

Håkon K. Gjessing

Professor/Principal Investigator

Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo

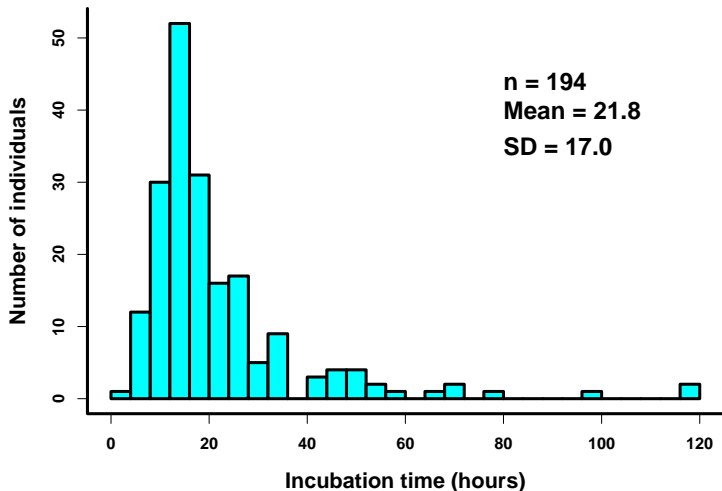
Department of Global Public Health and Primary Care, University of Bergen

Makerere

Wednesday, 7 June 2023



HISTOGRAM OF INCUBATION TIMES FOR 194 INDIVIDUALS



CONFIDENCE INTERVALS (CI)

Confidence interval based on the Normal distribution:

- \bar{X} is approximately normally distributed ... so the interval

$$(\bar{X} - 1.96 \cdot \text{SE}(\bar{X}), \bar{X} + 1.96 \cdot \text{SE}(\bar{X}))$$

has a 95% chance of covering the “true” μ .

Lower limit:

$$\bar{X} - 1.96 \cdot \text{SE}(\bar{X}) = 21.8 - 1.96 \cdot 1.22 = 19.4$$

Upper limit:

$$\bar{X} + 1.96 \cdot \text{SE}(\bar{X}) = 21.8 + 1.96 \cdot 1.22 = 24.2$$

95% confidence interval: (19.4, 24.2)

HYPOTHESIS TESTING, NORMAL DISTRIBUTION

One-sided test:

- Null hypothesis $H_0 : \mu \leq 18$ hours
 - Alternative hypothesis $H_A : \mu > 18$ hours
-

- If $\mu = 18$ happened to be true, we should have $\bar{X} \approx 18$
- We actually *observe* $\bar{X} = 21.8$
- Remember: $SE(\bar{X}) \approx 1.22$

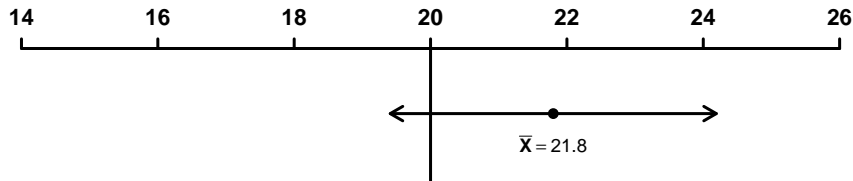
P-VALUE

The probability of observing \bar{X} *at least* 21.8
if $H_0 : \mu = 18$ were true

So, is \bar{X} larger than just random variation away from $\mu = 18$?

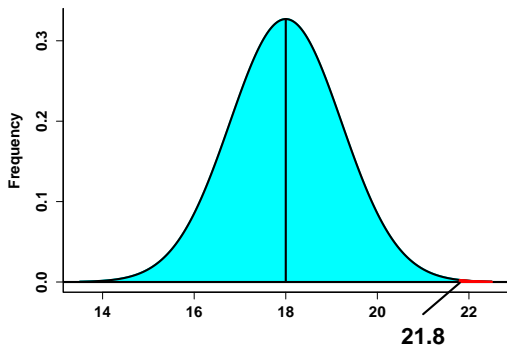
CONFIDENCE INTERVALS (95% CI)

Remember from previous lecture:



- It seems clear that 18 is unlikely given the 95% confidence interval.
- But can we be more precise in *how* unlikely??

