STAT 110 Practice Examination

Annotated Solutions

1. The sample mean $\bar{y}$ is closest to

   A. 41.3
   Just one observation, not the average of all five.

   B. 39.7
   Again only a single value.

   C. **42.0**

   D. 52.5
   Mean can't be larger than every data point.

   E. 209.9
   That is the sum *of the numbers, not their mean.*

2. The sample standard deviation s is closest to

   A. 1.0
   Would imply the five numbers are almost identical.

   B. 8.2
   Needs values about 8 away from the mean—none are.

   C. 94.0
   Mixes up variance and standard deviation.

   D. **2.9**

   E. 1.7
   Used n in the denominator instead of $n-1$; too small.

3. Watching one T20 cricket match that totals 325 runs, choose the best statement:

   A. It is unlikely totals ever exceed 500.
   One match can't tell us that.

   B. 325 is "typical" because the game was random.
   Need a league-wide distribution.

   C. **We can't tell if 325 is unusual without knowing likely score ranges.**

   D. More than 300 runs is normal.
   No supporting evidence.

E. Every game has exactly 325 runs.
Plainly impossible.

4. Are variation and uncertainty important in statistics?

A. **Yes—models describe variation and CIs/SEs quantify uncertainty.**

B. No; parameters are known exactly.
Population parameters are unknown.

C. Yes; it makes analysis look sophisticated.
Not a scientific reason.

D. No; big data eliminate uncertainty.
Large n reduces but never removes it.

E. No; once data are collected there's no uncertainty.
Randomness persists.

5. $\Pr(Y = 3 \text{ or } 4)$ for the gene-copy study

A. 0.30
Omits the 4-copy probability.

B. **0.80**

C. 0.20
Only the missing 0.20 probability.

D. 0.50
Counts 3-copies but not 4-copies.

E. 1.00
Would mean everyone has 3 or 4 copies.

6. $E[Y]$

A. 2.25
Dropped one probability weight.

B. **2.85**

C. 2.50
Used wrong weights.

D. 3.00
Added instead of weighting.

E. 2.80
Rounded from a wrong calculation.

7. Best description of $E[Y]$

A. A person will have exactly $E[Y]$ copies.
It's a long-run average, not a guarantee.

B. **It's the expected (average) number of copies for one randomly chosen person.**

C. Half the people have fewer than $E[Y]$.
That defines the median.

D. Exactly $E[Y]$ copies means higher risk.
Mean alone says nothing about risk.

E. Copy number is normal with mean $E[Y]$.
Normality wasn't given.

8. Meaning of $\Pr(B \mid V^c)$

   A. **Probability a customer chooses a plant-based burger *given they are not vegetarian.***

   B. Probability a customer is not vegetarian, given they chose the burger.
   Condition reversed.

   C. Probability a vegetarian chooses the burger.
   Condition is the opposite group.

   D. Joint probability of non-vegetarian & burger.
   Not conditional.

   E. Probability of choosing a burger regardless of diet.
   Ignores the condition.

9. $\Pr(V \mid B)$

   A. **0.632**

   B. 0.600
   That's $\Pr(B \mid V)$.

   C. 0.180
   Only one piece of Bayes' rule.

   D. 0.285
   Another single piece.

   E. 0.105
   Yet another single piece.

10. $\Pr(V \cup B)$

   A. **0.405**

   B. 0.180
   Just vegetarians who choose burgers.

   C. 0.300
   $\Pr(V)$ *only.*

   D. 0.285
   $\Pr(B)$ *only.*

   E. 1.000
   Would imply every customer is vegetarian or buys a burger.

11. $\Pr(B)$

A. 0.180
Only vegetarians choosing burgers.

B. 0.300
$\Pr(V)$, *not* $\Pr(B)$.

C. **0.285**

D. 0.105
Only non-vegetarians choosing burgers.

E. 1.000
Would mean everyone buys a burger.

12. Best definition of a random variable

A. Summarises both population and sample.
It's defined before any sample is taken.

B. Normally distributed variable.
Distribution can be anything.

C. **A random process with a numerical outcome.**

D. Fixed but unknown value.
The value is random before observing.

E. Depends on the observed sample.
Definition doesn't rely on realised data.

13. Which is best modelled as continuous?

A. Eggs in a nest.
Count data, discrete.

B. Purchasing visitors.
Count data, discrete.

C. **pH value of seawater**

D. Voters supporting a candidate.
Discrete count.

E. Tasks completed.
Discrete count.

14. $E[\,2X - 3Y\,]$ with $E[X] = 10,\ E[Y] = 4$

A. 4
Used $X - 3Y$ *instead of* $2X - 3Y$.

B. 20
Multiplied rather than added/subtracted.

C. 12
Dropped one coefficient.

D. **8**

E. 0
Incorrectly cancelled the terms.

15. sd($2X - 3Y$) (independent X, Y)

