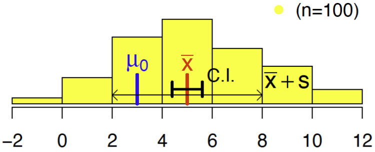
Surrounding the one-sample t-test

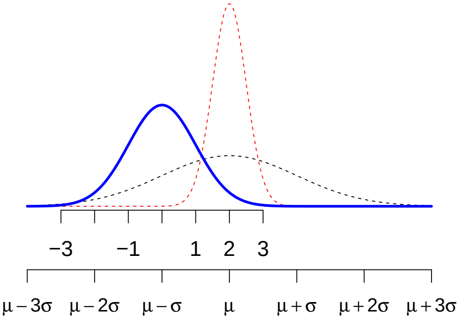


Slides were developed together with Dr. Michael Messer, former ASTAT member
Please send your comments to: T. Levajković, tijana.levajkovic@tuwien.ac.at

Reminder

- How is the mean distributed under normal distribution?
- Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables and $X_1 \sim N(\mu, \sigma^2)$
- For the mean it holds $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$
  - $\bar{X}$ is also normally distributed
  - $\bar{X}$ has expectation $\mu_{\bar{X}} = \mu$ (equal to the expectation of X_i)
  - $\bar{X}$ has standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ (decrease of factor $1/\sqrt{n}$)

    Interpretation: the typical deviation of the mean from its expectation is $\sigma/\sqrt{n}$
- Standardization: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$



From the z-test to the t-test

- How is the mean distributed under normal distribution?
- Let $X_1, \ldots, X_n$ be i.i.d. random variables and $X_1 \sim N(\mu, \sigma^2)$
- For the mean it holds $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$
  - $\bar{X}$ is also normally distributed
  - $\bar{X}$ has expectation $\mu_{\bar{X}} = \mu$ (equal to the expectation of X_i)
  - $\bar{X}$ has standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ (decrease of factor $1/\sqrt{n}$)
    Interpretation: the typical deviation of the mean from its expectation is $\sigma/\sqrt{n}$
- Point of view in statistics: μ and σ are unknown population parameters
- Concept in z-Test:
  - μ fixed via null hypothesis: $H_0 : \mu = \mu_0$ holds true
  - σ assumed as known. But unknown in practice (Problem!)
    Way out: *Estimate* σ through the empirical standard deviation S

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- Plug in estimator S

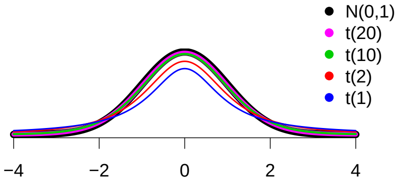
$$\boxed{\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)}$$

t-distributed with $n-1$ degrees of freedom. What is the t-distribution?

t-distribution

- Let $X_1, \ldots, X_n$ be i.d.d. random variables and $X_1 \sim N(\mu, \sigma^2)$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \text{vs} \qquad \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$



The $t(n)$-distribution ($n \in \{1, 2, \ldots\}$)

- Definition: $X \sim t(n) :\Leftrightarrow X$ has density $f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\cdot\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, x \in \mathbb{R}$
- $t(n)$ has a one parameter: degrees of freedom n, Γ denotes the Gamma function
- $t(n)$ symmetric around 0
- $t(n)$ has heavier tails (polynomial) than the $N(0,1)$-distribution (exponential)
  - Intuition: estimation of σ via S increases the variability of the rescaled mean
- the density of $t(n)$ converges pointwise to the density of $N(0,1)$
  - Intuition: the estimation of σ via S gets more precise (law of large numbers)
- in R via `rt(...,df=n)`, `dt(...,df=n)`, etc. with `df` denoting 'degrees of freedom'

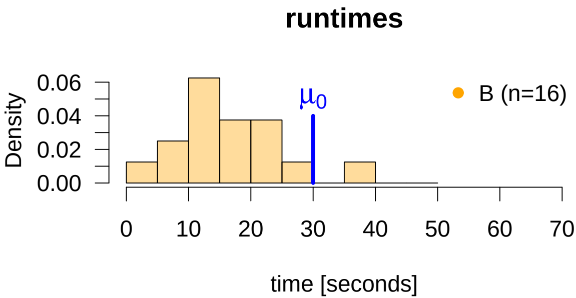
From what is said, it is (hopefully) plausible, but not rigorously proven, that $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.

