# Topic 2. Distributions, hypothesis testing, and sample size determination

## 2. 1. The Student - t distribution (ST&D p.56 and p.77)

Consider a repeated drawing of samples of size r = 5 from a normal distribution. For each sample compute $\bar{Y}$, s, $s_{\bar{Y}}$, and another statistic, *t*, where $t = (\bar{Y} - \mu)/s_{\bar{Y}}$. The *t* statistics is the deviation of a normal variable $\bar{Y}$ from its hypothesized mean measured in standard error units, or it is the number of standard deviations that separate $\bar{Y}$ and $\mu_\circ$. This statistic is distributed according to the Student's t distribution with r-1 degrees of freedom.

Now imagine organizing the large population of sampled *t* values in a frequency distribution. The frequency curve will be more variable than the normal distribution (Fig. 1) because $s_{\bar{Y}}$ is quite variable from sample to sample. The larger the sample size, the closer *t* approaches a normal distribution. When *t* values are based on samples containing 60 or more variates, they are approximately normally distributed. With infinite degrees of freedom $t = z$.

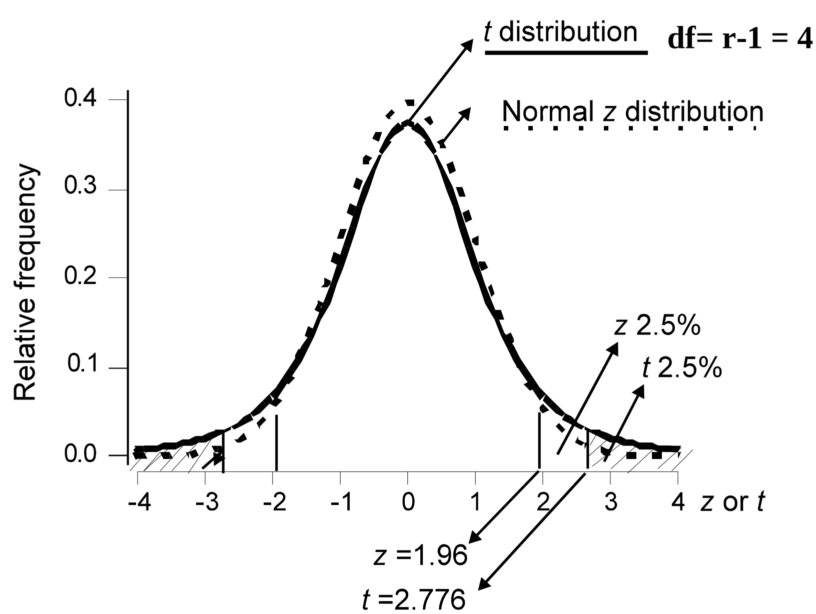


Fig. 1. Distribution of *t* (df=5-1=4) compared to *z*. The curve of *t* is symmetric and somewhat flatter than the *z* distribution, lying under it at the center and above it in the tails. The increase in the *t* value relative to *z* is the price we pay for being uncertain about the population variance

## 2. 2. Confidence limits [S&T p. 77]

Suppose we have a sample $\{Y_1, ..., Y_r\}$ with mean $\bar{Y}$ drawn from a population with unknown mean μ, and we want to estimate μ. If we manipulate the definition for the *t* statistic we get:

$$\bar{Y} = \mu \pm t_{r-1}\, s_{\bar{Y}}$$

Note that μ is a fixed but unknown parameter while $\overline{Y}$ is a known but random statistic. The statistic $\overline{Y}$ is distributed about μ according to the *t* distribution, that is, it satisfies

$$\Pr\{\overline{Y} - t_{\alpha/2,\, r\text{-}1}\ s_{\overline{Y}} \le \mu \le \overline{Y} + t_{\alpha/2,\, r\text{-}1}\ s_{\overline{Y}}\} = 1 - \alpha$$

Note that for a confidence interval of size α, you must use a *t* value corresponding to an upper percentile of α /2 since both the upper and lower percentiles must be included (see Fig. 1). Therefore the confidence interval is

$$\overline{Y} - t_{\alpha/2,\, r\text{-}1}\ s_{\overline{Y}} \le \mu \le \overline{Y} + t_{\alpha/2,\, r\text{-}1}\ s_{\overline{Y}}$$

These two terms represent the lower and upper 1- α **confidence limits** of the mean. The interval between these terms is called **confidence interval.** For example, in Barley Data Set (see SAS example below), $\overline{Y}$ = 75.94 and $s_{\overline{Y}}$ = 1.227 / √14 = 0.3279. A table gives the $t_{0.025,13}$ value of 2.16, which we multiply by 0.3279 to conclude at the 5% level that μ = 75.94 ± 0.72. It is incorrect to say that there is a probability of 95% that the true mean is within 75.94 ± 0.72. If we repeatedly obtained samples of size 14 from the population and constructed these limits for each, we could expect 95% of the intervals to contain the true mean.