    A. 10.50
    That's the variance, not the s.d.

    B. 1.22
    Square-rooted the wrong total.

    C. **3.24**

    D. 2.12
    Left out a variance component.

    E. 1.50
    Left out two variance components.

16. A healthy adult is 1.5 s.d. below the mean ($\mu = 40,\ \sigma = 8$). Score?

    A. -1.5
    z-score itself, not the raw score.

    B. 38.5
    Moved in the wrong direction.

    C. 52.0
    Moved above *the mean.*

    D. **28.0**

    E. 40.0
    That's the mean ($z = 0$).

17. Code to find $P(\text{score} > 48)$

    A. **1-pnorm(1.0)**

    B. 1-pnorm(48)
    Treats raw score as a z-score.

    C. 1-pnorm(-1.0)
    Upper tail of the wrong z.

    D. pnorm(1.0)
    Gives lower tail.

    E. pnorm(48)
    Again uses raw score as z.

18. Sampling distribution of $\bar{y}$ for $n = 64$

    A. Normal$(40, 8)$
    Forgot to divide by $\sqrt{64}$.

    B. **Normal**$(40, 1)$

C. Normal$(0, 1)$
Standardised already.

D. t_{63}
σ is known, so we use z.

E. Unknown.
Central Limit Theorem gives it.

19. Same set-up as 18 but population not normal (still $n = 64$)

A. Normal$(40, 8)$
Same issue—must divide by $\sqrt{n}$.

B. **Approximately normal**$(40, 1)$

C. Normal$(0, 1)$
Already standardised.

D. t_{63}
σ known.

E. Unknown.
CLT still applies for large n.

20. Which formula gives a 95 % confidence interval for a *population mean* when the population standard deviation is *unknown*?

A. $\bar{y} \pm z^{\star}\sqrt{\frac{p(1-p)}{n}}\%$
This is the CI for a proportion, *not a mean.*

B. $\bar{y} \pm z^{\star}\frac{\sigma}{\sqrt{n}}\%$
Uses the unknown σ instead of the sample s.

C. $\bar{y} \pm t^{\star}_{n-1}\frac{\sigma}{\sqrt{n}}\%$
σ is unknown; we should plug in s, not σ.

D. $\bar{y} \pm z^{\star}\, s\%$
Forgot to divide by $\sqrt{n}$ and *uses z instead of t.*

E. $\bar{y} \pm t^{\star}_{n-1}\frac{s}{\sqrt{n}}$

21. If we increase the confidence level from 90 % to 99 % while keeping the same sample,

A. The point estimate changes.
The sample mean stays the same.

B. The standard error changes.
SE depends only on s and n.

C. The sample size must change.
Sample size is fixed.

D. **The interval widens to guarantee the higher confidence.**

E. A higher confidence level always narrows the interval.
It actually widens.

22. A 99 % confidence interval for the mean content of coffee jars is (201.2 g, 203.0 g). Which interpretation is correct?

A. 99 % of individual jars weigh between 201.2 g and 203.0 g.
CI is about the mean*, not single jars.*

B. 99 % of *future* jars will fall in this range.
Again, CI targets the mean.

C. There is a 99 % chance the true mean lies in this very interval.
The probability refers to the method*, not to μ once the interval is drawn.*

D. If we took many samples, 99 % of the *sample means* would be in this range.
CIs vary from sample to sample; means do not cluster in one fixed interval.