For details see e.g., Messer, M. and Schneider, G. *Statistik: Theorie und Praxis im Dialog*, Springer Berlin

Example from last lecture

Reminder:

- Runtimes of an algorithm implemented by $n = 16$ students that took a certain programming class.



- Delicate assertion from a colleague of the lecturer: " The course was held by the lecturer a couple of times before. If all participants that have ever taken the course had implemented this algorithm, then the mean runtime would have been $\mu_0 = 30$

From z-test to the t-test

- Assertion about a huge 'population'
- the null hypothesis, that the mean runtime was $\mu_0 = 30$, was judged by the z-test (and it was rejected)
- now same procedure, but replace the z-statistic

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

  (σ assumed as known)
  through the t-statistic

$$\boxed{T := \frac{\bar{X} - \mu_0}{S / \sqrt{n}}}$$

  (unknown σ estimated via $S \to$ practically applicable)

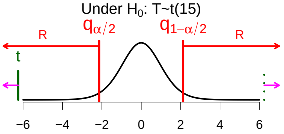
  We then speak about the (one-sample) t-test

- Note that T is a proper *statistic* in the sense that it is a bare function of the random variables / the data. Particularly, it does not depend on unknown parameters

The (one-sample) t-Test according to 'Student' (google: W.S. Gosset)

- *Set significance level*: Choose (e.g.,) $\alpha = 5\%$
- *Model assumption*: $X_1, \dots, X_n$ i.i.d. RVs, with $X_1 \sim N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$ ($n = 16$)
  (The data $x_1, \dots, x_n$ are assumed to be realizations of i.i.d. normal distributed RVs with unknown expectation μ and unknown variance σ^2)
- *Null hypothesis*: $H_0 : \mu = 30$
  Describes the assertion: the claimed expectation is $\mu_0 = 30$)
- *Test statistic* for the evaluation of the data (measures discrepancy). Now t-statistic

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \approx \frac{16.5 - 30}{8.7/4} \approx -6.2$$



- Distribution of the test statistic if H_0 holds true

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t(15)$$

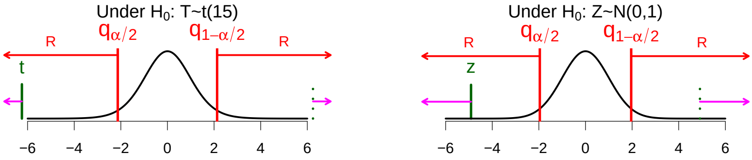
- *p-value:* quantifies discrepancy (judge t according to the distribution of T)

$$p = \mathbb{P}_{H_0}(|T| \geqslant |t|) \approx 1.5 \cdot 10^{-5}$$

  Probability to make an observation which is at least as extreme as in the data, if H_0 holds true (here: two-sided Test)
- *Decision:* Reject the null hypothesis, because $p \leqslant \alpha \Leftrightarrow t \in R$
  Say: the observed discrepancy was *significant* ($p < 10^{-4}$)
- *Interpretation*: If H_0 holds true, then something very unlikely was observed. In that sense, the data are hardly compatible with H_0.

t-test vs. z-test

- 'Result', interpretation etc. are in the t-test the 'same' as in the z-test. Differences are only in the the test statistics and in the distributions
- We take a closer look:
  - $s = 8.7$ (t-test) has <u>underestimated</u> $\sigma = 11$ (z-test)
  - Thus, $t \approx -6.2$ more extreme than $z \approx -4.9$ (as s and σ in the denominator)
  - However: p-value in t-test ($\approx 1.5 \cdot 10^{-5}$) larger than p in z-test ($\approx 9 \cdot 10^{-7}$)
  - No contradiction, because the tails of the t-distribution are heavier than those of $N(0,1)$ (visually not observable in the graphics)
  - For the same reason: in the t-test slightly smaller rejection area R: for $\alpha = 5\%$ it holds first regarding $t(15)$ that $q_{\alpha/2} = -2.13$, and second regarding $N(0,1)$ that $q_{\alpha/2} = -1.96$ (while $q_{\alpha/2}$ denotes the $\alpha/2$-quantiles)



