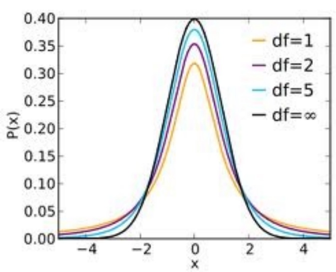
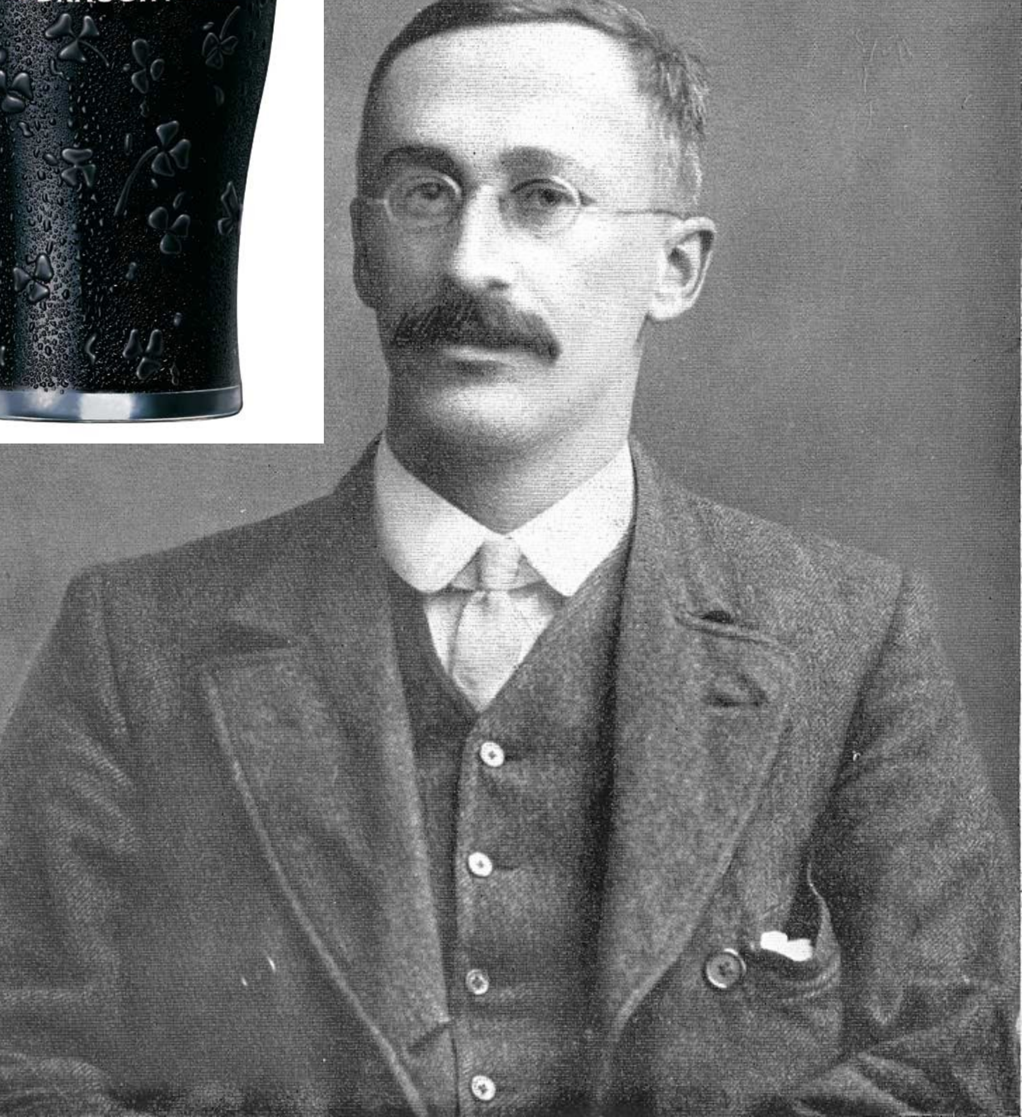
William Sealy Gosset (June 13, 1876–October 16, 1937)

Another researcher at Guinness had previously published a paper containing trade secrets of the Guinness brewery. To prevent further disclosure of confidential information, Guinness prohibited its employees from publishing any papers regardless of the contained information. However, after pleading with the brewery and explaining that his mathematical and philosophical conclusions were of no possible practical use to competing brewers, he was allowed to publish them, but under a pseudonym ("Student"), to avoid difficulties with the rest of the staff.[1] Thus his most famous achievement is now referred to as [Student's t-distribution, which might otherwise have been Gosset's t-distribution.]