E. **In the long run, 99 % of such intervals built this way would contain the true mean μ.**

23. Which statement about a p-value is correct?

A. A large p-value proves the null hypothesis is true.
"Prove" is too strong.

B. $p = 0.03$ means there is a 3 % chance the alternative is false.
p*-value is calculated under H_0, not H_A.*

C. **$p = 0.03$ means that, if H_0 were true, we would see data this extreme only 3 % of the time.**

D. p-value is the probability calculations were wrong.
It measures data extremeness, not mistakes.

E. A small p-value guarantees practical importance.
Statistical $\neq$ practical significance.

24. Which step is genuine *statistical estimation* of a population mean?

A. Just replace μ by $\bar{y}$ in a sentence.
Pure description, no inference.

B. Compute the standard error only.
SE alone isn't an estimate of μ.

C. Use the full population data.
Then it's no longer estimation.

D. Guess a number that "looks right."
Not statistical.

E. **Build a confidence interval or carry out a test based on the sample.**

25. In the penguin data, the first line of R code calculates `mean(bill_length_mm)`. What does this value summarise?

A. **The average bill length for the sampled penguins.**

B. The average flipper length.
Wrong variable.

C. The standard deviation of bill length.
Mean, not spread.

D. The median bill length.
Mean $\neq$ median.

E. The sample size.
Mean is not a count.

26. The next line of code asks for `sd(flipper_length_mm)`. What statistic is that?

A. Mean flipper length.
It's the spread, not the centre.

B. Median flipper length.
Again, centre vs spread.

C. Range of flipper length.
Range is max–min.

D. **The standard deviation of flipper length (how much they vary).**

E. Sample size.
Standard deviation isn't a count.

27. A one-sample t-test compares pollution *before* and *after* a policy. Which null hypothesis is correct?

A. $\mu_{\textbf{after}} - \mu_{\textbf{before}} = 0$

B. $\mu_{\text{after}} = 0$
Should compare difference*, not absolute level.*

C. A small p-value means results are "not unusual."
It actually means they are *unusual under H_0.*

D. The test measures correlation between before & after.
It tests mean difference.

E. Significance alone decides policy.
Statistical vs practical issues.

28. You have paired "before/after" data. Which test is appropriate?

A. **Paired-sample t-test on the differences.**

B. Two-sample independent t-test.
Ignores pairing.

C. Two-proportion z-test.
Data are numeric, not proportions.

D. Chi-square test.
Counts, not means.

E. ANOVA.
More than two groups required.

29. Statistical *power* is

   A. The chosen α level.
   That's Type I error rate.

   B. Probability of rejecting H_0 when H_0 is true.
   That's Type I error again.

   C. **Probability of rejecting H_0 when a specified alternative is true.**

   D. The observed p-value.
   p-*value and power differ.*

   E. Guaranteed by p ¡ 0.05.
   Low p *doesn't ensure high power.*

30. Which change *increases* the power of a test (all else fixed)?

   A. Call the data "paired" when they are not.
   Inflates Type I error, not power properly.

   B. Reduce the sample size.
   Smaller n lowers power.

   C. Lower α from 0.05 to 0.01.
   Harder to reject, so power drops.

   D. Replicate the same small study many times with no changes.
   Each run keeps low power.

   E. **Increase the sample size.**

31. Standard error for the difference of two independent means: which value is correct? ($s_1 = 4.5$, $n_1 = 25$; $s_2 = 6.0$, $n_2 = 36$)

