 & 6 & 23 & 28 & 41 \\
14 & 29 & 56 & 24 & 75 & 60 & -48 &
\end{array}
$$

Questions:

1) does crossing tend to produce taller plants?

2) how much average difference?

Q 1 asks for a hypothesis test.

Q 2 asks for a confidence interval.

Interpret Q 1 as question about population mean:

Let μ be the population average difference for cross minus self in such pairs.

Treat our 15 observations as sample of size $n = 15$ from population.

Q 1 becomes: is $\mu = 0$?

Mechanics of confidence interval:

1) work out summary statistics, $\bar{x} = 20.93$ and $s = 37.74$.

2) compute the interval

$$\bar{x} \pm t_{n-1,\alpha} \frac{s}{\sqrt{n}}$$

What is $t_{n-1,\alpha}$?

Multiplier found from "Student's t table" — table C in text.

Called t^* in text.

To find the multiplier:

a) select desired confidence level.

b) compute α from this level by subtracting level from 1 (or 100%) and dividing by 2.

c) look up t^* in Table C in line for $df = n - 1$. Go to column for α (called p in table).

Jargon: $n - 1$ is called the "degrees of freedom".

Example: for 95% Conf Int. Find $\alpha = 0.025$.

For our data set: $n - 1 = 14$.

In Table C find $t^* = 2.145$.

In our example, then, the CI is

$$20.93 \pm 2.145 \frac{37.74}{\sqrt{15}} = 20.93 \pm 20.90$$

which runs from 0.03 to 41.83.

We are 95% confident that true gain in average height due to cross-fertilization is in the range 0.03/8 to 41.83/8 inches.

Corresponding hypothesis test:

1) compute sample mean and sd as before.

2) compute t-statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

NOTE: exactly the same as large sample z formula.

3) Get P value from t table.

In our case:

$$t = \frac{20.93 - 0}{37.74/\sqrt{15}} = 2.148$$

How to find P-value:

A) from tables:

Look at row $df = 14$ in Table C.

Want area outside -2.148 to 2.148.

In Table see that area from -2.145 to 2.145 is 0.9500 and area from -2.264 to 2.264 ix 0.96

So

$$0.05 > P > 0.04$$

Of course, 2.145 is very close to 2.148 so P is very close to 0.05.

Interpretation:

P is on the small side so there is "significant" evidence against $\mu = 0$.

Jargon: say difference is **significant** if $P < 0.05$.

Now go back to the beginning:

Confidence interval for a parameter:

An interval, calculated from data, usually

$$\text{estimate} \pm \text{multiplier} \times \text{standard error}$$

with an associated **confidence level** C.

$$Prob(\text{interval includes parameter}) = C.$$

Intervals discussed so far:

1) for population mean, σ known, population distribution normal.

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

2) for population mean, σ known, sample size large.

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$$

Exactly the same formula as (1).

3) for population proportion, sample size large:

$$\widehat{p} \pm z \frac{\sqrt{\widehat{p}(1 - \widehat{p})}}{\sqrt{n}}$$

where $\widehat{p}$ is sample proportion.

4) for population mean, σ unknown, n small, population distribution normal:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

Not discussed yet but closely related:

5) for population mean, σ unknown, n large:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

or

$$\bar{x} \pm z \frac{s}{\sqrt{n}}$$

(For reasonably large n no perceptible difference. Simplest to always use t^* when σ is unknown.)

Case not covered:

If σ is not known and you doubt that the population distribution is normal: need to investigate **non-parametric methods**.

Jargon: the standard deviation of an estimate (guess calculated from the sample) is called the **standard error of the estimate**.

Example is $\sigma/\sqrt{n}$.

If σ is unknown we usually use the data to guess σ and get

Estimated standard error of $\bar{x}$ is $s/\sqrt{n}$.

Theory underlying confidence intervals.

Start with population: mean μ, sd σ.

1) if you sample with replacement

a) if population distribution is normal then sampling distribution of $\bar{x}$ is normal and sd of $\bar{x}$ is $\sigma/\sqrt{n}$

b) if n is large same is true approximately by central limit theorem.

