ISE 315: Engineering Statistics

Lecture 14: Major Exam 1 Practice Problems

Instructor: Mansur M. Arief, PhD
Industrial and Systems Engineering, KFUPM

Office: 22-219 — Email: mansur.arief@kfupm.edu.sa

Based on Montgomery & Runger, Applied Statistics and Probability for Engineers, 6th Ed.

Lecture 14

Six Fully Worked Practice Problems + Concept Checks (Chapters 3–5)

How to Use This Session

Attack plan: Procedure 5.3; master table: Table 5.6

- **Six full problems**, one from each core problem type:
  - Paired t / proportion CI + sample size / Z-test with β and power
  - proportion test / pooled t / variance CI + χ^2-test

How to Use This Session

Attack plan: Procedure 5.3; master table: Table 5.6

- **Six full problems**, one from each core problem type:
  - Paired t / proportion CI + sample size / Z-test with β and power
  - proportion test / pooled t / variance CI + χ^2-test
- **Pause at each problem statement.** Attempt it fully before watching the solution.

How to Use This Session

Attack plan: Procedure 5.3; master table: Table 5.6

- **Six full problems**, one from each core problem type:
  - Paired t / proportion CI + sample size / Z-test with β and power
  - proportion test / pooled t / variance CI + χ^2-test
- **Pause at each problem statement.** Attempt it fully before watching the solution.
- Every solution follows Procedure 5.3: parameter $\to$ conditions $\to$ master-table row $\to$ hypotheses $\to$ compute $\to$ conclude in context.

How to Use This Session

Attack plan: Procedure 5.3; master table: Table 5.6

- **Six full problems**, one from each core problem type:
  - Paired t / proportion CI + sample size / Z-test with β and power
  - proportion test / pooled t / variance CI + χ^2-test
- **Pause at each problem statement.** Attempt it fully before watching the solution.
- Every solution follows Procedure 5.3: parameter $\to$ conditions $\to$ master-table row $\to$ hypotheses $\to$ compute $\to$ conclude in context.
- **Ten concept checks** at the end — one-line answers, common traps.

Problem 1: Guardrail Fine-Tuning (Before/After)

Example 5.7 in the textbook

An AI safety team tests whether fine-tuning improves the accuracy (% of harmful prompts correctly blocked) of a guardrail model. Eight evaluation suites are scored before and after fine-tuning:

Suite	1	2	3	4	5	6	7	8
Before (x_{1j})	85	78	92	88	76	95	82	90
After (x_{2j})	88	82	90	94	80	93	87	94

(a) Compute the differences $d_j = \text{After} - \text{Before}$ and find $\bar{d}$ and s_D.

(b) Test at $\alpha = 0.05$ whether fine-tuning **improved** the mean accuracy. Follow Procedure 4.2.

(c) Bound the P-value from the t-table (Procedure 4.3).

(d) Construct a **95% CI** for μ_D via (5.9). Consistent with (b)?

Problem 1 Solution (a): Differences

Same suites, two conditions $\Rightarrow$ **paired**, *(5.8)*

Problem 1 Solution (a): Differences

Same suites, two conditions $\Rightarrow$ **paired**, *(5.8)*

Suite	1	2	3	4	5	6	7	8	Sum
d_j	3	4	-2	6	4	-2	5	4	22

Problem 1 Solution (a): Differences

Same suites, two conditions $\Rightarrow$ **paired**, *(5.8)*

Suite	1	2	3	4	5	6	7	8	Sum
d_j	3	4	-2	6	4	-2	5	4	22

$$\bar{d} = \frac{22}{8} = 2.75$$

Problem 1 Solution (a): Differences

Same suites, two conditions $\Rightarrow$ **paired**, *(5.8)*

Suite	1	2	3	4	5	6	7	8	Sum
d_j	3	4	-2	6	4	-2	5	4	22

$$\bar{d} = \frac{22}{8} = 2.75$$

$$\sum d_j^2 = 9 + 16 + 4 + 36 + 16 + 4 + 25 + 16 = 126$$

Problem 1 Solution (a): Differences

Same suites, two conditions $\Rightarrow$ **paired**, *(5.8)*

Suite	1	2	3	4	5	6	7	8	Sum
d_j	3	4	-2	6	4	-2	5	4	22

$$\bar{d} = \frac{22}{8} = 2.75$$

$$\sum d_j^2 = 9 + 16 + 4 + 36 + 16 + 4 + 25 + 16 = 126$$

$$s_D^2 = \frac{\sum d_j^2 - n\bar{d}^2}{n-1} = \frac{126 - 8(2.75)^2}{7} = \frac{126 - 60.5}{7} = \frac{65.5}{7} = 9.357$$

Problem 1 Solution (a): Differences

Same suites, two conditions $\Rightarrow$ **paired**, *(5.8)*

Suite	1	2	3	4	5	6	7	8	Sum
d_j	3	4	-2	6	4	-2	5	4	22

$$\bar{d} = \frac{22}{8} = 2.75$$

$$\sum d_j^2 = 9 + 16 + 4 + 36 + 16 + 4 + 25 + 16 = 126$$

$$s_D^2 = \frac{\sum d_j^2 - n\bar{d}^2}{n-1} = \frac{126 - 8(2.75)^2}{7} = \frac{126 - 60.5}{7} = \frac{65.5}{7} = 9.357$$

$$s_D = \sqrt{9.357} = 3.059$$

Problem 1 Solution (b): Paired t-Test

Steps 1–3: Parameter: μ_D (mean improvement). Paired data.
$H_0 : \mu_D = 0 \quad \text{vs} \quad H_1 : \mu_D > 0$ (right-tailed: "improved"), $\quad \alpha = 0.05$

Problem 1 Solution (b): Paired t-Test

Steps 1–3: Parameter: μ_D (mean improvement). Paired data.
$H_0 : \mu_D = 0 \quad \text{vs} \quad H_1 : \mu_D > 0$ (right-tailed: "improved"), $\quad \alpha = 0.05$

Steps 4–5: Paired t (5.8), df $= n - 1 = 7$.
Reject if $T_0 > t_{0.05,7} = 1.895$.

Problem 1 Solution (b): Paired t-Test

Steps 1–3: Parameter: μ_D (mean improvement). Paired data.
$H_0 : \mu_D = 0 \quad \text{vs} \quad H_1 : \mu_D > 0$ (right-tailed: "improved"), $\quad \alpha = 0.05$

Steps 4–5: Paired t (5.8), df $= n - 1 = 7$.
Reject if $T_0 > t_{0.05,7} = 1.895$.

Step 6: $T_0 = \dfrac{\bar{d} - 0}{s_D/\sqrt{n}} = \dfrac{2.75}{3.059/\sqrt{8}}$

Problem 1 Solution (b): Paired t-Test

Steps 1–3: Parameter: μ_D (mean improvement). Paired data.
$H_0 : \mu_D = 0 \quad \text{vs} \quad H_1 : \mu_D > 0$ (right-tailed: "improved"), $\quad \alpha = 0.05$

Steps 4–5: Paired t (5.8), df $= n - 1 = 7$.
Reject if $T_0 > t_{0.05,7} = 1.895$.

Step 6: $T_0 = \dfrac{\bar{d} - 0}{s_D/\sqrt{n}} = \dfrac{2.75}{3.059/\sqrt{8}} = \dfrac{2.75}{1.082} = 2.543$

Problem 1 Solution (b): Paired t-Test

Steps 1–3: Parameter: μ_D (mean improvement). Paired data.
$H_0 : \mu_D = 0$ vs $H_1 : \mu_D > 0$ (right-tailed: "improved"), $\alpha = 0.05$

Steps 4–5: Paired t (5.8), $\text{df} = n - 1 = 7$.
Reject if $T_0 > t_{0.05,\,7} = 1.895$.

Step 6: $T_0 = \dfrac{\bar{d} - 0}{s_D/\sqrt{n}} = \dfrac{2.75}{3.059/\sqrt{8}} = \dfrac{2.75}{1.082} = 2.543$

Step 7: $2.543 > 1.895 \Rightarrow$ **Reject H_0**.

