

## Outline

### Chapter 5: Inference for a single population

Introductory Statistics for Engineering Experimentation

Peter R. Nelson, Marie Coffin and Karen A.F. Copeland

Slides by Douglas Bates

5.1 Central Limit Theorem

5.2 A confidence interval for  $\mu$

5.3 Prediction and tolerance intervals

5.4 Hypothesis tests

5.5 Inference for Binomial Populations

### The Central Limit Theorem

- ▶ The central limit theorem is one of the most important results in mathematical statistics. It says that the sample means from a *random sample* (meaning independent samples from a stable process) will be normally distributed, regardless of what the original distribution was, when  $n$  is sufficiently large.
- ▶ Formally, if  $\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_n$  is a random sample from a distribution with  $\sigma^2 < \infty$  then for large samples,  $\bar{\mathcal{Y}}$  is approximately normally distributed.
- ▶ This is a remarkably powerful result; first, because it is very general and secondly because it is a description of the asymptotic or “limiting” distribution but it holds for quite small values of  $n$ .

### Other properties of the distribution of the sample mean

- ▶ If the random variables  $\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_n$  are a random sample (sometimes also described as a “independent and identically distributed” or i.i.d. sample) from a distribution with mean  $\mu$  and variance  $\sigma^2$  then  $E(\bar{\mathcal{Y}}) = \mu$  and  $\text{Var}(\bar{\mathcal{Y}}) = \sigma^2/n$ .
- ▶ So the central limit theorem states that, for large  $n$ ,

$$\bar{\mathcal{Y}} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

- ▶ Exactly how large  $n$  must be depends on the form of the original distribution. If it is continuous and reasonably symmetric then  $n = 15$  may be large enough. If it is skewed but continuous we may need  $n = 30$  or more. For discrete and skewed we may need as much as  $n = 100$ .
- ▶ Although in practice we only have one sample and one average,  $\bar{y}$  we can use computer simulation to consider the sorts of samples we could have gotten and the distribution of the statistic.

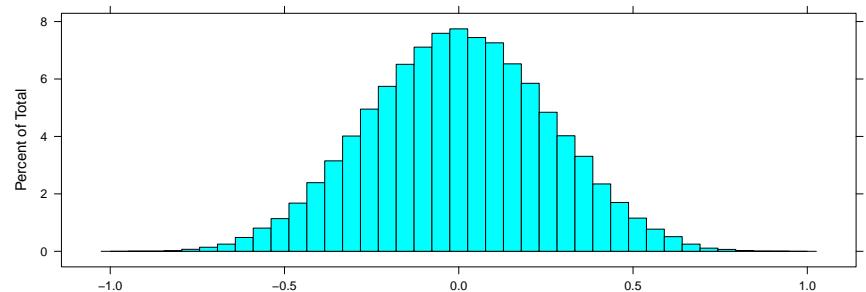
## Conducting a simulation study (not part of the course)

- ▶ Suppose we wish to simulate the value of a statistic (e.g. mean or median or variance or standard deviation) from samples of size  $n$  drawn from a certain distribution. Let  $K$  be the number of replicates we want to obtain.
- ▶ The **sample size**,  $n$ , is typically small. The number of replicates,  $K$ , can be very large. The larger the value of  $K$ , the more accurately we can determine the distribution of the statistic. With modern computers we can afford to use values of  $K$  in the hundreds of thousands or more.
- ▶ First determine how to evaluate the statistic from a single sample of size  $n$  then use the **replicate** function to repeat the process  $K$  times.

## Mean of samples of size 5 from $U(-1,1)$

What is the shape of the distribution of the mean of a sample of size  $n = 5$  from a  $U(-1, 1)$  distribution?

```
> mns5 <- replicate(50000, mean(runif(5, min = -1, max = 1)))
> histogram(~mns5, breaks = seq(-1, 1, len = 40))
```



## Sampling densities of statistics

- ▶ The idiom

```
replicate(K, <statfn>(r<distab>(n, <pars>)))
```

produces  $K$  replicates of the statistic calculated by **<statfn>** (examples are **mean**, **median**, **var** and **sd**) on samples of size **n** from distribution **<distab>** with parameter(s) **<pars>**.

- ▶ Typically  $K$  is large and  $n$  is small. Values of 10,000 or 100,000 are used for  $K$  on modern computers. The larger the value of  $K$  the smoother the approximation to the sampling density.  $n$  is the size of the actual sample you can afford to collect.

## Effect of changing the sample size, $n$

- ▶ Performing multiple simulations allows us to see how characteristics of the distribution of  $\bar{Y}$  depends on  $n$ .