Fig. 2. Vertical lines represent 20 95% confidence intervals. One out of 20 intervals does not include the true mean (horizontal line). The confidence level represents the percentage of time the interval covers the true (unknown) parameter value [ST&D p.61]

## 2. 3. Hypothesis testing [ST&D p. 94]

Another use of the *t* distribution, more in the line of experimental design and ANOVA, is in *hypothesis testing*. Results of experiments are usually not clear-cut and therefore need statistical tests to support decisions between alternative hypotheses. Recall that in this case a *null hypothesis* $H_o$ is established and an alternative $H_1$ is tested against this. For example, we can use the Barley data (ST&D p. 30) to test the $H_o$: μ = 78 against the $H_1$: μ ≠ 78.

```
DATA barley;
INPUT extract @@;
TEST78 = extract - 78.0;
CARDS;
77.7 76.0 76.9 74.6 74.7 76.5 74.2 75.4 76.0 76.0 73.9 77.4 76.6 77.3
;
PROC UNIVARIATE;
var test78;
PROC CHART;
Vbar test78;
RUN; QUIT;
```

```
Mean:75.943 SD: 1.227  SE: 0.328
```

The formula for $t$ is as before:

$t = (\bar{Y} - \mu)/\ s_{\bar{Y}}$ With $s_{\bar{Y}} = \sqrt{s^2/n}$

For our example: $t = (75.94 - 78) / 0.3279 = -6.28$.

We decide to reject $H_o$ if the probability that we could have obtained this sample from a population with $H_o$: $\mu = 78$ (satisfying $H_o$) is less than some pre-assigned number such as 0.05. Since $t$ is larger than the critical value $t_{0.025,13}$ $|-6.28| > 2.160$ we conclude that the probability of incorrectly rejecting $H_o$ is less than 0.05. This is equivalent to calculate a 95% confidence interval for the mean ($\bar{Y} \pm t_{0.025,13} * s_{\bar{Y}}$).

In this example the lower and upper 95% confidence limits for $\mu$ under $H_o$ are [75.23 and 76.65]. $H_o$ (78) is not within the confidence interval and we reject $H_o$. A value of $\bar{Y}$ = 75.94 is not expected in a sample of 14 values from a population with $H_o$ $\mu = 78$ and SD: 1.23 unless the particular random sample is an unusual one.

The value 0.05 is called the *significance level* of the test and it is denoted $\alpha$. It represents the probability of incorrectly rejecting $H_o$ when it is actually true, a **Type I** error. The other possible error, a **Type II** error, is to incorrectly accept $H_o$ when it is false [S&T p. 118]. The probability of this event is denoted $\beta$ and is not unique; it depends on the true value of $\mu$. The relationships between hypotheses and decisions can be summarized as follows:

| | | Null hypothesis | |
|---|---|---|---|
| | | **Accepted** | **Rejected** |
| Null hypothesis | **True** | Correct decision | **Type I error** |
| | **False** | **Type II error** | Correct decision |

### 2.3.1. Power of a test for a single sample. $H_0$: $\mu_1=\mu_0$

The magnitude of $\beta$ depends on how far the alternative parametric mean is from the parametric mean of the null hypothesis. As the alternative mean approaches the parametric mean, $\beta$ increases to a maximum value of $1-\alpha$. An important concept in connection with hypothesis testing is the **power** of a test (ST&D p.119).

$$Power = 1-\beta = P(Z > Z_{\alpha/2} - \frac{|\mu_1 - \mu_0|}{\sigma_{\bar{Y}}}) - or - P(t > t_{\alpha/2} - \frac{|\mu_1 - \mu_0|}{s_{\bar{Y}}})$$

The first equation is for a population with **known** $\sigma^2$ and de $2^{nd}$ one for a population with **unknown** $\sigma^2$. The term that is subtracted is the difference between the two means expressed in standard error units.

