18.05 Problem Set 8, Spring 2025 Solutions

Problem 1. (35: 5, 5, 5, 5, 5, 5, 5 pts.) **Spinning gold**
(a) Solution: H_0: $\theta = 0.5$

Test statistic: x = number of heads in 250 spins.

Data: $x = 140$.

The probability of getting a result at least as extreme as seen is the $\boxed{p\text{-value.}}$

We'll show how to compute this value and decide if the test is one or two-sided at the same time.

A one-sided test would have alternative hypothesis H_A: $\theta > 0.5$. In this case, data at least as extreme means $x \geq 140$. Using R we compute the one-sided p-value:

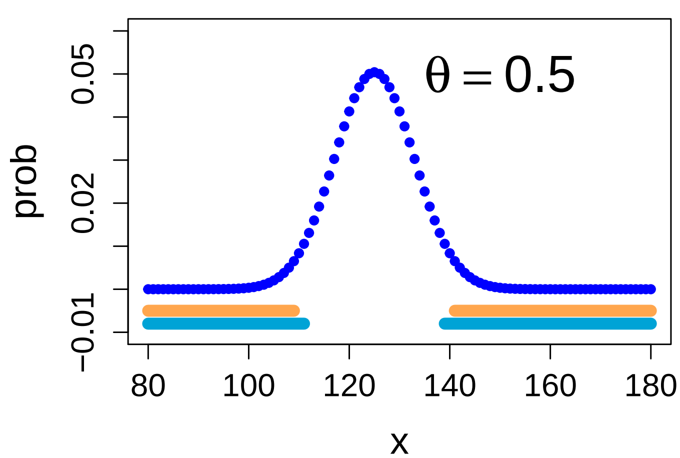
$$p = P(x \geq 140 | H_0) = \texttt{1 - pbinom(139, 250, 0.5)} = 0.03321$$

This is not the 7% value we were looking for. So let's consider the two-sided test.

A two-sided test would have alternate hypothesis H_A: $\theta \neq 0.5$. Since the null distribution, binomial(250,0.5), is symmetric around 0, each tail in the rejection region will have probability $\alpha/2$ and the two-sided p-value is computed by doubling the smaller of the one-sided p-values. We computed the right tail p-value just above. This is the smaller of the two p-values so our two-sided p-value is $2 \times 0.03321 = 0.06642$. This rounds to 0.07, so the figure of 7% is the two-sided p-value.

Note: we could have used the normal approximation binomial$(250, 0.5) \approx \text{N}(125, 250/4)$, and the z-statistic $z = \frac{x-125}{\sqrt{250/4}} \approx \text{N}(0, 1)$. In this case, our p-values would be: one-sided: $p = P(z \geq \dfrac{15}{\sqrt{250/4}}) \approx 0.02889$ and two-sided: $p = P(|z| \geq \dfrac{15}{\sqrt{250/4}}) \approx 0.05778$.

(b) Solution: As instructed, we use a two-sided rejection region as in part (a). The exact p-value was $p = 0.066$. Since $0.05 < p < 0.1$ we reject H_0 at significance $\alpha = 0.1$ and don't reject at $\alpha = 0.05$.



The figure shows the null distribution, the $\alpha = 0.1$ rejection region (blue-green) and the $\alpha = 0.05$ rejection region (orange). Notice that the data $x = 140$ is in the 0.1 rejection region but not the 0.05 rejection region.

(c) Solution: Again, we use two-sided rejection region. The problem asks us to find the rejection region for $\alpha = 0.01$. We use R to find the endpoints for the rejection region (called critical values):

$$\texttt{criticalValue.left = qbinom(0.005,250,0.5) - 1} = 104$$
$$\texttt{criticalValue.right = qbinom(0.995,250,0.5) + 1} = 146$$

Note: we added or subtracted one to the value returned by `qbinom`. For a discrete distribution like the binomial there is not an exact critical value. So `qbinom(x, n, p)` returns the smallest integer with more than x probability in its left tail. Since the rejection region must have at most $\alpha/2$ in either tail we have to move the R answer by one towards the tail.

Conclusion: we reject for greater than or equal to 146 heads or less than or equal to 104 heads.

