23. Inferences about means

- Central Limit Theorem revisited.
- Student's t distribution
- Confidence interval: One-sample t-interval for the mean.
- Test of hypotheses: One-sample t-test for the mean.

- Coffee cups.

  A coffee machine dispenses coffee into paper cups.
  Supposed to give 10 ounces, but amount varies.
  Random sample of 20 cups (sorted)

9.5	9.5	9.6	9.6	9.7	9.7	9.8	9.8	9.8	9.9
9.9	9.9	9.9	9.9	10.0	10.0	10.0	10.1	10.1	10.2

  Summary statistics: $\bar{y} = 9.845, s = 0.1986.$
  Evidence that the machine is shortchanging customers?

- Recall the one-proportion z-test. What is the same? What is different?

- Student's t distribution:
- Given n independent random variables from a $N(\mu, \sigma)$ model, the distribution of
$$\frac{\bar{y} - \mu}{s/\sqrt{n}}$$
is Student's t with $n-1$ degrees of freedom ($df = n - 1$).
- Symmetric, bell-shaped distribution, centered at zero.
- Distribution is very close to standard Normal when df is large, longer tailed when df small.
- Discovered by William Gosset via simulation and published in 1908 under pseudonym "Student". Mathematical derivation provided later by Fisher.

- Table T in Appendix D has critical values t^*_{df}.

	0.20	0.10	0.050	0.02	0.010	Two-tail probability
	0.10	0.05	0.025	0.01	0.005	One-tail probability
df						
1	3.078	6.314	12.71	31.82	63.657	
2	1.886	2.920	4.303	6.965	9.925	
3	1.638	2.353	3.182	4.541	5.841	
5	1.476	2.015	2.571	3.365	4.032	
10	1.372	1.812	2.228	2.764	3.169	
20	1.325	1.725	2.086	2.528	2.845	
30	1.310	1.697	2.042	2.457	2.750	
60	1.296	1.671	2.000	2.390	2.660	
1000	1.282	1.646	1.962	2.330	2.581	
∞	1.282	1.645	1.960	2.326	2.576	$t^*_{df} = z^*$ for $df = \infty$
	80%	90%	95%	98%	99%	Confidence levels

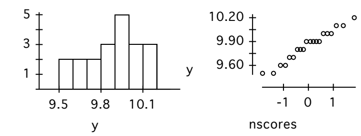
- One-sample t-test for the mean.

- Hypotheses. Interpretation of μ. Test $H_0 : \mu = \mu_0$ versus $H_A : \mu < \mu_0$ with $\mu_0 = 10$ here.

- Model. $t = \frac{\bar{y} - \mu_0}{SE(\bar{y})}$ where $SE(\bar{y}) = \frac{s}{\sqrt{n}}$.

  and reject H_0 if $t < t^*_{df}$. If conditions hold, use the t distribution (with $df = n - 1$) for the sampling distribution of t under H_0.

  1. Independent observations.
  2. Random sample or a randomized experiment.
  3. Sample represents at most 10% of the population.
  4. Sample sufficiently large so that Central Limit Theorem effect is present. Small n suffices given Nearly Normal Condition. Larger n required if outliers or skewness are present. Histogram and Normal probability plots recommended. See discussion in text.



- Mechanics: calculate t statistic and P-value.

  When using the t table, we can only trap the P-value between two critical values.

  In Data Desk, you can use the function $CumTDistr(t, df)$ to calculate probabilities and $InvCumTDistr(p, df)$ for critical values. A t calculator can also be used; see STAT 141 website. Or Google.

- Conclusion. Decision and interpretation.

- One-sample t-interval for the mean.

- Derivation: the following two statements are equivalent.

$$-t^*_{n-1} \leq \frac{\bar{y} - \mu}{SE(\bar{y})} \leq t^*_{n-1}$$

$$\mu \text{ is in the interval } \bar{y} \pm t^*_{n-1} \times SE(\bar{y})$$

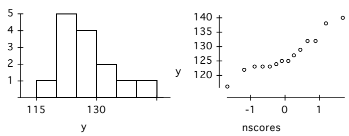
- Get the critical value t^*_{n-1} from a table or other technology. Remember to check conditions (same as those for t-test).