$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \approx \frac{16.5 - 30}{8.7/4} \approx -6.2 \qquad z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \approx \frac{16.5 - 30}{11/4} \approx -4.9$$

The standard error of the mean

- Let $X_1, \ldots, X_n$ be i.i.d. RVs and $\sigma^2 = \mathbb{V}ar(X_1) \in (0, \infty)$
- Mean $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$
- Standard deviation of the mean

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- In the context of statistics: σ unknown population parameter
- Estimate σ via S

$$\boxed{\underline{\text{Definition:}} \qquad SEM := \frac{S}{\sqrt{n}}}$$

Standard Error of the Mean

- Meaning: the estimated variability of the mean
- The t-statistic measures the discrepancy $\bar{X} - \mu_0$ in the units SEM

$$\boxed{T = \frac{\bar{X} - \mu_0}{SEM}}$$

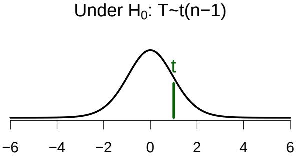
The standard error of the mean

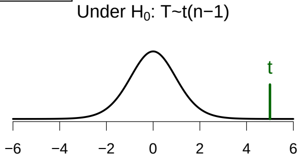
- Let $X_1, \dots, X_n$ be i.i.d. RVs and $\sigma^2 = \mathbb{V}ar(X_1) \in (0, \infty)$

$$\boxed{SEM = \frac{S}{\sqrt{n}}}$$

- The t-statistic measures the discrepancy $\bar{X} - \mu_0$ in the units SEM

$$\boxed{T = \frac{\bar{X} - \mu_0}{SEM}}$$



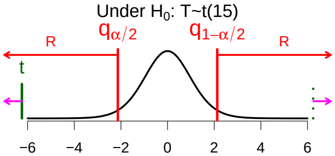


- Intuition: If $H_0 : \mu = \mu_0$ holds true
  - $|t| = 1 \Leftrightarrow |\bar{x} - \mu_0| = 1 \cdot sem$
    it is not unlikely to observe a discrepancy of <u>one</u> *sem*
  - $|t| = 5 \Leftrightarrow |\bar{x} - \mu_0| = 5 \cdot sem$
    it is very unlikely to observe a discrepancy of <u>five</u> *sem*
    'chance has a hard time realizing this'

From the standard error to the confidence interval

- Let $X_1, \ldots, X_n$ be i.i.d. RVs and $X_1 \sim N(\mu, \sigma^2)$

$$\boxed{T = \frac{\bar{X} - \mu_0}{SEM}} \qquad SEM = \frac{S}{\sqrt{n}}$$



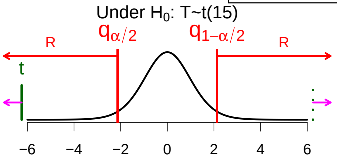
- In our example: $t = \frac{\bar{x} - \mu_0}{sem} \approx \frac{16.5 - 30}{8.7/4} \approx -6.2$

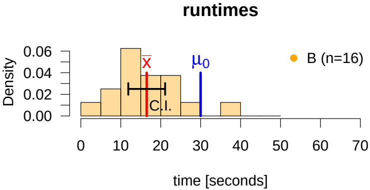
  it is very unlikely to observe a discrepancy $|\bar{x} - \mu_0|$ of $6.2 \cdot sem$ under H_0.

From the standard error to the confidence interval

- Let $X_1, \dots, X_n$ be i.i.d. RVs and $X_1 \sim N(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$

$$\boxed{T = \frac{\bar{X} - \mu_0}{SEM}} \qquad SEM = \frac{S}{\sqrt{n}}$$





$$(*) \qquad \alpha = \mathbb{P}_{H_0}(T \in R)$$
$$= \mathbb{P}_{H_0}(|\bar{X} - \mu_0| \geqslant q_{1-\alpha/2} \cdot SEM)$$
$$= \mathbb{P}_{H_0}((\bar{X} - q_{1-\alpha/2} \cdot SEM, \bar{X} + q_{1-\alpha/2} \cdot SEM) \not\ni \mu_0)$$

($q_{1-\alpha/2}$ is the $(1 - \alpha/2)$-quantile of $t(n-1)$)

- Meaning: Under H_0, the interval

$$I := (\bar{X} - q_{1-\alpha/2} \cdot SEM, \bar{X} + q_{1-\alpha/2} \cdot SEM)$$

overlaps the parameter μ_0 with probability $1 - \alpha$

- I is called a $(1 - \alpha)$-*confidence interval* for μ (abbreviate: C.I.)