**The power of a test is the probability of rejecting the null hypothesis when it is false and the alternative hypothesis is correct. Or in other words a measure of the ability of the test to detect $\mu_1$ (e.g. the ability to differentiate 2 treatments that are different)**

Suppose for example that the true mean is the same as our calculated $\overline{Y} = 75.94$. What is the power of a test for $H_0$: $\mu = 74.88$. See Fig 3. Since $\alpha = 0.05$, r = 14 ($t_{0.025,13} = 2.160$), and $s_{\overline{Y}} = 0.32795$. Using the previous formula for **unknown $\sigma^2$**:

$$Power = 1 - \beta = P(t > 2.160 - \frac{|75.94 - 74.88|}{0.32795}) = P(t > -1.072) = 0.85$$

The probability of the Type II error $\beta$ we are looking for is the shaded area to the left of the lower curve.

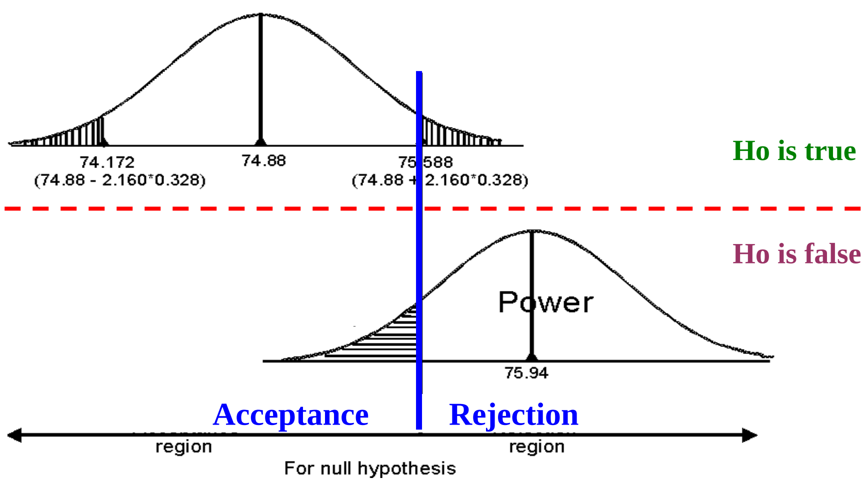


**Fig. 3**. Type I and Type II errors in the Barley data set.

## 2.3.2. Power of the test for the difference between the means of two samples (T-test).

$H_0$: $\mu_1 - \mu_2 = 0$ versus $H_1$: $\mu_1 - \mu_2 \neq 0$ (two tail test) or
$H_1$: $\mu_1 - \mu_2 > 0$ (one tail test) depending of the particular situation

*The power formula used above* **for a single sample with unknown $\sigma^2$** *should be modified for the difference* **between two means**.

The *general* power formula for both **equal** and **unequal** sample sizes reads as:

$$Power = P(t > t_{\frac{\alpha}{2}} - \frac{|\mu_1 - \mu_2|}{s_{\overline{Y}1-\overline{Y}2}}) = P(t > t_{\frac{\alpha}{2}} - \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{s^2_{pooled}}{N}}}),$$

where $s^2_{pooled}$ is a weighted variance given by: $s^2_{pooled} = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$

and $N = \dfrac{n_1 n_2}{n_1 + n_2}$.

Notice in the case where **$n_1 = n_2 = n$** (**equal sample sizes**) that the formulas reduce to:

$$s^2_{pooled} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)} = \frac{(n-1)(s_1^2+s_2^2)}{2(n-1)} = \frac{s_1^2+s_2^2}{2}$$

$$N = \frac{n_1 n_2}{n_1+n_2} = \frac{n^2}{2n} = \frac{n}{2}$$

$$Power = P(t > t_{\frac{\alpha}{2}} - \frac{|\mu_1-\mu_2|}{s_{\bar{Y}1-\bar{Y}2}}) = P(t > t_{\frac{\alpha}{2}} - \frac{|\mu_1-\mu_2|}{\sqrt{\frac{2s^2_{pooled}}{n}}})$$

The variance of the difference between two random variables is the sum of the variances (error are always added) (ST&D 113-115). If the variances are the same, then the variance of the difference is 2 * average $s^2$, explaining the multiplication by 2 in the formula of the standard error of the difference between the two means

$$\sqrt{\frac{2s^2_{pooled}}{n}}$$

The degrees of freedom for the critical $t_{\alpha/2}$ are

**Equal sample size**: 2*(n-1)

**Unequal sample size**: $(n_1\text{-}1) + (n_2\text{-}1)$.

**Summary of hypothesis testing** [S&T p. 92]

- Formulate a meaningful hypothesis for which a test statistic can be computed.
- Choose a probability α and keep in mind the power 1-β of the test.
- Compute the sample value of the test statistic and find the probability of obtaining, by chance, a value more extreme than that observed.
- If the tests statistic is > than the critical value reject $H_o$.