2) if you sample without replacement and n is not too large compared to population size but n is large enough for central limit approx to be ok then same conclusion holds.

So: what is chance $\bar{x}$ comes out within z standard errors of μ?

Answer: want chance $\bar{x}$ is in interval

$$\mu - z\frac{\sigma}{\sqrt{n}} \text{ to } \mu + z\frac{\sigma}{\sqrt{n}}$$

Convert limits to standard units: subtract μ (mean of sampling distribution of $\bar{x}$) and divide by sd $(\sigma/\sqrt{n})$

Get $-z$ to z. Look up area in normal tables from $-z$ to z.

If we pick z in advance to make that area C the chance comes out to C.

So, for instance, with $z = 1.96$ we find chance is 0.95.

Key point: every time that $\bar{x}$ comes out within z standard errors of μ, our confidence interval includes μ.

So: chance confidence interval includes μ is C

More theory:

Start with population: mean μ, sd σ.

3) if n is large then s will be close to σ so chance using $s/\sqrt{n}$ will be almost the same as chance using $\sigma/\sqrt{n}$.

Start with population of whom proportion having trait is p.

4) if n is large then number X in sample with trait is approximately normally distributed: mean is np and sd is $\sqrt{np(1-p)}$.

So sample proportion $\hat{p} = X/n$ has approximately normal distribution.

Mean is $np/n = p$.

SD is $\sqrt{np(1-p)}/n = \sqrt{p(1-p)/n}$.

Compute chance sample proportion is within z times $\sqrt{p(1-p)/n}$ of p.

Limits are
$$p - z\sqrt{p(1-p)/n} \text{ to } p + z\sqrt{p(1-p)/n}$$
Convert to standard units: subtract p, divide by standard error.

Get $-z$ to z.

As in previous case area area from $-z$ to z is C as designed.

5) if n large then replacing p by $\hat{p}$ in standard error makes little difference.

So: chance sample proportion is within z times $\sqrt{\hat{p}(1-\hat{p})/n}$ of p is also about C.

So chance conf int includes p is about C.

More theory:
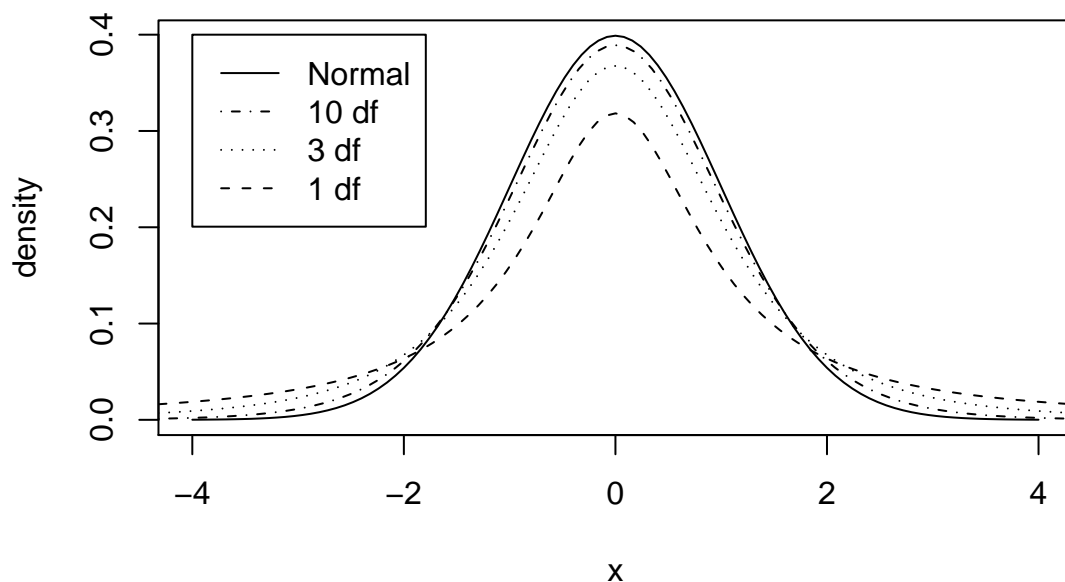
Population distribution normal, mean μ, sd σ.