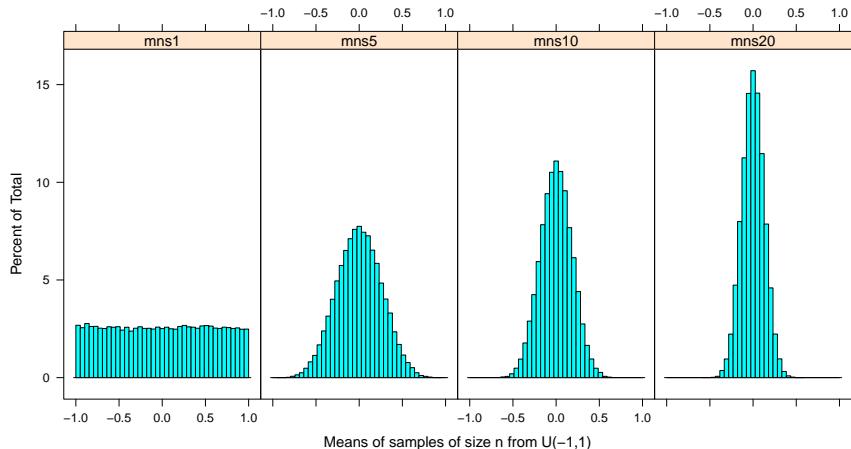
```
> mns1 <- runif(50000, -1, 1)
> mns10 <- replicate(50000, mean(runif(10, -1, 1)))
> mns20 <- replicate(50000, mean(runif(20, -1, 1)))
> sapply(list(mns1, mns5, mns10, mns20), mean)
[1] -0.0032447496  0.0003987894  0.0002170505 -0.0002594118
```

```
> sapply(list(mns1, mns5, mns10, mns20), var)
[1] 0.33458433 0.06670560 0.03375242 0.01666237
```

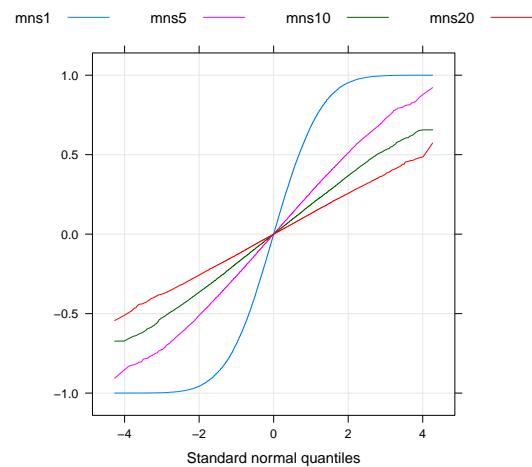
- ▶ As  $n$  increases the expected value of the sample mean stays near 0.
- ▶ As  $n$  increases the variance of the sample mean decreases. Roughly,  $V(\bar{X}_n) = \frac{1}{3} \cdot \frac{1}{n}$

## Shape of distribution of $\bar{X}_n$

- As  $n$  increases, the shape of the distribution of  $\bar{X}_n$  tends to the “bell-curve” or Gaussian shape and it has less variability. That is, it tends to a “central limit”.



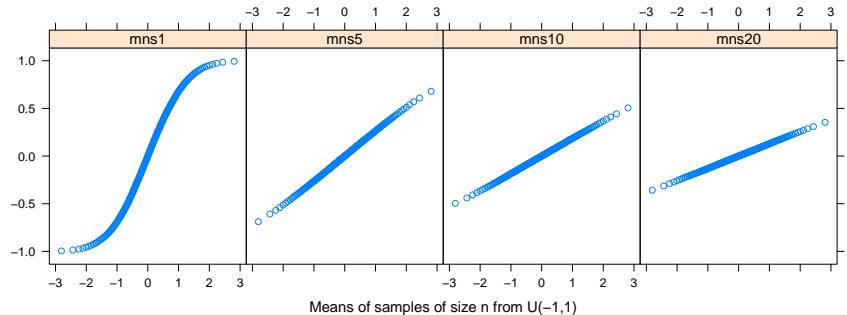
## Overlaid normal probability plots for $\bar{Y}_n$



The conclusion is that the distribution of means from an i.i.d. sample of a uniform distribution is very close to a normal, even for  $n = 5$ .