(d) Solution: (i) Again we use a two-sided rejection region. For $\alpha = 0.05$ the rejection region is given by the critical values

$$\texttt{criticalValue.left = qbinom(0.025,250,0.5) - 1} = 109$$
$$\texttt{criticalValue.right = qbinom(0.975,250,0.5) + 1} = 141$$

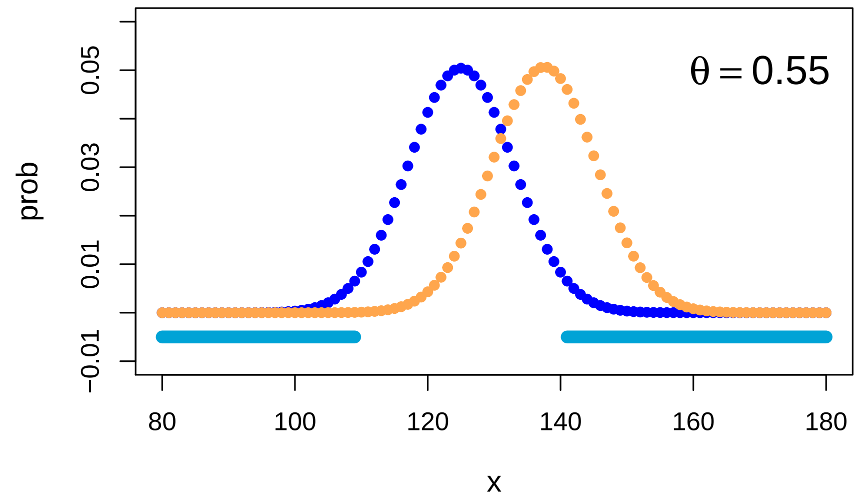
$$\begin{aligned}\text{power when } \theta = 0.55 &= P(\text{reject} \mid \theta = 0.55) \\ &= P(x \leq 109 \text{ or } \; x \geq 141 \mid \theta = 0.55) \\ &= \texttt{sum(dbinom(0:109, 250, 0.55)) + sum(dbinom(141:250, 250,0.55))} = 0.35237\end{aligned}$$

Likewise

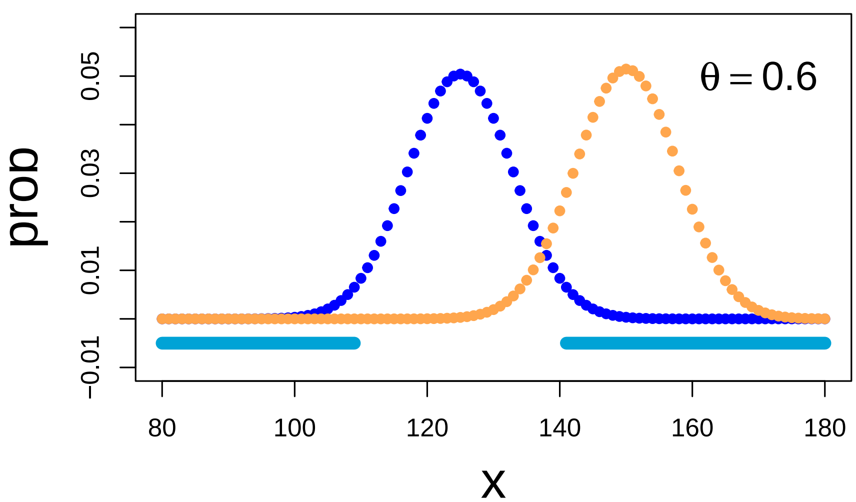
$$\begin{aligned}\text{power when } \theta = 0.6 &= P(\text{reject } \vert \theta = 0.6) \\ &= P(x \leq 109 \text{ or } \; x \geq 141 \mid \theta = 0.6) \\ &= \texttt{sum(dbinom(0:109, 250, 0.6)) + sum(dbinom(141:250, 250,0.6))} = 0.88963\end{aligned}$$

Note: We could have used `pbinom`. Doing the sums with `dbinom` was an easy way of avoiding off-by-one errors with `pbinom`.

(ii) The two plots below show the null distribution and the distribution of H_A for $\theta = 0.55$ and $\theta = 0.6$ The blue line below the graphs shows the rejection region. The greater power when $\theta = 0.6$ is explained by its greater separation from H_0. Most of the probability of $p(x|\theta = 0.6)$ is over the right side of the rejection region.



Distributions of H_0 and H_A with $\theta = 0.55$



Distributions of H_0 and H_A with $\theta = 0.6$

(e) Solution: The answer is $n = 1055$ with H_A giving a power of 0.9003.