*Practical considerations*

(1) If $H_o$ is rejected, the test gives no information as to *how far* the hypothesis $H_o$ is from being true.

(2) $H_o$ is almost always rejected if the sample size is too large, and is almost always not rejected if the sample size is too small.

## 2. 4. Determining the sample size

### 2. 4. 1. Factors affecting replications

There are many factors that affect the decision of how many replications of each treatment should be applied in an experiment. Some of the factors can be put into statistical terms and be determined statistically. Others are non-statistical and depend on experience and knowledge of the subject research matter or available resources for research.

**i.** ***Nonstatistical factors***: cost and availability of experimental material, the restraints and cost of measurements, the difficulty and applicability of the experimental procedure, the knowledge or experience of the subject matter of research, the objective of the study or intended inference populations.

Example: Use existing information on the experimental units for better sampling selections. Suppose several new varieties are to be tested at different locations. Instead of randomly selecting say 10 locations from 100 possible testing sites, one should study these locations first. Maybe it is possible to classify these 100 locations into a few distinct classes of sites based on various physical or environmental attributes of these locations. Thus, it may be appropriate to choose one location randomly from each class to represent certain environmental characteristics for the varietal testing experiment. This example illustrates the point that in order to determine a suitable number of replications, it is not sufficient only to find a number but also the properties of replicated material ought to be considered.

Example: The desired biological response dictates the sample size. In an experiment of testing treatment differences, the important biological difference that needs to be detected must be clearly understood and defined. A relevant sample size can only be determined statistically to assure success with high probability, to detect this significant difference. Thus, in testing a herbicide effect on weed control, what is the meaningful biological effect? 60% weed control or 90% weed control is satisfactory? Between two competitive herbicides, is 5% difference in controlling weeds important enough to be detected, or 10% should be detected in an experiment? It should be pointed out, the smaller difference to be detected, the large sample size is required. And a positive difference can always be shown significant statistically no matter how small it is by simply increasing sample size. But a negative hypothesis or no difference between treatments can never be proved with certainty, no matter how large a sample size is used in the study.

**ii.** ***Statistical factors***: Statistical procedures that are developed to determine the number of replications are primarily based on considerations of the required **precision** of an estimator or required **power** of a test. Some commonly used procedures are described in the next section. Note that for a given $\mu$ and $s_{\bar{Y}}$, if 2 of the 3 quantities $\alpha$, $\beta$, or the number of observations r are specified then the third one can be determined. In choosing a sample size to detect a particular difference, one must admit the possibility of either Type I or Type II error and choose the sample size accordingly.

We will analyze examples for sample size estimation for confidence intervals of certain length (2.4.2 and 2.4.3), for comparisons of two means using Type I and II errors (2.4.4), and for confidence intervals for standard deviations using the Chi-square distribution (2.4.5).

### 2. 4. 2 Sample size for the estimation of the mean with known $\sigma^2$. Using the z statistic

Where attention is concentrated on estimation rather than testing, a procedure for determining the required number of observations is available for continuous data. The problem is to determine the necessary sample size to estimate a mean by a confidence interval guaranteed to be no longer than a prescribed length.

If the population variance $\sigma^2$ is known, or if it is desired to estimate the confidence interval in terms of the population variance, the z statistic for the standard normal distribution can be used. No initial sample is required to estimate the sample size. The sample size can be computed as soon as the required confidence interval length is decided upon. Recall that

$Z = \frac{\overline{Y} - \mu}{\sigma_{\overline{y}}}$ and the confidence interval $\overline{Y} \pm z_{\alpha/2}\, \sigma_{\overline{Y}}$

The formula for the confidence interval using the z statistic is

$$d = z_{\alpha/2}\sigma / \sqrt{r}$$

With d= half-length of the confidence interval. This can be rearranged to give

$$r = z^2_{\alpha/2}\, \sigma^2 / d^2$$

### 2. 4. 3 Sample size for the estimation of the mean.

**Stein's Two-Stage Sample [**S&T p. 124].

When the variance is unknown and the SE of the mean is used instead of the variance, the *t* distribution should replace the normal distribution. The additional complication generated by the use of *t* is that this distribution also depends on the number of replications. Therefore an iterative approach is required.