Sufficient evidence at the 5% level that fine-tuning improved the guardrail's mean blocking accuracy.

Problem 1 Solution (c) P-value Bounds, (d) 95% CI

(c) From Table B.3, row $\nu = 7$: $t_{0.025,\,7} = 2.365$ and $t_{0.01,\,7} = 2.998$.

Problem 1 Solution (c) P-value Bounds, (d) 95% CI

(c) From Table B.3, row $\nu = 7$: $t_{0.025,7} = 2.365$ and $t_{0.01,7} = 2.998$.

Since $2.365 < T_0 = 2.543 < 2.998$ (one-tailed test):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

Problem 1 Solution (c) P-value Bounds, (d) 95% CI

(c) From Table B.3, row $\nu = 7$: $t_{0.025,7} = 2.365$ and $t_{0.01,7} = 2.998$.

Since $2.365 < T_0 = 2.543 < 2.998$ (one-tailed test):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

(d) 95% CI for μ_D via (5.9), $t_{0.025,7} = 2.365$:

$$\bar{d} \pm t_{0.025,7} \frac{s_D}{\sqrt{n}}$$

Problem 1 Solution (c) P-value Bounds, (d) 95% CI

(c) From Table B.3, row $\nu = 7$: $t_{0.025,7} = 2.365$ and $t_{0.01,7} = 2.998$.

Since $2.365 < T_0 = 2.543 < 2.998$ (one-tailed test):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

(d) 95% CI for μ_D via (5.9), $t_{0.025,7} = 2.365$:

$$\bar{d} \pm t_{0.025,7}\,\frac{s_D}{\sqrt{n}} = 2.75 \pm 2.365\,(1.082) = 2.75 \pm 2.559$$

Problem 1 Solution (c) P-value Bounds, (d) 95% CI

(c) From Table B.3, row $\nu = 7$: $t_{0.025,\,7} = 2.365$ and $t_{0.01,\,7} = 2.998$.

Since $2.365 < T_0 = 2.543 < 2.998$ (one-tailed test):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

(d) 95% CI for μ_D via (5.9), $t_{0.025,\,7} = 2.365$:

$$\bar{d} \pm t_{0.025,\,7}\,\frac{s_D}{\sqrt{n}} = 2.75 \pm 2.365\,(1.082) = 2.75 \pm 2.559$$

$$\boxed{(0.191,\ 5.309)\ \text{percentage points}}$$

Problem 1 Solution (c) P-value Bounds, (d) 95% CI

(c) From Table B.3, row $\nu = 7$: $t_{0.025,7} = 2.365$ and $t_{0.01,7} = 2.998$.

Since $2.365 < T_0 = 2.543 < 2.998$ (one-tailed test):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

(d) 95% CI for μ_D via (5.9), $t_{0.025,7} = 2.365$:

$$\bar{d} \pm t_{0.025,7}\,\frac{s_D}{\sqrt{n}} = 2.75 \pm 2.365\,(1.082) = 2.75 \pm 2.559$$

$$\boxed{(0.191,\ 5.309)\ \text{percentage points}}$$

Problem 2: Late Shipments at a Distribution Hub

A logistics manager audits a distribution hub. In a random sample of $n = 250$ outbound shipments, $x = 40$ arrived **late** to the customer.

(a) Check the large-sample conditions and construct a **95% CI** for the true late-delivery proportion p via (3.7).

(b) Management wants a follow-up audit with margin of error at most 0.04 at 95% confidence. Using part (a) as a prior, find the minimum sample size (3.8).

(c) With **no** prior estimate, what sample size is required for $E = 0.04$? Use (3.9).

(d) Why is the answer in (c) larger than in (b)?

Problem 2 Solution (a): 95% CI for p

$$\hat{p} = \frac{40}{250} = 0.16$$

Problem 2 Solution (a): 95% CI for p

$$\hat{p} = \frac{40}{250} = 0.16$$

Large-sample check: $n\hat{p} = 40 \geq 5 \;\checkmark$ and $n(1-\hat{p}) = 210 \geq 5 \;\checkmark$

Problem 2 Solution (a): 95% CI for p

$$\hat{p} = \frac{40}{250} = 0.16$$

Large-sample check: $n\hat{p} = 40 \geq 5$ ✓ and $n(1-\hat{p}) = 210 \geq 5$ ✓

Margin of error with $z_{0.025} = 1.96$:

$$E = 1.96\sqrt{\frac{0.16\,(0.84)}{250}}$$

Problem 2 Solution (a): 95% CI for p

$$\hat{p} = \frac{40}{250} = 0.16$$

Large-sample check: $n\hat{p} = 40 \geq 5 \ \checkmark$ and $n(1 - \hat{p}) = 210 \geq 5 \ \checkmark$

Margin of error with $z_{0.025} = 1.96$:

$$E = 1.96\sqrt{\frac{0.16\,(0.84)}{250}} = 1.96\,(0.02319) = 0.0454$$

Problem 2 Solution (a): 95% CI for p

$$\hat{p} = \frac{40}{250} = 0.16$$

Large-sample check: $n\hat{p} = 40 \geq 5 \ \checkmark$ and $n(1-\hat{p}) = 210 \geq 5 \ \checkmark$

Margin of error with $z_{0.025} = 1.96$:

$$E = 1.96\sqrt{\frac{0.16\,(0.84)}{250}} = 1.96\,(0.02319) = 0.0454$$

95% CI (3.7):

$$\boxed{0.16 \pm 0.0454 = (0.1146,\ 0.2054)}$$

Problem 2 Solution (a): 95% CI for p

$\hat{p} = \dfrac{40}{250} = 0.16$

Large-sample check: $n\hat{p} = 40 \geq 5 \ \checkmark$ and $n(1-\hat{p}) = 210 \geq 5 \ \checkmark$

Margin of error with $z_{0.025} = 1.96$:

$$E = 1.96\sqrt{\frac{0.16\,(0.84)}{250}} = 1.96\,(0.02319) = 0.0454$$

95% CI (3.7):

$$\boxed{0.16 \pm 0.0454 = (0.1146,\ 0.2054)}$$

We are 95% confident that between about 11% and 21% of shipments arrive late.

Problem 2 Solution (b)–(d): Sample Sizes

(b) With prior $\hat{p} = 0.16$ (3.8):

$$n = \left\lceil \frac{z_{0.025}^{2}\,\hat{p}(1-\hat{p})}{E^2} \right\rceil = \left\lceil \frac{(1.96)^2(0.16)(0.84)}{(0.04)^2} \right\rceil$$

Problem 2 Solution (b)–(d): Sample Sizes

(b) With prior $\hat{p} = 0.16$ (3.8):

$$n = \left\lceil \frac{z_{0.025}^2\,\hat{p}(1-\hat{p})}{E^2} \right\rceil = \left\lceil \frac{(1.96)^2(0.16)(0.84)}{(0.04)^2} \right\rceil = \left\lceil \frac{0.5163}{0.0016} \right\rceil = \lceil 322.7 \rceil = \boxed{323}$$

Problem 2 Solution (b)–(d): Sample Sizes

(b) With prior $\hat{p} = 0.16$ (3.8):

$$n = \left\lceil \frac{z_{0.025}^2\,\hat{p}(1-\hat{p})}{E^2} \right\rceil = \left\lceil \frac{(1.96)^2(0.16)(0.84)}{(0.04)^2} \right\rceil = \left\lceil \frac{0.5163}{0.0016} \right\rceil = \lceil 322.7 \rceil = \boxed{323}$$