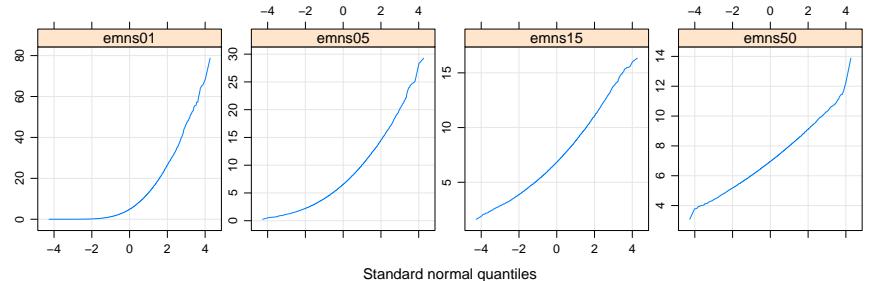
## More detail on the shape of the distribution of $\bar{Y}$

- In addition to the histogram we can use normal probability plots to evaluate the deviations of the distribution of  $\bar{Y}$  from normality.



## Sample means from an exponential distribution

```
> emns01 <- replicate(50000, mean(rexp(1, rate = 1/7)))
> emns05 <- replicate(50000, mean(rexp(5, rate = 1/7)))
> emns15 <- replicate(50000, mean(rexp(15, rate = 1/7)))
> emns50 <- replicate(50000, mean(rexp(50, rate = 1/7)))
```



Even for  $n = 50$  there is noticeable skewness in the distribution (although we would not be far wrong in assuming normality at  $n = 50$ ).

## Elementary uses of the C.L.T.

- If we have plausible values of the variance of our process, perhaps from a pilot study, we can use the normal distribution and the Central Limit Theorem (C.L.T.) to evaluate probabilities regarding the sample mean.
- Example 5.1.3 discusses product lifetimes that have an unknown mean and a variance of approximately 8 years. The number of products to sample so that we are 95% certain that  $\bar{y}$  will be within 1 year of the true mean is derived from

$$0.95 = P(|\bar{Y} - \mu| < 1)$$

The distribution of  $\bar{Y}$  will be approximately normal with mean  $\mu$  and standard deviation  $\sigma/\sqrt{n}$ . For a standard normal, 95% of the probability is within “2” standard deviations of the mean (the actual multiple is `qnorm(0.025) = -1.95996`) so we want  $1 = \text{qnorm}(0.025)^2 \frac{8}{n}$ . That is,  $n >$

```
> 8 * qnorm(0.025)^2
```

```
[1] 30.73167
```

## Confidence intervals

- Our “best guess” at a parameter is called a *point estimate*. For example, we usually use the sample mean,  $\bar{y}$ , as the point estimate of  $\mu$ .
- An *interval estimate* or *confidence interval* is an interval of plausible values for the parameter. Values outside the interval are “unreasonable” and values inside are “not unreasonable”.
- To calibrate the meaning of “unreasonable” we assign a value  $\alpha$  to the probability of getting data like we did or even more extreme when the parameter is outside. This corresponds to the “p-value” in a hypothesis test.
- The *coverage probability* or *confidence level* is  $1 - \alpha$ . Typically we set  $\alpha = 0.05$  or  $\alpha = 0.01$  resulting in 95% or 99% confidence intervals.
- Formally, the coverage probability is the probability that an interval constructed in this way will cover the true parameter value.

## Approximations for binomial or Poisson distributions

- The text describes approximations of the probabilities for a binomial or Poisson distribution based on the normal distribution.
- These are interesting from the point of view of understanding that these distributions will tend to have a “bell-curve” shape when  $n$  is large and  $p$  is moderate for the binomial or  $\lambda t$  is large for the Poisson.
- In practice, though, you can evaluate probabilities for such distributions exactly so there is no need to use approximations.

## A confidence interval on $\mu$

- In the unlikely event that someone were to tell us what the standard deviation,  $\sigma$ , of the population was but somehow not know much about the mean,  $\mu$ , we could create a  $(1 - \alpha)$  confidence interval as

$$\bar{y} \pm z(\alpha/2) \frac{\sigma}{\sqrt{n}}$$

where  $z(\alpha/2)$  is the **upper**  $\alpha/2$  quantile of the standard normal distribution.

- For example, the upper 0.025 quantile of the standard normal is

```
> qnorm(0.025, low = FALSE)
```

so a 95% confidence interval on  $\mu$  for this artificial, “known sigma” case is

$$\bar{y} \pm 1.959964 \frac{\sigma}{\sqrt{n}}$$

## Use of Student's T distribution

- ▶ In the real world no one tells us what  $\sigma$  is and we must estimate it as  $s$ . A statistician named William Gossett, who wrote under the pseudonym "A Student", derived the distribution of the shifted, scaled sample mean when the scale is based on the estimate,  $s$ , not the theoretical value  $\sigma$ .
- ▶ This distribution is called the "Student's t distribution". It is similar to the standard normal distribution but a bit more spread out. The spreading depends on the number of "degrees of freedom" in the estimate of  $\sigma^2$ . The degrees of freedom are written as  $\nu$ . For a single sample  $\nu = n - 1$ .
- ▶ As  $\nu$  increases the T distribution approaches the standard normal. If we were using tables we would call anything with  $\nu > 30$  a standard normal. When using a computer we don't bother.
- ▶ Notation: the t distribution with  $\nu$  degrees of freedom is written  $t(\nu)$ . The corresponding R functions are `dt`, `pt`, `qt` and `rt`. The upper  $\alpha$  quantile is written  $t(\alpha; \nu)$ .