Consider a (1 - α)% confidence interval of a mean, μ.

$$\overline{Y} - t_{\alpha/2,\, r-1}\ s_{\overline{Y}} \le \mu \le \overline{Y} + t_{\alpha/2,\, r-1}\ s_{\overline{Y}}$$

The length of the (1 - α) % **half-length (d)** of the confidence interval is

$$d = t_{\alpha/2,\, r-1}\ s_{\overline{Y}} = t_{\alpha/2,\, r-1}\ \ s/\sqrt{r}$$

This formula can be rearranged to estimate necessary sample size r to estimate a mean by a confidence interval no longer than 2d,

$$r = t^2_{\alpha/2,\, r-1} \; s^2 / d^2 \;\; (\approx \; z^2 \; s^2 / d^2)$$

Stein's procedure is to use a pilot study to estimate $s^2$. Then compute r from this formula.

The equation may be expressed in terms of the **coefficient of variation**, (CV= s / $\overline{Y}$, ST&D p. 26) and the population mean as

$$r = \frac{t^2_{\alpha/2,n-1} CV^2}{(d/\mu)^2}$$

Here d/μ is the confidence interval as a fraction of the population mean. For example d/μ = 0.1 means that the length of d (half-length of the confidence interval) should not be larger than one tenth of the population mean.

***Example 1*** [ST&D, p. 125). The desired length of the confidence interval for the estimate μ is at most 10 mm, and the desired confidence interval is 95%. A preliminary sample gives values of 22, 19, 13, 22, and 23 mm. With $(t_{0.025,\,4})^2 = 7.71$, $s^2 = 16.7$, and $d^2 = 25$ we get

$$r = \frac{16.7 * 7.71}{25} = 5.15$$

Therefore we need six observations. If the result is >> than the r of the pilot study, r will be too far away from the number of degrees of freedom used in the *t* statistic. We must iterate until the equation is satisfied with the same *r* values on both sides of the equal sign.

***Example 2***: An experimenter wants to estimate the mean height of certain mature plants. From a pilot study of 5 plants, he finds that *s* = 10 cm. What is the required sample size, if he wants to have the length of a 95% confidence interval of the mean shorter than 5 cm?

Using $r = t^2_{\alpha/2,\, r-1} \, s^2 / d^2$, the sample size is estimated **iteratively**,

| initial-n | $t_{\alpha/2=2.5\%,\, df}$ | n |
|---|---|---|
| 5 | 2.776 | $(2.776)^2 (10)^2 / 2.5^2 = 123$ |
| 123 | 1.96 | $(1.96)^2 (10)^2 / 2.5^2 = 62$ |
| 62 | 2.00 | 64 |
| 64 | 2.00 | 64 |

Thus, with 64 observations he could have an estimated average height within 5 cm deviation of the true mean. Note that if we start with "Z" approximation, then: $n = Z^2 \; s^2 / d^2 = (1.96)^2 (10)^2 / 2.5^2 = 62$, which is not too far from the more exact estimate, 64.

### 2. 4. 4 Sample size estimation for a comparison of two means

In testing the hypothesis $H_o$: $\mu = \mu_o$, we can also take into account the possibility of a Type II error and the power of a test (1-$\beta$). To calculate $\beta$ we need to known the alternative $\mu_1$ or at least the minimum difference we want to detect between the means $\delta = \mu_o - \mu_1$. The appropriate formula for computing r, the number of observations on **each** treatment is given by equation:

$$r = 2\,(\sigma / \delta)^2\,(Z_{\alpha/2} + Z_\beta)^2$$

For $\alpha$= 0.05 and $\beta$= 0.20, $z_\beta = 0.8416$, $z_{\alpha/2} = 1.96$, and $(Z_{\alpha/2} + Z_\beta)^2 = 7.849$.

So, to discriminate two means that are 2 $\sigma$ apart approximately 4 samples per treatment are required ($2 * (½)^2 * 7.849$). If the two means are one or one half standard deviations apart that number increases to 16 and 63 respectively.

