Quiz 5 — Solutions

STAT 515 · Hypothesis Testing for a Mean (Homework 7 material)

Total Points: 40

Problem 1: Setting Up Hypotheses (8 points)

(a) (4 points) Let μ = the **true mean weight of all cereal boxes**, in grams.

$$H_0 : \mu = 500 \qquad \text{vs.} \qquad H_a : \mu < 500$$

This is a **one-sided (lower-tailed)** test. The consumer group's suspicion — boxes are underfilled — is the research claim and goes in H_a. (Writing $H_0 : \mu \geq 500$ is also acceptable; the equality is what the test is carried out at.)

Grading: 1 point parameter in words and symbols, 2 points for both hypotheses with the correct direction, 1 point for "one-sided."

(b) (4 points) Let μ = the **true mean wait time at the clinic after the scheduling change**, in minutes.

$$H_0 : \mu = 22 \qquad \text{vs.} \qquad H_a : \mu \neq 22$$

This is a **two-sided** test, because "changed" allows a decrease or an increase.

Grading: 1 point parameter, 2 points hypotheses, 1 point for "two-sided."

Problem 2: Types of Error (10 points)

(a) (3 points) A **Type I error** is rejecting H_0 when H_0 is true. Here: concluding the machine is **out of calibration** (mean fill $\neq$ 500 mL) when in fact it is filling correctly at 500 mL. Consequence: the line is stopped and the machine readjusted unnecessarily.

Grading: 1 point for the general definition, 2 points for the correct contextual statement.

(b) (3 points) A **Type II error** is failing to reject H_0 when H_a is true. Here: concluding there is not enough evidence of a calibration problem when the machine **is** in fact off target. Consequence: mis-filled bottles keep being produced and shipped.

Grading: 1 point definition, 2 points context.

(c) (4 points) Lowering α from 0.05 to 0.01 **decreases** the probability of a Type I error (it is α by definition). With the sample size held fixed, it **increases** the probability of a Type II error, β, because the rejection region shrinks and it becomes harder to reject H_0 even when H_0 is false. Power $= 1 - \beta$ therefore goes down. The only way to reduce both error probabilities at once is to **increase the sample size**.

Grading: 1 point Type I decreases, 2 points Type II increases with reasoning, 1 point for the trade-off / sample-size remark.

Problem 3: Large-Sample z-Test (16 points)

$n = 36$, $\bar{x} = 242.0$, $s = 24.0$, $\mu_0 = 250$.

(a) (3 points) Let μ = true mean simple reaction time of athletes, in ms.

$$H_0 : \mu = 250 \qquad \text{vs.} \qquad H_a : \mu \neq 250 \qquad \text{(two-sided)}$$

(b) (4 points) The sample size $n = 36$ exceeds 30, so the Central Limit Theorem makes $\bar{x}$ approximately normal regardless of the shape of the reaction-time distribution, and s is a reliable enough estimate of σ to be used in its place. The large-sample z test therefore applies.

$$\text{SE} = \frac{s}{\sqrt{n}} = \frac{24.0}{\sqrt{36}} = \frac{24.0}{6} = 4.000$$

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{242.0 - 250}{4.000} = \frac{-8.0}{4.000} = \boxed{-2.000}$$

Grading: 1 point for the justification ($n > 30$ / CLT), 1 point SE, 2 points test statistic. Sign must be negative. A student who uses t with $df = 35$ gets essentially the same number ($t = -2.000$, $p = 0.0532$) — accept the arithmetic but deduct the justification point.

(c) (4 points) Since the p-value $= 0.0455 < \alpha = 0.05$, we **reject** H_0.

Conclusion in context: At the 5% significance level there is sufficient evidence to conclude that the mean simple reaction time of athletes **differs from** 250 ms; the sample suggests it is faster (lower) than the general-population mean.

Grading: 2 points for the comparison and decision, 2 points for a conclusion written in context. A bare "reject H_0" with no context earns 2 of 4.

(d) (3 points) $z_{0.025} = 1.96$:

$$242.0 \pm 1.96(4.000) = 242.0 \pm 7.84 \implies \boxed{(234.16,\ 249.84) \text{ ms}}$$

Grading: 1 point $z^* = 1.96$, 2 points interval.

(e) (2 points) The hypothesized value $\mu_0 = 250$ falls **outside** the 95% confidence interval (just above the upper endpoint 249.84). A two-sided test at $\alpha = 0.05$ rejects H_0 exactly when μ_0 lies

outside the 95% interval, so the interval and the test agree. Note how close the call is — consistent with a p-value of 0.0455, only barely below 0.05.

Grading: 1 point for noting 250 is outside the interval, 1 point for stating the general equivalence.

Problem 4: Interpreting a p-Value (6 points)

(a) (3 points) The p-value is the probability, **assuming** H_0 is true ($\mu = 250$), of obtaining a sample result at least as extreme as the one observed — that is, a sample mean at least as far from 250 in either direction as 242.0 is. Here that probability is 0.0455.

Grading: 1 point for the required phrase "assuming H_0 is true", 1 point for "at least as extreme as observed", 1 point for mentioning both directions (two-sided).

(b) (3 points) False. A p-value is a probability computed **about the data, given that** H_0 is true — it is not a probability about H_0 itself. In this framework H_0 is either true or false; it is not assigned a probability.

Grading: 1 point for "false", 2 points for the justification (conditioning direction / H_0 is not random).

R verification

```
n <- 36; xbar <- 242.0; s <- 24.0; mu0 <- 250

se <- s / sqrt(n); se                    # 4.000
zstat <- (xbar - mu0) / se; zstat        # -2.000
2 * pnorm(-abs(zstat))                   # 0.04550
qnorm(0.975)                             # 1.95996
xbar + c(-1, 1) * qnorm(0.975) * se      # 234.16  249.84
```

Note on the table value: a standard normal table gives $P(Z > 2.00) = 0.0228$, so a student working from the table reports $p = 2(0.0228) = 0.0456$. Accept either 0.0455 or 0.0456.