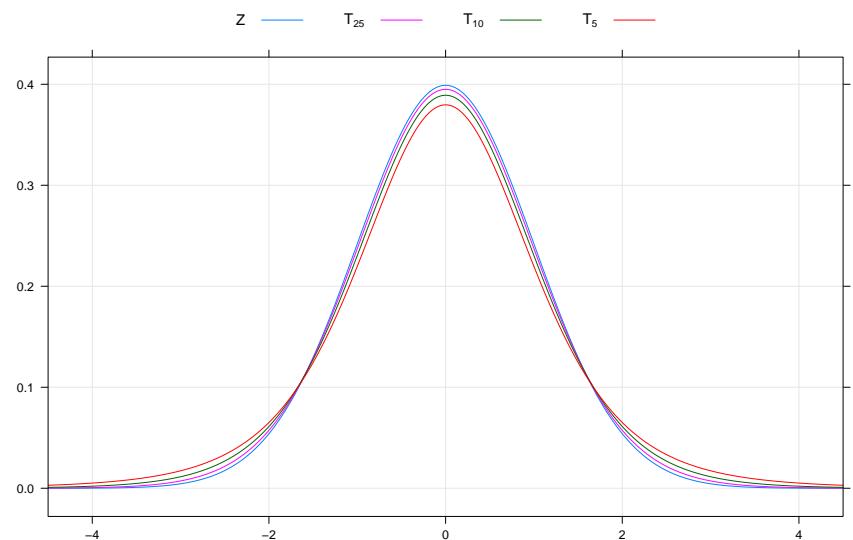
## General form of the confidence interval

- ▶ The general form of the confidence interval on  $\mu$  is

$$\bar{y} \pm t\left(\frac{\alpha}{2}, n - 1\right) \frac{s}{\sqrt{n}}$$

- ▶ We can use this formula for any values of  $n$ . If  $n$  is large we don't need strong assumptions on the shape of the original distribution. If  $n$  is small we must assume that the original distribution is close to the normal (but, of course, we can't check this with a small sample - a "Catch 22" situation).
- ▶ The R function to create this interval is `t.test`. The name comes from the corresponding hypothesis test, which we will discuss later.

## Graphical comparison of $t(\nu)$ and $Z$



## Example 5.2.2

The example provides (probably fictitious) discharge times for a particular electric vehicle

```
> sd(charge <- c(5.11,2.1,4.27,5.04,4.47,3.73,5.96,6.21))
```

```
[1] 1.3108
```

```
> summary(charge)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.10	4.14	4.76	4.61	5.32	6.21

```
> t.test(charge)
```

```
One Sample t-test
```

```
data: charge
```

```
t = 9.9502, df = 7, p-value = 2.211e-05
```

```
alternative hypothesis: true mean is not equal to 0
```

```
95 percent confidence interval:
```

```
3.5154 5.7071
```

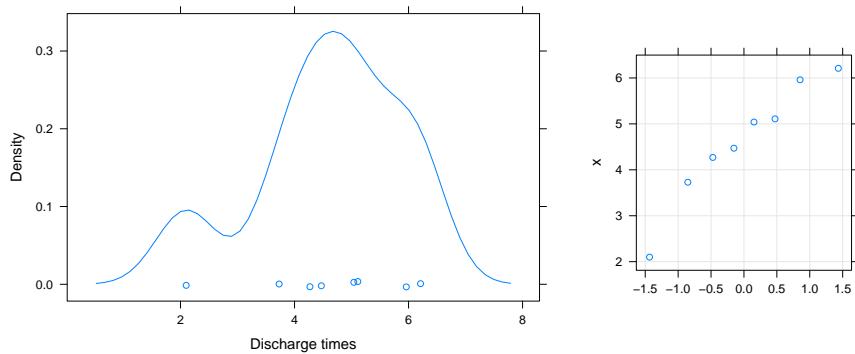
```
sample estimates:
```