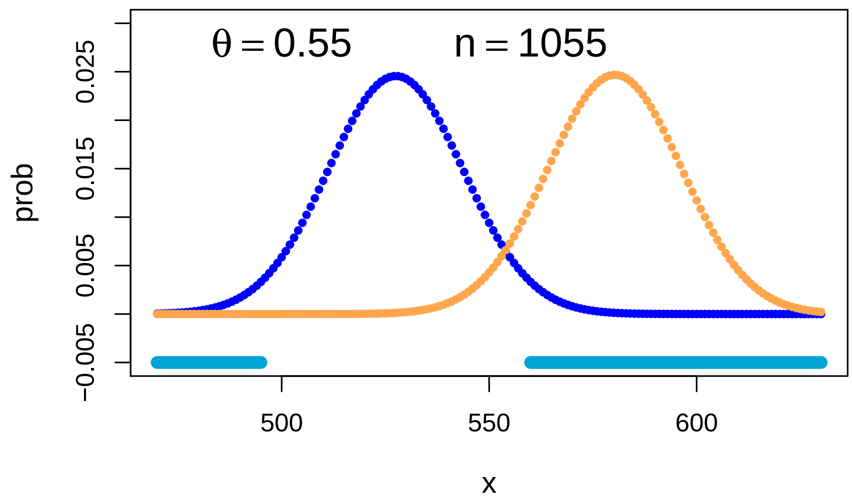
To get this we need to compute the power for various values of n. The steps for each n are:

1. Find the rejection region.
2. Compute the power.

Here is the R-code for one value of n. Code with a loop to check through all values of n until we find the first with power = 0.9 is in ps8.2024-sol.r, which is posted in the usual place.

```
theta = 0.55
n = 300;
# Find critical values for rejection region (based on theta=0.5)
criticalValue.left = qbinom(0.025,n,0.5) - 1;
criticalValue.right = qbinom(0.975,n,0.5) + 1;
rejectionRegion = c(0:criticalValue.left, criticalValue.right:n)
power = sum(dbinom(rejectionRegion, n, theta))
print(power)
```

See the two plots with part (d): power increases as n increases because the distributions become more separated.



Plot for $n = 1055$ of the H_0 and $H_A : \theta = 0.55$ distributions. The blue lines show the rejection region.

An alternative approach approximating the exact answer with normal distributions is given at the end of these solutions.

(f) Solution: We use the usual Bayesian update table.

Hypothesis	prior	likelihood	posterior
$\theta = 0.5$	1/2	$c_1(0.5)^{250}$	$c_2(0.5)^{250} = 0.14757$
$\theta = 0.55$	1/2	$c_1(0.55)^{140}(0.45)^{110}$	$c_2(0.55)^{140}(0.45)^{110} = 0.85243$

The normalizing factor $c_2 = \dfrac{1}{(0.5)^{250} + (0.55)^{140}(0.45)^{110}}$.

The posterior probability that $\theta = 0.55$ is 0.85.

(g) Solution: If we use the Beta(1, 1) (flat) prior on θ in [0,1]. Then the posterior for θ is a Beta(141, 111) distribution. With this posterior

$$P(\theta > 0.5|\text{data}) = 1 - \texttt{pbeta}(.5, 141, 111) \approx 0.97.$$

This is 97%.

Problem 2. (10: 5, 5 pts.) **Polygraph analogy.**

(a) Solution: Type I error is rejecting the null-hypothesis when it is indeed true. This corresponds to thinking someone is lying when they are in fact being truthful. The experiment had $\frac{9}{140}$ type I errors. This is our estimate of the probability of a type I error.

Type II error is not rejecting the null-hypothesis when it is indeed false. This corresponds to thinking someone is telling the truth when they are in fact lying. Based on the data our estimate of the probability of a Type II error is $\frac{15}{140}$.

(b) Solution: Significance = P(type I error) = P(reject $H_0 \mid H_0$).
Power = 1 - P(type II error) = P(reject $H_0 \mid H_A$).

Problem 3. (25: 5, 10, 10 pts.) **z-test**
(a) Solution: Let μ be the actual speed of a given driver. We are given that

$$x_i \sim \mathrm{N}(\mu, 5^2) \;\Rightarrow\; \bar{x} \sim \mathrm{N}(\mu, 5^2/3).$$

The most natural hypotheses are:

H_0: the driver is not speeding, i.e. $\mu \le 40$.
H_A: the driver is speeding, i.e. $\mu > 40$.

Both are composite.

Note: we will work with H_0: $\mu = 40$, which is simple.

(b) Solution: (i) Giving a ticket to a non-speeder is a type I error (rejecting H_0 when it is true). H_0 is composite, but we can do all our computations with the most extreme value $\mu = 40$ because the one-sided rejection region will have its largest significance level when $\mu = 40$.