Gosset had almost all of his papers including *The probable error of a mean published in Pearson's journal [Biometrika using the pseudonym Student.]*

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{\mathrm{s.e.}(\hat{\beta})},$$





Two-Sample or Unpaired t-test

- The unpaired test (aka independent sample or unrelated test) arises when the data in two groups are not connected, as in:
  - <u>Parallel group clinical trial</u>: one group receives the test treatment, and one group receives the control treatment (randomized)
  - <u>Unmatched case-control study</u>: the control group not designed to “look like” the diseased population (e.g., med student control)

Two-Sample or Unpaired t-test

- Example from literature: Larochelle et al. (1987) give the plasma atrial natriuretic factor concentration in blood taken from the aorta in **7** patients with essential hypertension as **25.0 ng/L (SE=6.0)** and in **8** patients with renovascular hypertension as **46.5 ng/L (SE=10.2)**. Does plasma atrial natriuretic factor differ significantly in the two groups?

Two-Sample or Unpaired t-test

- FYI: plasma atrial natriuretic factor = ANF, hormone produced by the heart that helps regulate blood pressure and kidney function.
- Hypertension = systolic pressure > 140 mmHg

  or diastolic pressure > 90 mmHg
- Essential hypertension = high blood pressure with no identifiable cause.
- Renovascular hypertension = caused by narrowing of the arteries that carry blood to the kidneys.

Two-Sample or Unpaired t-test

- With small samples, it leads to a more powerful test if we can assume that $\sigma_A = \sigma_B$.
- If populations have the same std. dev. then s_A and s_B both estimate the same quantity σ.
- Best estimate of σ is s_P (pooled).

Two-Sample t-test

- Suppose we wish to test the null hypothesis that the means from two populations, estimated from two independent samples, are equal but the sample sizes are small.
- Sample 1: number of subjects n_1, mean $\overline{x}_1$, std. dev. s_1
- Sample 2: number of subjects n_2, mean $\overline{x}_2$, std. dev. s_2

Two-Sample t-test

- Assumptions:
  1. Data are normally distributed
  2. Data are independent
  3. Standard deviations from the two populations are equal. (~less than factor of two difference between samples)
- Calculate a pooled std. dev. as

$$s_p = (((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2)/(n_1 + n_2 - 2))^{1/2}$$

Two-Sample t-test

- Standard error of the difference is:

$$SE(\overline{x}_1 - \overline{x}_2) = s_p * sqrt(1/n_1 + 1/n_2)$$

- and

$$t = (\overline{x}_1 - \overline{x}_2)/SE(\overline{x}_1 - \overline{x}_2)$$

- Under the null hypothesis, this is distributed as the Student's t-distribution with $n_1 + n_2 - 2$ degrees of freedom.
- A 95% confidence interval for $\overline{x}_1 - \overline{x}_2$ is:

$$(\overline{x}_1 - \overline{x}_2) \pm t_{0.05} \text{ x } SE(\overline{x}_1 - \overline{x}_2)$$

Two-Sample t-test

- Short example:
- Asthmatics' FEV_1: $n_1=5$, $\overline{x}_1=1.86$, $s_1=0.378$
- Controls' FEV_1: $n_2=6$, $\overline{x}_2=2.51$, $s_2=0.210$
- FEV_1 = forced expiratory volume after 1 second (in liters)
- Then, $s_p=0.297$, $SE(\overline{x}_1 - \overline{x}_2)=0.180$, $t=0.65/0.180=3.62$, $df=5+6-2=9$.

Two-Sample t-test

- From table, with 9 deg. of freedom, $t_{0.01}$ = 3.25, therefore $p<0.01$.
- Also with 9 deg. of freedom $t_{0.05}$ = 2.262 and so 95% confidence interval for the difference is given by:

$$0.65 \pm 2.262 \times 0.180$$

$$0.24 – 1.06 \text{ L/s}$$

Two-sample t-test

- For unpaired test, $df=(n_A-1)+(n_B-1) = n_A+n_B-2$.
- Back to example of Larochelle et al. (1987).
- From data, we get $\overline{d} = 46.5 - 25.0 = 21.5$ ng/L.
- Using the fact that $SE(\overline{x}) = SD(\overline{x})/\mathrm{sqrt}(n)$ we can back-calculate std. devs. of 15.9 and 28.8 ng/L.

$$s_p = ((6 \times 15.9^2 + 7 \times 28.8^2)/(7+8-2))^{½}$$

$$= 23.7 \quad \text{(pooled)}$$

Two-sample t-test

- and $SE(\overline{d}) = (23.7^2/7 + 23.7^2/8)^{½} = 12.3$
- From table, with $\alpha = 0.05$ and df = 6+7=13, $t_{0.05} = 2.160$.
- Thus, the 95% confidence interval for δ is:

$$21.5 + (2.160 \times 12.3)$$

or $-5.1 – 48.1$ ng/L

- This confidence interval includes the null hypothesis value of zero difference between diagnostic groups. Therefore considerable uncertainty remains!