```
mean of x
```

```
4.6113
```

## Example 5.2.2 (cont'd)

Because the degrees of freedom,  $\nu = 7$ , are quite small we should check for normality.



## Another R evaluation of confidence intervals

- ▶ Another way of evaluating a confidence intervals on  $\mu$  is with the `confint` function, which provides confidence intervals on the parameters in a fitted model.
- ▶ To use `confint` we fit what we sometimes call the “trivial” model

$$\mathcal{Y}_i = \mu + \epsilon_i, \quad i = 1, \dots, n$$

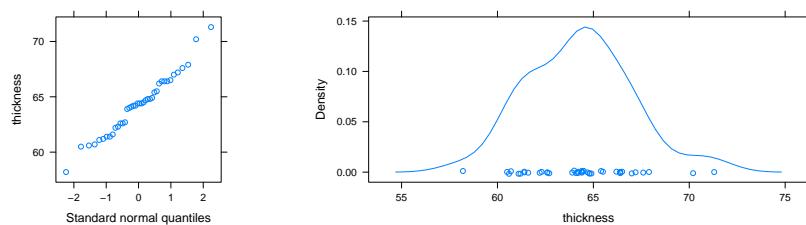
The estimate of  $\mu$ ,  $\hat{\mu} = \bar{y}$  will be called (`Intercept`) in the output. The formula for the model contains the constant term, `1`, as the only predictor.

```
> confint(fm1 <- lm(charge ~ 1))
```

2.5 % 97.5 %

(Intercept) 3.5154 5.7071

## Clear-coat thickness (example 5.2.4)



```
> with(ccthickn, summary(thickness))
Min. 1st Qu. Median Mean 3rd Qu. Max.
58.2   62.3   64.4   64.3   66.2   71.3
> sd(ccthickn$thickness)
[1] 2.7176
> confint(fm2 <- lm(thickness ~ 1, ccthickn))
2.5 % 97.5 %
(Intercept) 63.391 65.129
```

## Clear-coat thickness (cont'd)

The summary of the fit of the “trivial” model includes many of the statistics from the data.

```
> summary(fm2)
Estimate Std. Error t value Pr(>|t|)
(Intercept) 64.2600    0.4297 149.5 <2e-16
Residual standard error: 2.718 on 39 degrees of freedom
```

The values in this summary include

$\bar{y} = 64.26$  The parameter estimate,  $\hat{\mu}$ .

$s = 2.718$  The sample standard deviation,  $\hat{\sigma}$

$n - 1 = 39$  The degrees of freedom,  $\nu$ , for the variance estimate,  $s^2$ .

$\frac{s}{\sqrt{n}} = 0.4297$  The standard error of the mean,  $\sqrt{\text{Var}(\mathcal{Y})}$

## Sample sizes

- The half-width of a confidence interval, also called the *margin of error* depends on
  - The **confidence level** Higher confidence levels require wider intervals
  - The **standard deviation** More variability in the original distribution results in wider intervals.
  - The **sample size** Larger samples produce narrower intervals.
- Given a working value for  $\sigma$  we can determine the sample size needed to attain a given margin of error.
- If we are willing to assume that  $n$  is large we can use  $z(\alpha/2)$  in the calculation. For small  $n$  it gets tricky because  $\nu = n - 1$  determines the multiplier which, in turn, affects the sample size. We must solve a nonlinear equation but computers are good at that.

## Sample size calculations

- Example 5.2.5 shows calculations for the sample size from the formula  $n = \left[ \frac{t(\alpha/2; \infty)s}{d} \right]^2$  when the desired margin of error,  $d$ , is 0.2, the working value of  $s$  is 0.4 and  $\alpha$  is 5% and we round the answer to the next largest integer.  
`> ceiling(qnorm(0.025)*0.4/0.2)^2`

[1] 16

- Because this is a small value of  $n$  we should instead solve for  $n$  in  $n = \left[ \frac{t(\alpha/2; n-1)s}{d} \right]^2$