The alternatives are two-tailed because we do not know if the alternative $\mu_1$ is larger or smaller than $\mu_o$. This solution has several obvious difficulties. The first is that we rarely know $\sigma^2$ and we cannot honestly use *Z* table. An approximate solution is to multiply

r * (error d.f. +3) / (error d.f. +1) [ST&D p. 123];

where error d.f. is (number of pairs -1) for meaningfully paired samples [ST&D p. 106] and 2(n-1) for independent samples. The same equation and correction can be used to estimate the number of blocks in a randomized complete block design analysis of variance [ST&D p. 241]

Another option is to estimate $\sigma$ with the pooled sample standard deviation *s* (ST&D p100 Eq.5.8) and replace *z* by *t*. Using the equation:

$$r = 2\,(s_{(pooled)} / \delta)^2\,(t_{\alpha/2,\, n1+n2-2} + t_{\beta,\, n1+n2-2})^2$$

the sample size, is estimated **iteratively**, as in 2. 4. 3. If r is known this equation can be used to estimate the power of the test (find first $t_{\beta,\, n1+n2-2}$).

If no estimate of *s* is available, the equation may be expressed in terms of the coefficient of variation, and the difference $\delta$ between means as a proportion of the mean:

$$r \approx 2\,[(\sigma/\mu) / (\delta/\mu)]^2\,(Z_{\alpha/2} + Z_\beta)^2 \approx 2(CV/\delta\%)^2(Z_{\alpha/2} + Z_\beta)^2$$

The problem may be also obviated by defining $\delta$ in terms of $\sigma$. For example we might want to detect a difference between means of one standard deviation in size. In this case $r = 2\,(Z_{\alpha/2} + Z_\beta)^2$

***Example***: Two varieties will be compared for yield, with a previously estimated sample variance, $s^2 = 2.25$. How many replications should be used, so that these varieties can be compared with an $\alpha = 5\%$, and $\beta = 20\%$ to detect a difference greater than 1.5 tons/acre?

Approximate $\boldsymbol{r} = 2\,(\sigma/\delta)^2\,(Z_{\alpha/2} + Z_\beta)^2 = 2\,(1.5/1.5)^2(1.96+0.8416)^2 = 15.7$ and then use $\boldsymbol{r} = 2\,(s\,/\,\delta)^2\,(t_{\alpha/2} + t_\beta)^2$ to estimate the sample size iteratively.

| Guesstimate n | df = 2(r - 1) | $t_{0.025}$ | $t_{0.20}$ | estimated n |
|---|---|---|---|---|
| 16 | 30 | 2.0423 | 0.8538 | 16.8 |
| 17 | 32 | 2.0369 | 0.8530 | 16.7 |

The answer is that each variety should be planted in 17 plots. In general, 17 samples per treatment are necessary to discriminate two means that are one standard deviation apart with $\alpha = 5\%$, and $\beta = 20\%$.

## 2. 4. 5. Sample size to estimate population standard deviation

The Student-t distribution is used to establish confidence intervals around the sample mean as an estimate for the population mean of a normally distributed random variable. The chi-squared distribution is used in a similar way to establish confidence intervals around the sample variance as an estimate for the population variance.

### 2. 4. 5. 1. The Chi- square distribution [ST&D p. 55]

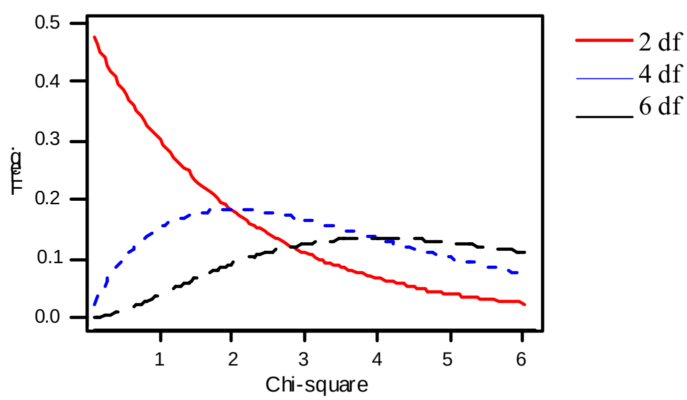


Fig. 4. Distribution of $\chi^2$, for 2, 4, and 6 degrees of freedom.