Q1: Choose the FALSE statement.

In a <u>double-blinded</u> study…

A. conscious or subconscious bias is removed.

B. the name derives from literally “blindfolding” someone.

C. each experiment is repeated twice and outlying data is discarded.

D. the placebo effect is lessened.

E. subjects are randomly assigned to the experimental or control group.

Q2: The significance level α is:

A. The threshold value of p for rejection of the null hypothesis.

B. Typically equal to 0.10 by convention.

C. The probability of making a Type II error.

D. All of the above.

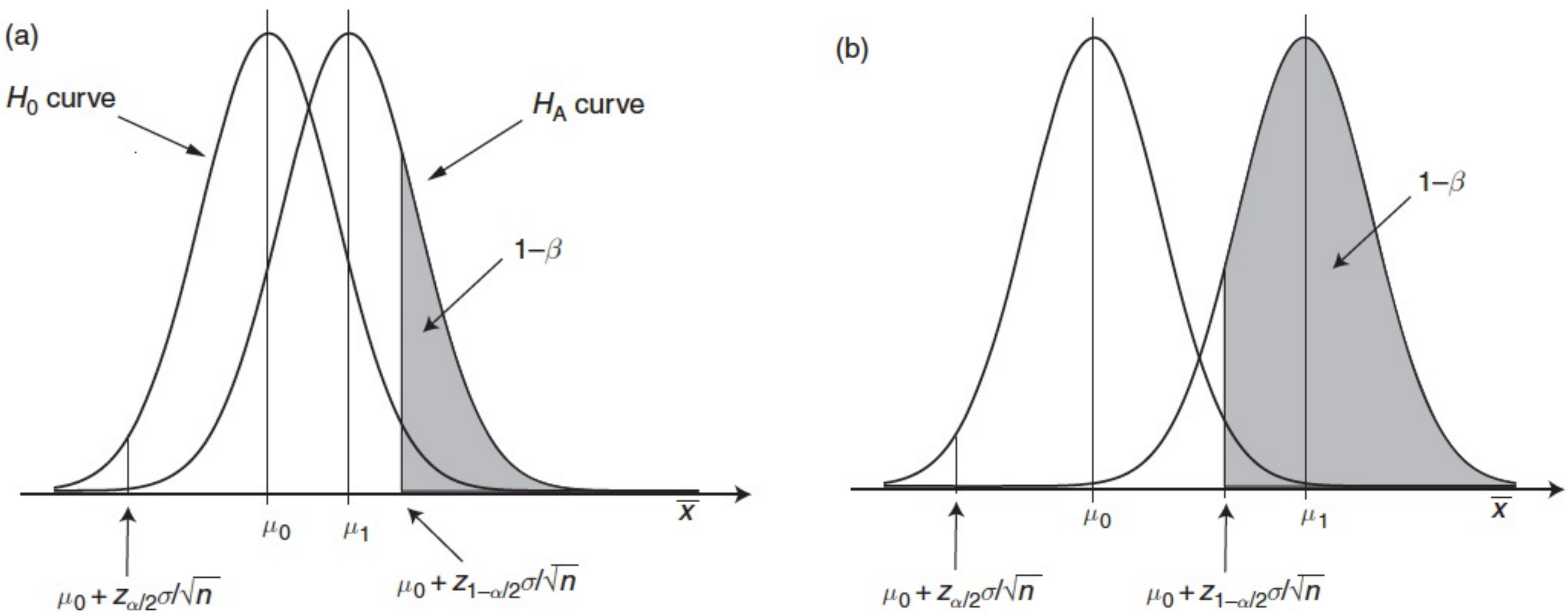
Q3: A nondirectional one-sample Z-test is portrayed. The shaded region represents what quantity?