```
> ceiling(uniroot(function(x) x-(qt(.025,x-1)*0.4/0.2)^2,  
+ c(2,100))$root)
```

[1] 18

## Section 5.3: Prediction and tolerance intervals

- A confidence interval on  $\mu$  provides a measure of the precision of the information regarding the unknown population parameter. It does not directly tell us about bounds on where we expect a future observation to fall.
- A *prediction interval* indicates where a single future observation is likely to be.
- A *tolerance interval* indicates where a large proportion of the population is likely to be.
- Unlike the confidence interval on  $\mu$ , prediction intervals and tolerance intervals depend strongly on the shape of the distribution of the data.
- In theory one can make a confidence interval arbitrarily narrow by taking a sufficiently large sample. You can't do this for a prediction interval.

## Prediction intervals on a future observation

- If it is reasonable to assume that the *data* (i.e.  $\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_n$ ) are from normal distribution then we could say that a model for the data is

$$\mathcal{Y}_i = \mu + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- Our estimate  $\hat{\mu} = \bar{\mathcal{Y}}_n$  is independent of  $\epsilon_{n+1}$ . The variability in the difference between  $\mathcal{Y}_{n+1}$  and  $\bar{\mathcal{Y}}_n$  is the sum of the variability in  $\bar{\mathcal{Y}}_n - \mu$  ( $\frac{\sigma^2}{n}$ ) and the variability in  $\epsilon_{n+1}$  ( $\sigma^2$ ).
- Because we estimate  $\sigma^2$  the  $(1 - \alpha)$  prediction interval becomes

$$\bar{y} \pm t(\alpha/2; n-1)s\sqrt{1 + \frac{1}{n}}$$

## Evaluating a prediction interval

- ▶ The prediction interval could be evaluated according to the formula. For the clear-coat thickness data the 95% prediction interval on a future thickness measurement is

```
> with(ccthickn, mean(thickness) + c(-1,1) * qt(0.975, 39) *  
+      sd(thickness) * sqrt(1 + 1/40))  
  
[1] 58.69478 69.82522
```

- ▶ An alternative is to use the `predict` function applied to the trivial model and with the optional argument `interval = "pred"`. This produces a matrix with n rows that are identical so I just look at the first row.

```
> predict(fm2, int = "pred")[1,]  
  
   fit      lwr      upr  
64.26000 58.69478 69.82522
```

## Tolerance intervals

- ▶ A tolerance interval is more difficult to describe and to calculate than is a prediction interval.
- ▶ Methods for tolerance intervals are given in the text but we will not cover this topic in this course.

## Section 5.4 Hypothesis tests

- ▶ A hypothesis test is a procedure for deciding if a particular value of a parameter is reasonable, given the observed data.
- ▶ We have a probability model (e.g. our sample is a random sample from a normal distribution with mean  $\mu$  and variance  $\sigma^2$ ), the observed data,  $y_1, y_2, \dots, y_n$ , and a particular value of the parameter in mind (e.g. the mean clear-coat thickness,  $\mu$ , should be 65 microns).
- ▶ We consider two competing claims called the *null hypothesis*, written  $H_0$ , and the *alternative hypothesis*, written  $H_a$ .
- ▶  $H_0$  is the “no change” assumption. For our example, it is  $\mu = 65$ .  $H_a$  is the result we are trying to establish. It is also the result indicated by the data. In our example  $\bar{y} = 64.26$  microns. If we are interested only in whether we are “off target” then  $H_a : \mu \neq 65$ . If we are interested in whether the clear coats are systematically too thin then  $H_a : \mu < 65$ .
- ▶ These are called “two-tailed” and “one-tailed” alternatives, respectively.

## The p-value for a hypothesis test

- ▶ We would like to establish  $H_a$  directly but, because of the variability in the data, we can't.
- ▶ Instead we try to “rule out”  $H_0$ . Again, because of the variability we can't rule it out completely. What we do is to calculate “the probability of seeing the data that we did, or something even more unusual, assuming that  $H_0$  is true”. This is called the *p-value* for the test.
- ▶ Because the p-value is a probability it must be between 0 and 1. If the p-value is small (i.e. close to 0) we reject  $H_0$  in favor of  $H_a$ . If the p-value is not small we fail to reject  $H_0$ .
- ▶ Note that we never confirm  $H_0$ . We either reject it (i.e. rule it out) or fail to reject it (i.e. are unable to rule it out). The latter conclusion represents “no decision”.
- ▶ The p-value is evaluated assuming  $H_0$  is true. The form of the evaluation depends on  $H_a$ .

## Performing a hypothesis test

- ▶ To set up a hypothesis test you must first decide on  $H_0$  and  $H_a$ . See pages 190–191 in the text for examples. Determine if you want a one-sided or two-sided alternative. If two-sided you are done. If one-sided then the only claim that makes sense as  $H_a$  is the one indicated by the data.
- ▶ The text describes methods based on rejection regions and critical values of test statistics. These are used when you can't calculate probabilities for distributions like the T distribution. We can do that so we use the more direct approach.

## Examples 5.4.5, 5.4.7 and 5.4.10

- ▶ The situation described in examples 5.4.5 and 5.4.7 involves drums of material that should have a mass of 230 kg. A random sample of size 25 yields  $\bar{y} = 229$  and  $s = 4$ .
- ▶ The sample mean is less than the nominal value of 230 kg. We wish to determine if it is “significantly less” than 230 kg. Our test is of the form

$$H_0 : \mu = 230 \quad \text{vs.} \quad H_a : \mu < 230$$

- ▶ The observed t statistic, assuming that  $H_0$  is true, is

$$t_{\text{obs}} = \frac{229 - 230}{4/\sqrt{25}} = -1.25$$

- ▶ The probability of seeing values like this, or even more extreme, when  $H_0$  is true, is  $P(T_{24} \leq -1.25)$  which we evaluate as `pt(-1.25, df = 24) = 0.1117`. This value is not sufficiently small for us to “rule out” or reject  $H_0$ .

## Use of the R function `t.test`

- ▶ The `t.test` function in *R* is used for one- or two-sample t tests on the population mean,  $\mu$ .
- ▶ In the one-sample form you must specify the variable name as the first argument. Use `with` or the `$` operator to access a variable in a data frame.
- ▶ You should specify the nominal value  $\mu_0$  of the population mean as the `mu` argument.
- ▶ The default alternative is “`two.sided`”. Specify `alt = "greater"` or `alt = "less"` for one-sided alternatives.

## Clear-coat thickness data

The `ccthickn` data are from a process with a target thickness of 65 microns but the sample mean,  $\bar{y} = 64.24$  microns. Is this “significant evidence” the  $\mu < 65$ ?