(c) Without a prior (3.9):

$$n = \left\lceil \frac{(1.96)^2(0.25)}{(0.04)^2} \right\rceil$$

Problem 2 Solution (b)–(d): Sample Sizes

(b) With prior $\hat{p} = 0.16$ (3.8):

$$n = \left\lceil \frac{z_{0.025}^2\, \hat{p}(1-\hat{p})}{E^2} \right\rceil = \left\lceil \frac{(1.96)^2(0.16)(0.84)}{(0.04)^2} \right\rceil = \left\lceil \frac{0.5163}{0.0016} \right\rceil = \lceil 322.7 \rceil = \boxed{323}$$

(c) Without a prior (3.9):

$$n = \left\lceil \frac{(1.96)^2(0.25)}{(0.04)^2} \right\rceil = \left\lceil \frac{0.9604}{0.0016} \right\rceil = \lceil 600.25 \rceil = \boxed{601}$$

Problem 2 Solution (b)–(d): Sample Sizes

(b) With prior $\hat{p} = 0.16$ (3.8):

$$n = \left\lceil \frac{z_{0.025}^2\, \hat{p}(1-\hat{p})}{E^2} \right\rceil = \left\lceil \frac{(1.96)^2(0.16)(0.84)}{(0.04)^2} \right\rceil = \left\lceil \frac{0.5163}{0.0016} \right\rceil = \lceil 322.7 \rceil = \boxed{323}$$

(c) Without a prior (3.9):

$$n = \left\lceil \frac{(1.96)^2(0.25)}{(0.04)^2} \right\rceil = \left\lceil \frac{0.9604}{0.0016} \right\rceil = \lceil 600.25 \rceil = \boxed{601}$$

(d) The conservative formula sets $\hat{p} = 0.5$, which **maximizes** $\hat{p}(1-\hat{p})$ at 0.25. Our prior gives $0.16 \times 0.84 = 0.1344 < 0.25$, so (c) plans for the worst-case variability and demands a larger sample.

Problem 3: Battery Runtime of Inventory-Scanning Drones

Example 5.8 in the textbook

A warehouse operator deploys a new thermal management system intended to **increase** the mean battery runtime μ of its inventory-scanning drones beyond the historical average of $\mu_0 = 100$ minutes. Runtime is normal with known variance $\sigma^2 = 144$ min^2.

(a) With $n = 36$ drones you find $\bar{x} = 104$ min. Test at $\alpha = 0.05$ whether the system increased μ (use (4.3) and Procedure 4.2).

(b) Give a practical interpretation of the result.

(c) If the true mean is now $\mu = 106$ min, find the Type II error probability β (Procedure 4.4, (4.9)–(4.10)).

(d) Find the sample size n to detect $\mu = 106$ with power 0.90 at $\alpha = 0.05$ (one-sided version of (4.11)).

Problem 3 Solution (a)–(b): One-Sided Z-Test

Steps 1–3: Parameter: μ; $\sigma = 12$ known $\Rightarrow$ Z-test (4.3).

$H_0 : \mu = 100 \quad \text{vs} \quad H_1 : \mu > 100$ (right-tailed), $\quad \alpha = 0.05$

Problem 3 Solution (a)–(b): One-Sided Z-Test

Steps 1–3: Parameter: μ; $\sigma = 12$ known $\Rightarrow$ Z-test (4.3).

$$H_0 : \mu = 100 \quad \text{vs} \quad H_1 : \mu > 100 \text{ (right-tailed)}, \quad \alpha = 0.05$$

Steps 4–5: Reject if $Z_0 > z_{0.05} = 1.645$.

Problem 3 Solution (a)–(b): One-Sided Z-Test

Steps 1–3: Parameter: μ; $\sigma = 12$ known $\Rightarrow$ Z-test (4.3).

$$H_0 : \mu = 100 \quad \text{vs} \quad H_1 : \mu > 100 \text{ (right-tailed)}, \quad \alpha = 0.05$$

Steps 4–5: Reject if $Z_0 > z_{0.05} = 1.645$.

Step 6: $Z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{104 - 100}{12/\sqrt{36}}$

Problem 3 Solution (a)–(b): One-Sided Z-Test

Steps 1–3: Parameter: μ; $\sigma = 12$ known $\Rightarrow$ Z-test (4.3).
$H_0 : \mu = 100 \quad \text{vs} \quad H_1 : \mu > 100$ (right-tailed), $\quad \alpha = 0.05$

Steps 4–5: Reject if $Z_0 > z_{0.05} = 1.645$.

Step 6: $Z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{104 - 100}{12/\sqrt{36}} = \dfrac{4}{2} = 2.00$

Problem 3 Solution (a)–(b): One-Sided Z-Test

Steps 1–3: Parameter: μ; $\sigma = 12$ known $\Rightarrow$ Z-test (4.3).
$H_0 : \mu = 100 \quad \text{vs} \quad H_1 : \mu > 100$ (right-tailed), $\quad \alpha = 0.05$

Steps 4–5: Reject if $Z_0 > z_{0.05} = 1.645$.

Step 6: $Z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{104 - 100}{12/\sqrt{36}} = \dfrac{4}{2} = 2.00$

P-value (4.5): $1 - \Phi(2.00) = 1 - 0.9772 = 0.0228 < 0.05$

Problem 3 Solution (a)–(b): One-Sided Z-Test

Steps 1–3: Parameter: μ; $\sigma = 12$ known $\Rightarrow$ Z-test (4.3).
$H_0 : \mu = 100 \quad \text{vs} \quad H_1 : \mu > 100$ (right-tailed), $\quad \alpha = 0.05$

Steps 4–5: Reject if $Z_0 > z_{0.05} = 1.645$.

Step 6: $Z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{104 - 100}{12/\sqrt{36}} = \dfrac{4}{2} = 2.00$

P-value (4.5): $1 - \Phi(2.00) = 1 - 0.9772 = 0.0228 < 0.05$

Step 7: $2.00 > 1.645 \Rightarrow$ **Reject** H_0.

Problem 3 Solution (a)–(b): One-Sided Z-Test

Steps 1–3: Parameter: μ; $\sigma = 12$ known $\Rightarrow$ Z-test (4.3).
$H_0 : \mu = 100 \quad$ vs $\quad H_1 : \mu > 100$ (right-tailed), $\quad \alpha = 0.05$

Steps 4–5: Reject if $Z_0 > z_{0.05} = 1.645$.

Step 6: $Z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \dfrac{104 - 100}{12/\sqrt{36}} = \dfrac{4}{2} = 2.00$

P-value (4.5): $1 - \Phi(2.00) = 1 - 0.9772 = 0.0228 < 0.05$

Step 7: $2.00 > 1.645 \Rightarrow$ **Reject** H_0.

(b) At the 5% level, the data supports that the thermal system extended mean runtime beyond 100 minutes. Whether 4 extra minutes justifies the retrofit cost is a separate business decision.

Problem 3 Solution (c): β at $\mu = 106$

Procedure 4.4

Step 1: Fail-to-reject region in $\bar{x}$ units (one-sided form of (4.9)):

$$\bar{x}_c = \mu_0 + z_{0.05} \frac{\sigma}{\sqrt{n}}$$

Problem 3 Solution (c): β at $\mu = 106$

Procedure 4.4

Step 1: Fail-to-reject region in $\bar{x}$ units (one-sided form of (4.9)):

$$\bar{x}_c = \mu_0 + z_{0.05}\,\frac{\sigma}{\sqrt{n}} = 100 + 1.645\,(2) = 103.29$$

Problem 3 Solution (c): β at $\mu = 106$

Procedure 4.4

Step 1: Fail-to-reject region in $\bar{x}$ units (one-sided form of (4.9)):

$$\bar{x}_c = \mu_0 + z_{0.05}\,\frac{\sigma}{\sqrt{n}} = 100 + 1.645\,(2) = 103.29$$

We fail to reject when $\bar{X} \leq 103.29$.

Problem 3 Solution (c): β at $\mu = 106$

Procedure 4.4

Step 1: Fail-to-reject region in $\bar{x}$ units (one-sided form of (4.9)):

$$\bar{x}_c = \mu_0 + z_{0.05} \frac{\sigma}{\sqrt{n}} = 100 + 1.645\,(2) = 103.29$$

We fail to reject when $\bar{X} \leq 103.29$.