*Relation between the normal and chi-square distributions:* The reason that the chi square distribution provides a confidence interval for $\sigma^2$ is that the normal and chi square distributions have the following relationship. If $Z_1, Z_2, \ldots, Z_n$ are random variables from a *standard* normal distribution then the sum $Z^2_1 + \ldots + Z^2_n$ has a $\chi^2$ distribution with n degrees of freedom. Chi square is defined as the sum of squares of independent, normally distributed variables with zero means and unit variances.

$$\boxed{\chi^2_{1,\alpha} = Z^2_{(0,1)\ \alpha/2} = \tau^2_{\propto\ \alpha/2}}$$

$\chi^2_{1,\ 0.05} = 3.84$ and $Z^2_{(0,1),\ 0.025} = \tau^2_{\propto,\ 0.025} = 1.96^2 = 3.84$

Note that $Z$ vales from both tails of the $N$ distribution go into the upper tail of the $\chi^2$ for 1 d.f. because of the disappearance of the minus sign in the squaring.

$$\sum_{i=1}^{n} Z_i^2 = \sum \frac{(Y_i - \mu)^2}{\sigma^2} = \frac{1}{\sigma^2} \sum (Y_i - \mu)^2$$

and when we change the parametric μ to a sample means this expression becomes:

$$\frac{1}{\sigma^2} \sum (Y_i - \overline{Y})^2 = \frac{(n-1)s^2}{\sigma^2}$$ that has a $\chi^2_{n-1}$ distribution.

This expression is important because it provides a relationship between the sample variance and the parametric variance.

### 2. 4. 5. 2. Confidence interval formulas

Based on the formula given in the previous paragraph, it follows that a (1-α)% confidence interval for the variance can be derived as follows: Suppose $Y_1, Y_2, \ldots, Y_n$ are random variables drawn from a normal distribution with mean μ and variance $\sigma^2$. We can make the following statement about the ratio $(n-1)\ s^2/\sigma^2$:

$$P\{\ \chi^2_{1-\alpha/2,\ n-1} \le (n-1)\ s^2/\sigma^2 \le \chi^2_{\alpha/2,\ n-1}\} = 1 - \alpha$$

Simple algebraic manipulation of the quantities in the inequality within brackets yields

$$P\ \{(n-1)\ s^2/\ \chi^2_{\alpha/2,\ n-1} \le \sigma^2 \le (n-1)\ s^2/\ \chi^2_{1-\alpha/2,\ n-1}\} = 1 - \alpha$$

$$\text{or}\ \ P\ \{\chi^2_{1-\alpha/2,\ n-1}\ /(n-1) \le s^2/\sigma^2 \le \chi^2_{\alpha/2,\ n-1}/(n-1)\} = 1 - \alpha$$

The 2nd form is particularly useful when the precision of $s^2$ can be expressed in terms of the percent of $\sigma^2$. For example, in the Barley SAS data above, $s^2 = 1.5057$, and n = 14. If we let α = 0.05 then from a table we find $\chi^2_{0.975,,\ 13} = 5.01$ and $\chi^2_{0.025,,\ 13} = 24.7$. Therefore the 95% confidence interval for $\sigma^2$ is [0.79 - 3.91].

Example: An experimenter wants to estimate σ with a 90% confidence that the estimated s is within 20% difference from σ, what is the required sample size?

$p\,(0.8 \le S/\sigma \le 1.2) = .90$ or $p\,(0.64 \le S^2/\sigma^2 \le 1.44) = 0.90$

thus $X^2_{(1-\alpha/2,\, n-1)} / (n-1) = 0.64$ or $X^2_{(\alpha/2,\, n-1)} / (n-1) = 1.44$

Since $\chi^2$ is not symmetrical, the above two solutions may not be exactly equal if sample size is small. However, usually small sample cannot provide good estimate of $\sigma^2$. Thus the required sample size would be large enough not to have any problem. The actual computation involves the following iterative process:

Guesstimate n

| Size | $1 - \alpha/2 = 95\%$ | | $\alpha/2 = 5\%$ | |
|---|---|---|---|---|
| (n-1) | $\chi^2_{(n-1)}$ | $\chi^2_{(n-1)}/(n-1)$ | $\chi^2_{(n-1)}$ | $\chi^2_{(n-1)}/(n-1)$ |
| 20 | 10.90 | 0.545 | 31.4 | 1.57 |
| 30 | 18.50 | 0.616 | 43.8 | 1.46 |
| 40 | 26.50 | 0.662 | 55.8 | 1.40 |
| **35** | **22.46** | **0.642** | **49.8** | **1.42** |

Thus a rough estimate of the required sample size is approximately 36 (n-1= 35).