   A. **2.11**

   B. 4.44
   Forgot to divide by sample sizes.

   C. 0.61
   Only used one group's variance.

   D. 2.92
   Miscalculated the square-root step.

   E. 16.32
   Added variances, then squared again.

32. Which R command correctly performs an *independent* two-sample t-test?

   A. `t.test(groupA, groupB, var.equal = FALSE)`

   B. `t.test(groupA, groupB, paired = TRUE)`
   Sets a paired test you don't have.

C. `chisq.test(groupA, groupB)`
Chi-square uses counts.

D. `prop.test(groupA, groupB)`
Proportions, not means.

E. `wilcox.test(groupA, groupB)`
Non-parametric alternative.

33. Interpreting a 95 % CI *for the difference* (drug A minus placebo) of $(-4.3,\ 1.2)$:

A. CI gives the mean reduction for drug A alone.
It describes the difference.

B. Drug A is definitely better.
CI includes zero—no guarantee.

C. A single patient will improve by 1.2 units.
CI is about the mean, not an individual.

D. The CI proves no reduction is possible.
Includes zero but doesn't prove equality.

E. **Because the CI crosses 0, we lack strong evidence that drug A outperforms placebo.**

34. A regression-output question asked, "Which statement is *not* correct?" Only one option is wrong.

A. The errors are normally distributed.
This is a valid model assumption.

B. Homoscedasticity was assessed.
Also true.

C. Linearity appears reasonable.
Also true.

D. **There is no random error term in the fitted model.**
Every regression has ε.

E. Independence of observations was checked.
Also true.

35. With $\hat{y} = 150 - 0.75x$, the fitted mean at $x = 20$ is

A. **135**

B. 148.5
Added instead of subtracted 0.75×20.

C. 135 *for every bird.*
Mean model; individuals vary around it.

D. 150
Forgot the slope term.

E. The model gives no spread.
Residual SD gives the spread.

36. Interpret the slope $\hat{\beta}_1 = -0.75$ in $\hat{y} = 150 - 0.75x$.

   A. **On average, y decreases by 0.75 units for each 1-unit increase in x.**

   B. The line rises by 0.75 units per x.
   Sign is negative.

   C. One bird's y always drops 0.75.
   Slope describes the mean, not individuals.

   D. When x changes 10 units, y always drops 7.5.
   Again, average not guarantee.

   E. -0.75 is the intercept.
   Intercept is 150.

37. A scatterplot gives $r = -0.64$ between age and aptitude.

   A. **The variables have a moderate negative linear relationship.**

   B. Positive correlation.
   The sign is negative.

   C. No linear trend.
   $|r|$ is well away from 0.

   D. Relationship is curved.
   Nothing suggests curvature.

   E. Outliers dominate.
   No evidence shown.

38. Least-squares line–fitting uses which criterion?

   A. Draw a line by eye.
   Subjective, not least squares.

   B. Minimise the *sum* of residuals.
   They cancel to 0 anyway.

   C. Maximise correlation.
   Not the fitting rule.

   D. **Minimise the sum of *squared* residuals.**

   E. Find a perfect fit.
   Rarely possible with real data.

39. Which equation gives the *predicted* aptitude from age?

   A. $\hat{x} = 150 - 0.75y$
   Swapped variables.

   B. $\hat{y} = 150 - 0.75x + \varepsilon$
   ε is not part of the prediction.

   C. $\hat{y} = 150 - 0.75x$

   D. $y = 150 - 0.75x$
   Drops the "hat"—not clearly prediction.

E. $x = 150 - 0.75y$
Predicts age from aptitude.

40. The regression output gives $p = 0.00177$ for the slope. Which option matches this?

A. 0.020
Too large.

B. 0.15
Too large.

C. 0.051
Too large.

D. 0.50
Far too large.

E. **0.0018 (rounded).**

41. The standard error for $\hat{\beta}_1$ is 0.310. Which description is correct?

A. The standard error tells us the average error made when predicting $\hat{y}$%
That is the residual *standard error, not SE($\hat{\beta}_1$).*

B. It is the mean of the sampling distribution of $\hat{\beta}_1$%
The mean of that distribution is the true β_1, not the SE.