HYPOTHESIS TESTING: COMPUTE THE Z-SCORE AND P-VALUE

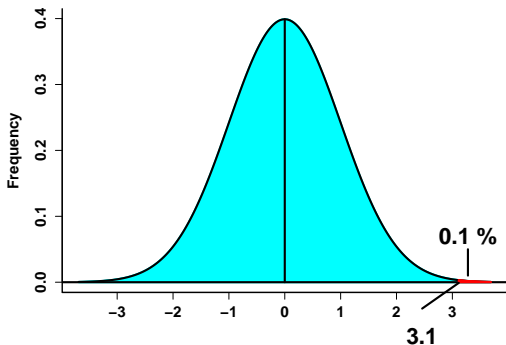


Z-score:

$$Z = \frac{21.8 - 18}{1.22} = 3.1$$

That is, \bar{X} is more than 3 SE away from $\mu = 18$

HYPOTHESIS TESTING: COMPUTE THE Z-SCORE AND P-VALUE

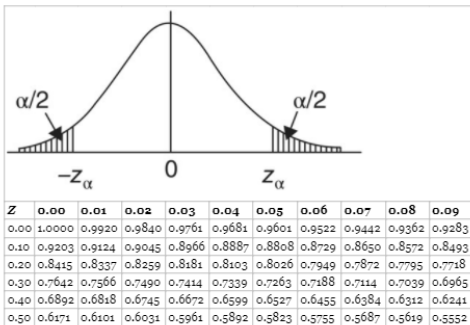


P-value = probability that $Z \geq 3.1$ in a standard Normal distribution

TEXTBOOK: THE NORMAL DISTRIBUTION

Appendix Statistical Tables

Table T1 The Standard Normal distribution. The value tabulated is the probability, α , that a random variable, Normally distributed with mean zero and standard deviation one, will be



2.40	0.0164	0.0160	0.0155	0.0151	0.0147	0.0143	0.0139	0.0135	0.0131	0.0128
2.50	0.0124	0.0121	0.0117	0.0114	0.0111	0.0108	0.0105	0.0102	0.0099	0.0096
2.60	0.0093	0.0091	0.0088	0.0085	0.0083	0.0080	0.0078	0.0076	0.0074	0.0071
2.70	0.0069	0.0067	0.0065	0.0063	0.0061	0.0060	0.0058	0.0056	0.0054	0.0053
2.80	0.0051	0.0050	0.0048	0.0047	0.0045	0.0044	0.0042	0.0041	0.0040	0.0039
2.90	0.0037	0.0036	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028
3.00	0.0027	0.0026	0.0025	0.0024	0.0024	0.0023	0.0022	0.0021	0.0021	0.0020
3.10	0.0019	0.0019	0.0018	0.0017	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014
3.20	0.0014	0.0013	0.0013	0.0012	0.0012	0.0012	0.0011	0.0011	0.0010	0.0010
3.30	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.40	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005	0.0005

P-VALUE

From tables of the Normal distribution (see textbook):

$$\mathbf{P\text{-value}} \approx \mathbf{0.0020/2 = 0.001}$$

Not likely to be a coincidence: **H_0 rejected!**

- P-value is the probability of an extreme result *if* H_0 is true.
- But where should we draw the line for “coincidence”??

LEVEL OF THE TEST

Typical limit for “extreme”: **5%**

... this level makes sense..., but mostly chosen because of tradition.

TWO-SIDED TEST

- Null hypothesis $H_0 : \mu = 18$ hours
- Alternative hypothesis $H_A : \mu \neq 18$ hours

Two-sided p-value = 2 * One-sided p-value

$$\text{P-value} = 2 * 0.001 = 0.002$$

We reject H_0 since p-value is less than 5%

But there is a **DILEMMA**:

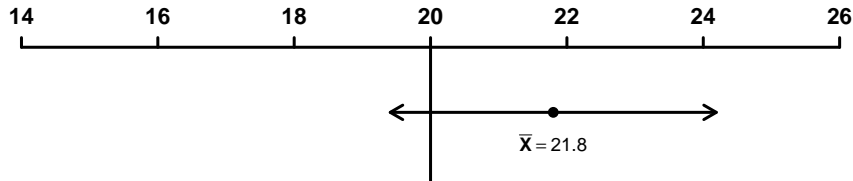
Two-sided tests *a/ways* reject H_0 if enough data!

(A bit weird... but still sometimes useful)

CONFIDENCE INTERVALS FOR TESTING?

- Can confidence intervals be used for hypothesis testing?
- *Should* confidence intervals be used for hypothesis testing?

CONFIDENCE INTERVALS (95% CI)



- Test by checking if H_0 is outside 95% CI.
- This corresponds to a *two-sided* test with level 5%.
- **But** test and CI can sometimes give different conclusions.

Recommendation:

- Use confidence intervals to measure precision!
- Use hypothesis tests (with p-value) for testing!