A. The median

B. The power

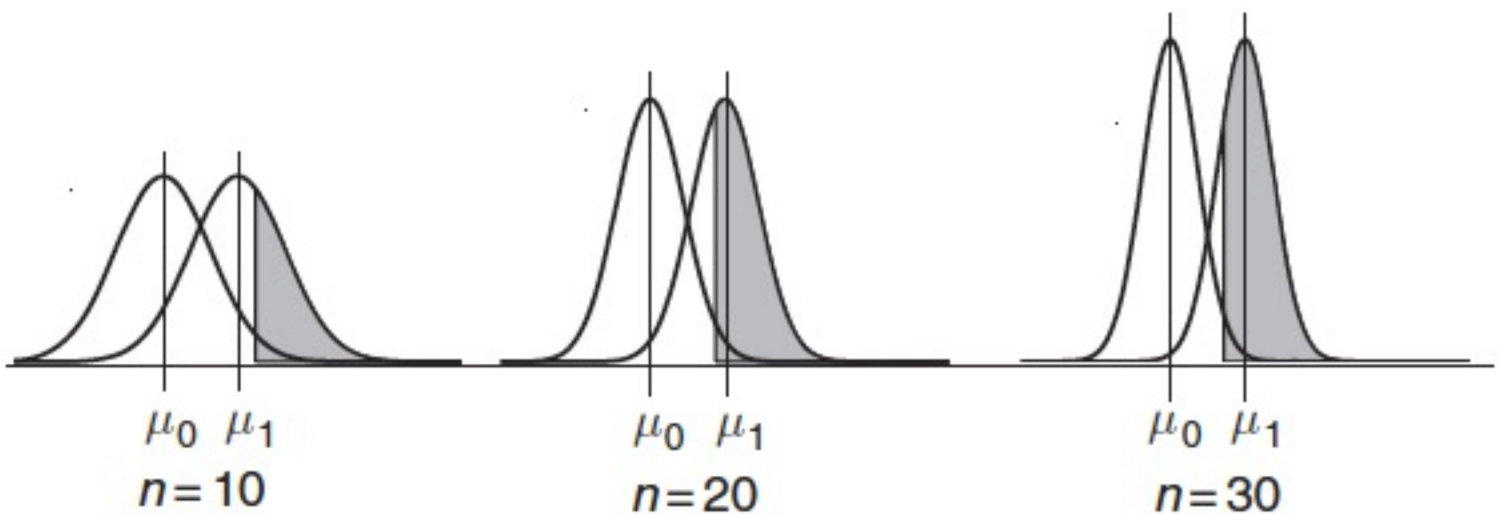
C. Upper quartile

D. Skewness factor



The **power** of the hypothesis test is given by $(1 - \beta)$. It is the probability of rejecting a false null hypothesis.

When we do not reject a false null hypothesis, we make a **type II error**. If the null hypothesis is false, the probability of making a type II error is β. In other words, β is the probability of obtaining a statistically insignificant result, even when the null hypothesis is incorrect.



Power Depends on Sample Size

- Power = $1-\beta$ = $P(\text{ reject } H_0 \mid H_1 \text{ true })$
  - "Probability of rejecting the null hypothesis if the alternative hypothesis is true."
- More subjects ➔ higher power

Power is Effected by…..

- Variation in the outcome (σ^2)
  - $\downarrow \sigma^2 \quad \rightarrow$ power $\uparrow$
- Significance level (α)
  - $\uparrow \alpha \rightarrow$ power $\uparrow$
- Difference (effect) to be detected (δ)
  - $\uparrow \delta \rightarrow$ power $\uparrow$
- One-tailed vs. two-tailed tests
  - Power is greater in one-tailed tests than in comparable two-tailed tests

Power Changes

- 2n = 32, 2 sample test, 81% power, δ=2, σ = 2, α = 0.05, 2-sided test
- Variance/Standard deviation
  - σ: 2 → 1 Power: 81% → 99.99%
  - σ: 2 → 3 Power: 81% → 47%
- Significance level (α)
  - α : 0.05 → 0.01 Power: 81% → 69%
  - α : 0.05 → 0.10 Power: 81% → 94%

Power Changes

- $2n = 32$, 2 sample test, 81% power, $\delta=2$, $\sigma = 2$, $\alpha = 0.05$, 2-sided test
- Difference to be detected (δ)
  - δ : 2 → 1 Power: 81% → 29%
  - δ : 2 → 3 Power: 81% → 99%
- Sample size (n)
  - n: 32 → 64 Power: 81% → 98%
  - n: 32 → 28 Power: 81% → 75%
- One-tailed vs. two-tailed tests
  - Power: 81% → 88%

Power should be….?

- Phase III: industry minimum = 80%
- Some say Type I error = Type II error
- Many large “definitive” studies have power around 99.9%
- Proteomics/genomics studies: aim for high power because Type II error a bear!

What is δ?