Step 2: Standardize under the **truth** $\mu = 106$ (this is (4.10) at work):

$$\beta = P(\bar{X} \leq 103.29 \mid \mu = 106) = P\left(Z \leq \frac{103.29 - 106}{2}\right)$$

Problem 3 Solution (c): β at $\mu = 106$

Procedure 4.4

Step 1: Fail-to-reject region in $\bar{x}$ units (one-sided form of (4.9)):

$$\bar{x}_c = \mu_0 + z_{0.05} \frac{\sigma}{\sqrt{n}} = 100 + 1.645\,(2) = 103.29$$

We fail to reject when $\bar{X} \leq 103.29$.

Step 2: Standardize under the **truth** $\mu = 106$ (this is (4.10) at work):

$$\beta = P(\bar{X} \leq 103.29 \mid \mu = 106) = P\left(Z \leq \frac{103.29 - 106}{2}\right) = P(Z \leq -1.355)$$

Problem 3 Solution (c): β at $\mu = 106$

Procedure 4.4

Step 1: Fail-to-reject region in $\bar{x}$ units (one-sided form of (4.9)):

$$\bar{x}_c = \mu_0 + z_{0.05}\,\frac{\sigma}{\sqrt{n}} = 100 + 1.645\,(2) = 103.29$$

We fail to reject when $\bar{X} \leq 103.29$.

Step 2: Standardize under the **truth** $\mu = 106$ (this is (4.10) at work):

$$\beta = P(\bar{X} \leq 103.29 \mid \mu = 106) = P\left(Z \leq \frac{103.29 - 106}{2}\right) = P(Z \leq -1.355)$$

$$\beta = \Phi(-1.355) \approx \boxed{0.0877}$$

Problem 3 Solution (c): β at $\mu = 106$

Procedure 4.4

Step 1: Fail-to-reject region in $\bar{x}$ units (one-sided form of (4.9)):

$$\bar{x}_c = \mu_0 + z_{0.05}\,\frac{\sigma}{\sqrt{n}} = 100 + 1.645\,(2) = 103.29$$

We fail to reject when $\bar{X} \leq 103.29$.

Step 2: Standardize under the **truth** $\mu = 106$ (this is (4.10) at work):

$$\beta = P(\bar{X} \leq 103.29 \mid \mu = 106) = P\left(Z \leq \frac{103.29 - 106}{2}\right) = P(Z \leq -1.355)$$

$$\beta = \Phi(-1.355) \approx \boxed{0.0877}$$

Problem 3 Solution (d): Sample Size for Power 0.90

$\text{Power} = 0.90 \Rightarrow \beta = 0.10 \Rightarrow z_\beta = z_{0.10} = 1.282; \quad z_\alpha = z_{0.05} = 1.645$

Problem 3 Solution (d): Sample Size for Power 0.90

Power $= 0.90 \Rightarrow \beta = 0.10 \Rightarrow z_\beta = z_{0.10} = 1.282; \quad z_\alpha = z_{0.05} = 1.645$

Shift to detect: $\delta = 106 - 100 = 6$ min

Problem 3 Solution (d): Sample Size for Power 0.90

Power $= 0.90 \Rightarrow \beta = 0.10 \Rightarrow z_\beta = z_{0.10} = 1.282; \quad z_\alpha = z_{0.05} = 1.645$

Shift to detect: $\delta = 106 - 100 = 6$ min

One-sided version of (4.11) (replace $z_{\alpha/2}$ with z_α):

$$n = \left\lceil \frac{(z_\alpha + z_\beta)^2 \, \sigma^2}{\delta^2} \right\rceil = \left\lceil \frac{(1.645 + 1.282)^2 \, (144)}{6^2} \right\rceil$$

Problem 3 Solution (d): Sample Size for Power 0.90

Power $= 0.90 \Rightarrow \beta = 0.10 \Rightarrow z_\beta = z_{0.10} = 1.282; \quad z_\alpha = z_{0.05} = 1.645$

Shift to detect: $\delta = 106 - 100 = 6$ min

One-sided version of (4.11) (replace $z_{\alpha/2}$ with z_α):

$$n = \left\lceil \frac{(z_\alpha + z_\beta)^2 \, \sigma^2}{\delta^2} \right\rceil = \left\lceil \frac{(1.645 + 1.282)^2 \, (144)}{6^2} \right\rceil = \left\lceil \frac{(2.927)^2 (144)}{36} \right\rceil$$

Problem 3 Solution (d): Sample Size for Power 0.90

Power $= 0.90 \Rightarrow \beta = 0.10 \Rightarrow z_\beta = z_{0.10} = 1.282; \quad z_\alpha = z_{0.05} = 1.645$

Shift to detect: $\delta = 106 - 100 = 6$ min

One-sided version of (4.11) (replace $z_{\alpha/2}$ with z_α):

$$n = \left\lceil \frac{(z_\alpha + z_\beta)^2\,\sigma^2}{\delta^2} \right\rceil = \left\lceil \frac{(1.645 + 1.282)^2\,(144)}{6^2} \right\rceil = \left\lceil \frac{(2.927)^2(144)}{36} \right\rceil$$

$$= \left\lceil \frac{1233.6}{36} \right\rceil = \lceil 34.27 \rceil = \boxed{35 \text{ drones}}$$

Problem 3 Solution (d): Sample Size for Power 0.90

Power $= 0.90 \Rightarrow \beta = 0.10 \Rightarrow z_\beta = z_{0.10} = 1.282; \quad z_\alpha = z_{0.05} = 1.645$

Shift to detect: $\delta = 106 - 100 = 6$ min

One-sided version of (4.11) (replace $z_{\alpha/2}$ with z_α):

$$n = \left\lceil \frac{(z_\alpha + z_\beta)^2\,\sigma^2}{\delta^2} \right\rceil = \left\lceil \frac{(1.645 + 1.282)^2\,(144)}{6^2} \right\rceil = \left\lceil \frac{(2.927)^2(144)}{36} \right\rceil$$

$$= \left\lceil \frac{1233.6}{36} \right\rceil = \lceil 34.27 \rceil = \boxed{35 \text{ drones}}$$

Round **up**, always. Note $n = 35 < 36$: our study in (a) was already adequately powered for a 6-minute shift.

Problem 4: False-Flag Rate of a Prompt Filter

Example 5.9 in the textbook

A vendor claims its prompt filter falsely flags at most $p_0 = 0.10$ of benign prompts. As the evaluation engineer, you suspect the true false-flag rate is **higher**. You run a random sample of $n = 400$ benign prompts and find $x = 52$ falsely flagged.

(a) State your hypotheses and conduct the test at $\alpha = 0.05$ via (4.8).

(b) Find the P-value (4.5).

(c) Would the conclusion change at $\alpha = 0.01$? Justify.

(d) Construct a 95% CI for p via (3.7).

(e) Is the CI consistent with the test in (a)? Explain briefly.

Problem 4 Solution (a)–(b): Proportion Z-Test

$\hat{p} = 52/400 = 0.13.$ **Steps 1–3:** Parameter: p.

$H_0 : p = 0.10$ vs $H_1 : p > 0.10$ (right-tailed), $\alpha = 0.05$

Problem 4 Solution (a)–(b): Proportion Z-Test

$\hat{p} = 52/400 = 0.13$. **Steps 1–3:** Parameter: p.

$H_0 : p = 0.10 \quad \text{vs} \quad H_1 : p > 0.10$ (right-tailed), $\alpha = 0.05$

Steps 4–5: Z-test (4.8); reject if $Z_0 > z_{0.05} = 1.645$.

Note: the standard error uses $p_0 = 0.10$, **not** $\hat{p}$.