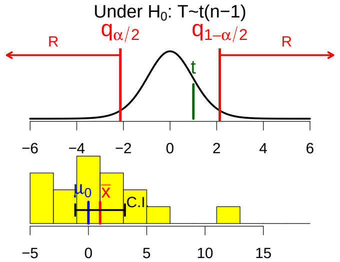
Confidence interval

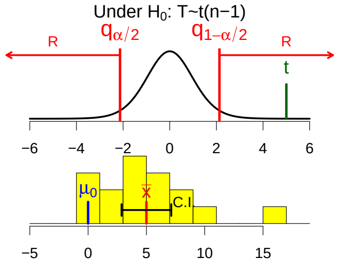
- Let $X_1, \ldots, X_n$ be i.i.d. RVs and $X_1 \sim N(\mu, \sigma^2)$,with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, and let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of $t(n-1)$

Under $H_0 : \mu = \mu_0$, the confidence interval

$$I := \left(\bar{X} - q_{1-\alpha/2} \cdot SEM,\ \bar{X} + q_{1-\alpha/2} \cdot SEM\right)$$

overlaps the parameter μ_0 with probability $1-\alpha$





- $t \in R \Leftrightarrow i \not\ni \mu_0$ (see $(*)$ in previous slide)

  Reject H_0 if and only if μ_0 is not overlapped by the confidence interval i
- Meaning: Equivalence of test and confidence interval ($\rightarrow$ 'Student's C.I.')

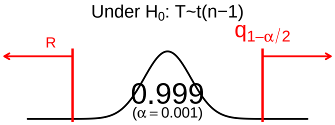
Confidence interval

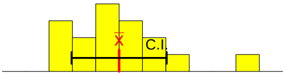
- Let $X_1, \ldots, X_n$ be i.i.d. RVs and $X_1 \sim N(\mu, \sigma^2)$,with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, and let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of $t(n-1)$

Under $H_0 : \mu = \mu_0$, the confidence interval

$$I := \left(\bar{X} - q_{1-\alpha/2} \cdot SEM,\ \bar{X} + q_{1-\alpha/2} \cdot SEM\right)$$

overlaps the parameter μ_0 with probability $1 - \alpha$

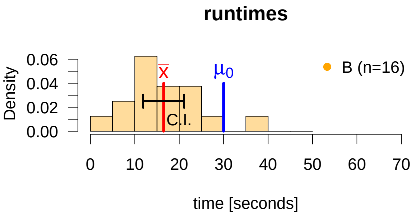




- Decrease of α
  $\leftrightarrow$ increase of $q_{1-\alpha/2}$
  $\leftrightarrow$ increase of the width of the confidence interval
  Plausible: 'If μ_0 shall be overlapped with large probability, then I must be large'

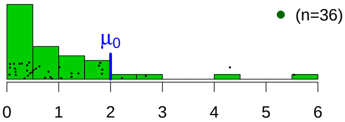
Interpretation

- Delicate assertion from a colleague of the lecturer: " The course was held by the lecturer a couple of times before. If all participants that have ever taken the course had implemented this algorithm, then the mean runtime would have been $\mu_0 = 30$'



- The 95% confidence interval (here $\alpha = 5\%$) does <u>not</u> overlap μ_0
- If the colleague is right, then the data describe an unlikely observation. In fact, one of the those observations that appear only in 5% of cases, if he is right.
- In that sense the data are barely compatible with the assertion
- Intuitively: $q_{1-\alpha/2} \approx 2.13 \rightarrow$ the c.i. has a diameter of about 4*sem* $\rightarrow$ the distance $|\bar{x} - \mu_0|$ is about 6*sem*. That is a large distance if we take in account that the typical deviation of $\bar{X}$ from μ_0 is about 1*SEM*