(Sounds logical, doesn't it)

CAUSES OF DEATH IN THREE NORWEGIAN COUNTIES¹

- Data collection 1974–78
- All men and women aged 35–49 years
- Three rural Norwegian counties:
Oppland, Sogn og Fjordane, and Finmark
- Cardiovascular health screening examination
- More than 90% participation
- Self-report on past and current **smoking habits**
- Mortality in cohort followed-up to the end in year 2000
- Linked to **cause of death** registry at Statistics Norway
- A subset: 4000 (random) out of about 50000
(<http://folk.uio.no/borgan/abg-2008/data/data.html>)
- **Focus: Mortality 40-70 years**

1: Vollset SE, Tverdal A, Gjessing HK. Smoking and deaths between 40 and 70 years of age in women and men. *Annals of internal medicine*. 2006;144(6):381.



TOTALLY INDECENT SELF-PROMOTION

Statistics for Biology and Health

Statistics for Biology and Health

Odd O. Aalen, Ørnulf Borgan and Håkon K. Gjessing
Survival and Event History Analysis

Aalen • Borgan • Gjessing

Odd O. Aalen
Ørnulf Borgan
Håkon K. Gjessing

Time-to-event data are ubiquitous in fields such as medicine, biology, demography, sociology, economics and reliability theory. Recently, a need to analyse more complex event histories has emerged. Examples are individuals that move among several states, trials that include some units (e.g. before others), internal time-dependent covariation, and the estimation of causal effects from observational data.

The aim of this book is to bridge the gap between standard textbook models and a range of models where the dynamic structure of the data comes to its full right. The consistent demonstration of such models in stochastic processes, the authors show how counting processes, martingales, and stochastic integrals fit very nicely with empirical data. Beginning with standard analyses such as Kaplan-Meier plots and Cox regression, the presentation progresses to the additive hazard model and recurrent event data. Stochastic processes are also used in natural models for individual frailty; they allow sensible interpretations of a number of surprising artefacts seen in population data.

The stochastic process framework is naturally connected to causality. The authors show how dynamic path analysis can incorporate many modern causality ideas in a framework that takes the time aspect seriously.

To make the material accessible to the readers, a large number of practical examples, mainly from medicine, are developed in detail. Stochastic processes are introduced in an intuitive and non-technical manner. The book is aimed at investigators who use event history methods and want a better understanding of the statistical concepts. It is suitable as a textbook for graduate courses in statistics and biostatistics.

Odd O. Aalen is professor of medical statistics at the University of Oslo, Norway. He PhD from the University of California, Berkeley in 1975 introduced counting processes and martingales in event history analysis. He has also contributed to numerous other areas of event history analysis, such as additive hazards regression, frailty, and causality through dynamic modeling.

Ørnulf Borgan is professor of statistics at the University of Oslo, Norway. Since his PhD in 1984 he has contributed extensively to event history analysis. He is co-author of the monograph *Statistical Models Based on Counting Processes*, and is editor of *Scandinavian Journal of Statistics*.

Håkon K. Gjessing is professor of medical statistics at the Norwegian Institute of Public Health and the University of Bergen, Norway. Since his PhD in probability in 1985, he has worked on a broad range of theoretical and applied problems in biostatistics.

STATISTICS | LIFE SCIENCES,
MEDICINE, HEALTH SCIENCES



springer.com



Survival and Event History Analysis

Survival and Event History Analysis

A Point Process View

Springer

(Includes the example data analyzed using survival analysis)

SINGLE PROPORTION

- Risk of dying age 40–70, all causes (ignoring length of follow-up)
- $X = 586$ total number of deaths
- $n = 4000$ total number of individuals
- p is the proportion (risk) of dying, estimated by \hat{p} :

$$\hat{p} = \frac{X}{n} = \frac{586}{4000} = 0.1465 = 14.7\%$$

SINGLE PROPORTION

- Risk of dying age 40–70, all causes (ignoring length of follow-up)
- $X = 586$ total number of deaths
- $n = 4000$ total number of individuals
- p is the proportion (risk) of dying, estimated by \hat{p} :

$$\hat{p} = \frac{X}{n} = \frac{586}{4000} = 0.1465 = 14.7\%$$

What is the **uncertainty** of this estimate? **Standard Error:**

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{0.1465(1 - 0.1465)}{4000}} = 0.00559$$

SINGLE PROPORTION

- Risk of dying age 40–70, all causes (ignoring length of follow-up)
- $X = 586$ total number of deaths
- $n = 4000$ total number of individuals
- p is the proportion (risk) of dying, estimated by \hat{p} :

$$\hat{p} = \frac{X}{n} = \frac{586}{4000} = 0.1465 = 14.7\%$$

What is the **uncertainty** of this estimate? **Standard Error:**

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = \sqrt{\frac{0.1465(1 - 0.1465)}{4000}} = 0.00559$$