- Coffee machine.

- Connection between confidence interval and test.

- Mean corpuscular volume. (Sanders & Schmidt).
  - In a 1997 class project, J. Sauer discussed mean corpuscular volume (MCV) or the average volume of red blood cells (measured in femtoliters fl)
  - The mean MCV for a general population without known medical problems is 90 fl. It has been suspected that many chemotherapy drugs used to treat cancer cause elevated MCV levels.
  - MCV levels were recorded for a sample of 92 cancer patients at Los Palos Medical Center ($\bar{y} = 92.66, s = 7.66$). Do the data support the idea that chemotherapy drugs cause an elevated MCV?

- Systolic blood pressure measured on the same subject each day at 11:00 pm for two weeks (serial order): Summary statistics: $\bar{y} = 127.1, s = 6.53$. Construct 90% confidence interval for μ. Conditions and interpretation?
  123 140 132 138 125 116 127 123 129 132 123 122 125 124



- Sample size. Suppose I want to measure my current average late-evening systolic blood pressure . For how many days should I take measurements so that a 95% confidence interval has margin of error ≈ 2.0?

- The confidence interval has margin of error

$$ME = t^*_{n-1} \frac{s}{\sqrt{n}} \quad \text{so} \quad n = \left(\frac{t^*_{n-1} s}{ME} \right)^2 .$$

  Replace s by a prior estimate of σ; e.g., from a pilot study.
  Replace t^*_{n-1} by z^*.
  If resulting (rounded) value of n is large (say $n \geq 60$), then use this.
  Otherwise, use this n in t^*_{n-1}, then calculate new n.

- Degrees of freedom.

  1. Given observations $y_1, \ldots, y_n$, what value a minimizes $\sum (y_i - a)^2$?
     Answer: $a = \bar{y}$.
  2. Suppose $Y_1, \ldots, Y_n$ are independent $N(\mu, \sigma)$ random variables.
  3. Note that $E\left\{\sum (Y_i - \mu)^2\right\} = n\sigma^2$.
  4. Point 1 implies that $\sum (Y_i - \bar{Y})^2 \leq \sum (Y_i - \mu)^2$.
  5. So we must have $E\left\{\sum (Y_i - \bar{Y})^2\right\} \leq n\sigma^2$.
  6. In fact, we have $E\left\{\sum (Y_i - \bar{Y})^2\right\} = (n-1)\sigma^2$.
  7. Consequently, $E\left\{s^2\right\} = \sigma^2$; i.e., s^2 is an unbiased estimator of σ^2.

- Blair Witch Project (Weiss). Sholl et al., "The relation of sex and sense of direction to spatial orientation in an unfamiliar environment", J. of Environmental Psychology, 2000, 20:17-28.

  As part of this study, a group of students were each asked to rate their sense of direction as either good or poor. The students were then taken to a wooded area and asked to point south by moving a pointer attached to a $360°$ protractor.

  The following (sorted) values are the absolute errors in direction for the 15 female students who rated their sense of direction as good. ($\bar{x} = 55.4, s = 42.24$)

  8 12 14 18 20 27 31 36 68 69 78 91 109 122 128

  Is there evidence that these students are better at estimating direction than purely random guessing?

Hypotheses: Under random guessing, the population distribution of absolute errors is Uniform (0, 180). So under H_0 we have

$$\mu = 90 \quad \text{and} \quad \sigma = \frac{180}{\sqrt{12}} = 51.96 \ .$$

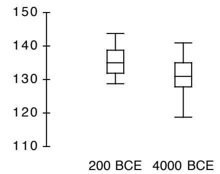
This is a (rare) example where both μ and σ are specified under H_0.

Can you carry out a test based on a sample of size $n = 1$; e.g. $\{10\}$?

24. Comparing means.

- Condition: Two independent samples or a completely randomized experiment to compare two treatments.
- Two-sample t-test for the difference between means.
- Two-sample t-interval for the difference between means.
- Pooled t methods (very brief).