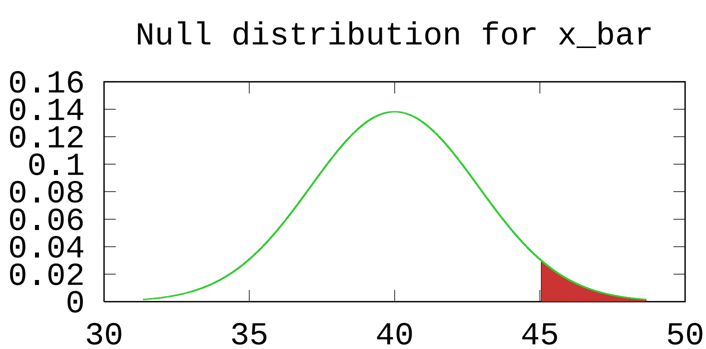
So the null distribution is $\bar{x} \sim \mathrm{N}(40, 5^2/3)$. The critical value is

$$c_{0.04} = \texttt{qnorm(0.96,40,5/sqrt(3)) = 45.054}$$

(Equivalently $c_{0.04} = 40 + z_{0.04}\frac{5}{\sqrt{3}} = 45.054$.)

That is, they should issue a ticket if the average of the three guns is more than 45.054.

(ii) Here is a plot of the null distribution $\mathrm{N}(40, 5^2/3)$. The rejection region with probability of 0.04 is shown.



(iii) We cannot compute this posterior probability without a prior probability that a random driver is speeding.

(iv) If no one is speeding then 100% of tickets are given in error.

(c) Solution: (i) Power = $P(\text{rejection} \mid H_A)$. So to find the power we first must find the rejection region. For $n = 3$ this was done in part (b): rejection region = $[45.054, \infty)$. So

$$\text{power } = P(\text{rejection} \mid \mu = 45) = \texttt{1 - pnorm(45.054, 45, 5/sqrt(3)) = } \boxed{0.493}$$

(ii) With n cameras (guns) let's write $\overline{x}_n$ for the sample mean. The null distribution is

$$\bar{x}_n \sim \text{N}(40, 5^2/n)$$

The critical value (left endpoint of the rejection region) and power depend on n.

We use R to compute the power for $n = 3$ using the code below: `n = 3`

```
mu = 40
sigma = 5/sqrt(n)
alpha = 0.04
xcrit = qnorm(1 - alpha, mu, sigma)
power = 1 - pnorm(xcrit, 45, sigma)
```

By increasing n, we find the first value of n gives power greater than 0.9 is $\boxed{n = 10}$.

(iii) In order to do the computations algebraically we need to write everything in terms of standard normal values.

$$c_{0.04} = \texttt{qnorm(0.96, 40, } 5/\sqrt{n}) = 40 + z_{0.04}\frac{5}{\sqrt{n}}$$

where $z_{0.04}$ is the standard normal critical value

$$z_{0.04} = \texttt{qnorm(0.96, 0, 1) = 1.751}.$$

We want

$$\text{power} = P(\overline{x} \geq c_{0.04} \mid \mu = 45) = 0.9$$

Standardizing and doing some algebra we get

$$P\left(\frac{\overline{x} - 45}{5/\sqrt{n}} \geq \frac{c_{0.04} - 45}{5/\sqrt{n}}\right) = 0.9 \quad \Rightarrow \quad P\left(z \geq \frac{-5}{5/\sqrt{n}} + z_{0.04}\right) = 0.9$$

Thus $\dfrac{-5}{5/\sqrt{n}} + z_{0.04} = z_{0.9}$. We get

$$n = (z_{0.04} - z_{0.9})^2 = (1.7507 - (-1.2816))^2 = 9.1945.$$

Setting n to be the next biggest integer we get $n = 10$.

Problem 4. (25: 5, 5, 5, 5, 5 pts.) **Climate change in Massachusetts**

(a) Solution: The formula for sample variance is $s^2 = \frac{1}{19}\sum_{i=1}^{20}(x_i - \overline{x})^2$.

For 1895 to 1914, $\overline{x} = 46.060$ and $s^2 = 1.017$.

For 2004 to 2023, $\overline{x} = 49.415$ and $s^2 = 1.501$.

(b) Solution: The pooled variance is

$$s_p^2 = \frac{19 * 1.017 + 19 * 1.501}{38} \cdot \left(\frac{1}{20} + \frac{1}{20}\right) = 0.1259.$$

So

$$t = \frac{49.415 - 46.060}{\sqrt{0.1259}} = 9.455.$$

The null distribution is $t(38)$ so

$$p = 1 - \texttt{pt(9.455, 38)} = 7.92 \cdot 10^{-12}.$$

The p-value is much less than $\alpha = 0.001$, so we reject the null.