C. It measures variability of the y-values around the regression line%
That quantity is σ_ε, not SE($\hat{\beta}_1$).

D. It is the difference between observed and true slope%
That difference is one realisation*, not the spread.*

E. **It describes the variability in the sampling distribution of $\hat{\beta}_1$.**

42. With multiplier 2.093, the 95 % CI for β_1 is closest to

A. $(-1.44, -0.82)$
Half-width too small (used wrong SE).

B. $(99.26, 120.48)$
Uses the intercept *scale.*

C. $(-1.78, -0.48)$

D. $(-11.74, 9.48)$
Uses residual SD, not slope SE.

E. $(-2.43, 0.17)$
Wrong half-width and sign range includes 0.

43. Should we interpret $\hat{\beta}_0$ in this study?

A. Yes; it gives aptitude when age = 0 months.
Age = 0 never occurs and is outside data range.

B. Yes; intercept is always meaningful.
Not if $x = 0$ is implausible.

C. No; R^2 is too low.
R^2 doesn't govern interpretability of β_0.

D. No; SE is large.
Uncertainty $\neq$ scientific meaning.

E. **No; age = 0 is impossible here, so the intercept lacks scientific meaning.**

44. Two intervals from `predict(..., interval = "|A|")` and `"|B|"`:

A. Interval A: CI for mean, Interval B: PI
Their widths are opposite.

B. Both are prediction intervals.
One is clearly narrower.

C. Interval A for low aptitude, B for high.
Aptitude not in code.

D. **Interval A is a 95 % *prediction* interval; Interval B a 95 % CI for the mean response.**

E. Interval B is narrower due to bigger n.
Same n; difference is PI vs CI.

45. Predicted aptitude for a child who first speaks at 60 months:

A. 177
Plugged 60 into a wrong formula.

B. 6591
Multiplied instead of subtracting.

C. 104
Used +1.13 rather than –1.13.

D. **42**

E. 60
Simply echoed the age.

46. Can the prediction in Question 45 be trusted?

A. Yes; $R^2 > 0.3$ guarantees reliability.
R^2 alone never guarantees it.

B. Yes; residual SD is small.
Still outside data range.

C. No; negative correlation forbids prediction.
Sign isn't the issue.

D. **No; age = 60 is an extrapolation beyond the observed 7–42 months.**

E. No; linear models are never useful.
Too extreme.

47. Multiple linear regression allows us to

A. Have at most two predictors.
Any number is allowed.

B. Exclude categorical variables.
They enter via dummies.

C. Abandon `lm` for `mlm`.
`lm` fits multiple predictors.

D. **Assess each predictor's effect while *controlling* for the others.**

E. Interpret only one predictor's effect.
Each coefficient has meaning.

48. Interpreting $\hat{\beta}_0$ from the model with temperature only

A. **Temperature = 0 °C lies within the data (-4.5 – 24.8 °C), so $\hat{\beta}_0 = 1873.9$ is the estimated mean metabolic rate at 0 °C.**

B. We mustn't interpret it; 0 °C is biologically impossible.
0 °C is observed.

C. 0 °C is outside data – so extrapolation.
It's inside.

D. Intercept gives change per degree.
Slope does that.

E. Intercepts are never interpreted.
Context determines interpretability.

49. From the model with temperature + activity, which pair is $(\hat{\beta}_2,\, s_{\hat{\beta}_2})$?

A. 1623.69, 19.38
That's the intercept.

B. **52.64, 2.95**

C. –18.63, 1.05
Those are for temperature.

D. 52.64, 19.38
SE mismatched to estimate.

E. –18.63, 2.95
Estimate and SE from different predictors.

50. Best wording of the 95 % CI for β_2 (activity)

A. 46.8–58.5 change; temp ignored.
Must say "holding temperature fixed."

B. 46.8–58.5 *level* at 1 h.
CI is for change.

C. –20.7—16.6 for temperature.
That's the other predictor.

D. **We are 95 % confident the *mean* metabolic rate increases by 46.8–58.5 Kcal/day for each extra hour of activity, *holding temperature constant.***

E. Mixed effects of both predictors.
CI isolates activity's slope.

51. $R^2 = 84.8\%$. Which statement is *not* correct?

A. R^2 is the squared correlation between y and $\hat{y}$.
True.

B. R^2 ranges from 0 to 1.
True.

C. R^2 is the proportion of variance explained.
True.

D. The label "Multiple R-squared" shows R^2 in R output.
True.

E. **A model is useless unless $R^2 > 0.5$.**

52. Residual-plot shows a funnel shape. Which assumption fails?

A. Linearity
Trend looks straight.

B. Independence
Plot gives no info on independence.

C. Outliers
None obvious.

D. **Constant variance (homoscedasticity).**

E. No assumptions violated
We see heteroscedasticity.

53. ANOVA: hypotheses being tested

A. Difference of two means.
Only two groups.

B. Test of independence.
Different procedure.

C. All four means unequal.
Too strict.

D. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$; H_A**: at least one mean differs.**