Data not bell-shaped

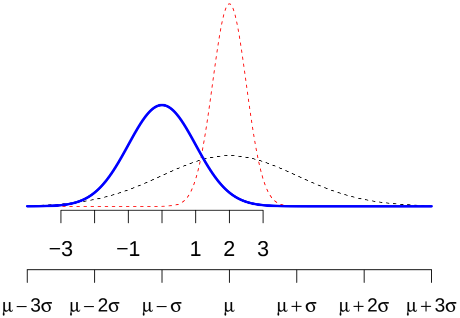


- Here: $n = 36$ data
- $H_0 : \mu = \mu_0$
- but distribution of the data is asymmetric
- Particularly, not approximately bell-shaped
- Previous model assumption, that the data are sampled from the normal distribution is not appropriate
- Way out: asymptotic normality of the mean

Reminder

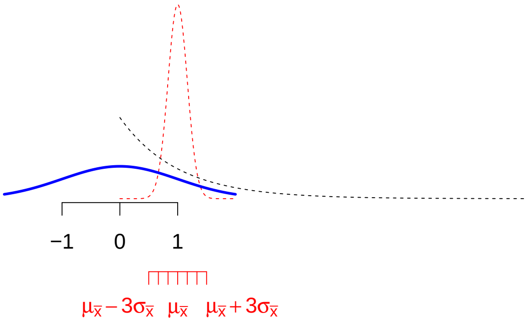
- How is the mean distributed under normal distribution?
- Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables and $X_1 \sim N(\mu, \sigma^2)$
- For the mean it holds $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$
  - $\bar{X}$ is also normally distributed
  - $\bar{X}$ has expectation $\mu_{\bar{X}} = \mu$ (equal to the expectation of X_i)
  - $\bar{X}$ has standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ (decrease of factor $1/\sqrt{n}$)

    Interpretation: the typical deviation of the mean from its expectation is $\sigma/\sqrt{n}$
- Standardization: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$



Asymptotic normality of the mean

- How is the mean distributed for a large sample size n?
- Let $X_1, \ldots, X_n$ i.i.d. random variables and $\mathbb{V}ar(X_1) \in (0, \infty)$ (not necessarily normally distributed!)
- Then $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \overset{d}{\approx} N(\mu, \sigma^2/n)$ approximately
    - $\bar{X}$ is approximately normally distributed
    - $\bar{X}$ has expectation $\mu_{\bar{X}} = \mu$
    - $\bar{X}$ has standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$
- Standardization: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$ for $n \to \infty$ (central limit theorem)



Asymptotic one-sample test and confidence interval

- Let $X_1, \ldots, X_n$ be i.i.d. RVs with $\mu \in \mathbb{R}$ and $\mathbb{V}ar(X_1) \in (0, \infty)$, and let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of $N(0,1)$

Under $H_0 : \mu = \mu_0$ it approximatively holds for large n

$$T = \frac{\bar{X} - \mu_0}{SEM} \overset{d}{\approx} N(0,1)$$

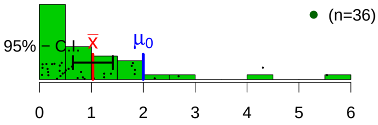
and equivalently: the confidence interval

$$I := \left(\bar{X} - q_{1-\alpha/2} \cdot SEM,\ \bar{X} + q_{1-\alpha/2} \cdot SEM\right)$$

overlaps the parameter μ_0 with probability about $1 - \alpha$

- i.e.,: test and confidence interval are constructed according to 'Student'. Only difference: Use the quantiles of $N(0,1)$, instead of $t(n-1)$

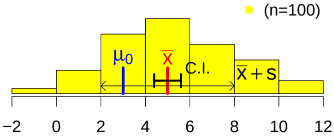
Approximate procedure



- Here: $n = 36$ data
- $H_0 : \mu = \mu_0$
- Approximate 95%-confidence interval does not overlap μ_0
  $\rightarrow$ reject H_0 on the 5% level

Question

Can $H_0 : \mu = \mu_0 = 3$ be rejected on the 5%-level?