```
> with(ccthickn, t.test(thickness, mu = 65, alt = "less"))

One Sample t-test
data: thickness
t = -1.7221, df = 39, p-value = 0.04648
alternative hypothesis: true mean is less than 65
95 percent confidence interval:
-Inf 64.98398
sample estimates:
mean of x
64.26
```

## Sample sizes

- ▶ Our conclusion in a hypothesis test is either to “rule out” or reject  $H_0$  in favor of  $H_a$  or to fail to rule  $H_0$  out.
- ▶ We don’t know if  $H_0$  is true or not. If it is true and we reject it, we have made one type of “error” or mis-classification. If it is false and we fail to reject it we have made another type of “error”. These are called Type I and Type II, respectively.
- ▶ For a test on a particular set of data the probability of a Type I error in rejecting  $H_0$  is the p-value. For planning purposes we control for this type of error by saying it should not exceed some value,  $\alpha$ . Typically  $\alpha = 0.05$  or  $\alpha = 0.01$ .
- ▶ To evaluate the probability of a Type II error we must specify a specific alternative, such as  $\mu_1$ , instead of a general alternative like  $\mu \neq \mu_0$ . We also need a working value of  $\sigma$ .
- ▶ The *power* of a test is a function of  $\mu_1$ . It is the probability of rejecting  $H_0 : \mu = \mu_0$  when  $\mu = \mu_1$ . Sometimes it is written as  $1 - \beta$ . In the text it is  $\gamma$ .

## Example 5.4.13

- ▶ This example involves a test of  $H_0 : \mu = 5$  versus  $H_a : \mu > 5$  (i.e. a one-sample test with a one-sided alternative) with  $\sigma = 1$ ,  $\alpha = 0.05$ ,  $\mu_1 = 6$  and a power of 80% or 0.8.

```
> power.t.test(delta = 1, sd = 1, sig = 0.05, power = 0.8,
+                 type = "one", alt = "one")
```

One-sample t test power calculation  
n = 7.727622  
delta = 1  
sd = 1  
sig.level = 0.05  
power = 0.8  
alternative = one.sided
- ▶ We round fractional sample sizes up so we would use a sample of size  $n = 8$ . Note that the sample size calculated in this example in the text, assuming a “known” value of  $\sigma$  and not an estimated value, is  $n = 7$ .
- ▶ The values `sig = 0.05` and `sd = 1` are the defaults and could be omitted.

## Use of power.t.test in R

- ▶ The function `power.t.test` can be used to evaluate the power or the sample size for one- and two-sample t-tests.
- ▶ Two out of the three arguments `n`, `delta` ( $= \mu_1 - \mu_0$ ) and `power` must be specified and the third is calculated.
- ▶ By default `delta` is in standard deviation units (i.e.  $\delta = \frac{\mu_1 - \mu_0}{\sigma}$ ). If you want to specify `delta` in the units of the response you should also give a value for the optional argument `sd` (= working value of  $\sigma$ ).
- ▶ The default test type is "`two.sample`" (the next chapter). For this chapter specify `type = "one.sample"`.
- ▶ The default alternative is "`two.sided`". For one-sided alternatives specify `alt = "one"`.

## Example 5.4.14

- ▶ This example is like the previous one except that the magnitude of  $\delta = \mu_1 - \mu_0$  is now 1.5 mg and the desired power is 90%.