E. Equality of full distributions.
Focus is on means.

54. F-value for that ANOVA

A. **27890**

B. 355 921
That is SS_{type}.

C. 118 640
That is MS_{type}.

D. 4
Residual MS.

E. 14 455
SS_{res}

55. Which statement is *not* correct about ANOVA?

A. ANOVA compares between-group to within-group variance.
Correct.

B. Under H_0, the F-value follows an F-distribution.
Correct.

C. Large F suggests group means explain much variation.
Correct.

D. df in the table choose the reference F-distribution.
Correct.

E. **The ANOVA table itself tells us which group means differ.**

56. Which reference distribution gives the p-value?

A. Standard normal.
Wrong family.

B. χ^2_3.
Not used here.

C. $F_{-}(3, 3398)$

D. χ^2_{3398}
Wrong.

E. t_{3398}
Wrong family.

57. Interpreting the tiny p-value from ANOVA ($\alpha = 0.01$)

A. **Because p ¡ α, the data are very unlikely if all pitch-type means were equal.**

B. p ¡ α proves slider = change-up speed.
ANOVA is omnibus.

C. All four means differ.
Post-hoc needed to know.

D. Shows Kershaw is "hard to face."
Beyond the scope of speed.

E. Identifies slowest and fastest pitch.
Need pairwise tests.

58. `TukeyHSD(a_baseball)` does what?

    A. Fits separate regressions.
    No, it's post-hoc on the ANOVA.

    B. Tests equality of variances.
    Levene/Bartlett do that.

    C. **Compares every pair of pitch-type means with a multiple-comparison adjustment.**

    D. Gives an updated ANOVA table.
    Already have one.

    E. Tests each pitch vs overall mean.
    Not Tukey's HSD.

59. Interpreting the row `FF-CU` in Tukey output

    A. More fastballs than curveballs per 100 pitches.
    Counts, not speeds.

    B. Faster than "non-fastballs."
    CU only.

    C. More strikes per 100 pitches.
    Not strike data.

    D. **Fastballs average 30.04 – 30.60 km/h *faster* than curveballs (95 % confidence).**

    E. 95 % probability statement.
    Confidence, not probability about a fixed quantity.

60. Which is *not* an example of binary data?

    A. **The *number* of eggs in a nest**
    Takes many integer values – a count, not yes/no.

    B. Presence / absence of a gene.
    Two outcomes.

    C. Dolphin breeding status (breeder / non-breeder).
    Two outcomes.

    D. Task completed within time? (yes / no)
    Two outcomes.

    E. Penguin age status (adult / juvenile).
    Two outcomes.

61. Which *binomial* assumption is *violated* in this bird-nest study?

    A. Two possible outcomes (success / failure) for each trial hold.
    This one is satisfied.

B. The probability of success p is the same for every trial.
Also satisfied.

C. Successive trials are independent.
No evidence of dependence was given.

D. Each trial outcome is recorded accurately.
No hint of mis-classification.

E. **The number of trials n is fixed in advance.**
Here the count of nests keeps growing, so n is not predetermined.

62. What proportion of visits (6 430 of 7 729) resulted in a sale?

A. **0.83**

B. 0.53
6 430 ÷ 12 200 was used—wrong denominator.

C. 0.94
Divided 7 729 by 8 200.

D. 0.75
Rounded from 5 800 ÷ 7 729, which is not the right count.

E. 0.78
Used an intermediate rounded figure, not the exact ratio.

63. What fraction of tagged birds later returned? (419 of 791)

A. 0.33
419 ÷ 1 270—wrong denominator.

B. **0.53**

C. 0.91
791 ÷ 870—swapped numerator and denominator.

D. 0.41
327 ÷ 791—wrong numerator.

E. 0.80
Rounded 634 ÷ 791—again wrong numerator.

64. In R, `prop.test(x1, x2)` reports a 95 % CI labelled "`p1 - p2`". What parameter does that interval estimate?

A. p_1
Only the first group's proportion.

B. p_2
Only the second group's proportion.

C. $p_1 - p_2$ **(the *difference* between the two proportions).**

D. $\frac{p_1}{p_2}$
That would be a risk ratio.

E. The pooled overall proportion.
Not what the CI targets.

65. By default, `prop.test` for a 2×2 table tests the null hypothesis

A. $p_1 = 0.5$
Only one group.

B. $p_2 = 0.5$
Only one group.

C. $p_1 + p_2 = 1$
Adds instead of compares.

D. **$p_1 = p_2$ (equivalently, $p_1 - p_2 = 0$).**

E. $p_1 p_2 = 0$
Product is irrelevant here.

66. For a 2×2 table, `prop.test` and `chisq.test` (with Yates correction) will always give

A. **The same χ^2 statistic and therefore the same p-value.**

B. Different χ^2 but same p-value.
Statistic itself is identical.

C. Same statistic but a different reference distribution.
Both use χ^2_1.

D. An F-test in one case.
Neither uses the F-family here.

E. Completely unrelated results.
They address the same hypothesis.

67. Expected count (cancer cases, "light smokers" cell) is calculated as (row total$\times$column total)/grand total. Which option matches that formula?