Here t-test naively:

- $\bar{x} \approx 5$ (balance in equilibrium)
- $s \approx 3$ (bell-shaped distribution, 2/3 of the data captured)
- $sem \approx 3/10$ (100 data points)
- $q \approx 2$ (n large, 97.5%-quantile of $N(0,1)$ is ≈ 1.96)
- Together: C.I. $i \approx (4.4, 5.6)$ ($\bar{x} \pm q \cdot sem$)
- Hence: Reject H_0 (C.I. does not overlap μ_0)

Of course, this is not 'precise', but the important message is that $\bar{x}$ is further than $6sem$ away from μ_0. This is extremely far! More precise estimates would have yielded the same massage!

t-test in R

```
# Enter data
x <- c(...)
# perform t-test
t.test(x,mu=3,...)
# Output

        One Sample t-test

data:  x
t = 8.3512, df = 99, p-value = 4.22e-13
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 4.745578 5.833572
sample estimates:
mean of x
5.289575
```

- Our naive estimates were plausible
  - Mean and C.I. fit well
  - $\bar{x}$ is even further than 8*sem* apart from μ_0

Multiple-choice questions

(1) In the latest issue of Consumer Reports, some data on the calorie content of beef hot dogs is given. The sample mean of the numbers of calories in 20 different hot dog brands was 156 and the sample standard deviation 23. Assume that the calorie content of beef hot dogs follow a normal distribution (μ, σ^2), with both μ and σ unknown. Obtain a 90% confidence interval for the mean number of calories μ.

a. (147.11, 164.89)
b. (146.90, 165.10)
c. (153.91, 158.09)
d. (149.17, 162.83)

Multiple-choice questions

(2) In the situation of a two-sided one-sample t-test we find $\bar{x} = 10$, $s^2 = 36$ and $n = 9$. For a given significance level we find the rejection region $R = (-\infty, -2.2] \cup [2.2, \infty)$. Then for the null hypothesis $H_0 : \mu = 5$ it holds

a. we reject H_0, and we would also reject for any smaller significance level
b. we reject H_0, and we would also reject for any larger significance level
c. we do not reject H_0, but we would reject if only the significance level was chosen large enough
d. we do not reject H_0, but we would reject if only the significance level was chosen small enough.

Multiple-choice questions

(3) A fast food chain advertises that their large bag of french fries has a weight of 150 grams. Some high school students, who enjoy french fries at every lunch, suspect that they are getting less than the advertised amount. With a scale borrowed from their physics teacher, they weigh a random sample of 15 bags. Assuming the level of significance $\alpha = 10\%$, what would be the conclusion if the sample mean is 145.8 g and standard deviation is 12.81 g? (Assume that all conditions for inference are met.) The values from the table of t-distribution should be used.

		cum. prob.					
		$t_{0.75}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
	14	0.692	1.345	1.761	2.145	2.624	2.977
df	15	0.691	1.341	1.753	2.145	2.602	2.947
	16	0.691	1.337	1.746	2.145	2.583	2.921

t-Table

a. There is sufficient evidence to prove the fast food chain advertisement is true.
b. There is sufficient evidence to prove the fast food chain advertisement is false.
c. The students have sufficient evidence to reject the fast food chain's claim.
d. The students do not have sufficient evidence to reject the fast food chain's claim.

Multiple-choice questions

(4) A 90% confidence interval for a population mean based on a sample of size 500 was $(35, 38)$. Which of the following is the best interpretation of the interval?

a. Across many samples, 90% of sample means should lie within an interval made by this method.
b. 90% of data points should lie within this interval.
c. Across many samples, 90% of intervals created using this method would capture the true population mean.
d. There is a 90% probability that the true value of the population mean is in (35, 38).

Multiple-choice questions

(5) A company selling home appliances claims that the accompanying instruction guides are written at a 6th grade reading level. An English teacher believes that the true figure is higher and with the help of an AP Statistics student runs a hypothesis test. The student randomly picks one page from each of 25 of the company's instruction guides, and the teacher subjects the pages to a standard readability test. The reading levels of the 25 pages are given in the following table:

Reading grade level	5	6	7	8	9	10
Number of pages	6	10	4	2	2	1

Assuming that the conditions for inference are met, is there statistical evidence to support the English teacher's belief?

a. No, because the p-value is greater than 0.10.
b. Yes, the p-value is between 0.05 and 0.10 indicating some evidence for the teacher's belief.
c. Yes, the p-value is between 0.01 and 0.05 indicating evidence for the teacher's belief.
d. Yes, the p-value is between 0.001 and 0.01 indicating strong evidence for the teacher's belief.

Thank you for your attention!