Standard 95% Confidence Interval for p

(Normal approximation, Wald interval):

$$\hat{p} \pm 1.96 \cdot SE(\hat{p}) = 0.1465 \pm 1.96 \cdot 0.00559 = (0.136, 0.157) = (13.6\%, 15.7\%)$$

TWO PROPORTIONS: COMPARE MALES AND FEMALES

- Overall death risk, by gender:

	Males	Females
Population proportions:	p_M	p_F
X (deaths)	398	188
n (total)	2086	1914
\hat{p}	19.1%	9.8%
$SE(\hat{p})$	0.00860	0.00680
95% CI each group:	(17.4%, 20.8%)	(8.6%, 11.2%)

TWO PROPORTIONS: COMPARE MALES AND FEMALES

- Non-overlapping confidence intervals means a significant difference (**almost** always...)
 - **Warning:** overlapping does NOT mean non-significant
-
- BUT: Would also like to **measure** gender difference (not only whether significant or non-significant)

RISK DIFFERENCE (RD)

	Males	Females
X (deaths)	398	188
n (total)	2086	1914
\hat{p}	19.1%	9.8%
$SE(\hat{p})$	0.00860	0.00680

Risk Difference:

$$RD = \hat{p}_M - \hat{p}_F = 19.1\% - 9.8\% = 9.3\%$$

RISK DIFFERENCE (RD)

	Males	Females
X (deaths)	398	188
n (total)	2086	1914
\hat{p}	19.1%	9.8%
$SE(\hat{p})$	0.00860	0.00680

Risk Difference:

$$RD = \hat{p}_M - \hat{p}_F = 19.1\% - 9.8\% = 9.3\%$$

- Uncertainty of RD (i.e. standard error of RD):

$$SE(RD) = \sqrt{SE(\hat{p}_M)^2 + SE(\hat{p}_F)^2} = \sqrt{0.00860^2 + 0.00680^2} = 0.011$$

NOTE ON TEXTBOOK FORMULAS

- SE for confidence intervals:

$$\begin{aligned}SE(\text{RD}) &= \sqrt{SE(\hat{p}_1)^2 + SE(\hat{p}_2)^2} \\ &= \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}\end{aligned}$$

- SE for hypothesis testing:

$$SE(\text{RD}) = \sqrt{\hat{p}(1 - \hat{p}) \left\{ \frac{1}{n_1} + \frac{1}{n_2} \right\}}$$

You can use the first BOTH for confidence intervals AND hypothesis testing. (But it illustrates that Standard Errors can be calculated in many different ways, depending on assumptions)

RISK DIFFERENCE (RD): CONFIDENCE INTERVAL

$$RD = \hat{p}_M - \hat{p}_F = 19.1\% - 9.8\% = 9.3\%$$

$$SE(RD) = \sqrt{SE(\hat{p}_M)^2 + SE(\hat{p}_F)^2} = \sqrt{0.00860^2 + 0.00680^2} = 0.011$$

95% Confidence Interval of RD (Wald interval):

$$RD \pm 1.96 \cdot SE(RD) = 0.093 \pm 1.96 \cdot 0.011$$

(7.1%, 11.4%)

RISK DIFFERENCE (RD): SIGNIFICANCE TEST

$$RD = \hat{p}_M - \hat{p}_F = 19.1\% - 9.8\% = 9.3\%$$

$$SE(RD) = \sqrt{SE(\hat{p}_M)^2 + SE(\hat{p}_F)^2} = \sqrt{0.00860^2 + 0.00680^2} = 0.011$$

Null hypothesis:

$$H_0 : p_M = p_F \quad \text{that is,} \quad RD_0 = p_M - p_F = 0$$

Test statistic:

$$\mathbf{Z\text{-score}} = \frac{RD - RD_0}{SE(RD)} = \frac{0.093 - 0}{0.011} \approx 8.5$$

- Is RD larger than random variation away from $RD_0 = 0$ should allow?

TEST OF SIGNIFICANCE, CONCLUSION

- RD is 8.5 standard deviations away from zero.
- Extremely unlikely by chance (WAY outside table in book).
- H_0 rejected.
- Higher mortality among men than women.

IS SIGNIFICANCE REALLY SIGNIFICANT?

- P-value measures degree of chance.
- NOT the actual **importance** of the result.
- **Statistical** significance is not the same as **biological** significance.
- Confidence intervals better for deciding biological importance.

In our example, more than 9 percentage points higher mortality is both statistically and biologically significant!