A. Adds row to column totals.
Should multiply.

B. **Multiplies the row total by the column total, then divides by the grand total.**

C. Uses the larger of the two totals.
Not the expected-count rule.

D. Divides column by row total.
Reverses the formula.

E. Grand total divided by both subtotals.
Upside-down.

68. Using that rule, the expected number of cancer cases in the "never-smoked" group is approximately

A. 4.6
Halved the correct value.

B. 18.4
Doubled the correct value.

C. 12.0
Rounded a mid-step mistakenly.

D. **9.2**

E. 25.0
Used column total instead of grand total.

69. The χ^2 test on that 4×2 table checks

A. Whether the four row totals equal the two column totals.
Totals are fixed.

B. Goodness-of-fit against a theoretical Poisson.
Wrong framework.

C. **Independence between smoking status (4 levels) and cancer (yes/no).**

D. Equal variances across rows.
Different concept.

E. Equality of four separate proportions at once.
Independence phrasing is preferred.

70. The degrees of freedom for a 4×2 contingency table are

$$(r-1)\,(c-1) \;=\; (4-1)(2-1) =$$

A. 1
Forgot $(r-1)$.

B. 2
Computed $(c-1)$ only.

C. **3**

D. 5
Added instead of multiplied.

E. 6
Used $r \times (c-1)$.

71. The test yielded $p \approx 0.30$. At $\alpha = 0.05$ we therefore

A. Declare smoking *causes* cancer.
Causality not established and $p > 0.05$.

B. Prove H_0 is true.
Failing to reject $\neq$ proof.

C. Need a bigger sample: $n < 500$.
$n > 6\,000$ already.

D. **Fail to reject H_0; evidence of association is weak.**

E. Automatically switch to Fisher's exact test.
Expected counts are large enough.

72. What does the Mann–Whitney (Wilcoxon rank-sum) test compare?

A. Equality of *means*.
It compares distributions via ranks.

B. Variances of two groups.
That's Levene's/Bartlett's test.

C. **Whether one distribution tends to give higher (or lower) values than the other.**

D. Medians *only*.
Shift in distribution, not just the median.

E. Counts of observations above 0.
Not its statistic.

73. Its test statistic is the

A. Sum of the *signed* residuals.
Signed ranks belong to the Wilcoxon signed-rank *test for paired data.*

B. Mean of raw values in group A.
Raw data not used directly.

C. Variance of the pooled sample.
Not relevant.

D. Largest observation's rank.
Uses all ranks, not just one.

E. **Sum of the ranks for one of the two groups.**

74. Which statement about a non-parametric rank test is *true*?

A. They are always *more* powerful than t-tests.
Often less *powerful.*

B. They provide easy CIs for a difference of means.
Mean-based CIs are awkward.

C. **They require no normality assumption for the underlying data.**

D. Significance implies causation.
Design, not p-values, yields causality.

E. Their p-value is guaranteed to be larger than in a parametric test.
Could be smaller or larger.

75. A p-value of 0.004 from a Mann–Whitney test means

A. The average of the ranks equals the average of raw data.
Ranks are not raw values.

B. 0.4 % probability the null is true.
$p \neq Pr(H_0 \text{ true})$.

C. **If the two distributions were identical, we would see a rank-sum this extreme only 0.4 % of the time.**

D. There is no evidence of a difference.
$p = 0.004$ is strong evidence against H_0.

E. Effect size must be large.
Significance $\neq$ effect magnitude.

76. A stratified random sample

A. Might ignore one stratum entirely.
By definition, each stratum is sampled.

B. Picks exactly the same # units from each stratum.
Proportional or unequal allocation is common.

C. **Draws separate random subsamples within every stratum.**

D. Always uses systematic sampling inside strata.
Can be simple random or other methods.

E. Guarantees lower variance than an SRS.
Often but not always.

77. In the cannabis-use survey, the biggest threat to validity was

A. Selection bias from non-response.
High response rate reported.

B. Confounding by age.
Age was adjusted for.

C. **Information (reporting) bias: people may under-report cannabis use.**

D. Instrument calibration error.
Self-report, not instruments.

E. Over-powered sample inflating small effects.
Not the main concern.

78. The MAO-A gene paper comparing Māori and non-Māori was criticised mainly because

A. It involved no ethics approval.
Approval was obtained.

B. **Researchers failed to involve Māori sufficiently in study design and interpretation.**

C. The statistical model assumed independence.