Problem 4 Solution (a)–(b): Proportion Z-Test

$\hat{p} = 52/400 = 0.13$. **Steps 1–3:** Parameter: p.
$H_0 : p = 0.10$ vs $H_1 : p > 0.10$ (right-tailed), $\alpha = 0.05$

Steps 4–5: Z-test (4.8); reject if $Z_0 > z_{0.05} = 1.645$.
Note: the standard error uses $p_0 = 0.10$, **not** $\hat{p}$.

Step 6: $Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \dfrac{0.13 - 0.10}{\sqrt{0.10\,(0.90)/400}}$

Problem 4 Solution (a)–(b): Proportion Z-Test

$\hat{p} = 52/400 = 0.13.$ **Steps 1–3:** Parameter: p.

$H_0 : p = 0.10 \quad \text{vs} \quad H_1 : p > 0.10$ (right-tailed), $\alpha = 0.05$

Steps 4–5: Z-test (4.8); reject if $Z_0 > z_{0.05} = 1.645$.

Note: the standard error uses $p_0 = 0.10$, **not** $\hat{p}$.

Step 6: $Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \dfrac{0.13 - 0.10}{\sqrt{0.10\,(0.90)/400}} = \dfrac{0.03}{0.015} = 2.00$

Problem 4 Solution (a)–(b): Proportion Z-Test

$\hat{p} = 52/400 = 0.13$. **Steps 1–3:** Parameter: p.

$H_0 : p = 0.10 \quad \text{vs} \quad H_1 : p > 0.10$ (right-tailed), $\quad \alpha = 0.05$

Steps 4–5: Z-test (4.8); reject if $Z_0 > z_{0.05} = 1.645$.
Note: the standard error uses $p_0 = 0.10$, **not** $\hat{p}$.

Step 6: $Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \dfrac{0.13 - 0.10}{\sqrt{0.10\,(0.90)/400}} = \dfrac{0.03}{0.015} = 2.00$

Step 7: $2.00 > 1.645 \Rightarrow$ **Reject** H_0. Evidence at 5% that the false-flag rate exceeds the claimed 10%.

Problem 4 Solution (a)–(b): Proportion Z-Test

$\hat{p} = 52/400 = 0.13$. **Steps 1–3:** Parameter: p.
$H_0 : p = 0.10$ vs $H_1 : p > 0.10$ (right-tailed), $\alpha = 0.05$

Steps 4–5: Z-test (4.8); reject if $Z_0 > z_{0.05} = 1.645$.
Note: the standard error uses $p_0 = 0.10$, **not** $\hat{p}$.

Step 6: $$Z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{0.13 - 0.10}{\sqrt{0.10\,(0.90)/400}} = \frac{0.03}{0.015} = 2.00$$

Step 7: $2.00 > 1.645 \Rightarrow$ **Reject** H_0. Evidence at 5% that the false-flag rate exceeds the claimed 10%.

(b) P-value $= P(Z > 2.00) = 1 - 0.9772 = \boxed{0.0228}$

Problem 4 Solution (c)–(e): Changing α, and the CI

(c) At $\alpha = 0.01$: $z_{0.01} = 2.326$ and $Z_0 = 2.00 < 2.326$ (equivalently $0.0228 > 0.01$).
Fail to reject — the conclusion **changes** at the stricter level.

Problem 4 Solution (c)–(e): Changing α, and the CI

(c) At $\alpha = 0.01$: $z_{0.01} = 2.326$ and $Z_0 = 2.00 < 2.326$ (equivalently $0.0228 > 0.01$).
Fail to reject — the conclusion **changes** at the stricter level.

(d) 95% CI (3.7) (now the standard error uses $\hat{p} = 0.13$):

$$0.13 \pm 1.96\sqrt{\frac{0.13\,(0.87)}{400}}$$

Problem 4 Solution (c)–(e): Changing α, and the CI

(c) At $\alpha = 0.01$: $z_{0.01} = 2.326$ and $Z_0 = 2.00 < 2.326$ (equivalently $0.0228 > 0.01$).
Fail to reject — the conclusion **changes** at the stricter level.

(d) 95% CI (3.7) (now the standard error uses $\hat{p} = 0.13$):

$$0.13 \pm 1.96\sqrt{\frac{0.13\,(0.87)}{400}} = 0.13 \pm 1.96\,(0.01682) = 0.13 \pm 0.0330$$

Problem 4 Solution (c)–(e): Changing α, and the CI

(c) At $\alpha = 0.01$: $z_{0.01} = 2.326$ and $Z_0 = 2.00 < 2.326$ (equivalently $0.0228 > 0.01$).
Fail to reject — the conclusion **changes** at the stricter level.

(d) 95% CI (3.7) (now the standard error uses $\hat{p} = 0.13$):

$$0.13 \pm 1.96\sqrt{\frac{0.13\,(0.87)}{400}} = 0.13 \pm 1.96\,(0.01682) = 0.13 \pm 0.0330$$

$$\boxed{(0.0970,\ 0.1630)}$$

Problem 4 Solution (c)–(e): Changing α, and the CI

(c) At $\alpha = 0.01$: $z_{0.01} = 2.326$ and $Z_0 = 2.00 < 2.326$ (equivalently $0.0228 > 0.01$).
Fail to reject — the conclusion **changes** at the stricter level.

(d) 95% CI (3.7) (now the standard error uses $\hat{p} = 0.13$):

$$0.13 \pm 1.96\sqrt{\frac{0.13\,(0.87)}{400}} = 0.13 \pm 1.96\,(0.01682) = 0.13 \pm 0.0330$$

$$\boxed{(0.0970,\ 0.1630)}$$

(e) The CI **contains** $p_0 = 0.10$, yet we rejected in (a). No contradiction: the test is **one-sided** with p_0 in its standard error, while the CI is **two-sided** with $\hat{p}$

Problem 5: Two Warehouse Picking Methods

Example 5.10 in the textbook

A warehouse compares the picking rate (order lines per hour) of a zone-picking method (A) against batch picking (B). Independent teams of pickers are sampled; population variances are assumed equal.

	Method A	**Method B**
Sample size	$n_1 = 10$	$n_2 = 10$
Sample mean	$\bar{x}_1 = 85$	$\bar{x}_2 = 80$
Sample std. dev.	$s_1 = 4$	$s_2 = 6$

(a) Test at $\alpha = 0.05$ whether Method A has a significantly **higher** mean picking rate. Use (5.3) and (5.4).

(b) Bound the P-value from the t-table (Procedure 4.3).

(c) Construct a **95% CI** for $\mu_1 - \mu_2$ via (5.5). Consistent with (a)?

Problem 5 Solution (a): Pooled t-Test

Steps 1–3: Parameter: $\mu_1 - \mu_2$; variances unknown, assumed equal; independent samples.

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 > 0 \text{ (right-tailed)}, \quad \alpha = 0.05$$

Problem 5 Solution (a): Pooled t-Test

Steps 1–3: Parameter: $\mu_1 - \mu_2$; variances unknown, assumed equal; independent samples.
$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 > 0$ (right-tailed), $\quad \alpha = 0.05$

Step 6a — pooled variance (5.3):

$$s_p^2 = \frac{9\,(16) + 9\,(36)}{18}$$

Problem 5 Solution (a): Pooled t-Test

Steps 1–3: Parameter: $\mu_1 - \mu_2$; variances unknown, assumed equal; independent samples.