Fact: sampling distribution of

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

follows curve called: **Student's t distribution on $n-1$ degrees of freedom**.

Some t curves with $N(0, 1)$ curve:

Essential point: chance that t comes out between $-t^*$ and t^* is area under t curve with $n-1$ df.

If t is between $-t^*$ and t^* then μ is between

$$\bar{x} - t^* \frac{s}{\sqrt{n}} \text{ and } \bar{x} + t^* \frac{s}{\sqrt{n}}$$

So: chance confidence interval works is area from $-t^*$ to t^*.

Table C shows values of t^* which give variety of areas.

Get middle areas along bottom; areas to right on top.

Summary of confidence intervals:

Steps to go through:

1) What parameter do we want a confidence interval for?

A: Population proportion, p, or B: population mean μ?

A: If population proportion is parameter of interest:

i) find correct multiplier z to give desired confidence level.

ii) compute sample proportion $\hat{p} = X/n$.

Here X is number of successes in n independent trials.

Can be used if sampling without replacement but n small compared to population size N.

iii) Compute interval:

$$\widehat{p} \pm z\sqrt{\frac{\widehat{p}(1 - \widehat{p})}{n}}$$

B: if population mean is parameter of interest. Identify case in list:

Case 1: population standard deviation σ known and sample size n large.

Case 2: population standard deviation σ known and sample size n is small and population distribution is normal.

Case 3: population sd unknown, sample size n is large.

Case 4: population sd unknown, sample size n is small, population distribution is normal.

Remaining cases: n small, population distribution not normal. Need non-parametric procedures or other help.

Now finish cases:

Cases 1 or 2: when σ is known get multiplier z from normal tables. Compute

$$\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$$

WARNING: not valid if n small and population distribution not normal.

Cases 3 or 4: when σ is unknown get multiplier t from t tables. Compute

$$\bar{x} \pm t\frac{s}{\sqrt{n}}$$

Note: if n is quite large can get away with finding multiplier in normal tables. But no real point in thinking about whether n is large enough. Just use t; that is what software does.

Hypothesis Testing

General framework:

Considering making a yes/no decision about a parameter.

Parameters considered so far:

Population mean μ or proportion p.

What kind of decision?

Is $p = 1/2$? Is $\mu \geq 1.5$? is $\mu = 0$?

Standard Jargon:

We have two hypotheses to choose between.

Examples: $p = 1/2$ or $p \neq 1/2$? $\mu \geq 1.5$ or $\mu < 1.5$.

Neyman Pearson method of doing hypothesis testing.

Pick one of two choices as **null hypothesis**.

Other is **alternative hypothesis**.

Assess evidence against null hypothesis:

Work out value of **test statistic** (t or z so far) which measures discrepancy between data and null hypothesis.

Assume null true (just barely) and compute:

Chance of obtaining a statistic as extreme as the one you did get if you did experiment again and null were true.

Last quantity is called P-value.

To make firm decision if needed:

Set acceptable error rate, α: **type I error rate** or **level** of the test.

"Reject null at level α if $P < \alpha$."

Standard values of α: 0.05, 0.01, 0.001.

If $P < 0.05$ say "results are significant"

If $P < 0.01$ say "results are highly significant"

If $P < 0.001$ say "results are very highly significant"

If no firm decision needed: just report P and describe result in words.

If P is really small say **strong** or **very strong** evidence against null.

(Evidence against null assumed to be evidence for alternative: not always right.)

Steps in doing a hypothesis testing problem:

1) Identify what parameter is being investigated. So far: population mean μ or population proportion p.

2) Formulate null hypothesis: one of

$$\mu \leq \mu_0, \quad \mu = \mu_0, \quad \mu \geq \mu_0$$

(or with p, p_0 for proportions).

3) Formulate alternative hypothesis:

a) if null is $\mu \leq \mu_0$ then alternative is $\mu > \mu_0$.

b) if null is $\mu \geq \mu_0$ then alternative is $\mu < \mu_0$.

c) if null is $\mu = \mu_0$ then alternative is one of $\mu > \mu_0$, $\mu < \mu_0$ or $\mu \neq \mu_0$.

Which one? Depends on scientific problem of interest!