- Egyptian skulls, Ex 24.16. Some archaeologists theorize that ancient Egyptians interbred with several different immigrant populations over thousands of years. To see if there is any indication of changes in body structure that might have resulted, they measured 30 skulls of male Egyptians dated from 4000 BCE and 30 others dated from 200 BCE. The variable considered here is maximum skull breadth (mm). Is there evidence of a difference in the population means for the two groups?
  4000 BCE, $n_1 = 30, \bar{y}_1 = 131.4, s_1^2 = 26.3, s_1 = 5.13$.
  200 BCE, $n_2 = 30, \bar{y}_2 = 135.6, s_2^2 = 16.3, s_2 = 4.04$.



- Hypotheses.
  Parameters $\mu_1, \sigma_1, \mu_2, \sigma_2$. Interpretation?
  $H_0 : \mu_1 - \mu_2 = \Delta_0$
  $H_A : \mu_1 - \mu_2 \neq \Delta_0$
  Here $\Delta_0 = 0$
- Model.
  - Conditions:
    - Independent observations in each sample
    - Independent groups
    - E.g., two random samples or a completely randomized experiment
    - Nearly Normal condition or large n for each sample
    - 10% condition
  - If $\Delta_0 = 0$, can we pool when calculating $SE(\bar{y}_1 - \bar{y}_2)$ like we did for $SE(\hat{p}_1 - \hat{p}_2)$?

– Test statistic and criterion: $t = \dfrac{(\bar{y}_1 - \bar{y}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

– Sampling distribution model under H_0 is approximated by Student's t distribution with a complicated formula for df (cf text, page 619):

$$\frac{1}{df} = \frac{c_1^2}{n_1 - 1} + \frac{c_2^2}{n_2 - 1} \quad \text{where} \quad c_1 = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{and} \quad c_2 = 1 - c_1 \ .$$

$df = n_1 - 1$ if $c_1 = 1$ and $df = n_2 - 1$ if $c_2 = 1$

$df = n_1 + n_2 - 2$ if $c_1 = \frac{n_1 - 1}{n_1 + n_2 - 2}$ i.e., if $\frac{s_1^2}{n_1(n_1 - 1)} = \frac{s_2^2}{n_2(n_2 - 1)}$

Lower and upper bounds for df:

$$\min\{n_1 - 1,\ n_2 - 1\} \ \leq \ df \ \leq \ n_1 + n_2 - 2$$

- Mechanics:

$$t = \frac{131.4 - 135.6}{1.19} = -3.52, \quad df \approx 55, \quad P\text{-value} \approx 0.001 \ .$$

– In the real world, use software for data analysis.
– For homework (given summary statistics), use technology such as the Data Desk t calculator available on STAT 141 website.
– For my exams, use the upper bound for df if $n_1 \approx n_2$ and $s_1 \approx s_2$; otherwise, use the lower bound.

- Conclusion: If the data sets were simple random samples, then we would have strong evidence that the population means differ. However, the samples were likely not random and so the observed difference in the means might be due to bias.

Observational study – cannot infer cause-and-effect. Lurking variables?

- Egyptial skulls example continued.

95% confidence interval for difference in means $\mu_1 - \mu_2$.

$$(\bar{y}_1 - \bar{y}_2) \pm ME$$

$$ME = t^*_{df} \times SE(\bar{y}_1 - \bar{y}_2) = t^*_{df}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

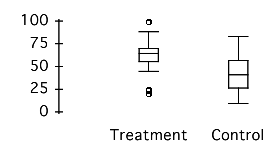
$t^*_{df} = 2.01$ with $df = 55$ calculated as above.

Evaluate:

- Sleeping beauty. (Pappenheimer & Karnovsky 1982)

Does a "sleep potion" work? Rabbits randomly assigned to treatment or control group; i.e., given compound containing potion or given a similar compound without potion. Response y = percentage of time asleep.