$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 > 0$ (right-tailed), $\quad \alpha = 0.05$

Step 6a — pooled variance (5.3):

$$s_p^2 = \frac{9\,(16) + 9\,(36)}{18} = \frac{144 + 324}{18} = \frac{468}{18} = 26$$

Problem 5 Solution (a): Pooled t-Test

Steps 1–3: Parameter: $\mu_1 - \mu_2$; variances unknown, assumed equal; independent samples.
$H_0 : \mu_1 - \mu_2 = 0$ vs $H_1 : \mu_1 - \mu_2 > 0$ (right-tailed), $\alpha = 0.05$

Step 6a — pooled variance (5.3):

$$s_p^2 = \frac{9\,(16) + 9\,(36)}{18} = \frac{144 + 324}{18} = \frac{468}{18} = 26, \qquad s_p = \sqrt{26} = 5.099$$

Problem 5 Solution (a): Pooled t-Test

Steps 1–3: Parameter: $\mu_1 - \mu_2$; variances unknown, assumed equal; independent samples.
$H_0 : \mu_1 - \mu_2 = 0 \quad$ vs $\quad H_1 : \mu_1 - \mu_2 > 0$ (right-tailed), $\quad \alpha = 0.05$

Step 6a — pooled variance (5.3):

$$s_p^2 = \frac{9\,(16) + 9\,(36)}{18} = \frac{144 + 324}{18} = \frac{468}{18} = 26, \qquad s_p = \sqrt{26} = 5.099$$

Step 6b — test statistic (5.4):

$$T_0 = \frac{85 - 80}{5.099\sqrt{\frac{1}{10} + \frac{1}{10}}}$$

Problem 5 Solution (a): Pooled t-Test

Steps 1–3: Parameter: $\mu_1 - \mu_2$; variances unknown, assumed equal; independent samples.

$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 > 0$ (right-tailed), $\quad \alpha = 0.05$

Step 6a — pooled variance (5.3):

$$s_p^2 = \frac{9\,(16) + 9\,(36)}{18} = \frac{144 + 324}{18} = \frac{468}{18} = 26, \qquad s_p = \sqrt{26} = 5.099$$

Step 6b — test statistic (5.4):

$$T_0 = \frac{85 - 80}{5.099\sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{5.0}{5.099\,(0.4472)} = \frac{5.0}{2.280} = 2.193$$

Problem 5 Solution (a): Pooled t-Test

Steps 1–3: Parameter: $\mu_1 - \mu_2$; variances unknown, assumed equal; independent samples.
$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 > 0$ (right-tailed), $\quad \alpha = 0.05$

Step 6a — pooled variance (5.3):

$$s_p^2 = \frac{9\,(16) + 9\,(36)}{18} = \frac{144 + 324}{18} = \frac{468}{18} = 26, \qquad s_p = \sqrt{26} = 5.099$$

Step 6b — test statistic (5.4):

$$T_0 = \frac{85 - 80}{5.099\sqrt{\frac{1}{10} + \frac{1}{10}}} = \frac{5.0}{5.099\,(0.4472)} = \frac{5.0}{2.280} = 2.193$$

Steps 4–5, 7: df $= 18$; $t_{0.05,\,18} = 1.734$. Since $2.193 > 1.734$: **Reject** H_0.

Problem 5 Solution (b) P-value Bounds, (c) 95% CI

(b) From Table B.3, row $\nu = 18$: $t_{0.025,\,18} = 2.101$ and $t_{0.01,\,18} = 2.552$.

Problem 5 Solution (b) P-value Bounds, (c) 95% CI

(b) From Table B.3, row $\nu = 18$: $t_{0.025,\,18} = 2.101$ and $t_{0.01,\,18} = 2.552$.

Since $2.101 < T_0 = 2.193 < 2.552$ (one-tailed):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

Problem 5 Solution (b) P-value Bounds, (c) 95% CI

(b) From Table B.3, row $\nu = 18$: $t_{0.025,\,18} = 2.101$ and $t_{0.01,\,18} = 2.552$.

Since $2.101 < T_0 = 2.193 < 2.552$ (one-tailed):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

(c) 95% (two-sided) CI (5.5), $t_{0.025,\,18} = 2.101$:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.025,\,18}\; s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Problem 5 Solution (b) P-value Bounds, (c) 95% CI

(b) From Table B.3, row $\nu = 18$: $\quad t_{0.025,\,18} = 2.101$ and $t_{0.01,\,18} = 2.552$.

Since $2.101 < T_0 = 2.193 < 2.552$ (one-tailed):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

(c) 95% (two-sided) CI (5.5), $t_{0.025,\,18} = 2.101$:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.025,\,18}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 5.0 \pm 2.101\,(2.280) = 5.0 \pm 4.790$$

Problem 5 Solution (b) P-value Bounds, (c) 95% CI

(b) From Table B.3, row $\nu = 18$: $t_{0.025,\,18} = 2.101$ and $t_{0.01,\,18} = 2.552$.

Since $2.101 < T_0 = 2.193 < 2.552$ (one-tailed):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

(c) 95% (two-sided) CI (5.5), $t_{0.025,\,18} = 2.101$:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.025,\,18}\; s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 5.0 \pm 2.101\,(2.280) = 5.0 \pm 4.790$$

$$\boxed{(0.210,\ 9.790) \text{ lines per hour}}$$

Problem 5 Solution (b) P-value Bounds, (c) 95% CI

(b) From Table B.3, row $\nu = 18$: $t_{0.025,\,18} = 2.101$ and $t_{0.01,\,18} = 2.552$.

Since $2.101 < T_0 = 2.193 < 2.552$ (one-tailed):

$$\boxed{0.01 < \text{P-value} < 0.025}$$

(c) 95% (two-sided) CI (5.5), $t_{0.025,\,18} = 2.101$:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{0.025,\,18}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 5.0 \pm 2.101\,(2.280) = 5.0 \pm 4.790$$

$$\boxed{(0.210,\ 9.790) \text{ lines per hour}}$$

Problem 6: Variability of a Startup's Checkout Flow

A startup monitors the completion time (seconds) of its checkout flow. A random sample of $n = 20$ checkouts gives a sample variance of $s^2 = 8.0\ \mathrm{s}^2$. Assume completion times are normally distributed.

(a) Construct a **95% CI** for σ^2 via (3.6), and the 95% CI for σ.

(b) Using the CI from (a) at $\alpha = 0.05$, would you reject $H_0: \sigma^2 = 4$ vs $H_1: \sigma^2 \neq 4$?

(c) The product spec requires the variance **not to exceed** $5\ \mathrm{s}^2$. State the hypotheses and test at $\alpha = 0.05$ via (4.7).

(d) Bound the P-value for (c). Would the conclusion change at $\alpha = 0.01$?

Problem 6 Solution (a)–(b): CI for σ^2 and σ

$\nu = 19$. From Table B.4: $\chi^2_{0.025,\,19} = 32.852$ and $\chi^2_{0.975,\,19} = 8.907$.

Problem 6 Solution (a)–(b): CI for σ^2 and σ

$\nu = 19$. From Table B.4: $\chi^2_{0.025,\,19} = 32.852$ and $\chi^2_{0.975,\,19} = 8.907$.

(a) CI for σ^2 (3.6), with $(n-1)s^2 = 19 \times 8.0 = 152$:

$$\left(\frac{152}{32.852}, \frac{152}{8.907} \right)$$

Problem 6 Solution (a)–(b): CI for σ^2 and σ

$\nu = 19$. From Table B.4: $\chi^2_{0.025,\,19} = 32.852$ and $\chi^2_{0.975,\,19} = 8.907$.

(a) CI for σ^2 (3.6), with $(n-1)s^2 = 19 \times 8.0 = 152$:

$$\left(\frac{152}{32.852},\ \frac{152}{8.907}\right) = \boxed{(4.627,\ 17.066)\ \mathrm{s}^2}$$

Problem 6 Solution (a)–(b): CI for σ^2 and σ

$\nu = 19$. From Table B.4: $\chi^2_{0.025,\,19} = 32.852$ and $\chi^2_{0.975,\,19} = 8.907$.

(a) CI for σ^2 (3.6), with $(n-1)s^2 = 19 \times 8.0 = 152$:

$$\left(\frac{152}{32.852}, \frac{152}{8.907}\right) = \boxed{(4.627,\ 17.066)\ \mathrm{s}^2}$$

CI for σ: take square roots: $\boxed{(2.151,\ 4.131)\ \mathrm{s}}$

Problem 6 Solution (a)–(b): CI for σ^2 and σ

$\nu = 19$. From Table B.4: $\chi^2_{0.025,\,19} = 32.852$ and $\chi^2_{0.975,\,19} = 8.907$.