(c) Solution: We run

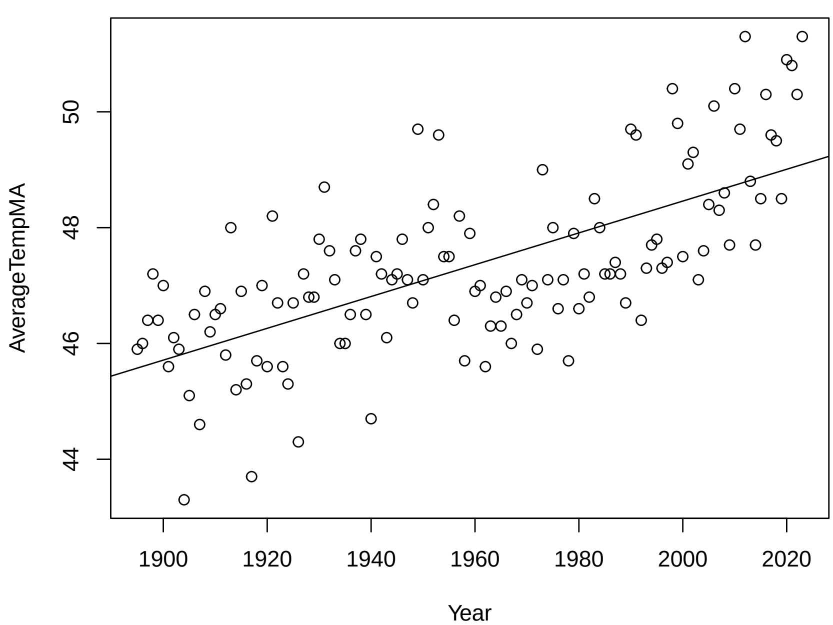
```
t.test(data[110:129,2], data[1:20,2], alternative="greater", var.equal=TRUE)
```

The output is:

```
t = 9.4543, df = 38, p-value = 7.931e-12
alternative hypothesis:  true difference in means is greater than 0
95 percent confidence interval:
2.756711 Inf
sample estimates:
mean of x mean of y
49.415 46.060
```

These values agree with (b) up to some rounding.

(d) Solution: Here is the plot:



The output of `summary(lm)` includes:

```
Estimate Std.  Error t value Pr(>|t|)
(Intercept) -6.472043 5.248783 -1.233 0.22
Year 0.027465 0.002679 10.253 <2e-16 ***
```

So the t statistic for the slope is 10.253 and the p-value is less than $2 \cdot 10^{-16}$. We certainly reject the null for any reasonable significance level.

(e) Solution: Using the best fit line, the prediction is $0.027465 \cdot 2100 - 6.472043 = \boxed{51.20^\circ}$.

Problem 5. (5 (extra credit): 2, 3 pts.) **Interpreting XKCD**

(a) Solution: We have hypotheses H_0 = 'the sun is okay' and H_A = 'the sun has gone nova'

The data is either 'yes' or 'no' from the detector.

We have the following likelihood table.

data	yes	no
$p(\text{data} \mid H_0)$	1/36	35/36
$p(\text{data} \mid H_A)$	35/36	1/36

The frequentist chooses the rejection region 'yes', which has significance 1/36. (Note: 1/36 is really $0.02777\ldots \approx 0.028$ not 0.027.)

The experimental data is 'yes', which is in the rejection region, so the frequentist correctly rejects H_0 in favor of H_A.

The Bayesian views this as silly, since, from their perspective, the posterior odds that the sun has gone nova are

$$\text{prior odds} \times \text{likelihood ratio} = \text{prior odds} \times \frac{p(\text{yes}|H_A)}{p(\text{yes}|H_0)} = \text{prior odds} \times \frac{35}{1}.$$

If we conservatively put the prior odds at $1/10^8$ then the posterior odds are still very small.

Besides, if the sun has gone nova, losing the bet is the least of the Bayesian's problem.

(b) Solution: The comic is pointing out the flaw of multiple testing or what's sometimes called data mining. (The bad type of data mining, there is also a good type.) A significance level of 0.05 means that in 20 experiments where H_0 is true we'd expect to reject it once. The scientists test 20 colors. So even if no jelly bean color causes cancer there is a high probability (well, 64%) that one of the tests will produce a test statistic in the rejection region.

The fix is to plan on doing n tests and set the significance level for any one test to α/n. Then, assuming H_0 is true for all the tests, the probability that at least one of them will reject is roughly $n \cdot \alpha/n = \alpha$. This is called the Bonferroni correction. (Actually, because of the possibility of multiple rejections the probability at least one will reject is less than or equal to α.)