Group	n	$\bar{y}$	s
Treatment	21	63	18.5
Control	21	43	18.1



$\bar{y}_1 - \bar{y}_2 = 20.0$, $SE = 5.65$, $df = 40.0$, $t^* = 2.02$, CI: 20.0 ± 11.4
$T = 3.54$, P-value $= 0.0005$ (one-sided test)

Completely randomized design. Practical considerations. Conclusion. Check assumptions: box plots, histograms, normal quantile plots.

- 2001 Edmonton marathon.

  Summary stats for samples from three age-groups, y = finishing time in minutes.

	n	$\bar{y}$	s	$s/\sqrt{n}$
1. F20-29	22	260.4	33.9	7.2
2. F30-39	33	263.4	36.6	6.4
3. F40-49	24	262.7	34.6	7.1

  On average, are women in their 20's significantly faster than women in their 30's?

  Do mean finishing times for women in their 30's differ significantly from those in their 40's?

- Pooled t test and confidence interval.
  - Additional condition: $\sigma_1^2 = \sigma_2^2$.
  - In the standard error, replace s_1^2 and s_2^2 by the weighted average

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

  - Replace complicated df formula by $df = n_1 + n_2 - 2$.
  - Minor advantages: Exact theory, simpler formula, slightly higher df.
  - Major disadvantages: Less reliable when $\sigma_1^2 \neq \sigma_2^2$ and usually no reason to assume equality.
  - Exception: Randomized experiment and test of "no treatment effect".
  - Beware: Pooled t methods were once presented as the default by most authors and software. Now most recommend the two-sample t methods (i.e., not pooled). Sometimes the terminology distinguishing the two approaches is not clear.

25. Paired samples and blocks.

- Paired data.
- Randomized experiment to compare two treatments in blocks of size two.
- Paired t-test and confidence interval.

- Gripping strength of sinister people (Johnson & Bhattacharia).

  The following are measurements of the left-hand and right-hand gripping strengths of 10 left-handed individuals.

Left-hand	140	90	125	130	95	121	85	97	131	110
Right-hand	138	87	110	132	96	120	86	90	129	100
Difference	2	3	15	–2	–1	1	–1	7	2	10

  Do the data provide clear evidence that left-handed people have on average greater gripping strength in the left hand than in the right?

  Why is the two-sample t-test not appropriate here?

- Hypotheses.
  μ_d = mean difference (left minus right) for the statistical model.
  $H_0 : \mu_d = \Delta_0$
  $H_a : \mu_d > \Delta_0$
  Here $\Delta_0 = 0$

- Model.
  - Conditions:
    * Paired data, so analysis is based on differences.
    * Differences are independent.
    * Pairs obtained from a random sample or a randomized experiment.
    * 10% condition.
    * Distribution of differences is nearly Normal or n is sufficiently large.

  - Test statistic and criterion is same as one-sample t-test.

$$t = \frac{\bar{d} - \Delta_0}{SE(\bar{d})} \quad \text{where} \quad SE(\bar{d}) = \frac{s_d}{\sqrt{n}} .$$

  - Sampling distribution model under H_0 is Student's t with $df = n - 1$.
  - P-value is the right tail probability.

- Mechanics: $\bar{d} = 3.60, s_d = 5.46, n = 10, SE(\bar{d}) = 1.727, t = 2.085$, P-value = 0.033.

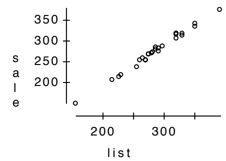
- Conclusion: Given the small P-value, I would reject the null hypothesis. Assuming the subjects were randomly sampled from a population of left-handed persons, there is moderate evidence that members of this population tend on average to have more gripping strength in the left hand compared with the right.

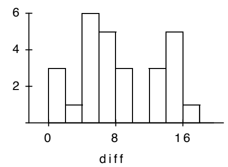
- Example. Belgravia real estate prices.

  The following data show the list and sale prices (thousands of dollars) for 27 houses in Belgravia (a neighbourhood just south of U of A) sold during 2003.