(a) CI for σ^2 (3.6), with $(n-1)s^2 = 19 \times 8.0 = 152$:

$$\left(\frac{152}{32.852},\ \frac{152}{8.907}\right) = \boxed{(4.627,\ 17.066)\ \mathrm{s}^2}$$

CI for σ: take square roots: $\boxed{(2.151,\ 4.131)\ \mathrm{s}}$

(b) $\sigma_0^2 = 4$ lies **outside** $(4.627, 17.066)$ $\Rightarrow$ by the CI–test duality, **reject** H_0 at $\alpha = 0.05$: the variance differs significantly from 4.

Problem 6 Solution (c)–(d): One-Sided χ^2-Test

(c) Steps 1–3: Parameter: σ^2. $H_0 : \sigma^2 = 5$ vs $H_1 : \sigma^2 > 5$ (right-tailed), $\alpha = 0.05$.

Problem 6 Solution (c)–(d): One-Sided χ^2-Test

(c) Steps 1–3: Parameter: σ^2. $H_0 : \sigma^2 = 5$ vs $H_1 : \sigma^2 > 5$ (right-tailed), $\alpha = 0.05$.

Steps 4–5: χ^2-test (4.7), df $= 19$; reject if $\chi_0^2 > \chi^2_{0.05,\,19} = 30.144$.

Problem 6 Solution (c)–(d): One-Sided χ^2-Test

(c) Steps 1–3: Parameter: σ^2. $H_0 : \sigma^2 = 5$ vs $H_1 : \sigma^2 > 5$ (right-tailed), $\alpha = 0.05$.

Steps 4–5: χ^2-test (4.7), df $= 19$; reject if $\chi_0^2 > \chi^2_{0.05,\,19} = 30.144$.

Step 6: $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{19 \times 8.0}{5} = \dfrac{152}{5}$

Problem 6 Solution (c)–(d): One-Sided χ^2-Test

(c) Steps 1–3: Parameter: σ^2. $H_0 : \sigma^2 = 5$ vs $H_1 : \sigma^2 > 5$ (right-tailed), $\alpha = 0.05$.

Steps 4–5: χ^2-test (4.7), df $= 19$; reject if $\chi_0^2 > \chi^2_{0.05,\,19} = 30.144$.

Step 6: $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{19 \times 8.0}{5} = \dfrac{152}{5} = 30.4$

Problem 6 Solution (c)–(d): One-Sided χ^2-Test

(c) Steps 1–3: Parameter: σ^2. $H_0 : \sigma^2 = 5$ vs $H_1 : \sigma^2 > 5$ (right-tailed), $\alpha = 0.05$.

Steps 4–5: χ^2-test (4.7), df $= 19$; reject if $\chi_0^2 > \chi^2_{0.05,\,19} = 30.144$.

Step 6: $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{19 \times 8.0}{5} = \dfrac{152}{5} = 30.4$

Step 7: $30.4 > 30.144 \Rightarrow$ **Reject** H_0 — the checkout flow does **not** meet the variability spec at the 5% level (just barely).

Problem 6 Solution (c)–(d): One-Sided χ^2-Test

(c) Steps 1–3: Parameter: σ^2. $H_0 : \sigma^2 = 5$ vs $H_1 : \sigma^2 > 5$ (right-tailed), $\alpha = 0.05$.

Steps 4–5: χ^2-test (4.7), df $= 19$; reject if $\chi_0^2 > \chi^2_{0.05,\,19} = 30.144$.

Step 6: $\chi_0^2 = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{19 \times 8.0}{5} = \dfrac{152}{5} = 30.4$

Step 7: $30.4 > 30.144 \Rightarrow$ **Reject** H_0 — the checkout flow does **not** meet the variability spec at the 5% level (just barely).

(d) From Table B.4, $\nu = 19$: $\chi^2_{0.05} = 30.144 < 30.4 < 32.852 = \chi^2_{0.025}$

$$\boxed{0.025 < \text{P-value} < 0.05}$$

At $\alpha = 0.01$: $\chi^2_{0.01,\,19} = 36.191 > 30.4 \Rightarrow$ **fail to reject** — the conclusion **changes**

Concept Checks

Ten quick questions — answer before the reveal

Concept Check 1

A logistics engineer computes a 95% one-sided upper confidence bound for the mean truck unloading time and obtains $\mu \leq 5.38$ hours. Which is a correct interpretation?

(a) The true mean is definitely less than 5.38 hours.

(b) There is a 95% probability that $\mu = 5.38$ hours.

(c) We are 95% confident the true mean does not exceed 5.38 hours.

(d) 95% of individual unloading times are below 5.38 hours.

(e) The sample mean is 5.38 hours.

Concept Check 1

A logistics engineer computes a 95% one-sided upper confidence bound for the mean truck unloading time and obtains $\mu \leq 5.38$ hours. Which is a correct interpretation?

(a) The true mean is definitely less than 5.38 hours.

(b) There is a 95% probability that $\mu = 5.38$ hours.

(c) We are 95% confident the true mean does not exceed 5.38 hours.

(d) 95% of individual unloading times are below 5.38 hours.

(e) The sample mean is 5.38 hours.

Answer: (c). Confidence describes the long-run reliability of the *procedure* (3.2). (a) claims certainty; (b) puts probability on a fixed parameter; (d) confuses the mean with individual values (that's a tolerance interval (3.11)); (e) is unsupported.

Concept Check 2

A startup computes both a 90% and a 99% CI for mean session length from the same data with σ known. Compared to the 90% interval, the 99% interval will be:

(a) Narrower with the same center.

(b) Wider with a different center.

(c) Wider with the same center.

(d) The same width but higher confidence.

(e) Narrower, because 99% requires more precision.

Concept Check 2

A startup computes both a 90% and a 99% CI for mean session length from the same data with σ known. Compared to the 90% interval, the 99% interval will be:

(a) Narrower with the same center.
(b) Wider with a different center.
(c) Wider with the same center.
(d) The same width but higher confidence.
(e) Narrower, because 99% requires more precision.

Answer: (c). Both center at $\bar{x}$; (3.1) uses $z_{0.005} = 2.576$ vs $z_{0.05} = 1.645$, so the 99% margin is wider. Higher confidence costs width.

Concept Check 3

An AI engineer wants a CI for the variance of a perception model's inference latency. The $n = 15$ measurements look clearly bimodal (non-normal). Which statement is most appropriate?

(a) The χ^2-interval is valid; $n = 15$ is enough for the CLT.
(b) The χ^2-interval requires normality and should not be used here.
(c) The χ^2-interval is robust to non-normality.
(d) Use a t-interval for the variance instead.
(e) Use a z-interval for σ^2.

Concept Check 3

An AI engineer wants a CI for the variance of a perception model's inference latency. The $n = 15$ measurements look clearly bimodal (non-normal). Which statement is most appropriate?

(a) The χ^2-interval is valid; $n = 15$ is enough for the CLT.
(b) The χ^2-interval requires normality and should not be used here.
(c) The χ^2-interval is robust to non-normality.
(d) Use a t-interval for the variance instead.
(e) Use a z-interval for σ^2.

Answer: (b). The χ^2 result behind (3.6) is derived *exactly* from normality and is **not robust** — unlike mean procedures, there is no CLT rescue. A bimodal sample is a clear violation.

Concept Check 4

A hypothesis test on port dwell times yields a P-value of 0.03. Which is the correct interpretation?

(a) H_0 is definitely false.
(b) There is a 3% probability that H_0 is true.
(c) If H_0 is true, there is a 3% chance of data this extreme or more extreme.
(d) The probability of a Type II error is 0.97.
(e) The power of the test is 97%.

Concept Check 4

A hypothesis test on port dwell times yields a P-value of 0.03. Which is the correct interpretation?

(a) H_0 is definitely false.

(b) There is a 3% probability that H_0 is true.

(c) If H_0 is true, there is a 3% chance of data this extreme or more extreme.

(d) The probability of a Type II error is 0.97.