4) Select test statistic: one of

$$z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \qquad z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \qquad t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

5) Use alternative to decide whether to look up **one tailed** or **two tailed** P value.

If alternative is $\neq$ it is "two tailed".

6) Get P from z tables or t tables.

7) Interpret P value.

Review of earlier examples:

1) could true weight be as low as 9.999580?

Parameter of interest is μ.

So null is either $\mu = 9.999580$ or $\mu \leq 9.999580$.

Alternative is $\mu > 9.999580$.

Notice $\mu_0 = 9.999580$.

Know σ so use

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Actual value was $z = 2.24$ (slide 164).

Alternative is one tailed and predicts big value of z. Look up P value as area to right of 2.24.

Found $P = 0.0125$. "The weight is significantly heavier than 5.999580.

NOTICE: not usual English meaning of "significant".

2) Is average dioxin/furan content above 1.5 ppt?

Parameter of interest is μ, $\mu_0 = 1.5$.

Trying to choose between < 1.5 and ≥ 1.5.

Could make either $\mu \leq 1.5$ or $\mu \geq 1.5$ be null.

Make null the one you want to prove is wrong.

I choose $\mu \geq 1.5$, alternative $\mu < 1.5$.

Compute test statistic
$$z = \frac{\bar{x} - 1.5}{s/\sqrt{n}} = -3$$
Alternative is one tailed; predicts large negative z values.

So look up P value as chance of large negative values: left tail.

Get $P = 0.0013$.

Highly statistically significant evidence that d/f concentration in this day's catch below FTAL (1.5 ppt).

3) Toss thumbtack: is chance of U equal to 1/2?

Parameter of interest is population proportion of U p.

(I.e. parameter of interest is $p =$ chance of U.)

Null hypothesis is $p = 1/2$ so $p_0 = 1/2$.

Alternative is naturally $p \neq 1/2$; no theory to suggest direction of departure.

Test statistic is
$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = -0.8$$
Since alternative is two tailed look up area in two tails!.

Area to left of -0.8 plus area to right of 0.8.

Get $P = 0.4237$.

Summary: no significant evidence that p is not 1/2.

4) Pairs of plants:

Parameter of interest is mean μ of population of differences.

Null hypothesis is $\mu = 0$ so μ_0 is just 0.

Alternative specified by science. Unless theory makes prediction: do 2 tailed test.

Alternative $\mu \neq 0$.

Test statistic: σ not known so
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = 2.148$$

Look up two sided area. As before P near but just below 0.05.

"Barely significant" evidence of a difference in average height between crossed and self fertilized plants.

Further topics in one sample tests and confidence intervals.

1) Interpretation of CIs: work out 2 95% conf intervals based on two independent experiments. Chance both intervals include target?

Solution: probability that two independent events both happen via multiplication rule:

Chance is 0.95*0.95=0.9025=90.25%

2) Sample size determination. To get margin of error equal to a target: set

$$\text{margin} = z\sigma/\sqrt{n}$$

Solve for n to get

$$n = \frac{z^2\sigma^2}{\text{margin}^2}$$

Problems in using this:

a) must know σ or design using tolerable guess.

b) inversely proportional to square of tolerable error; often comes out too big for experimenter to afford!

3) Caveats: (ways for tests not to make sense).

a) The method computes a chance based on the assumption that you have a sample from a population and the hypothesis is about that population.

You can't make a meaningful test if the data are just a convenience sample. (Such as the students in this class, e.g.)

b) Small P-value doesn't necessarily mean important difference. If n is really large then even a tiny difference will turn out to be "significant".

4) Jargon: **Type I**, **Type II** errors. **Power**.

When carrying out test there are two ways to be wrong:

a) the null hypothesis could be true but by bad luck you get a small P-value and reject H_o. Called a **Type I error**.

Fixing α places an upper limit on the Type I error rate.

b) the alternative hypothesis is true but by bad luck you get a larger P-value and do not reject H_o. Called a **Type II error**.

Theoretical statisticians try to pick procedures which make chance of Type II error small while fixing Type I error rate.

Power is name for 1 - chance of type II error.

Power depends on unknown parameter.

Power calculations used to set sample sizes.