Standard assumption, not unique flaw.

D. Results were not peer-reviewed.
They were published in a journal.

E. The gene was sequenced inaccurately.
Genotyping method wasn't the issue.

79. Which practice *violates* the principle of genuine co-design with communities?

A. Researchers present preliminary ideas at a community hui.
Involves community.

B. Community representatives sit on the advisory board.
Also involvement.

C. Community veto over data sharing is respected.
Aligns with co-design.

D. **Investigators finalise aims and methods *before* any community consultation.**

E. Draft results are fed back for comment.
Again involves community.

80. In the CARE principles for Indigenous data, the "C" stands for

A. **Collective benefit**

B. Consent
Important, but not the "C" in CARE.

C. Custodianship
Covered by the "A" and "R" elements.

D. Confidentiality
Not the first pillar here.

E. Collaboration
Embedded in the overall framework, but not the C-word.

81. In a clinical trial, what best describes a *placebo*?

A. Some control groups receive no treatment at all.
"No-treatment" ≠ placebo if nothing is given.

B. A placebo pill lets participants know they are in the control arm.
Blinding means they shouldn't *know.*

C. **An inert treatment that looks the same as the active drug, helping blind participants and researchers.**

D. Placebos are unethical because they deny therapy.
Ethics boards permit them when no proven therapy exists.

E. Placebos make the study single-blind only.
They enable single or *double blinding.*

82. Which variable qualifies as a *confounder* in an observational study?

A. Related only to the *predictor* but not the outcome.
Must relate to both.

B. Related only to the *outcome* but not the predictor.
Must relate to both.

C. Completely independent of both predictor and outcome.
Cannot confound.

D. **Associated with both predictor and outcome, potentially distorting the observed relationship.**

E. Any variable measured after the outcome.
Timing is wrong for confounding.

83. The "replication crisis" in science refers to

A. Journals publishing too *few* papers.
The issue is about reproducibility, not volume.

B. Choosing Bayesian methods over frequentist ones.
Both paradigms face replication issues.

C. **Many published findings failing to reproduce when the studies are repeated independently.**

D. A shortage of funding for statistics.
Not the core of the crisis.

E. Peer reviewers demanding larger samples.
Better power aids replication.

84. Testing dozens of hypotheses on one data set without adjustment mainly

A. Controls the Type I error at 5 % overall.
It inflates the family-wise error.

B. Leaves the false-positive rate unchanged.
Actual rate increases.

C. **Inflates the chance of at least one false positive (Type I error).**

D. Guarantees smaller p-values are true discoveries.
They may be false positives.

E. Eliminates the need for follow-up studies.
Replication is still needed.

85. "HARKing" (Hypothesising After Results are Known) is the practice of

A. **Formulating or changing hypotheses *after* seeing the data, then presenting them as if specified beforehand.**

B. Pre-registering study aims.
That is the opposite practice.

C. Randomising subjects after consent.
Unrelated to hypotheses.

D. Adding more predictors during peer review.
Could be data dredging, but not the definition of HARKing.

E. Publishing only significant outcomes.
That is publication bias.

86. Which statement about *maximum likelihood estimation* (MLE) is correct?

A. MLE requires a Bayesian prior.
MLE is purely likelihood based.

B. MLE only works for very large samples.
It is defined for any sample size.

C. Computes p-values directly.
MLE gives point estimates; tests use additional steps.

D. **MLE chooses parameter values that maximise the data's likelihood under the assumed model.**

E. Seldom used in real research.
It is standard in many fields.

87. In Bayesian analysis, which statement is *false*?

A. Bayesian inference combines prior information with the data likelihood.
Core principle—true.

B. The posterior distribution expresses updated beliefs about parameters.
True.

C. **Bayesian methods ignore the likelihood and rely only on the prior.**

D. Bayesian credible intervals are probability statements about parameters.
True.

E. Bayesian methods are widely applied in modern statistics.
True.