- δ is the minimum difference between groups that is judged to be clinically important
  - **Minimal effect which has clinical relevance in the management of patients**
  - **The anticipated effect of the new treatment (larger)**

The Choice of α and β depend on:

- the medical and practical consequences of the two kinds of errors
- prior plausibility of the hypothesis
- the desired impact of the results

The Choice of α and β

- $\alpha=0.10$ and $\beta=0.2$ for preliminary trials that are likely to be replicated.
- $\alpha=0.01$ and $\beta=0.05$ for the trial that are unlikely replicated.
- $\alpha=\beta$ if both test and control treatments are new, about equal in cost, and there are good reasons to consider them both relatively safe.

The Choice of α and β

- $\alpha > \beta$ if there is no established control treatment and test treatment is relatively inexpensive, easy to apply and is not known to have any serious side effects.
- $\alpha < \beta$ **(the most common approach 0.05 and 0,2**)if the control treatment is already widely used and is known to be reasonably safe and effective, whereas the test treatment is new,costly, and produces serious side effects.

Example 1

- An investigator wish to estimate the sample size necessary to detect a 10 mg/dl difference in cholesterol level in a diet intervention group compared to the control group. The variance from other data is estimated to be (50 mg/dl). For a two sided 5% significance level, $Z\alpha=1.96$, and for 90% power, $Z_\beta=1.282$.

- $2N=4(1.96+1.282)^2(\underline{\mathbf{50}})^2/10^2=\mathbf{1050}$

The χ^2 (chi-squared) Test

- 2x2 Contingency Tables: lets modify the test for a comparison of proportions to cover the situations of small samples.
- Recall the peptic ulcer study of Familiari et al. (1981).

Drug	Healed	Not Healed	Total	% healed
A: Pirenzipine	23 (a)	7 (c)	30 (m)	76.67
B: Trithiozine	18 (b)	13 (d)	31 (n)	58.06
Total	41 (r)	20 (s)	61 (N)	

The χ^2 (chi-squared) Test

- First, obtain a pooled estimate of the standard error of the difference in proportions, as done when comparing two means.
- If the null hypothesis is true, p_A and p_B both estimate a common parameter π, which is best estimated by $p=r/N = 41/61=0.6721$.
- (Note: p here is not the p-value! This is confusing but is common in the medical literature.)

The χ^2 (chi-squared) Test

- $SE(p_A - p_B) = (p(1-p)/m + p(1-p)/n)^{½} = 0.1202$
- (note use of pooled estimate)
- For a significance test we now calculate

  $Z = (p_A - p_B)/SE(p_A - p_B) = 0.1861/0.1202 = 1.548$
- From the Normal table, $p = 0.12$
- This calculation can be expressed in terms of the 2x2 contingency table…

The χ^2 (chi-squared) Test

- Thus, Z^2 is: (called χ^2 by convention):

  $\chi^2 = N(ad - bc)^2/(mnrs)$ exact!
- This test is termed the χ^2 or chi-squared test.
- For our example, the above formula gives χ^2 = 2.3940 (= 1.548^2)
- When dealing with a small number of subjects, the above expression for χ^2 is modified as…

The χ^2 (chi-squared) Test

- $\chi_c^2 = N(|ad - bc|) - ½N)^2/(mnrs)$
- The subtraction of ½ N causes χ_c^2 to be smaller than χ^2
- ½N is called the Yates correction for continuity, and arises because we are dealing with discrete data but the χ^2 distribution used to calculate the p-value is continuous (like the Normal distribution).
- For the peptic ulcer example, $\chi_c^2 = 1.62$.

The χ^2 (chi-squared) Test

- From χ^2 distribution table, α=0.2 → χ^2=1.64 for df=1 (close to χ_c^2=1.62)
- Thus significance level or p-value~0.2.
- Note that correlating sqrt(χ^2) on Normal table is only valid for df=1.

Chi-squared Test (χ^2) in 2x2 Tables

- In general,

		Factor A		
		Present	Absent	Total
Factor B	Present	a	c	m
	Absent	b	d	n
Total		r	s	N

- Essentially we calculate the values expected in the four cells of the table assuming the null hypothesis is true.

Chi-squared Test (χ^2) in 2x2 Tables

- $E_{ij} = R_i \times C_j/N$ where R_i is row total

  C_j is column total
- For example, $E_{11} = (a+c)(a+b)/N = mr/N$
- Then, $X^2 = \Sigma(O - E)^2/E$ O: observed

  E: expected
- With Yates correction,

$$X^2 = \Sigma(|O - E| - ½)^2/E$$

Chi-squared Test (χ^2) in 2x2 Tables

- This is equivalent to:

$$X_c^2 = N(|ad - bc| - \tfrac{1}{2}N)^2/(mnrs)$$

- Under the null hypothesis that the two factors are independent, X^2 has a χ^2 distribution with 1 degree of freedom.

- Next time: an example…