```
> power.t.test(delta = 1.5, power = 0.9, type = "one", alt = "one")
```

One-sample t test power calculation  
n = 5.471726  
delta = 1.5  
sd = 1  
sig.level = 0.05  
power = 0.9  
alternative = one.sided
- ▶ The larger tolerance decreases  $n$  and the higher power increases  $n$ . The net effect is to decrease  $n$ .

## Example 5.4.16

- As explained in the text, in practice we should use the T distribution and not the standard normal distribution for calculating sample sizes. This is what `power.t.test` does. The text provides an alternative based on tables but they are very coarse.
- The important quantity is written in the text as  $d = \frac{\mu - \mu_0}{\sigma}$ . In the *R* function it is `delta`
- For this example  $d = 0.4$  with 90% power on a one-tailed, one-sample test.

```
> power.t.test(del = 0.4, pow = 0.9, typ = "one", alt = "one")
```

```
One-sample t test power calculation
  n = 54.90553
  delta = 0.4
  sd = 1
  sig.level = 0.05
  power = 0.9
  alternative = one.sided
```

## Example 5.5.4

- We observe  $y = 18$  defectives in a sample of  $n = 1000$  from a process where the nominal defect rate is 1%. Test  $H_0 : p = 0.01$  versus  $H_a : p > 0.01$

```
> binom.test(18, 1000, p = 0.01, alt = "greater")

Exact binomial test
data: 18 and 1000
number of successes = 18, number of trials = 1000,
p-value = 0.01383
alternative hypothesis: true probability of success is greater than 0.0
95 percent confidence interval:
 0.01166572 1.00000000
sample estimates:
probability of success
          0.018

> sum(dbinom(18:1000, 1000, 0.01)) # explicit P(Y >= 18) for H0
[1] 0.01383258
```

## Section 5.5, Inference for Binomial Populations

- Results from a binomial simply consist of the number of “successes”,  $y$ , and the number of trials,  $n$ . This is already an i.i.d. sample,  $\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_n$ , if we consider each  $\mathcal{Y}_i$  as giving a binary (0/1) response.
- The parameter estimate is  $\hat{p} = \frac{y}{n}$ .
- Often approximate confidence intervals and hypothesis tests are formulated based on a normal distribution for  $\hat{P}$  using a mean of  $p$  and a variance of  $\frac{p(1-p)}{n}$ .
- However, we can do better. In particular, the p-value for a hypothesis test can be calculated exactly. We use `binom.test` for a one-sample binomial test and `prop.test` for comparing results from multiple samples. Sample sizes are calculated with `power.prop.test`.

## Comparison to results in the text

- In the text the p-values for tests on a binomial distribution are calculated through a normal approximation and not the binomial distribution.
- This results in p-values that are lower than they should be, which is dangerous. E.g. the text's p-value for Example 5.5.5 is 0.0055 versus the exact calculation of 0.01383. That could be important.
- Example 5.5.6 in the text gives a p-value of 0.0089 for a test of  $H_0 : p = 0.05$  versus  $H_a : p > 0.05$  for  $y = 8$  and  $n = 72$ . The more accurate value is

```
> binom.test(8, 72, p = 0.05, alt = "greater")$p.value
[1] 0.02715970

> sum(dbinom(8:72, size = 72, prob = 0.05))
[1] 0.02715970
```

## Confidence intervals on p

- ▶ The `binom.test` function produces a confidence interval based on “inverting” the exact hypothesis test. Usually we want to two-sided alternative when forming a confidence interval.
- ▶ The `conf.level` argument can be used to set the desired confidence level (default is 95%).
- ▶ These intervals are more reliable than those given in the book.
- ▶ Example 5.5.1 has  $y = 27$  out of  $n = 200$ . The 95% confidence interval should be

```
> binom.test(27, 200)$conf.int  
[1] 0.0908871 0.1903092  
attr(,"conf.level")  
[1] 0.95
```

Notice that it is wider on the side towards  $p = 0.5$  than on the other side. This is a characteristic of the binomial.

## Sample sizes for tests on the binomial

- ▶ Formulas (5.5.4), (5.5.7) and (5.5.8) in the text allow for calculating sample sizes for a confidence intervals with a given margin of error,  $d$ , or for tests with a null hypothesis of the form  $H_0 : p = p_0$ .
- ▶ All of these formulas are based on the normal approximation and will give values that are a bit too small.
- ▶ In the two-sample case (next chapter) we can use the function `power.prop.test` to determine the sample size for the two samples.