(e) The power of the test is 97%.

Answer: (c). That's the definition (4.4): a probability about the *data* computed under H_0 — never the probability that H_0 is true, and unrelated to β or power.

Concept Check 5

A founder tests $H_0: p = 0.4$ vs $H_1: p \neq 0.4$ (signup conversion) at $\alpha = 0.05$. The 95% CI for p is [0.433, 0.466]. What follows about the P-value?

(a) P-value $= \alpha$.

(b) P-value $= 1 - \alpha$.

(c) P-value $< \alpha$.

(d) P-value $> \alpha$.

(e) More information is required.

Concept Check 5

A founder tests $H_0: p = 0.4$ vs $H_1: p \neq 0.4$ (signup conversion) at $\alpha = 0.05$. The 95% CI for p is $[0.433,\ 0.466]$. What follows about the P-value?

(a) P-value $= \alpha$.

(b) P-value $= 1 - \alpha$.

(c) P-value $< \alpha$.

(d) P-value $> \alpha$.

(e) More information is required.

Answer: (c). The CI excludes $p_0 = 0.4$, so the two-sided test rejects at $\alpha = 0.05$; by the CI–test duality, P-value < 0.05.

Concept Check 6

Testing $H_0 : \mu = 50$ vs $H_1 : \mu > 50$ (fuel consumption per route, σ known), an engineer gets $Z_0 = 1.50$ at $\alpha = 0.05$. The correct conclusion is:

(a) Reject H_0 because $1.50 > 0.05$.

(b) Reject H_0 because $1.50 > 1.28$.

(c) Fail to reject H_0 because $1.50 < 1.645$.

(d) Fail to reject H_0 because $1.50 < 1.96$.

(e) Reject H_0 because $\mathrm{P} < 0.05$.

Concept Check 6

Testing $H_0 : \mu = 50$ vs $H_1 : \mu > 50$ (fuel consumption per route, σ known), an engineer gets $Z_0 = 1.50$ at $\alpha = 0.05$. The correct conclusion is:

(a) Reject H_0 because $1.50 > 0.05$.

(b) Reject H_0 because $1.50 > 1.28$.

(c) Fail to reject H_0 because $1.50 < 1.645$.

(d) Fail to reject H_0 because $1.50 < 1.96$.

(e) Reject H_0 because $\mathrm{P} < 0.05$.

Answer: (c). One-sided at $\alpha = 0.05$ uses $z_{0.05} = 1.645$, not 1.96 (two-sided). Also $\mathrm{P} = P(Z > 1.50) = 0.0668 > 0.05$, so (e) is false.

Concept Check 7

In a pooled t-test with $n_1 = 12$ and $n_2 = 15$, the degrees of freedom are:

(a) $\min(11, 14) = 11$

(b) $n_1 + n_2 = 27$

(c) $n_1 + n_2 - 2 = 25$

(d) $n_1 - 1 = 11$

(e) $(n_1 - 1)(n_2 - 1) = 154$

Concept Check 7

In a pooled t-test with $n_1 = 12$ and $n_2 = 15$, the degrees of freedom are:

(a) $\min(11, 14) = 11$

(b) $n_1 + n_2 = 27$

(c) $n_1 + n_2 - 2 = 25$

(d) $n_1 - 1 = 11$

(e) $(n_1 - 1)(n_2 - 1) = 154$

Answer: (c). The pooled t (5.4) spends one df per sample estimating the common variance (5.3): $\nu = 12 + 15 - 2 = 25$.

Concept Check 8

An engineer scores the *same* 10 evaluation suites before and after fine-tuning a model. The most appropriate test is:

(a) Two-sample Z-test

(b) Pooled t-test

(c) Welch's t-test

(d) Paired t-test

(e) F-test

Concept Check 8

An engineer scores the *same* 10 evaluation suites before and after fine-tuning a model. The most appropriate test is:

(a) Two-sample Z-test
(b) Pooled t-test
(c) Welch's t-test
(d) Paired t-test
(e) F-test

Answer: (d). Same suites in both conditions $\Rightarrow$ dependent observations. Analyze the differences with (5.8), $\mathrm{df} = 9$. Pooled and Welch assume *independent* samples.

Concept Check 9

A two-sided F-test for equality of two checkout-flow variances gives $F_0 = 2.10$ with $(u, v) = (9, 14)$; the critical value is $f_{0.025,\,9,\,14} = 3.21$. Conclusion?

(a) Reject H_0 because $F_0 > 1$.

(b) Fail to reject H_0; not enough evidence the variances differ.

(c) Reject H_0 because F_0 exceeds the critical value.

(d) The test is invalid because $F_0 < f_{0.025}$.

(e) More information is needed.

Concept Check 9

A two-sided F-test for equality of two checkout-flow variances gives $F_0 = 2.10$ with $(u, v) = (9, 14)$; the critical value is $f_{0.025,\,9,\,14} = 3.21$. Conclusion?

(a) Reject H_0 because $F_0 > 1$.

(b) Fail to reject H_0; not enough evidence the variances differ.

(c) Reject H_0 because F_0 exceeds the critical value.

(d) The test is invalid because $F_0 < f_{0.025}$.

(e) More information is needed.

Answer: (b). With the larger variance on top, reject only if $F_0 > 3.21$ (5.13). $F_0 > 1$ by construction — that alone never rejects.

Concept Check 10

Red-teaming two model versions: $n_1 = 100$ prompts with $x_1 = 30$ jailbreaks, and $n_2 = 100$ prompts with $x_2 = 20$. For testing $H_0 : p_1 = p_2$, the pooled proportion is:

(a) $\hat{p} = (0.30 + 0.20)/2 = 0.25$

(b) $\hat{p} = (30 + 20)/(100 + 100) = 0.25$

(c) $\hat{p} = \sqrt{0.30 \times 0.20} = 0.245$

(d) $\hat{p} = 0.30\,(0.5) + 0.20\,(0.5) = 0.25$

(e) (a) and (b) agree here, but (b) is the correct general formula.

Concept Check 10

Red-teaming two model versions: $n_1 = 100$ prompts with $x_1 = 30$ jailbreaks, and $n_2 = 100$ prompts with $x_2 = 20$. For testing $H_0: p_1 = p_2$, the pooled proportion is:

(a) $\hat{p} = (0.30 + 0.20)/2 = 0.25$

(b) $\hat{p} = (30 + 20)/(100 + 100) = 0.25$

(c) $\hat{p} = \sqrt{0.30 \times 0.20} = 0.245$

(d) $\hat{p} = 0.30\,(0.5) + 0.20\,(0.5) = 0.25$

(e) (a) and (b) agree here, but (b) is the correct general formula.

Answer: (e). The general rule pools *counts*: $\hat{p} = (x_1 + x_2)/(n_1 + n_2)$, used in (5.10). Averaging proportions only works when $n_1 = n_2$.

Before Major Exam 1

- Redo all six problems **closed-book**, then check against these slides.

Before Major Exam 1

- Redo all six problems **closed-book**, then check against these slides.
- Drill the master table Table 5.6 and the attack plan Procedure 5.3.

Before Major Exam 1

- Redo all six problems **closed-book**, then check against these slides.
- Drill the master table Table 5.6 and the attack plan Procedure 5.3.
- Statistical tables (t, χ^2, F) are provided with the exam — practice reading the correct row, column, and tail.

Before Major Exam 1

- Redo all six problems **closed-book**, then check against these slides.
- Drill the master table Table 5.6 and the attack plan Procedure 5.3.
- Statistical tables (t, χ^2, F) are provided with the exam — practice reading the correct row, column, and tail.
- Bring your questions to office hours before exam week.

Before Major Exam 1

- Redo all six problems **closed-book**, then check against these slides.
- Drill the master table Table 5.6 and the attack plan Procedure 5.3.
- Statistical tables (t, χ^2, F) are provided with the exam — practice reading the correct row, column, and tail.
- Bring your questions to office hours before exam week.

Good luck!