  Construct a 95% confidence interval for the mean difference between list and sale price. Interpretation?

list	sale	diff
155.0	151	4
214.9	207	7.9
226.0	214	12
229.5	220	9.5
255.0	238.5	16.5
259.9	253.9	6.0
264.9	260	4.9
269	255	14
269.9	255	14.9
274.9	270	4.9
275	270	5
279.8	272	7.8
279.9	273	6.9
284.9	280	4.9
285	285	0
289.9	283	6.9
289.9	275.5	14.4
296.5	287	9.5
318.8	318	0.8
319.5	319.5	0
319.5	307.5	12
319.9	316	3.9
329.5	315	14.5
329.5	320	9.5
349.7	344	5.7
349.9	335	14.9
390	377	13





Summaries
cases selected according to **sel27**

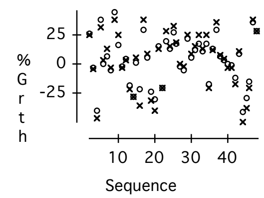
Variable	Count	Mean	StdDev
list	27	286.156	47.7643
sale	27	277.848	47.5080
diff	27	8.30741	4.84418

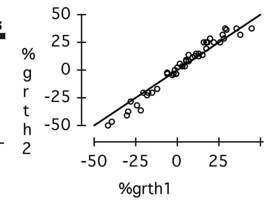
Comparison of two mutual funds: growth and volatility over 47 quarters ending Sept 30, 2009.

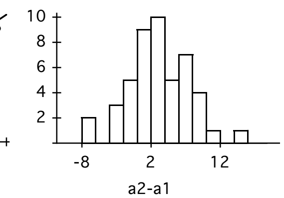
G = Growth over a quarter year
B = Balance at start of quarter
%grth = 4G/B
abs = absolute value of %grth

Summaries
cases selected according to **data**

Variable	Count	Mean	StdDev	StdErr
%grth1	47	3.36987	20.7371	3.02481
%grth2	47	3.30338	24.1300	3.51973
g2-g1	47	-0.066487	5.34192	0.779199
abs1	47	16.9091	12.2268	1.78346
abs2	47	20.0435	13.5253	1.97286
a2-a1	47	3.13436	4.29951	0.627147







- Blood pressure and heart rate.

  A subject monitored his BP and HR over a period of several weeks. On each occasion, the subject relaxed for about 5 minutes, then recorded Systolic, Diastolic, and HR measurements, then relaxed for another 5 minutes, then obtained a second set of measurements.

  One might expect that the additional rest would tend to lower BP and HR. Is there evidence that this is so? Results for one-sided tests listed below. Serial dependence is not a problem here – why?

Variable	n	$\bar{d}$	s_d	$s_d/\sqrt{n}$	t	P-value
Sys1 – Sys2	29	2.552	3.860	0.717	3.56	0.0007
Dia1 – Dia2	29	1.103	3.016	0.560	1.97	0.0294
HR1 – HR2	29	0.483	1.682	0.312	1.55	0.0662

- Seven farms (Ott). Yields (in bushels) for two new varieties of corn were compared. To deal with variability in yield from one farm to another, the investigators chose seven farms and two one-acre plots on each farm. The two varieties were randomly assigned to the two plots on each farm, the corn planted, and the crops harvested at maturity. Results: $\bar{d} = 4.429, s_d = 2.387$.

Variety A	48.2	44.6	49.7	40.5	54.6	47.1	51.4
Variety B	41.5	40.1	44.0	41.2	49.8	41.7	46.8
Difference	6.7	4.5	5.7	-0.7	4.8	5.4	4.6

- Suppose a study was carried out to compare gripping strength for left-handed and right-handed men. A sample of 10 men were selected from each group and a measurement of gripping strength was taken for the preferred hand; i.e., left for left-handers, right for right-handers. The sorted data from each sample are:

left	86	89	94	94	95	98	99	101	106	111
right	83	88	92	93	98	107	110	111	113	118
diff	3	1	2	1	–3	–9	–11	–10	–7	–7

  You wish to carry out a test to see if there is evidence that the means for the two populations differ. Which of the following represents the null distribution of your test statistic?
  A) t distribution with $df = 9$
  B) t distribution with $df = 15$
  C) standard normal distribution
  D) chi-square distribution with $df = 9$